

Dear Editor.

Thank you for your handling of our manuscript (tc-2018-286) by Hill et al. We are appreciative of the three very careful and helpful reviews. In the following pages, you will find our responses to the three reviews. We also have provided a revised manuscript. We look forward to hearing from you in the near future.

Kind Regards,
David Hill
Oregon State University

Reply to Matthew Sturm, Referee

Review of Converting Snow Depth to Snow Water 1 Equivalent Using Climatological Variables

February 18, 2019

Referee comments are left-justified, in black. Author replies are indented, in blue.

In this paper, the authors address the problem of converting more readily obtained snow depth measurements to snow water equivalent values. The problem is highly topical as airborne lidar and airborne and satellite-based photogrammetric snow depths become more readily available for widespread use. The authors primarily build on the method described by Sturm, Taras, Liston, Derksen, Jonas and Lea (2010), with the main difference in their method being the replacement of *climate classes of snow* by continuous climate variables (*mean annual precipitation and February mean temperature*) obtained from the PRISM data set. Though not explicitly stated, the authors also establish their regressions using a larger data set than the Sturm et al. study and most other studies of which I am aware. They reach the conclusion that their regressions show an improvement over the 2010 work.

As the lead author of that prior depth-to-SWE study, I find this a fine piece of work, clearly and honestly written, and useful to many practitioners. It should be published. That said, I am not sure that I fully agree with the conclusion of the authors as to the extent of the improvement, whether their improved method is more easily applied than the old, and I find the omission of any discussion of the well-known errors in the data set used to develop the regression equations troubling. I would like to see the authors grapple with this last issue explicitly in the paper before a version of the paper is published.

Thank you for this general overall assessment. Below, we provide point-by-point responses to your comments and we indicate where and how we plan to revise the manuscript prior to publication.

Examining the input data for this study (Table 1), 98.5% is essential SNOTEL snow pillow data; 1.5% comes from coring. Both types of data are known to contain biases. My personal experience for the latter (coring) is that it tends to undersample SWE (or produce low-biased density values), and across prior studies, there is agreement the method is no more accurate than about $\pm 10\%$.

We would like to point out that all of the data used to construct the regression model are snow pillow data. Table 1 summarizes the data used to build the model, and also the other independent data sets used to validate the model. We have clarified in the manuscript (beginning of section 2.2) and table (using bold font in the table) which data are used for what purposes.

It has been some time since I worked through the literature on snow pillow data, but I recall significant biases from these instruments as well. One source of error is due to snow bridging with, particularly, low biases during the melt when percolating meltwater can run off the pillow to the surrounding snowpack due to the shape of the pillows. Sonic sounders also can exhibit some measurement errors (in this case the ones near the SNOTEL site paired to the pillow SWE values, chiefly in not being representative of the snow depth on the pillow.

This is an important point. There are some studies¹ that show that SNOTEL sites can report SWE > accumulated precipitation, attributed to drifting snow. However, this would not bias snow density assuming that the SWE and Hs measurements are co-located. There are other studies² that have looked at the measurement bias in SWE depending on whether or not the snow pillow is steel vs. hypalon. One comprehensive study³ of biases notes a complex situation, where SWE is sometimes under-reported due to 'snow bridging', but over-reported at other times (see Table 1 of that paper). While that paper proposes methods for correcting SWE measurements, it is complex in practice, requiring continuous SWE, Hs, and near-ground temperature measurements. Please continue to our next remark below.

Given these potential sources of error, and the fact that the authors are attempting to develop general depth-SWE regressions, they should examine how these errors might cause their results to deviate from the "true" local conversion functions. For example, hypothetically, in a maritime regime, perhaps the natural snow packs retains frequent rain-on-snow water, but at the actual measurement sites it runs off from the pillows. Then there would be a consistent tendency in this February-warm location with high MAP (mean annual precipitation) to have light (or low) SWE vs. depth values. At least describing in what ways the modeled SWE values might diverge from the on-the-ground values would alert readers to limitations in the methodology.

This is a sensible suggestion. In the first draft of the paper, we did investigate the effect of measurement precision. In our revision, we now provide more discussion about potential errors in snow pillow measurements (to help alert readers, as you suggest). One complicating issue is that many studies that report on 'errors' in SWE from snow pillows define this error as the difference between the snow pillow and a coring measurement. The implicit assumption is that the coring measurement is the 'ground truth' but as you note, coring is good to +/- 10%. Given the lack of any consensus information on the distributions of errors in snow pillow measurements (we provide some citations to show the divergence of studies out there), we are unable to provide any good quantitative information on the effects of the pillow errors on the SWE estimates.

¹ <https://journals.ametsoc.org/doi/pdf/10.1175/JHM-D-12-066.1>

² https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs141p2_032059.pdf

³ <https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.5795>

As far as whether this study is an improvement over our 2010 study, it really comes down to which ancillary data set one wants to work with: a gridded data set of snow classes or of PRISM climate data. Each has advantages and disadvantages in terms of computational cost and hassle. Looking at Table 4 which compares our prior work to the new work, most of the statistical improvement comes from the taiga snow class, which, as the authors note (Line 415), is because in 2010 we assigned a fixed value to this class (e.g., a fixed value performed better than regressed values). This snow class was only 6% of our training set, and I suspect the sample we chose tended to be quite “stiff” because of the high percentage of depth hoar found in taiga snow, thus it did not tend to densify due to overburden stress (probably something of an Arctic bias we showed). The authors taiga sample set is deeper with greater SWE.

With regards to relative model performance. In our first draft, we tried to be as objective and factual as possible, in the sense of simply providing the comparative results (both figures and RMSE values). We feel that this is fair and appropriate. We also felt it appropriate to break out results by snow class so that readers could see how the comparison varied based on that. We have added information about how many data points in the aggregated CONUS, BC, and AK dataset are in each snow class (in section 3.1).

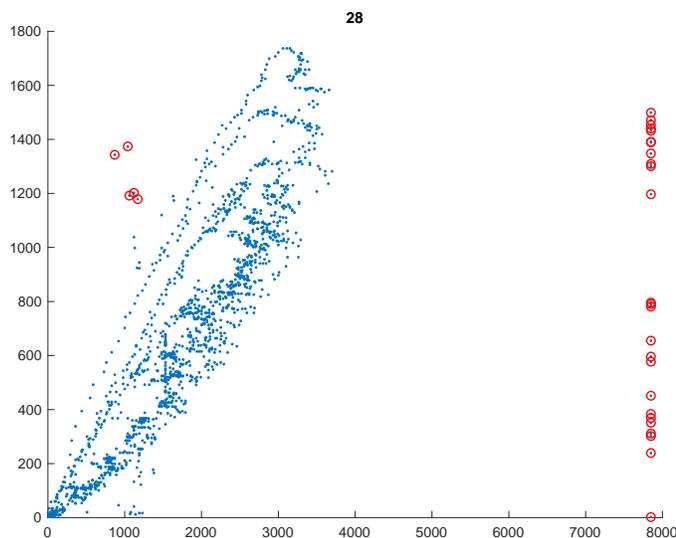
Regarding computational cost and hassle. The Sturm approach uses a straightforward equation and only requires access to the 1km snow-cover raster. Our approach uses more ‘data’ in the sense that numerous PRISM grids are required. However, we have packaged all necessary files into a freely available (will be released on GitHub upon acceptance of this paper) function that is very easy to use. By doing so, we alleviate any cost and hassle concerns.

One last substantive comment: The authors have an entire section on outlier detection and removal, but I would argue they have potentially removed real data. I applaud them for recognizing the hysteresis loop that is produced by depth-SWE seasonal evolution (Fig. 1) and their clever way of handling it in their regressions (Equation 5). We had actually during our work looked at using a rotated lemniscate to model this behavior, but dropped it because we could not make it work right. But if one recognizes that physically the bulk density increases during the melt during the Spring, then one also has to recognize that very early in the winter, deep fluffy snow will be found on some snow pillows....snow with bulk density values of that are less than 150 kg/m^3 . Figure 4 (clean version) has a lower depth-SWE line that at 2000 mm is about 350 mm SWE, a density of 175 kg/m^3 , and a density of 180 kg/m^3 at 3000 mm depth. I believe actual depth-SWE data on the low end has been removed, not erroneous data. Now one might argue we may in general seem to introduce a low bias when we do these sort of regressions, but that is not reason to label what may be accurate physical data as outliers. As further confirmation, the color of the removed data in Figure 4 is mostly blue (early season) and this removal would impact thin climate classes (e.g. taiga) more than thick classes.

This is a great point, and one which the authors have discussed at some length, following your review. Manual examination of many of the SNOTEL time series revealed

the presence of clearly wrong data (Figure 3 of the paper). We wanted to develop a wholly objective method for removing those data points. There is a lack of clarity and/or consensus in the literature about how to do this. The approach that we used seems like a good one, based upon the characteristics of the bivariate distribution. We recognize that some valid data points (mostly at low SWE-Hs values) are undoubtedly removed as well. Given the very low number (less than 1%; so the valid points removed are some small fraction of this 1%) of points that were removed in our process, we feel that this is acceptable.

This figure of the output of the data removal process illustrates things. SWE on vertical axis, h on horizontal. Removed points are in red



We acknowledge that referring to this process as ‘outlier’ detection is perhaps too strong and we have modified the language accordingly, notably re-titling Subsection 2.1.1.5. We also note that Anonymous Referee #3 had a similar comment and wanted to see a histogram of the DOY of the removed data points. We have gone back and looked at the distribution of DOY for all removed points. It turns out that the mean value of DOY was 160 and the standard deviation was 65. So, the bulk of the removed points comes from the middle of the snow season, not at the beginning or the end. This seems to alleviate a bit of the concern that you raise above.

One final comment, and this would be not only to the authors of this paper, but virtually every author out there. Please try to cite the seminal or original papers on a topic if possible...not the newest or easiest to cite. The authors here do well in citing Alford and Church, but when it comes to recognizing how snow depth and SWE are related in time and space, the seminal work of G. A. McKay should not be overlooked.

McKay, G. A., and B. F. Findlay, 1971: Variation of snow resources with climate and vegetation in Canada. Proc. 39th Western Snow Conf., Billings, MT, Western Snow Conference, 17–26.

McKay, G. A and D. M. Gray, 1981: The distribution of snowcover. Handbook of Snow, 1st ed. D. M. Gray and E. D. H. Male, Eds., Pergamon Press, 153–190.

We have added the former citation.

Detailed Comments:

Line 36: Surely someone before 2018 recognized that snow was important to hydrology...like Gerdel (see U.S. Army Corps of Engineers [1956] monograph on snow hydrology.

This citation has been added.

Lines 41-50: This paragraph is a little jumbled and doesn't address some of the well-known errors present in snow pillow measurements (see major comments), yet in the next paragraph, errors in SWE core values, which may actually be smaller in some cases, are identified. Perhaps here is where errors in the input data could be discussed in greater detail.

Yes, we have reworked the introduction a fair bit to bring in some information upfront about errors in coring and in snow pillows.

Lines 60-66: This little sections seems uneven, and given the huge literature on trying to extract SWE from remote sensing, particularly radar, very one-dimensional. Why even talk about snow remote sensing in the paper? I would simply say if falls outside of the scope of the work....and if there really is a reliable operational way to get SWE now from space, I don't know it.

We were trying to be comprehensive in laying out all of the options (in-situ vs. remote) for acquiring snow information. Your suggestion (huge literature that we don't do justice to) is on point and we have removed this section from the paper.

Line 74: Again, Goodison did the seminal work on the sonic sounders. Perhaps you could cite him.

This is a sensible suggestion and we will do so.

Line 96: I think this citation should be: Jonas, Tobias, Christoph Marty, and Jan Magnusson. "Estimating the snow water equivalent from snow depth measurements in the Swiss Alps." Journal of Hydrology 378, no. 1-2 (2009): 161-167.

Correct. We had it right in the references, but incorrect in the in-line citation.

Lines 117-118: I do not agree that *a priori* complexity produces more accuracy. What is really going on here is that proximity to high quality input data tends to produce better accuracy. But that may be true whether the model used with the data is complex or simple. Basically, in a very heterogeneous snow world, when we have local driving data, the results regardless of the model, get better. One might even be able to argue, given the difficulties of measuring radiation in snowy locations, that energy balance models can introduce errors. I don't think you need to work so hard on making a case for the type of statistical approach developed in this paper. Ease of use, and generally the lack of driving data most places, make the case for you.

This is a fair point and we revised our wording.

Line 282: I like this section on DOY, even though in the end you fix the value to 180. Just the fact that the regressions are insensitive to the DOY of peak SWE is interesting.

Agreed, thank you for noting this. It was an unexpected result.

Lines 336-337: I wish the authors would expand this section. It is the heart of why the regressions work, and it is how this study and our 2010 study are related. Climate classes tell us which snow is warm and deep and tends to densify rapidly; high MAP and high February temperatures tell us the same thing, perhaps as the authors claim, even better (or maybe it is just that the training set being larger is better?) This said, the authors I think are aware that there are several snow packs in which due to development of depth hoar and wind slab, there is very limited increase in density over time. Icy snow too can resist densification.

We have expanded this section somewhat. We fully recognize that all bulk-density methods that rely on simple inputs like DOY or climatological weather characteristics are unable to capture numerous features of snowpacks. That is a limitation of the emphasis on simplicity.

Line 342: Figure 6 is nice and clear.

Thank you. We have slightly modified this to show the data clouds as heat maps (2d histograms, essentially) at the suggestion of another reviewer.

Line 357: The model errors will have NO impact on the local snow regime...I think you mean the impact will be on the predicted results.

Correct, this was poorly stated, and we have reworded this.

Lines 405 – 410: I realize that the authors are fond of their Chugach results, undoubtedly obtained with much effort, but these data constitute 0.004% of the entire ensemble and could readily be omitted, with the space saved a deeper examination of why the systems is working, and where I might fail to work well.

Another referee was also lukewarm on the inclusion of this dataset. We have removed most of it, except for the useful information that it provides on the variability in Hs that is observed over short distances. That is a valuable point to retain.

Lines 431-432: Consider why this is: early in the winter, the addition of new snow to a thin pack makes a dramatic change in the bulk density (e.g., called here noise, but which is real) while later in the winter that noise dissipates because the addition is an increasingly small percentage of what is already on the ground. While a model using historical data cannot adjust for this effect, one could talk about how the uncertainty in the modeled result decreases with time. Does it then increase again after the DOY of peak SWE?

These are good points. Yes, our model is using only climatological weather data, which know nothing about individual snowfall events. We have added remarks on this issue and we have added a new figure that shows the errors as a function of the DOY.

Lines 447-460: This is much too cursory a discussion of precision and accuracy, and it sets up a false strawman: more stations or better precisions? The real question is how do we achieve better accuracy, and by this I suspect we mean better more accurate assessments of snow water resources. Given that 95% of the data being used is SNOTEL measurements, then this question has to start with whether the SNOTEL sites were actually designed to be “representative” or “index” sites....and I believe they were always meant to be the latter. Next it has to proceed to the issue of representativeness, as increasingly as we get depths from lidar or photogrammetry, we will be converting depths to SWE in locations not sampled by the SOTEL network. Are we moving into locations where the bulk density is likely to be higher or lower than at an index station? Why? I would rather see the authors just bypass the issue than trivialize the problem in a statistical experiment that doesn't tell us much about the core issue.

We have removed the commentary on whether or not future investments should be in more stations vs. better stations. We agree that SNOTEL stations are largely index stations in that their measurements are often directly regressed against downstream streamflow. However, for our purposes here (providing an equation to estimate SWE from h ‘anywhere anytime’) we do feel that it is valuable to discuss the effects of source data accuracy and precision on the estimated SWE. This will help the reader to understand how much uncertainty there will be in SWE derived from the current paper. So, we retain some of the statistical testing.

Reply to Adam Winstral, Referee

Referee comments are left-justified, in black. Author replies are indented, in blue.

The aim of this study is to develop a simple means of estimating snow densities to convert observed snow depths to snow-water-equivalent. The authors seek to use long-term climatological variables rather than station or modeled data so that snow depths garnered in remote locations without direct meteorological observations (e.g. crowd-sourced or Lidar data) can be easily and accurately converted to SWE. There is a growing need for improved means of characterizing snow density as greater amounts of snow depth data are becoming available (e.g. Lidar). Therefore this type of research is certainly warranted. Given that snow depths have always been more readily available than SWE or density data, other researchers have similarly produced methods of estimating densities. While not all of the previously developed approaches tackle the specific case presented here (i.e. meteorological data immediately preceding snow depth observation not required), the Sturm et al. (2010) and Jonas and Magnusson (2009) approaches do. The authors clearly acknowledge this and make a positive comparison of their method measured against the Sturm method. However, I find the presented Sturm comparison to be biased against the Sturm method (see further comments). They also hint at (lines 413, 493), but never provide evidence nor specifically claim to be, better than the Jonas approach. I would like to see the authors present in a more convincing manner how and why their method represents a substantial advancement over the previously published methods before I am ready to consider this manuscript worthy of publication.

This is my major concern:

The authors randomly split the aggregated CONUS, AK, and BC data into training and validation datasets (Section 2.2). They then use the “held-out” validation dataset to make the Sturm comparison (Section 3.1). So, essentially they have trained their model on data from the same locations with the same statistical metrics present in the comparison dataset. On the other hand, the comparison dataset is 100% independent of the Sturm training data. In order to present a fair comparison this needs to be done with a dataset that is totally independent from the derivation of both.

This is an important point. Our current approach aggregated all western North America snow pillow data (some ~2M points) and then randomly split it in two. So, for each station, some data at each station was used for model building, the other data at each station ended up being used for validation. We can see why it would be important to test a validation approach that separated the training and validation data either by location, or by time.

To address your concerns, we took all of the snow pillow data and we split up the stations randomly into two groups. We took all of the data from the first group and we used that to train the regression model. We then validated the regression model against the second group. We did several realizations of this process and found that the results were extremely close to those presented in the original manuscript. Anonymous

Referee #3 also raised a similar concern, and suggested an 80/20 cross validation (80% of the data used to train, 20% of the data used to validate) approach. This method also generated similar results. We believe this to be due to the very large N of our dataset.

Given how similar all of these approaches were, and given the lack of any clear 'preferred method' in the literature, we decided to retain our original approach.

We strongly agree with the referee that it would be ideal to have a perfect test between the two methods (our model and that of Sturm et al.). However, that would require that the two models be developed with the same training datasets and then validated using the same validation datasets. Unfortunately, we don't see a way to create this perfect 'laboratory test' for two models developed with different data.

The northeast dataset would be one ideal dataset for conducting this test and I'm not sure why this wasn't done. That said, it would certainly be more convincing if the inter-model comparisons were conducted over a wider range of conditions.

You are correct in that it would be ideal to have inter-model comparisons over a wide range of conditions. We believe that applying both models to the NE data set would not accomplish that. We prefer to keep our inter-model comparisons to the larger dataset from western North America snow pillow data, and we will retain the NE dataset for our model only.

I would also like to see direct comparisons to the Jonas method. As I stated in the above paragraph, the authors must present a convincing case that the new methodology represents an improvement over existing procedures. I just don't find that in the current manuscript.

With regards to Jonas et al. (J09). We specifically chose not to apply that model for the following reason. The J09 model has coefficients that depend upon month of year and elevation. In addition to this, there is a geographic 'offset' term that depends on boundaries drawn in the Swiss Alps. Therefore, the model cannot be applied in other regions (since we would have no idea what to use for an offset). We do not wish to ignore the offset and apply a 'partial model' since that is not what those authors constructed.

One thing that we have done is to apply the very simple Pistocchi¹ model which depends only on day of year (DOY). In Pistocchi's paper, he claims comparable performance to both Sturm and J09. We now include summary results (RMSE and bias only, no figures) for the Pistocchi model applied to the western North America snow pillow data.

We believe that the results for our model demonstrate an improvement (lower bias and RMSE than existing methods) and also a strength of our approach is that it allows for a

¹ <https://www.sciencedirect.com/science/article/pii/S2214581816300131>

continuously varying snow density in space rather than discontinuities due to discrete snow classes. Our plots below, provided in response to another comment, help illustrate this point.

Moderate concerns that need addressing:

I don't understand why rmse was normalized with respect to mean annual precipitation (Section 3 and Figure 8). This obviously biases the normalizations low where summer precipitation is more common. Artifacts of this can be seen in Figure 8 (e.g. low ratios in Arizona, New Mexico, Alaska where summer precipitation can be considerable compared to winter; high ratios in eastern Sierras where synoptic summer storms are rare). This type of normalization might be appropriate for annual or longer hydrologic studies, but for this snow-based, winter-focused research the normalization should be based on either mean wintertime precipitation or better yet, mean annual snowfall. Both mean wintertime precipitation and mean annual snowfall should be easily derivable from the PRISM data already used in this study.

This is a reasonable suggestion. Our intent was simply to provide some sort of 'relative' measure of the magnitude of the RMSE. We have actually redone this using the mean annual peak SWE to normalize the RMSE, which makes good sense.

Graphs. There are way too many data points in the scatter plots to understand what is really going on in Figures 6 and 9, and some of the plots in Figure 11. These should be presented as either heat plots or randomly select and plot a subset of these data. Additionally and partly due the aforementioned reason, the overlapping plots in Figure 9 are impossible to fully discern.

With regards to Figure 11 (Fig 12 in the revision), the symbols are colored by DOY in the left column, so we are unable to show that column as a heatmap. We have changed the center column to show the data as a heatmap.

With regards to Figure 6, we have changed the plot to a heatmap (which is just a 2d histogram). The 'footprint' or 'envelope' of the data cloud is unchanged of course.

With regards to Figure 9 (Fig 10 in the revision). The important point is how the 'width' of the data cloud is different between the two methods. The envelope that is closer to the 1:1 line indicates better performance. Our original approach was chosen since, in each case, our envelope was narrower (so we plotted ours on top). We cannot show two overlapping heatmaps. What we have done in the revision is to show Sturm's results as scatter symbols (as before) and to then plot our results as a transparent heat map on top.

I had difficulty accepting the reasoning for the residuals and mean biases apparent in the Figures 6b and d. I think these residuals, which are present in the validation dataset are also related to the choice of fitting a power law relationship rather than a linear least

squares one. Given that the training and validation data should maintain the same statistical metrics then these residuals should be present in the training data as well. If, in fact, this is the case then the combination of a power law fit and the predominance of accumulation season samples would be the reason. My suspicion is that if a linear least squares fit was chosen then there should be near zero mean biases in both the training and validation sets given that the two sets maintain the same characteristics. I would expect that in the linear scenario, there should be a wider spread in residuals (i.e. higher rmse) but very little change in mean bias. Of course, this would be entirely different if the validation set was truly independent.

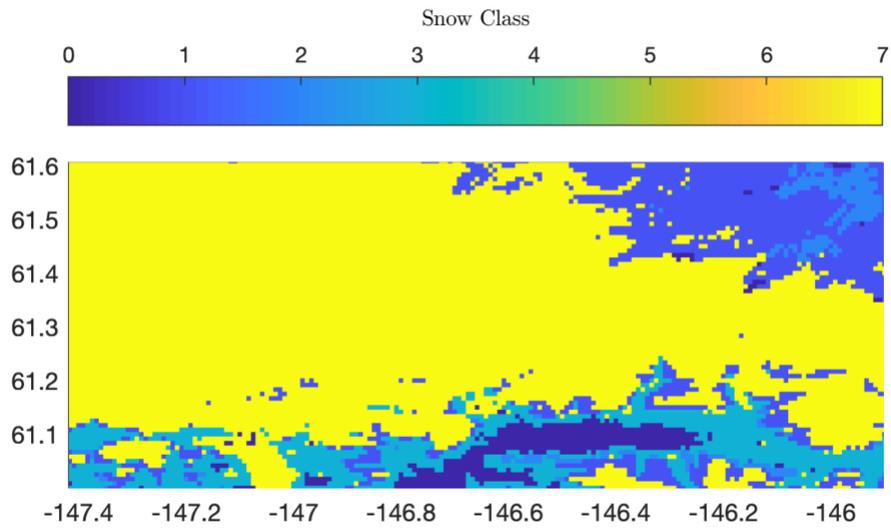
This is a fair comment, and our initial remarks may have been too speculative. We adopted a power law relationship based on the hysteresis loop (Figs 1 and 4) suggesting something other than a linear relationship between h and SWE. We feel that the best course of action is to remove our overly speculative comment.

How the different datasets were used needs better clarification. I didn't understand the purpose of the manually sampled Chugach data. As far as I can tell, these data were not included in the calibration nor the validation analyses. What do these data show? Why were they included? How do these data add anything new to the analysis? This should be clearly stated and incorporated into the story or leave the Chugach data out.

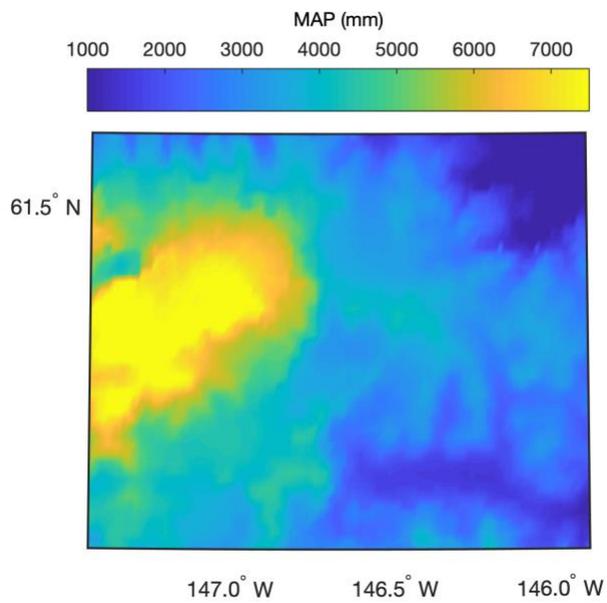
Two reviewers of this paper noted this. We have essentially dropped that dataset from the paper, with one exception. The large ensemble (80 or so) of collections (8 at each site) of probe measurements is valuable since it helps to quantify the variability in snow depth over small distances (in discussion section).

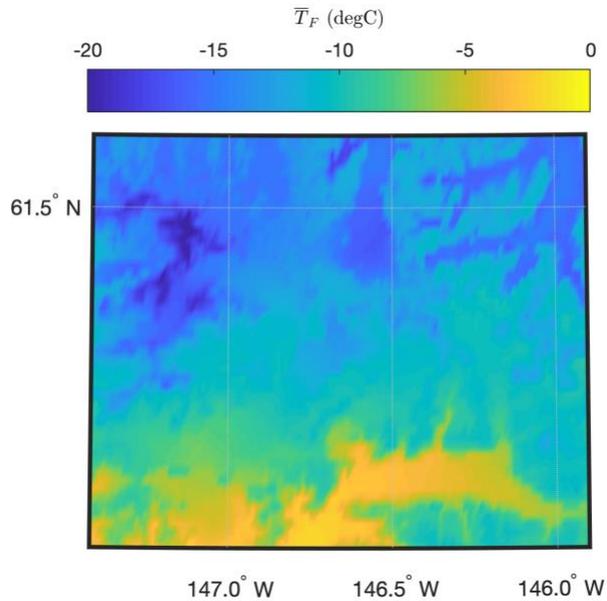
Section 2.1.2. Do these PRISM climatological variables, based on sparse station data and resolved at 800m, really pick up the heterogeneity you're aiming to capture as expressed on lines 132-37. It would be nice if you could show a spatially explicit example showing these capabilities.

We feel that the continuously variable PRISM data does a better job of capturing climate than 5 snow classes. Let us illustrate this with some sample figures. First, consider the map of snow class in the region just northeast of Valdez, Alaska.



Note that there are only a few snow classes and that the landscape is dominated by class 7 in this case. Now, for the exact same lat / lon bounding box, let us look at the MAP and Feb_T_Mean:





In both of these climatological rasters, we see very considerable variation over a region that is monolithic in snow class. These, we do feel that the use of 800m PRISM data will allow for smoother variability in snow density.

Tidbits:

The residuals (e.g. Figure 6) should be presented as modeled minus observed. In this manner the underestimations of SWE appear as negative residuals rather than the positive residuals currently presented. I find this much easier to understand.

Actually, the residuals are done correctly. Please see the new version of Fig 6, which has been much improved by showing it as a heatmap. Look at the top row. The residuals are indeed computed as model-observed. The vertical black line in the right column (panel (b)) is the mean residual. It is negative. And that makes sense since the cloud of data points appears to be, on average, below the 1:1 line. So, thank you for your suggestion, it was good for us to double check, but we do have the residuals defined correctly, we believe.

Lines 44-47 and 72-74. Each of these sentences contain two distinct thoughts that would perhaps be better if split into two sentences.

Thank you for the suggestion. We will improve the clarity of these lines.

Lines 120-22. I didn't think this sentence was necessary . . . unless you turn it into reasoning that this just adds a layer of computational costs / complexity that aren't necessary for your desired application.

We slightly adjusted the sentences there to improve the clarity.

Lines 141-2. Might want to add something about why you would also prefer to not use NWP data that could possibly substitute for the lack of observations (i.e. computational costs, errors in NWP data).

The purpose of this work is to provide a rapid, easy to use tool. Relying on external daily or sub-daily datasets and/or model output moves the work away from that goal and towards more sophisticated snow models. So, yes, it could be done, but at significant expense and effort.

Line 169. You also used snow pillow data from the northeast US. You might want to make that clear here . . . as in “Snow data for this project, aside from the aforementioned SNOTEL data, . . .”

Yes, thank you. We fixed this.

Section 2.1.1.5. Might want to mention that these issues are most common in summer when vegetation grows beneath the sensor.

We are not sure we fully understand this remark. Which particular issues are you referring to? The data that we considered was only winter time data, where snow was present.

Line 440. Roughness of underlying terrain is certainly one factor, but couldn't there be others as well (e.g. wind redistribution).

We have now noted this explicitly.

Reply to Anonymous Referee #3

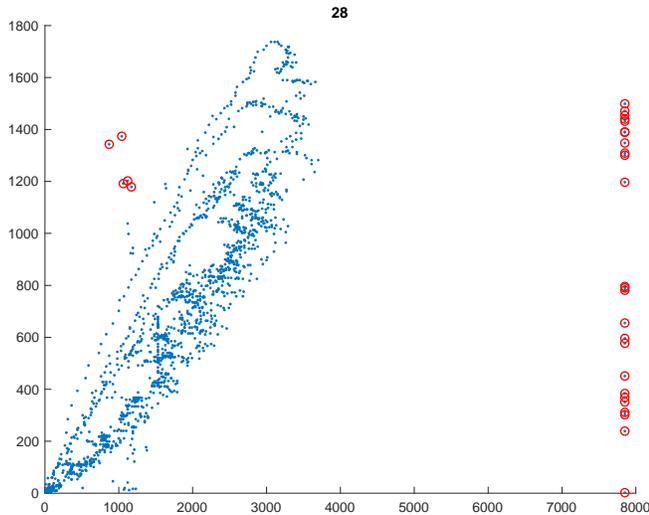
Received and published: 13 March 2019

Referee comments are left-justified, in black. Author replies are indented, in blue.

The authors address the issue of converting spatiotemporal snow depth measurements to estimates of snow water equivalent (SWE). This topic is relevant to many areas of research because of the relative ease of taking snow depth measurements over SWE. Framed in the context of citizen science or field work, snow depths collected by nonexperts and experts alike can be leveraged as a low-cost input to hydrological or climate analysis. In an era of high-availability altimetry (lidar or radar) and photogrammetry (structure from motion), an ensemble of methods to convert surface heights into SWE will be critical for both targeted basin studies (ASO) as well as future satellite missions. The authors develop three regression models to evaluate a snow depth to SWE conversion. Regression skill is evaluated using depth alone, depth separated by accumulation and ablation phases, and depth in combination with climate normal for precipitation, temperature as well as elevation. Their work differs from previous studies such as Sturm et al. 2010 in that the climate inputs are regressed as continuous variables. As such, any measurement of snow depth with coordinates could potentially be converted, independent of measurement scale. In general the paper is well written and clear in its advancements. The focus on estimates during the ablation phase is a clear contribution, where methods fail. Addressing that 'not all snow is equal' is a strength of the approach.

Prior to publication, I would like to see the outlier detection and validation portions of the paper revisited to reinforce the statistical analysis. While I agree that outlier detection is necessary, an enhanced description of where and when the outliers originate would help to identify potential seasonality or spatial clustering. For example, if many of the outliers are from the early snow season, does this preclude ability of the models to convert measurements that include fresh snow? There are artifacts in Figure 4 where SWE varies drastically but depth does not, are these melt events? A histogram of the outlier DOY or a table of the outlier properties may be all that is needed to address this. These additions could be used to reinforce the statement that the reduced dataset is physically plausible (Lines 229-230).

These are very good points. One other referee had similar remarks. Manual examination of many of the SNOTEL time series revealed the presence of clearly wrong data (Figure 3 of the paper). We wanted to develop a wholly objective method for removing those data points. The approach that we used seems like a good one, based upon the characteristics of the bivariate distribution. We recognize that some valid data points (mostly at low SWE-Hs values) are undoubtedly removed as well. Given the very low number (less than 1%; so the valid points removed are some small fraction of this 1%) of points that were removed in our process, we feel that this is acceptable. Here is a figure of the process at one particular station. SWE on vertical axis, h on horizontal. Red circles are removed points.



We particularly like your suggestion of looking at the characteristics of the removed points, and now include specific information on the DOY values of these points. It turned out that removed points were occurring throughout the snow season, and not just at the beginning and the end.

Your comment about events in Figure 4 where SWE changes a lot while Hs remains fixed is an interesting one. It is hard to understand how SWE could drop from 1 m to near zero while Hs remains fixed at 5 m. The lack of an accepted and easy to implement protocol for addressing snow pillow data quality control is an obstacle to analysis.

For the validation, it may be of benefit to use a cross-validation (CV) to determine if the model skill is overly optimistic. Using an N-folds CV with a 80/20 train/test split would be a simple approach to achieve this. In this regard, I'd also be interested to know if the non-SNOTEL datasets actually influence to the regression coefficients (What happens when the training datasets are SNOTEL only). The remainder of my comments addressed to specific lines or figure.

A few comments. First of all, the regression coefficients were constructed with snow pillow data only from the western United States and Canada. We have now tried to make this more obvious in the section discussing datasets. For example, in Table 1, we now use bold font to highlight which datasets were used to build the model.

Second, with regards to validation. We looked into this at some length before beginning this work, since we wanted to determine if there was some preferred way of doing validation in the snow density literature (or streamflow prediction, or any other discipline for that matter). We found no 'best' or 'preferred' method. We ended up doing a 50/50 split (aggregating all snow pillow data points and randomly dividing them up) in the first draft of the paper. Upon receiving the manuscript reviews, we also tested your suggested 80/20 split, and a 50/50 'station split' (divide up the stations, not the

aggregated data points). We found that all methods provided essentially the same results. We feel that this is likely due to the large N (number of observations) of our dataset. Given the lack of consensus in the literature, we feel that our approach is acceptable, and we are clear and upfront about our methods.

Lines 64-65: Are there additional references available to support this statement regarding L-band? The only cited application in the field is a conference proceeding.

Another referee though we should simply remove the remote sensing discussion and that seemed like a sensible change to us, so we did so.

Line 172: Each style of corer has its own associated bias. Could this be considered to bound or constrain errors for each region/dataset?

Corer data were not used to build the regression model. So, those biases would not affect the regression model coefficients. Any depth measurement that has a bias or random error and that is used to estimate a SWE value using the methods in this paper would propagate through into a bias or error in the SWE. We do try to present some discussion on this in the manuscript.

Line 185: I would expect readers to be unfamiliar with some of coring devices. For example, the Mt. Rose snow tube could be supported with Church, J. Improvement in Snow Survey Apparatus, TAGU, 1936.

Thank you for this suggestion, and we can add this citation.

Line 228: See concerns about outlier detection in the main comments. It would be important to describe the temporal aspect of the outlier detection.

Yes, as noted above, we provide information on this now in that section.

Line 228: uncleaned data -> source data

Good catch, we made this change.

Line 229: State how many outliers were removed from the other datasets via this process. Figure 4: An axis label is needed for the DOY color bar.

This has been handled with added parenthetical notes to column 4 of Table 1.

Line 231: How does this work for 'stations' where there are a very low number of observations, ie AK?

The process was objectively applied to all stations. Stations with low numbers of observations could still be processed, in terms of computing the characteristics of the bivariate distributions and then removing points that did not satisfy the criteria.

Table 1: Can this table be augmented with a % of retained points or an omission %? Is the BC survey missing the # of ultrasonic sites?

As per the remark just above re:line 229, yes we have done this. Regarding the BC comment. The first row of that table has two sets of numbers. One for the Western USA SNOTEL. One for the eastern USA SCAN. The BC row only has one set of numbers since we grouped all BC snow pillows together. In revised Table 1, we have split up the USA NRCS data into two rows to eliminate this confusion.

Line 250: Is this 50% of all measurements or 50% of each subset. If it is all of them, it could be such that the only ones removed are CONUS because of the low numbers elsewhere.

All of the aggregated snow pillow data were grouped (data points were grouped in one large bin) and then divided in two. Given the random nature of the division, each station should have ~50% of its data represented.

Line 256: Figure 3 is used as support for the outlier detection due to poor correlation (ie increasing h with no SWE) and but is referenced here as strongly correlated. It might be confusing to do both.

This is a good catch. We meant to refer to just the winter (snow present) portions of Figure 3. The noisy bits in that Figure are at times when there is no SWE. We will clarify our language.

Line 283: If this is an important consideration, why is the SCAN dataset not used to train the models?

There are several reasons. Foremost, we wanted to leave the northeastern USA data alone so that we could use those data as an independent test of the ability of the model to work in completely different regions / snow regimes. Second, the N (5 sites) of the northeastern USA dataset is a tiny fraction of the rest of the available data. Locations with multi-peak SWE curves may do better with a more complex model that is able to capture this behavior.

Line 290: Interesting that a static 180 works best as the DOY separator. Could a sentence on why this might occur be added to the discussion?

To be frank, we do not have a great explanation for this. When we discovered a fairly strong correlation between day of peak SWE and April temperature, we were confident

that the variable DOY approach would produce the best results. In this case, it appears that simpler is better.

Line 332: I see how it would not be possible to use an absolute value here but are snow-covered regions where the February normal is below -30C.

We chose this offset value based upon the lowest February temperature values observed at the snow pillow stations. This may limit our methods to not apply in some extremely cold regions.

Figure 6: Titles for each plot might make this easier to read if someone skips the caption.

We appreciate this stylistic suggestion. Our approach favors using the figure caption to provide details on the content in each figure panel, which is consistent with the approach of other papers in *The Cryosphere*. We are open to modifying this if the editors request it.

Table 5: Include the normalized errors for completeness of the table.

We are not able to normalize the errors for these datasets in the way that we do for the snow pillow sites (Figs 8-9 of new version of paper). For the snow pillow stations, we normalized the RMSE at each station based on (this is a change for the 2nd draft of this paper) the mean annual maximum SWE at that station. The information in table 5 is different. The RMSE values there are essentially being averaged 'spatially' over a distributed dataset, rather than being averaged temporally at a snow pillow station. Thus, we do not have a mean annual maximum SWE available for normalization in a consistent fashion. Note that in the 2nd paragraph of the discussion section, we do talk a bit about the east coast results and how they differ from western North America (smaller snowpack, etc.).

Line 423-430: Might be helpful to discuss measurement errors as a contributor.

We do discuss this (measurement errors) in lines 447 → 472 (numbering of original draft). In the specific context of the northeastern USA data, those data are generally high-quality coring data. Having not taken those data ourselves, it is hard to quantify the measurement errors. In some cases, the supporting documentation for those datasets is brief to non-existent. Also, note that, in response to another reviewer as well, we have added more general discussion of both coring error and snow pillow errors to the manuscript.

1 **Converting Snow Depth to Snow Water Equivalent Using** 2 **Climatological Variables**

3
4 David F. Hill¹, Elizabeth A. Burakowski², Ryan L. Crumley³, Julia Keon⁴, J. Michelle Hu⁵,
5 Anthony A. Arendt⁶, Katreen Wikstrom Jones⁷, Gabriel J. Wolken⁸

6
7 ¹Civil and Construction Engineering, Oregon State University, OR, USA

8 ²Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, NH, USA

9 ³Water Resources Graduate Program, Oregon State University, OR, USA

10 ⁴Civil and Construction Engineering, Oregon State University, OR, USA

11 ⁵Civil and Environmental Engineering, University of Washington

12 ⁶Applied Physics Laboratory, University of Washington

13 ⁷Alaska Division of Geological & Geophysical Surveys, Fairbanks, AK, USA

14 ⁸Alaska Division of Geological & Geophysical Surveys, Fairbanks, AK, USA; International Arctic Research Center,
15 University of Alaska Fairbanks, Fairbanks, AK, USA

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

17
18
19 **Abstract.** We present a simple method that allows snow depth measurements to be converted to snow water
20 equivalent (SWE) estimates. These estimates are useful to individuals interested in water resources, ecological
21 function, and avalanche forecasting. They can also be assimilated into models to help improve predictions of total
22 water volumes over large regions. The conversion of depth to SWE is particularly valuable since snow depth
23 measurements are far more numerous than costlier and more complex SWE measurements. Our model regresses
24 SWE against snow depth and climatological (30-year normal) values for mean annual precipitation (*MAP*) and mean
25 February temperature ($\bar{T}_{F_{mean}}$), producing a power-law relationship. Relying on climatological normals rather than
26 weather data for a given year allows our model to be applied at measurement sites lacking a weather station.
27 Separate equations are obtained for the accumulation and the ablation phases of the snowpack, which introduces
28 ‘day of water year’ (*DOY*) as an additional variable. The model is validated against a large database of snow pillow
29 measurements and yields a bias in SWE of less than 0.5 mm and a root-mean-squared-error (RMSE) in SWE of
30 approximately 65 mm. When the errors are investigated on a station-by-station basis, the average RMSE is about 5%
31 of the *MAP* at each station. The model is additionally validated against a completely independent set of data from
32 the northeast United States. Finally, the results are compared with other models for bulk density that have varying
33 degrees of complexity and that were built in multiple geographic regions. The results show that the model described
34 in this paper has the best performance for the validation data set.

35 **1 Introduction**

36 In many parts of the world, snow plays a leading-order role in the hydrological cycle (USACE, 1956; Mote et al.,
37 2018). Accurate information about the spatial and temporal distribution of snow water equivalent (SWE) is useful to
38 many stakeholders (water resource planners, avalanche forecasters, aquatic ecologists, etc.), but can be time
39 consuming and expensive to obtain.

40
41 Snow pillows (Beaumont, 1965) are a well-established tool for measuring SWE at fixed locations. Figure 1 provides
42 a conceptual sketch of the variation of SWE with time over a typical water year. A comparatively long accumulation
43 phase is followed by a short ablation phase. While simple in operation, snow pillows are relatively large in size and
44 they need to be installed prior to the onset of the season's snowfall. This limits their ability to be rapidly or
45 opportunistically deployed. Additionally, snow pillow installations tend to require vehicular access, limiting their
46 locations to relatively simple topography. Finally, snow pillow sites are not representative of the lowest or highest
47 elevation bands within mountainous regions (Molotch and Bales, 2005). In the western United States (USA), the
48 Natural Resources Conservation Service (NRCS) operates a large network of Snow Telemetry (SNOTEL) sites,
49 featuring snow pillows. The NRCS also operates the smaller Soil Climate Analysis Network (SCAN) which
50 provides the only, and very limited, snow pillow SWE measurements in the eastern USA.

51
52 SWE can also be measured manually, using a snow coring device that measures the weight of a known volume of
53 snow to determine snow density (Church, 1933). These measurements are often one-off measurements, or in the
54 case of 'snow courses' they are repeated weekly or monthly at a given location. The simplicity and portability of
55 coring devices expand the range over which measurements can be collected, but it can be challenging to apply these
56 methods to deep snowpacks due to the length of standard coring devices. Note that there are numerous different
57 styles of coring devices, including the Adirondack sampler and the Mt. Rose / Federal sampler (Church and Marr,
58 1937).

59
60 There are a number of issues that affect the accuracy of both snow pillow and snow coring measurements. With
61 coring measurements, if the coring device is not carefully extracted, a portion of the core may fall out of the device.
62 Or, snow may become compressed in the coring device during insertion. These effects have led to varying
63 conclusions, with some studies (e.g., Sturm et al., 2010) showing a low SWE bias and other studies (e.g., Goodison,
64 1978) showing a high SWE bias. As noted by Johnson et al. (2015) a good rule of thumb is that coring devices are
65 accurate to around $\pm 10\%$. Also, studies comparing different styles of snow samplers report statistically different
66 results, suggesting that SWE measurements are sensitive to the design of the specific coring device, such as the
67 presence of holes or slots, the device material, etc. (Beaumont and Work, 1963; Dixon and Boon, 2012). With snow
68 pillows, some studies (e.g., Goodison et al., 1981) note that ice bridging can lead to low biases in measured SWE,
69 with the snow surrounding the pillow partly supporting the snow over the pillow. Other studies (Johnson and Marks,
70 2004; Dressler et al., 2006; Johnson et al., 2015) note a more complex situation with SWE under-reported at times,

71 but over-reported at other times. Note that when snow pillow data are evaluated, they are most commonly compared
72 to coring measurements at the same location.

73
74 All methods of measuring SWE are challenged by the fact that SWE is a depth-integrated property of a snowpack.
75 This is why the snowpack must be weighed, in the case of a snow pillow, or a core must be extracted from the
76 surface to the ground. This measurement complexity makes it difficult to obtain SWE information with the spatial
77 and temporal resolution desired for watershed-scale studies. Other snowpack properties, such as the depth h , are
78 much easier to measure. For example, using a graduated device such as a meterstick or an avalanche probe to
79 measure the depth takes only seconds. Automating depth measurements at a fixed location can easily be done using
80 low-cost ultrasonic devices (Goodison et al., 1984; Ryan et al., 2008). High-spatial-resolution measurements of
81 snowpack depth are commonly made with Light Detection and Ranging (LIDAR). One example of this is the
82 Airborne Snow Observatory program (ASO; Painter et al., 2016). The comparatively high expense of airborne
83 LIDAR surveys typical limits measurements geographically (to a few basins) and temporally (weekly to monthly
84 interval).

85
86 Given the relative ease in obtaining depth measurements, it is common to use h as a proxy for SWE. Figure 1 shows
87 a conceptual sketch of the variation of SWE with h over a typical water year. Noting the arrows on the curve, we see
88 that SWE is multi-valued for each h . This is due to the fact that the snowpack increases in density throughout the
89 water year, producing a hysteresis loop in the curve. A large body of literature exists on the topic of how to convert
90 h to SWE. It is beyond the scope of this paper to provide a full review of these ‘bulk density equations,’ where the
91 density is given by $\rho_b = \text{SWE}/h$. Instead, we refer readers to the useful comparative review by Avanzi et al. (2015).
92 Here, we prefer to discuss a limited number of previous studies that illustrate the spectrum of methodologies and
93 complexities that can be used to determine ρ_b or SWE.

94
95 Many studies express ρ_b as an increasing function (often linear) of h . In some cases (e.g., Lundberg et al., 2006) a
96 second equation is added where ρ_b attains a constant value when a threshold h is exceeded. A single linear equation
97 captures the process of densification of the snowpack during the accumulation phase, but performs poorly during the
98 ablation phase, where depths are decreasing but densities continue to increase or approach a constant value.

99 Other approaches choose to parameterize ρ_b in terms of time, rather than h . Pistocchi (2016) provides a single
100 equation while Mizukami and Perica (2008) provide two sets of equations, one set each for early and late season.
101 Each set contains four equations, each of which is applicable to a particular ‘cluster’ of stations. This clustering was
102 driven by observed densification characteristics and the resulting clusters are relatively spatially discontinuous.
103 Jonas et al. (2009) take the idea of region- (or cluster-) specific equations and extend it further to provide
104 coefficients that depend on time and elevation as well. They use a simple linear equation for ρ_b in terms of h and the
105 slope and intercept of the equation are given as monthly values, with three elevation bins for each month (36 pairs of
106 coefficients). There is an additional contribution to the intercept (or ‘offset’) which is region-specific (one of 7
107 regions).

108

109 These classifications, whether based on region, elevation, or season, are valuable since they acknowledge that all
110 snow is not equal. McKay and Findlay (1971) discuss the controls that climate and vegetation exert on snow density,
111 and Sturm et al. (2010) address this directly by developing a snow density equation where the coefficients depend
112 upon the ‘snow class’ (5 classes). Sturm et al. (1995) explain the decision tree, based on temperature, precipitation,
113 and wind speed, that leads to the classification. The temperature metric is the ‘cooling degree month’ calculated
114 during winter months only. Similarly, only precipitation falling during winter months was used in the classification.
115 Finally, given the challenges in obtaining high quality, high-spatial-resolution wind information, vegetation
116 classification was used as a proxy. Using climatological values (rather than values for a given year), Sturm et al.
117 (1995) were able to develop a global map of snow classification.

118

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

130

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

137

138 The present paper seeks to generalize the ideas of Mizukami and Perica (2008), Jonas et al. (2009), and Sturm et al.,
139 (2010). Specifically, our goal is to regress physical and environmental variables directly into the equations. In this
140 way, environmental variability is handled in a continuous fashion rather than in a discrete way (model coefficients
141 based on classes). The main motivation for this comes from evidence (e.g., Fig. 3 of Alford, 1967) that density can
142 vary significantly over short distances on a given day. Bulk density equations that rely solely on time completely
143 miss this variability and equations that have coarse (model coefficients varying over either vertical bins or horizontal
144 grids) spatial resolution may not fully capture it either.

145
146 Our approach is most similar to Mizukami and Perica (2008), Jonas et al. (2009), and Sturm et al., (2010) in that a
147 minimum of information is needed for the calculations; we intentionally avoid approaches like Meloysund et al.
148 (2007) and McCreight and Small (2014). This is because our interests are in converting h measurements to SWE
149 estimates in areas lacking weather instrumentation. The following sections introduce the numerous data sets that
150 were used in this study, outline the regression model adopted, and assess the performance of the model.

151 **2 Methods**

152

153 **2.1 Data**

154

155 **2.1.1 Snow Depth and Snow Water Equivalent**

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

158

159 **2.1.1.1 USA NRCS Snow Telemetry and Soil Climate Analysis Networks**

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

166

167 **2.1.1.2 Canada (British Columbia) Snow Survey Data**

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

175

176 **2.1.1.3 Northeast USA Data**

177 In addition to the data from the SCAN sites, snow data for this project from the northeast US come from two
178 networks and three research sites (Figure 2b). The Maine Cooperative Snow Survey (MCSS, 2018) network
179 includes h and SWE data collected by the Maine Geological Survey, the United States Geological Survey, and
180 numerous private contributors and contractors. MCSS snow data are collected using the Standard Federal or

181 Adirondack snow sampling tubes typically on a weekly to bi-weekly schedule throughout the winter and spring,
182 1951-present. The New York Snow Survey network data were obtained from the National Oceanic and Atmospheric
183 Administration’s Northeast Regional Climate Center at Cornell University (NYSS, 2018). Similar to the MCSS,
184 NYSS data are collected using Standard Federal or Adirondack snow sampling tubes on weekly to bi-weekly
185 schedules, 1938-present.

186
187 The Sleepers River, Vermont Research Watershed in Danville, Vermont (Shanley and Chalmers, 1999) is a USGS
188 site that includes 15 stations with long-term weekly records of h and SWE collected using Adirondack snow tubes.
189 Most of the periods of record are 1981-present, with a few stations going back to the 1960s. The sites include
190 topographically flat openings in conifer stands, old fields with shrub and grass, a hayfield, a pasture, and openings in
191 mixed softwood-hardwood forests. The Hubbard Brook Experiment Forest (Campbell et al., 2010) has collected
192 weekly snow observations at the Station 2 rain gauge site, 1959-present. Measurement protocol collects ten samples
193 2 m apart along a 20 m transect in a hardwood forest opening about $\frac{1}{4}$ hectare in size. At each sample location along
194 the transect, h and SWE are measured using a Mt. Rose snow tube and the ten samples are averaged for each
195 transect. Finally, the Thompson Farm Research site includes a mixed hardwood forest site and an open pasture site
196 (Burakowski et al. 2013; Burakowski et al. 2015). Daily (from 2011-2018), at each site, a snow core is extracted
197 with an aluminum tube and weighed (tube + snow) using a digital hanging scale. The net weight of the snow is
198 combined with the depth and the tube diameter to determine ρ_b , similar to a Federal or Adirondack sampler.
199

200 **2.1.1.4 Chugach Mountains (Alaska) Data**

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

209 **2.1.1.5 Data Pre-Processing**

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

218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247

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 source 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 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 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)
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)

248

249 **2.1.2 Climatological Variables**

250 30-year climate normals at 800 m (nominal) resolution for CONUS and for the period 1981-2010 were obtained
251 from the PRISM website (Daly et al., 1994). PRISM normals for British Columbia (BC), Canada, were obtained
252 from the ClimateBC project (Wang et al., 2012), also for the 1981-2010 period. Finally, PRISM normals for Alaska
253 (AK) were obtained from the Integrated Resource Management Applications (IRMA) Portal run by the National
254 Park Service. The AK normals are for the 1971-2000 period and have a slightly coarser resolution (approximately
255 1.5 km). Figure 5 shows gridded maps of mean annual precipitation (MAP) and mean February Temperature (\bar{T}_F)
256 for these three climate products, plotted together. Other temperature products (max and min temperatures; other
257 months) were obtained as well, but are not shown.

258

259 **2.2 Regression Model**

260 In order to demonstrate the varying degrees of influence of explanatory variables, several regression models were
261 constructed. In each case, the model was built by randomly selecting 50% of the paired SWE- h measurements from
262 the aggregated CONUS, AK, and BC snow pillow datasets. The model was then validated by applying it to the
263 remaining 50% of the dataset and comparing the modeled SWE to the observed SWE for those points. Additional
264 validation was done with the northeast USA datasets (SCAN snow pillow and various snow coring datasets) which
265 were completely left out of the model building process.

266

267 **2.2.1 One-Equation Model**

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

271

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

273

274 This equation can be log-transformed into

275

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

277

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

281

282 2.2.2 Two-Equation Model

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

288

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

290

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

292

293 where DOY is the number of days from the start of the water-year (October 1 is the origin), and DOY^* is the critical
294 or dividing day-of-water-year separating the two phases. Put another way, DOY^* is the day of peak SWE.

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

304

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

307

308 (5) $SWE = SWE_{acc} \frac{1}{2} (1 - \tanh[0.01\{DOY - DOY^*\}]) +$
309 $SWE_{abl} \frac{1}{2} (1 + \tanh[0.01\{DOY - DOY^*\}])$

310

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

313

314 2.2.3 Two-Equation Model with Climate Parameters

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

321

$$322 \quad (6) \quad SWE = f(h, z, MAP, \bar{T}_{Jmin}, \bar{T}_{Jmean}, \bar{T}_{Jmax}, \bar{T}_{Fmin}, \bar{T}_{Fmean}, \bar{T}_{Fmax}, \bar{T}_{Mmin}, \bar{T}_{Mmean}, \bar{T}_{Mmax}, \bar{T}_{Amin}, \bar{T}_{Amean}, \bar{T}_{Amax})$$

323

324 where z is the elevation (m), MAP is the mean annual precipitation (mm) and the temperatures ($^{\circ}C$) represent the
 325 mean of minimum, mean, and maximum daily values for the months January through April (J, F, M, A). For
 326 example, \bar{T}_{Jmin} is the climatological normal of the average of the daily minimum temperatures observed in January.
 327 In the stepwise regression, explanatory variables were accepted if they improved the adjusted R^2 value by 0.001.

328 The result of the regression yielded

329

$$330 \quad (7) \quad SWE_{acc} = Ah^{a_1} MAP^{a_2} (\bar{T}_{Fmean} + 30)^{a_3}; \quad DOY < DOY^*$$

331

$$332 \quad (8) \quad SWE_{abl} = Bh^{b_1} MAP^{b_2} (\bar{T}_{Fmean} + 30)^{b_3}; \quad DOY \geq DOY^*$$

333

334 or, in log-transformed format,

335

$$336 \quad (9) \quad \log_{10}(SWE_{acc}) = \log_{10}(A) + a_1 \log_{10}(h) +$$

$$337 \quad \quad \quad a_2 \log_{10}(MAP) + a_3 \log_{10}(\bar{T}_{Fmean} + 30); \quad DOY < DOY^*$$

338

$$339 \quad (10) \quad \log_{10}(SWE_{abl}) = \log_{10}(B) + b_1 \log_{10}(h) +$$

$$340 \quad \quad \quad b_2 \log_{10}(MAP) + b_3 \log_{10}(\bar{T}_{Fmean} + 30); \quad DOY \geq DOY^*$$

341

342 indicating that only snow depth, mean annual precipitation and mean February temperature were relevant. Manual
 343 tests of model construction with other variables included confirmed that Eqns. (7-8) yielded the best results. In the
 344 above equations, note that an offset is added to the temperature in order to avoid taking the log of a negative
 345 number. These two SWE estimates for the individual (*acc* and *abl*) phases of the snowpack are then blended with
 346 Eqn. (5) to produce a single equation for SWE spanning the entire water year. The obtained regression coefficients

347 were $(A, a_1, a_2, a_3) = (0.0128, 1.070, 0.132, 0.506)$ and $(B, b_1, b_2, b_3) = (0.0271, 1.038, 0.201, 0.310)$. The
 348 physical interpretation of these coefficients is straightforward. If a_1 and b_1 were equal to unity, then the density,
 349 given by (SWE/h) , would be a constant at a given location. Since they are greater than unity, they capture the effect
 350 that snow density increases as depth increases. Turning to the coefficients on the climate variables, both a_2 and b_2
 351 are greater than zero. So, for two locations with equal depth, equal temperature characteristics, but different
 352 precipitation characteristics, the regression model predicts that the wetter location (larger MAP) will have a greater
 353 density. Finally, regarding temperature, both a_3 and b_3 are greater than zero. Therefore, for two locations with equal
 354 depth, equal precipitation characteristics, but different temperature characteristics, the regression model predicts that
 355 the warmer location (larger \bar{T}_{Fmean}) will have a greater density. These trends are similar in concept to Sturm et al.
 356 (2010), whose snow classes (based on climate classes) indicate which snow will densify more rapidly.

357 3 Results

358 A comparison of the three regression models (one-equation model, Eq. (2); two-equation model, Eqs. (3-5); multi-
 359 variable two-equation model, Eqs. (5, 7-8)) is provided in Figure 6. The left column shows scatter plots of modeled
 360 SWE to observed SWE for the validation data set with the 1:1 line shown in black. The right column shows
 361 histograms of the model residuals. The vertical lines in the right column show the mean error, or model bias.
 362 Visually, it is clear that the one-equation model performs relatively poorly with a large negative bias. This large
 363 negative bias is partially overcome by the two-equation model (middle row, Figure 6). The cloud of points is closer
 364 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
 365 see that the multi-variable two-equation model yields the best result by far. The residuals are now evenly distributed
 366 with a negligible bias. Several metrics of performance for the three models, including R^2 (Pearson coefficient), bias,
 367 and root-mean-square-error (RMSE), are provided in Table 2. Figure 7 shows the distribution of model residuals for
 368 the multi-variable two-equation model as a function of DOY.

369

370 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	0.972	-0.5	67

371

372 It is useful to also consider the model errors in a non-dimensional way. Therefore, an RMSE was computed at each
 373 station location and normalized by the mean annual maximum SWE (SWE_{max}) at that location. Figure 8 shows the
 374 probability density function of these normalized errors. The average RMSE is approximately 11% of SWE_{max} , with
 375 most falling into the range of 5-25%. The spatial distribution of these normalized errors is shown in Figure 9. For
 376 the SNOTEL stations, it appears there is a slight regional trend, in terms of stations in continental climates (northern
 377 Rockies) having smaller relative errors than stations in maritime climates (Cascades). The British Columbia stations
 378 also show higher relative errors.

379

380 **3.1 Results for Snow Classes**

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

384
 385 (11) $\rho_b = (\rho_{max} - \rho_0)[1 - e^{(-k_1h - k_2DOY)}] + \rho_0$

386
 387 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,
 388 and DOY in this case begins on January 1. This means that their DOY for October 1 is -92. The coefficients vary
 389 with snow class and the values determined by Sturm et al. (2010) are shown in Table 3.

390
 391 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

392
 393 To make a comparison, the snow class for each SNOTEL and British Columbia snow survey (Rows 1 and 3 of Table
 394 1) site was determined using a 1-km snow class grid (Sturm et al., 2010). The aggregated dataset from these stations
 395 was made up of 27% Alpine, 14% Maritime, 10% Prairie, 11% Tundra, and 38% Taiga data points. Equation (11)
 396 was then used to estimate snow density (and then SWE) for every point in the validation dataset described in Section
 397 2.2. Figure 10 compares the SWE estimates from the Sturm model and from the present multi-variable, two-equation
 398 model (Equations 5, 7-8). The upper left panel of Figure 10 shows all of the data, and the remaining panels show the
 399 results for each snow class. In all cases, the current model provides better estimates. Plots of the residuals by snow
 400 class are provided in Figure 11, giving an indication of the bias of each model for each snow class. Summaries of the
 401 model performance, broken out by snow class, are given in Table 4.

402
 403 Table 4: Comparison of model performance by Sturm et al. (2010) and the present study.

Model Snow Class	Sturm et al. (2010)			Multi-variable two-equation model		
	R ²	Bias (mm)	RMSE (mm)	R ²	Bias (mm)	RMSE (mm)
All Data	0.928	-29.2	111	0.972	-0.5	67
Alpine	0.973	10.1	55	0.971	-0.3	55
Maritime	0.968	-16.8	109	0.970	-4.5	105
Prairie	0.967	18.7	56	0.965	-0.2	51
Tundra	0.956	-10.5	82	0.969	-6.1	67
Taiga	0.943	-80.0	151	0.971	2.4	62

404
 405 **3.2 Comparison to Pistocchi (2016)**

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

408
 409 (12) $\rho_b = \rho_0 + K(DOY + 61)$,
 410

411 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
 412 November 1. Application of this model to the validation dataset yields a bias of 55 mm and an RMSE of 94 mm.
 413 These results are comparable to the Sturm et al. (2010) model, with a larger bias but smaller RMSE.

414
 415 **3.3 Results for Northeast USA**

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

422
 423 Figure 12 graphically summarizes the datasets and the performance of the multi-variable two-equation model of the
 424 current study. The RMSE values are comparable to those found for the western stations, but, given the
 425 comparatively thinner snowpacks in the northeast, represent a larger relative error (Table 5). The bias of the model
 426 is consistently positive, in contrast to the western stations where the bias was negligible.

427
 428 Table 5: Performance metrics for the multi-variable two-equation model applied to various northeastern USA
 429 datasets.

Dataset Name	R ²	Bias (mm)	RMSE (mm)
Maine Geological Survey, ME	0.91	8.9	33.3
Hubbard Brook (Station 2), NH	0.63	18.9	64.2
Thompson Farm, NH	0.85	7.1	21.6
NRCS SCAN	0.87	-1.8	38.7
Sleepers River, VT	0.93	14.0	29.7
New York Snow Survey	0.93	13.8	31.2

430

431 **4 Discussion**

432 The results presented in this study show that the regression equation described by equations (5, 7-8) is an
 433 improvement (lower bias and RMSE) over other widely used bulk density equations. The key advantage is that the
 434 present method regresses in relevant physical parameters directly, rather than using discrete bins (for snow class,
 435 elevation, month of year, etc.), each with its own set of model coefficients. The comparison (Figs. 10-11; Table 4) to
 436 the model of Sturm et al. (2010) reveals a peculiar behavior of that model for the Taiga snow class, with a large

437 negative bias in the Sturm estimates. Inspection of the coefficients provided for that class (Table 3) shows that the
438 model simply predicts that $\rho_b = \rho_{max} = 0.217$ for all conditions.

439
440 When our multi-variable two-equation model, developed solely from western North American data, is applied to
441 northeast USA locations, it produces SWE estimates with smaller RSME values and larger biases than the western
442 stations. When comparing the SWE- h curves of the SNOTEL data (Figure 4b) to those of the east coast data sets
443 (left column; Figure 12), it is clear that the northeast data generally have more scatter. This is confirmed by
444 computing the correlation coefficients between SWE and h for each dataset. It is unclear if this disparity in
445 correlation is related to measurement methodology or is instead a ‘signal to noise’ issue. Comparing Figures 4 and
446 12 shows the considerable difference in snowpack depth between the western and northeastern data sets. When the
447 western dataset is filtered to include only measurement pairs where $h < 1.5$ m, the correlation coefficient is reduced
448 to a value consistent with the northeast datasets. This suggests that the performance of the current (or other)
449 regression model is not as good at shallow snowpack depths. This is also suggested upon examination of the time
450 series of observed $\rho_b = SWE/h$ for a given season at a snow pillow site. Very early in the season, when the depths
451 are small, the density curve has a lot of variability. Later in the season, when depths are greater, the density curve
452 becomes much smoother. Very late in the season, when depths are low again, the density curve becomes highly
453 variable again.

454
455 Measurement precision and accuracy affect the construction and use of a regression model. Upon inspection of the
456 snow pillow data, it was observed that the precision of the depth measurements was approximately 25 mm and that
457 of the SWE measurements was approximately 2.5 mm. To test the sensitivity of the model coefficients to the
458 measurement precision, the depth values in the training dataset were randomly perturbed by +/- 25 mm and the SWE
459 values were randomly perturbed by +/- 2.5 mm and the regression coefficients were recomputed. This process was
460 repeated numerous times and the mean values of the perturbed coefficients were obtained. These adjusted
461 coefficients were then used to recompute the SWE values for the validation data set and the bias and RMSE were
462 found to be -10.5 mm and 72.7 mm. This represents a roughly 10% increase in RMSE, but a considerable increase in
463 bias magnitude (see Table 4 for the original values). This sensitivity of the regression analysis to measurement
464 precision underscores the need to have high-precision measurements for the training data set. Regarding accuracy,
465 random and systematic errors in the paired SWE - h data used to construct the regression model will lead to
466 uncertainties in SWE values predicted by the model. As noted in the introduction, snow pillow errors in SWE
467 estimates do not follow a simple pattern. Additionally, they are complicated by the fact that the errors are often
468 computed by comparing snow pillow data to coring data, which itself is subject to error. Lacking quantitative
469 information on the distribution of snow pillow errors, we are unable to quantify the uncertainty in the SWE
470 estimates.

471
472 Another important consideration has to do with the uncertainty of depth measurements that the model is applied to.
473 For context, one application of this study is to crowd-sourced, opportunistic snow depth measurements from

474 programs like the Community Snow Observations (CSO; Hill et al., 2018) project. In the CSO program,
475 backcountry recreational users submit depth measurements, typically taken with an avalanche probe, using a
476 smartphone in the field. The measurements are then converted to SWE estimates which are assimilated into
477 snowpack models. These depth measurements are ‘any time, any place’ in contrast to repeated measurements from
478 the same location, like snow pillows or snow courses. Most avalanche probes have cm-scale graduated markings, so
479 measurement precision is not a major issue. A larger problem is the considerable variability in snowpack depth that
480 can exist over short (meter scale) distances. The variability of the Chugach avalanche probe measurements was
481 assessed by taking the standard deviation of 8 h measurements per site. The average of this standard deviation over
482 the sites was 22 cm and the average coefficient of variation (standard deviation normalized by the mean) over the
483 sites was 15%. This variability is a function of the surface roughness of the underlying terrain, and also a function of
484 wind redistribution of snow. Propagating this uncertainty through the regression equations yields a slightly higher
485 (16%) uncertainty in the SWE estimates. CSO participants can do three things to ensure that their recorded depth
486 measurements are as representative as possible. First, avoid measurements in areas of significant wind scour or
487 deposition. Second, avoid measurements in terrain likely to have significant surface roughness (rocks, fallen logs,
488 etc.). Third, take several measurements and average them.

489
490 Expansion of CSO measurements in areas lacking SWE measurements can increase our understanding of the
491 extreme spatial variability in snow distribution and the inherent uncertainties associated with modeling SWE in
492 these regions. It could also prove useful for estimating watershed-scale SWE in regions like the northeastern USA,
493 which is currently limited to five automated SCAN sites with historical SWE measurements for only the past two
494 decades. Additionally, historical snow depth measurements are more widely available in the Global Historical
495 Climatology Network (GHCN-Daily; Menne et al. 2012), with several records extending back to the late 1800s.
496 While many of the GHCN stations are confined to lower elevations with shallower snow depths, the broader
497 network of quality-controlled snow depth data paired with daily GHCN temperature and precipitation measurements
498 could potentially be used to reconstruct SWE in the eastern US given additional model development and refinement.

499 **5 Conclusions**

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

505
506 We developed this model with data from paired SWE- h measurements from the western United States and British
507 Columbia. The model was tested against entirely independent data (primarily snow course; some snow pillow) from
508 the northeastern United States and was found to perform well, albeit with larger biases and root-mean-squared-
509 errors. The model was tested against other well-known regression equations and was found to perform better.

510

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

515 **6 Acknowledgements**

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

519 **7 Data Access**

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

- 521 1. NRCS Snow Telemetry: <https://www.wcc.nrcs.usda.gov/snow/SNOTEL-wedata.html>
- 522 2. NRCS Soil Climate Analysis Network: <https://www.wcc.nrcs.usda.gov/scan/>
- 523 3. British Columbia Automated Snow Weather Stations:
524 [https://www2.gov.bc.ca/gov/content/environment/air-land-water/water/water-science-data/water-data-](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)
525 [tools/snow-survey-data/automated-snow-weather-station-data](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)
- 526 4. Maine Cooperative Snow Survey: <https://mgs-maine.opendata.arcgis.com/datasets/maine-snow-survey-data>
- 527 5. New York Snow Survey: <http://www.nrcc.cornell.edu/regional/snowsury/snowsury.html>
- 528 6. Sleepers River Research Watershed. Snow data not available online; request data from contact at:
529 <https://nh.water.usgs.gov/project/sleepers/index.htm>
- 530 7. Hubbard Brook Experimental Forest: <https://hubbardbrook.org/d/hubbard-brook-data-catalog>
- 531 8. CONUS PRISM Data: <http://www.prism.oregonstate.edu/>
- 532 9. British Columbia PRISM Data: <http://climatebcdata.climatewna.com/>
- 533 10. Alaska PRISM Data: <https://irma.nps.gov/Portal/>

534

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

537 **References**

538

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

541

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

544

545 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](https://doi.org/10.1175/1520-0450(1965)004<0626:MHPPSG>2.0.CO;2) 1965.

546

547
548 Beaumont, R., and Work, R.: Snow sampling results from three samplers, *Hydrol. Sci. J.*, 8(4), 74-78,
549 <https://doi.org/10.1080/02626666309493359>, 1963.

550

551 Burakowski, E.A., Ollinger, S., Lepine, L., Schaaf, C.B., Wang, Z., Dibb, J.E., Hollinger, D.Y., Kim, J.-H., Erb, A.,
552 and Martin, M.E.: Spatial scaling of reflectance and surface albedo over a mixed-use, temperate forest landscape
553 during snow-covered periods, *Remote Sens. Environ.*, 158, 465-477, <https://doi.org/10.1016/j.rse.2014.11.023>,
554 2015.

555

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

559

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

563

564 Church, J.E.: Snow surveying: its principles and possibilities, *Geogr. Rev.*, 23(4), 529-563, DOI: 10.2307/209242,
565 1933.

566

567 Church, J.E., and Marr, J.C.: Further improvement of snow-survey apparatus, *Transactions of the American*
568 *Geophysical Union*, 18(2), 607-617, [10.1029/TR018i002p00607](https://doi.org/10.1029/TR018i002p00607), 1937.

569

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

572
573

574 Deeb, E., Marshall, H.-P., Forster, R., Jones, C., Hiemstra, C., and Siqueira, P.: Supporting NASA SnowEx remote
575 sensing strategies and requirements for L-band interferometric snow depth and snow water equivalent estimation,
576 Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, 1395-
577 1396, DOI: 10.1109/IGARSS.2017.8127224, 2017.

578

579 Dixon, D., and Boon, S.: Comparison of the SnowHydro sampler with existing snow tube designs, *Hydrol. Process.*,
580 26(17), 2555-2562, <https://doi.org/10.1002/hyp.9317>, 2012.

581

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

584

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

587

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

590

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

593

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

596

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

599

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

602

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

605

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

608

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

611
612 Jonas, T., Marty, C., and Magnusson, M.: Estimating the snow water equivalent from snow depth measurements, *J.*
613 *Hydrol.*, 378, 161-167, <https://doi.org/10.1016/j.jhydrol.2009.09.021>, 2009.
614
615 Kang, D.H., and Barros, A.P.: Full-system testing in laboratory conditions of an L-band snow sensor system for in
616 situ monitoring of snow-water content, *IEEE Trans. Geosci. Remote Sens.*, 49(3), 908-919,
617 10.1109/TGRS.2010.2072786, 2011.
618
619 Leinss, S., Wiesmann, A., Lemmetyinen, J., and Hajnsek, I.: Snow water equivalent measured by differential
620 interferometry, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 8(8), 3773-3790, 10.1109/JSTARS.2015.2432031,
621 2015.
622
623 Leys, C., Klein, O., Dominicy, Y., and Ley, C.: Detecting multivariate outliers: use a robust variant of the
624 Mahalanobis distance, *J. Exp. Soc. Psychol.*, 74, 150-156, <https://doi.org/10.1016/j.jesp.2017.09.011>, 2018.
625
626 Liang, X., Lettermaier, D., Wood, E., and Burges, S.: A simple hydrologically based model of land surface water
627 and energy fluxes for general circulation models, *J. Geophys. Res. Atmos.*, 99(D7), 14,415-14,428,
628 <https://doi.org/10.1029/94JD00483>, 1994.
629
630 Liston, G., and Elder, K.: A distributed snow evolution modeling system (SnowModel), *J. Hydrometeorol.*, 7, 1259-
631 1276, <https://doi.org/10.1175/JHM548.1>, 2006.
632
633 Lundberg, A., Richardson-Naslund, C., and Andersson, C.: Snow density variations: consequences for ground
634 penetrating radar, *Hydrol. Process.*, 20, 1483-1495, <https://doi.org/10.1002/hyp.5944>, 2006.
635
636 Maine Geological Survey: Maine Cooperative Snow Survey Dataset,
637 https://www.maine.gov/dacf/mgs/hazards/snow_survey/, 2018.
638
639 McKay, G., and Findlay, B., 1971: Variation of snow resources with climate and vegetation in Canada, *Proceedings*
640 *of the 39th Western Snow Conference*, Billings, MT, 17-26, 1971.
641
642 De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D.: The Mahalanobis distance, *Chemometr. Intell. Lab. Syst.*,
643 50(1), 1-18, [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7), 2000.
644
645 McCreight, J., and Small, E.: Modeling bulk density and snow water equivalent using daily snow depth
646 observations, *The Cryosphere*, 8, 521-536, <https://doi.org/10.5194/tc-8-521-2014>, 2014.
647

648 Meloyund, V., Leira, B., Hoiseith, K., and Liso, K.: Predicting snow density using meteorological data, *Meteorol.*
649 *Appl.*, 14, 413-423, <https://doi.org/10.1002/met.40>, 2007.

650

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

654

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

658

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

661

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

664

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

666

667 Pagano, T., Garen, D., Perkins, T., and Pasteris, P.: Daily updating of operational statistical seasonal water supply
668 forecasts for the western U.S., *J. Am. Water Resour. Assoc.*, 45(3), 767-778, [https://doi.org/10.1111/j.1752-](https://doi.org/10.1111/j.1752-1688.2009.00321.x)
669 [1688.2009.00321.x](https://doi.org/10.1111/j.1752-1688.2009.00321.x), 2009.

670

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

676

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

679

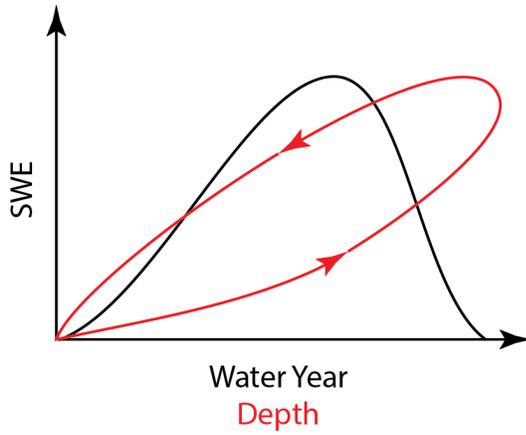
680 Rousseeuw, P.: Least Median of Squares Regression, *J. Am. Stat. Assoc.*, 79, 871-880, DOI:
681 [10.1080/01621459.1984.10477105](https://doi.org/10.1080/01621459.1984.10477105), 1984.

682

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

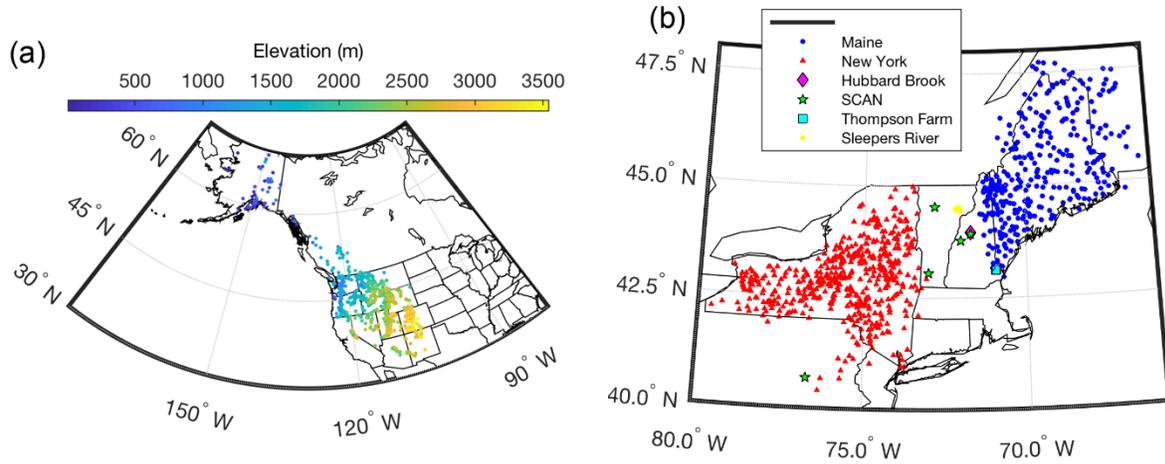
685
686 Schaefer, G., Cosh, M., and Jackson, T.: The USDA Natural Resources Conservation Service Soil Climate Analysis
687 Network (SCAN), *J. Atmos. Ocean. Technol.*, 24, 2073-2077, <https://doi.org/10.1175/2007JTTECHA930.1>, 2007.
688
689 Serreze, M., Clark, M., Armstrong, R., McGinnis, D., and Pulwarty, R.: Characteristics of the western United States
690 snowpack from snowpack telemetry (SNOTEL) data, *Water Resour. Res.*, 35(7), 2145-2160,
691 <https://doi.org/10.1029/1999WR900090>, 1999.
692
693 Shanley, J., and Chalmers, A.: The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont, *Hydrol.*
694 *Process.*, 13(12-13), 1843-1857, [https://doi.org/10.1002/\(SICI\)1099-1085\(199909\)13:12/13<1843::AID-](https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1843::AID-HYP879>3.0.CO;2-G)
695 [HYP879>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1843::AID-HYP879>3.0.CO;2-G), 1999.
696
697 Sokol, J., Pultz, T., and Walker, A.: Passive and active airborne microwave remote sensing of snow cover,” *Int.*
698 *J. Remote. Sens.*, 24, 5327-5344, <https://doi.org/10.1080/0143116031000115076>, 2003.
699
700 Sturm, M., Holmgren, J., and Liston, G.: A seasonal snow cover classification system for local to global
701 applications, *J. Clim.*, 8, 1261-1283, [https://doi.org/10.1175/1520-0442\(1995\)008<1261:ASSCCS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2), 1995.
702
703 Sturm, M., Taras, B., Liston, G.E., Derksen, C., Jonas, T., and Lea, J.: Estimating snow water equivalent using snow
704 depth data and climate classes, *J. Hydrometeorol.*, 11, 1380-1394, <https://doi.org/10.1175/2010JHM1202.1>, 2010.
705
706 U.S. Army Corps of Engineers: Snow hydrology: Summary report of the snow investigations of the North Pacific
707 Division, 437pp., 1956.
708
709 Vuyovich, C., Jacobs, J., and Daly, S.: Comparison of passive microwave and modeled estimates of total watershed
710 SWE in the continental United States, *Water Resour. Res.*, 50(11), 9088-9012,
711 <https://doi.org/10.1002/2013WR014734>, 2014.
712
713 Wang, T., Hamann, A., Spittlehouse, D.L., and Murdock, T.: ClimateWNA - High-Resolution Spatial Climate Data
714 for Western North America, *J. Appl. Meteorol. Climatol.*, 51, 16-29, <https://doi.org/10.1175/JAMC-D-11-043.1>,
715 2012.
716
717 Wigmosta, M.S., Vail, L., and Lettenmaier, D.: A distributed hydrology-vegetation model for complex
718 terrain, *Water Resour. Res.*, 30, 1665-1679, <https://doi.org/10.1029/94WR00436>, 1994

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

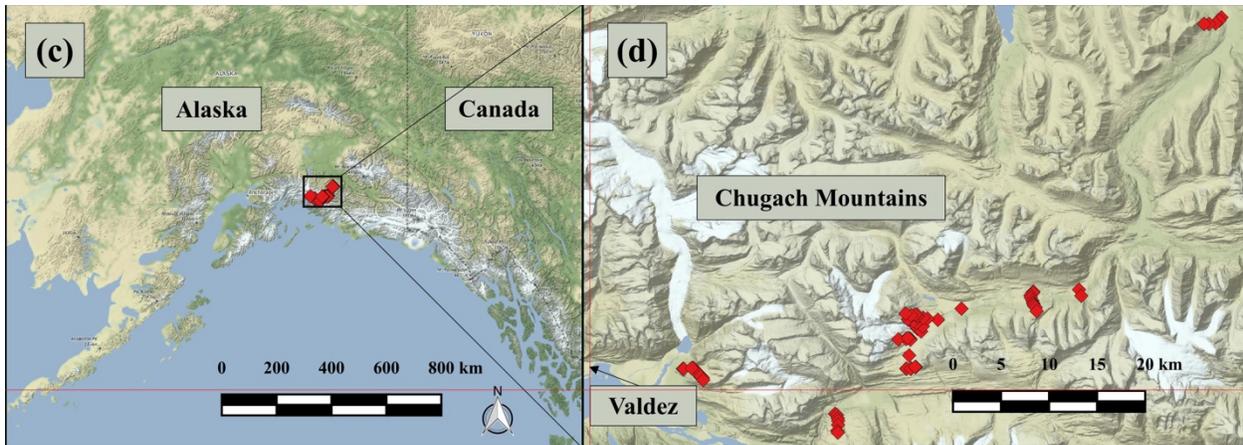


722

723 Figure 2: Distribution of measurement locations used in this study. (a) Western USA and Canada station locations,
724 with colors indicating station elevation in meters. (b) Northeast USA locations, with stations colored according to
725 data source. (c, d) Measurement sites in the Chugach Mountains, southcentral Alaska.
726



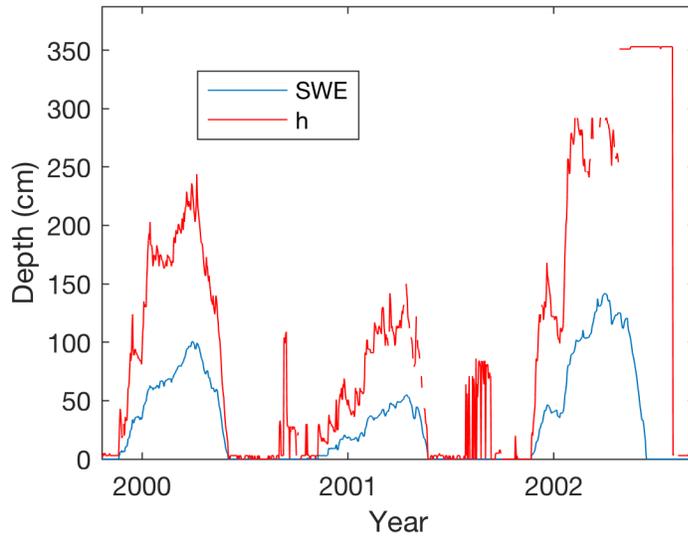
727



728

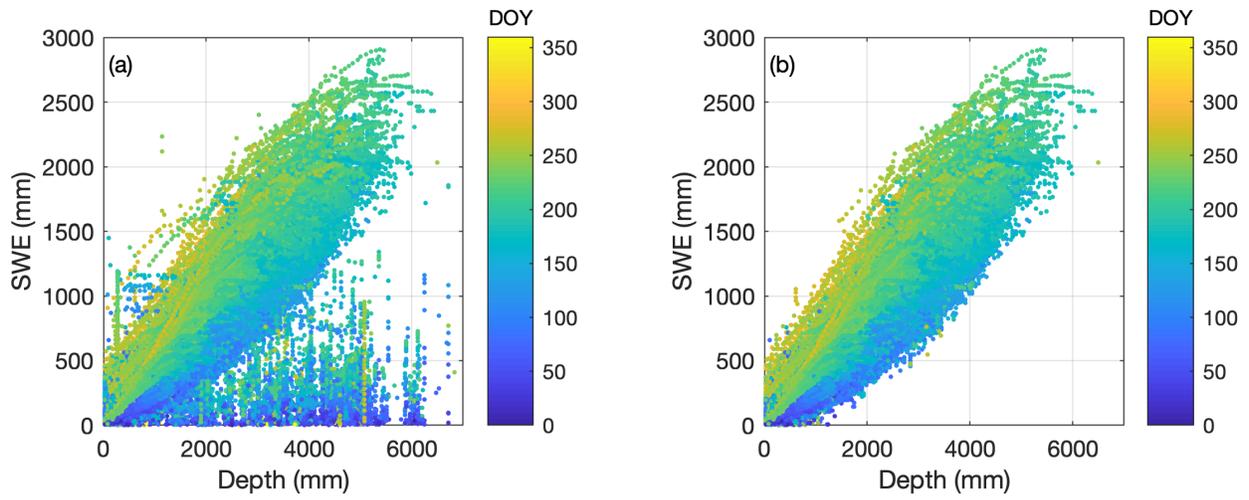
729

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



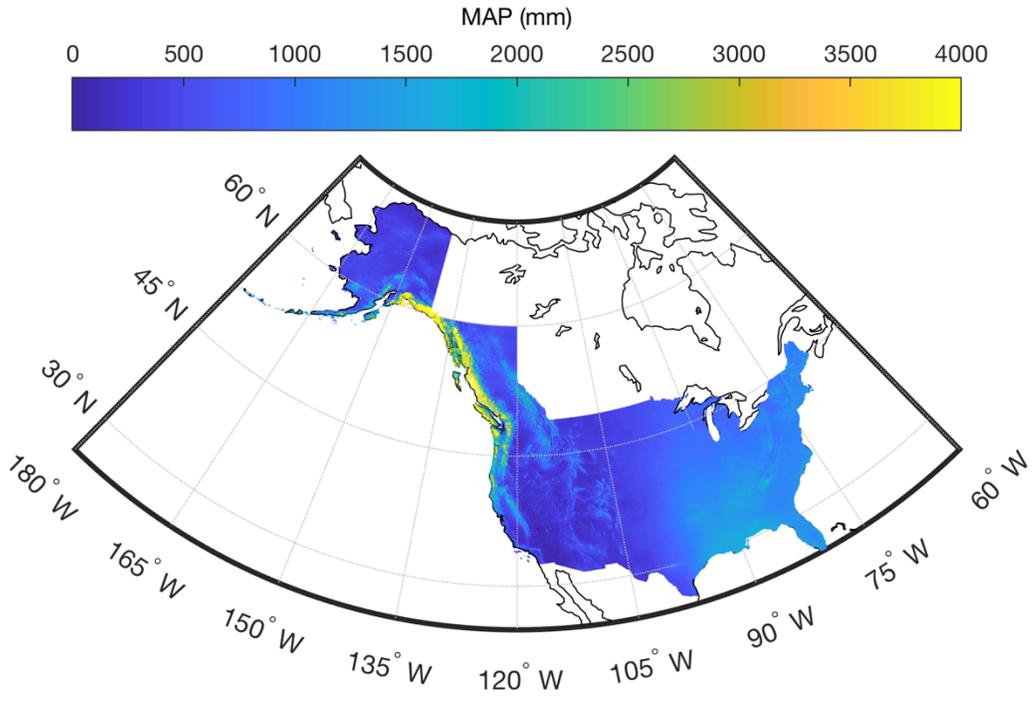
732

733 Figure 4: Scatter plot of SWE vs. h for the complete SNOTEL dataset before (a) and after (b) removing data points.
734 Symbols are colored by 'day of water year' (DOY ; October 1 is the origin).
735

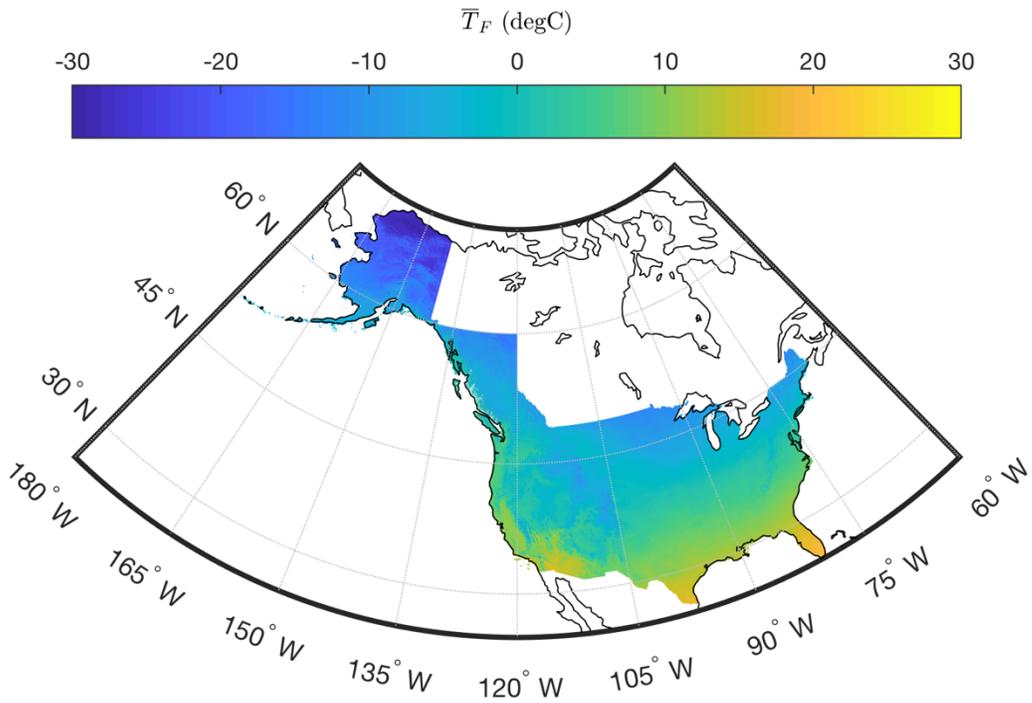


736

737 Figure 5: Gridded maps of mean annual precipitation (MAP) and mean February temperature ($\bar{T}_{F,mean}$) for the study
738 regions. Climate normals are from the PRISM data set (1981-2010 for CONUS and British Columbia; 1971-2000
739 for Alaska).

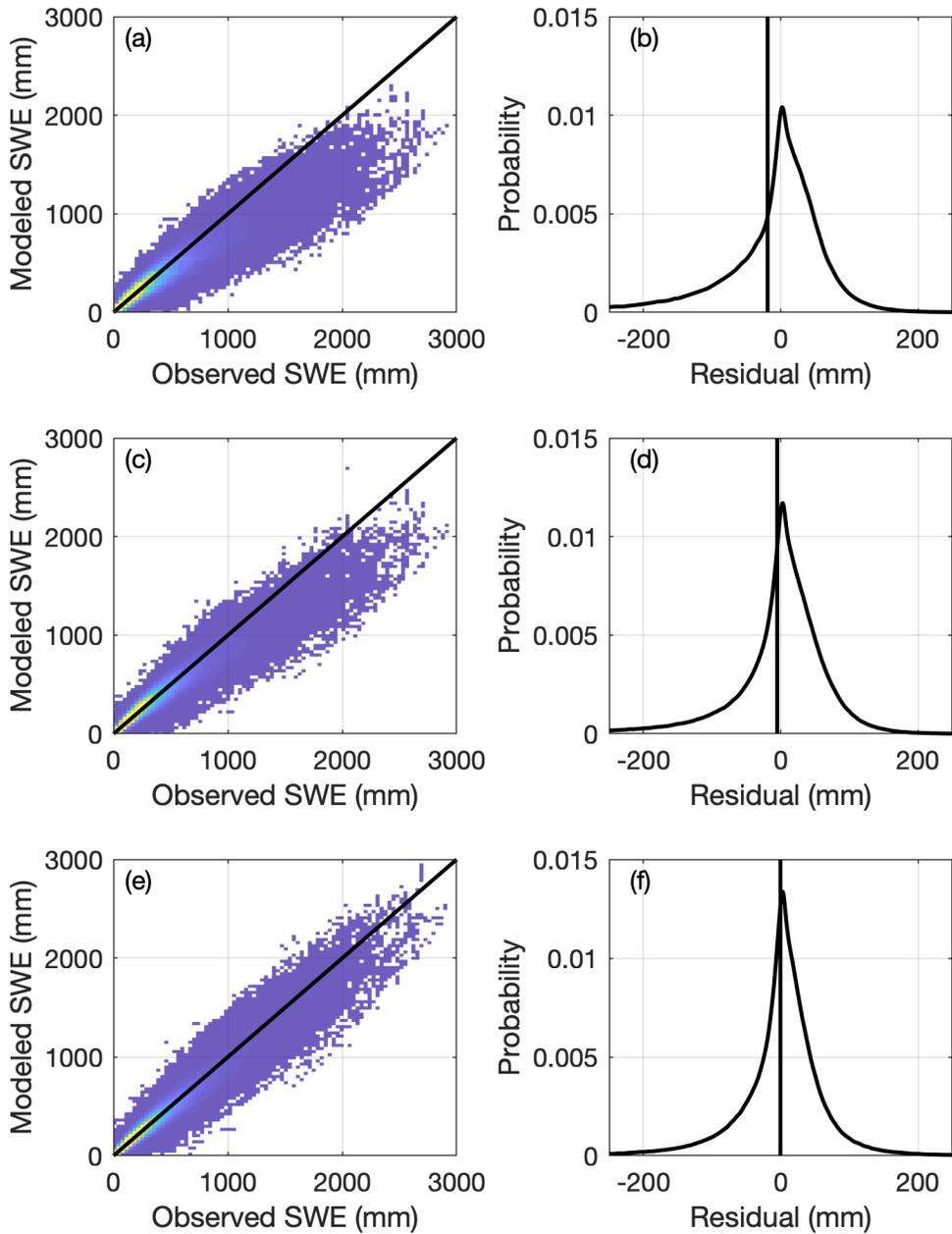


740



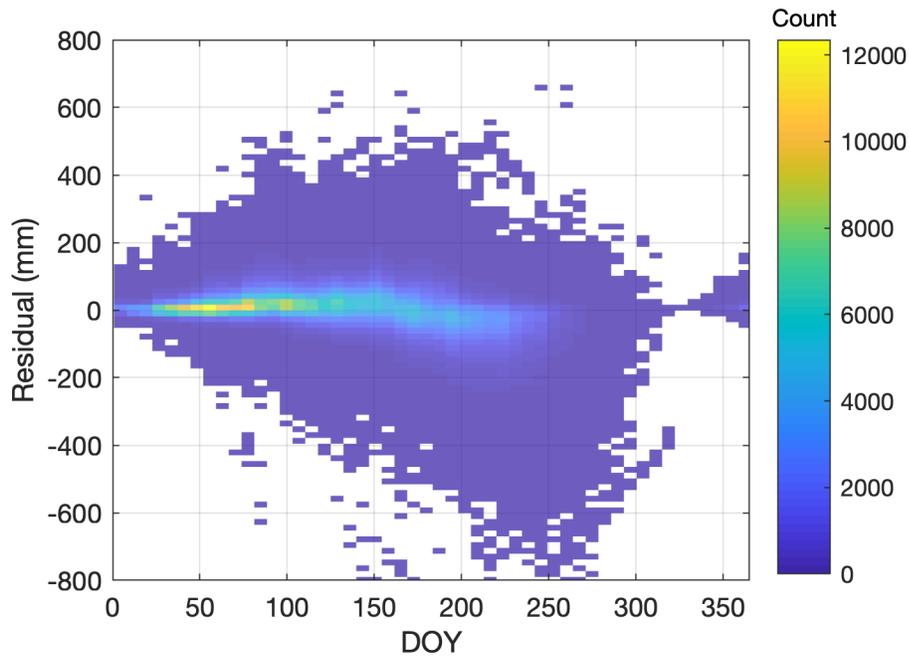
741
742

743 Figure 6: Two-dimensional histograms (heat maps; left column) of modeled vs. observed SWE and probability
744 density functions (right column) of the residuals for three simple models applied to the CONUS, AK, and BC snow
745 pillow data. Warmer colors in the heat maps indicate greater density. The vertical lines in the right column indicate
746 the location of the mean residual, or bias. Top row (a-b): One-equation model (Section 2.2.1). Middle row (c-d):
747 Two-equation model (Section 2.2.2). Bottom row (e-f): Multi-variable two-equation model (Section 2.2.3).
748



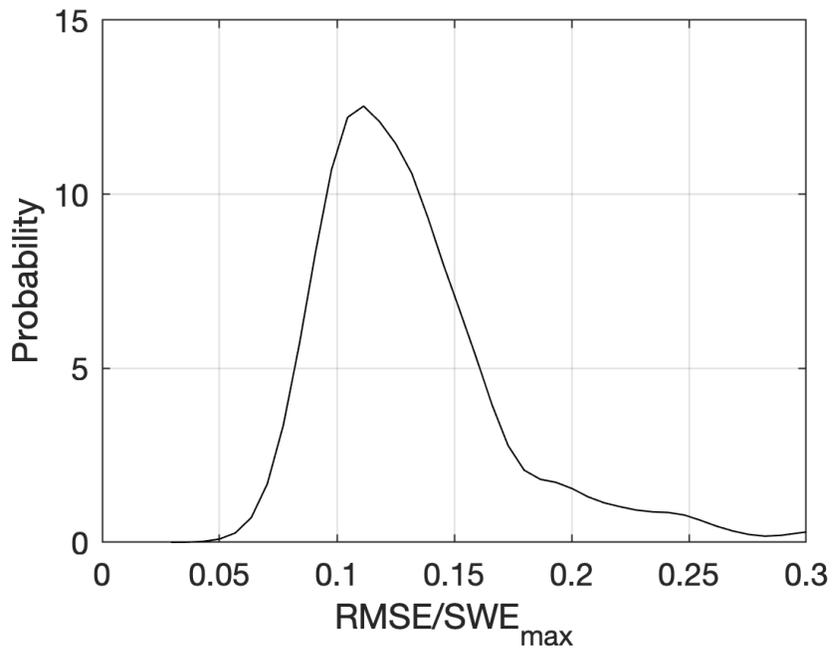
749

750 Figure 7: Heat map of SWE residuals as a function of *DOY*.
751



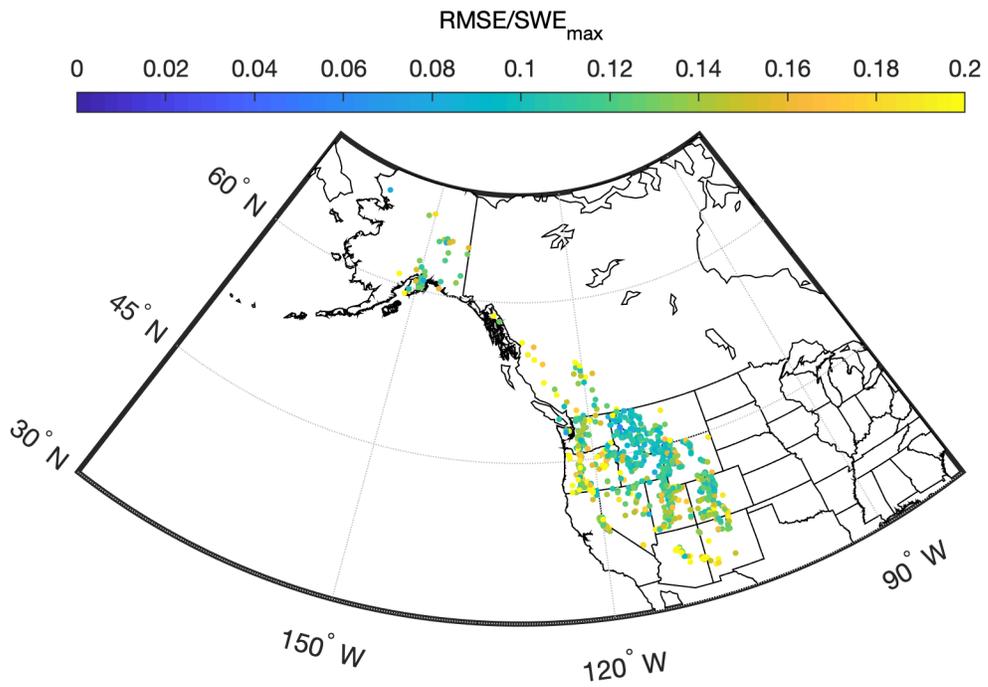
752

753 Figure 8: Probability density function of snow pillow station root-mean-square error (RMSE) normalized by station
754 mean annual maximum SWE.



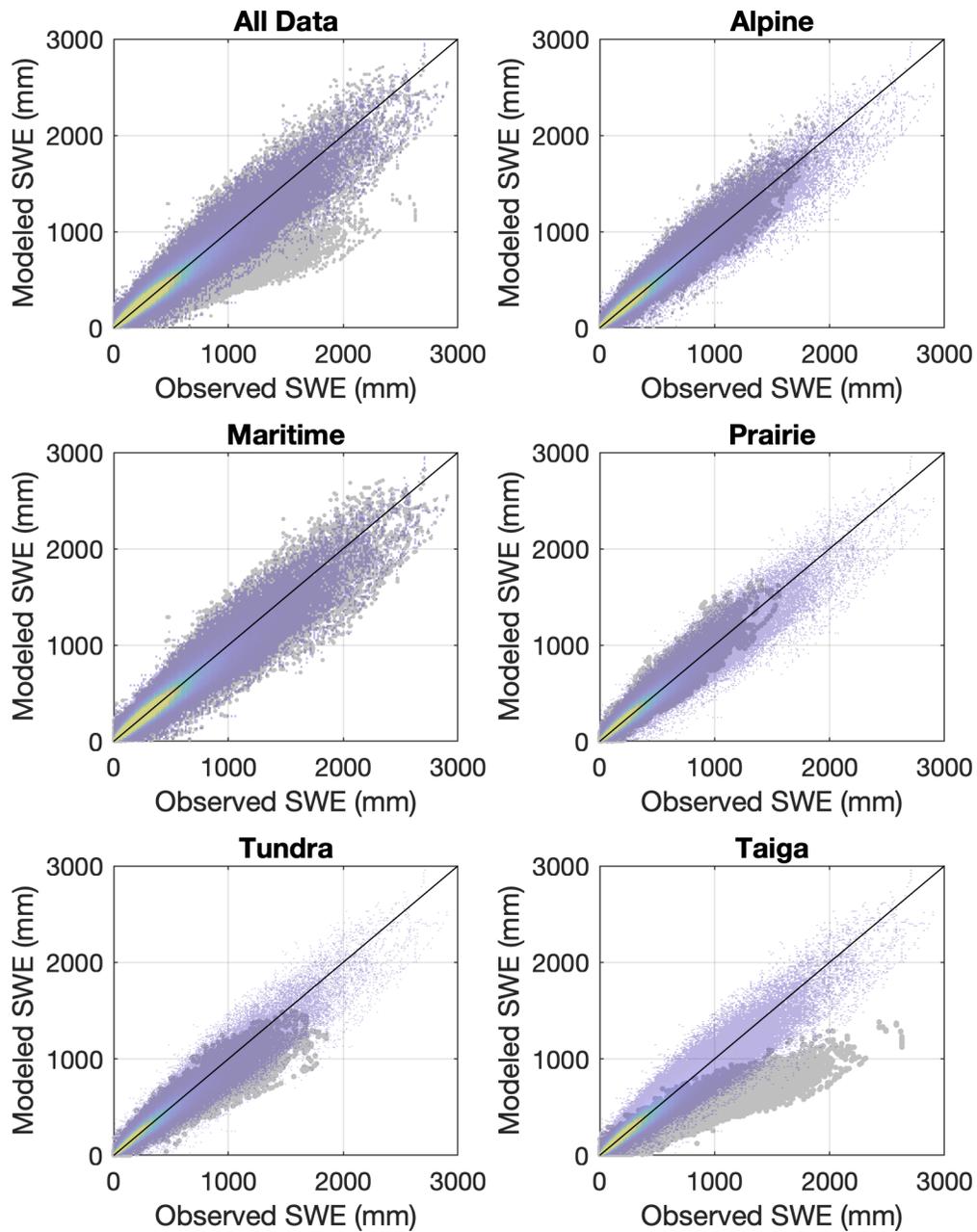
755

756 Figure 9: Spatial distribution of snow pillow station root-mean-square error (RMSE) normalized by station mean
757 annual maximum SWE.
758



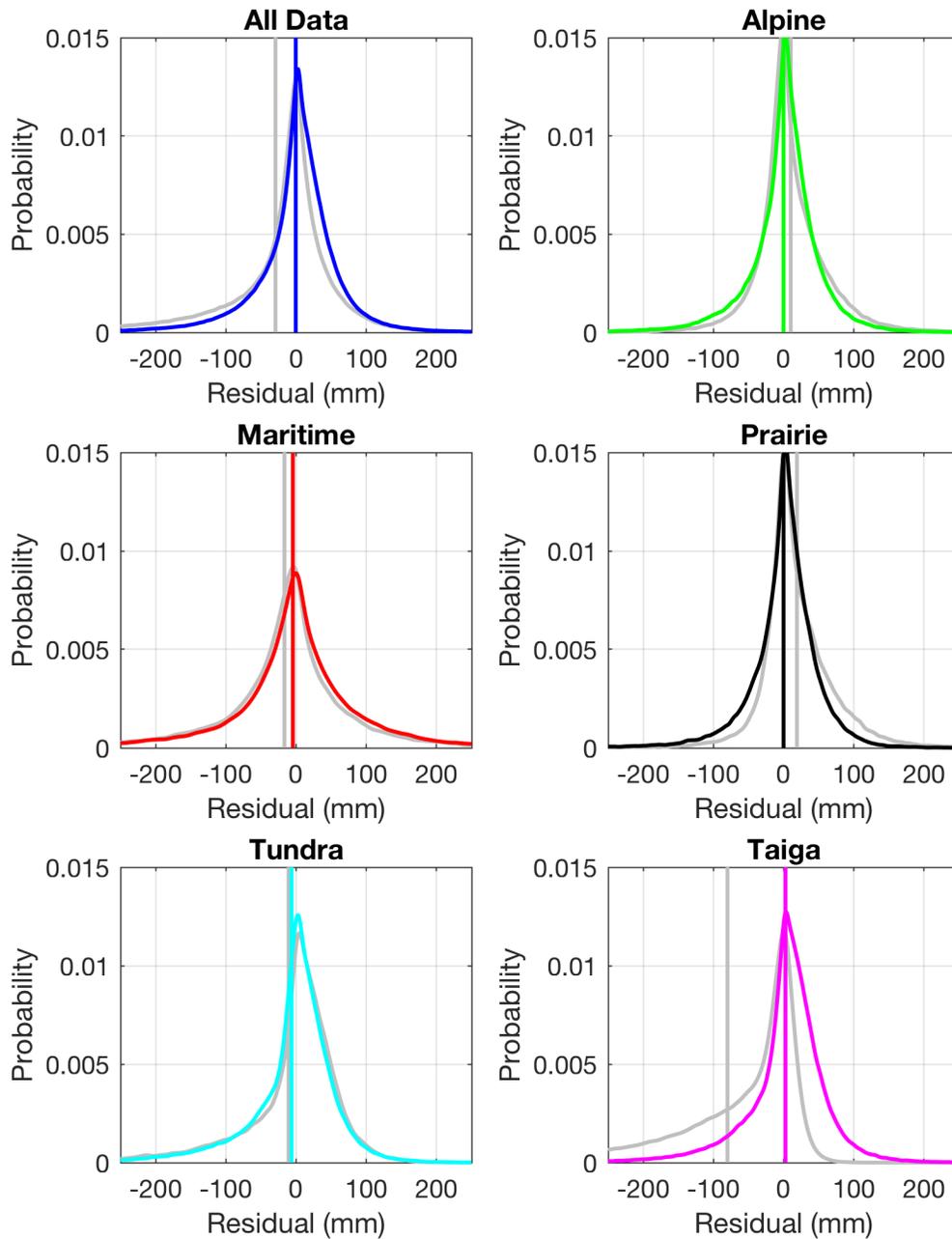
759

760 Figure 10: Comparison of the multi-variable, two-equation model of the present study with the model of Sturm et al.
761 (2010). The subpanels show modeled SWE vs. observed SWE for all of the data binned together, as well as for the
762 data broken out by the snow classes identified by Sturm et al. (1995). The gray symbols show the Sturm result and
763 the transparent heat maps (warmer colors indicate greater density) show the current result. The models are being
764 applied to the validation data set (50% of the aggregated snow pillow data for CONUS, AK, and BC).



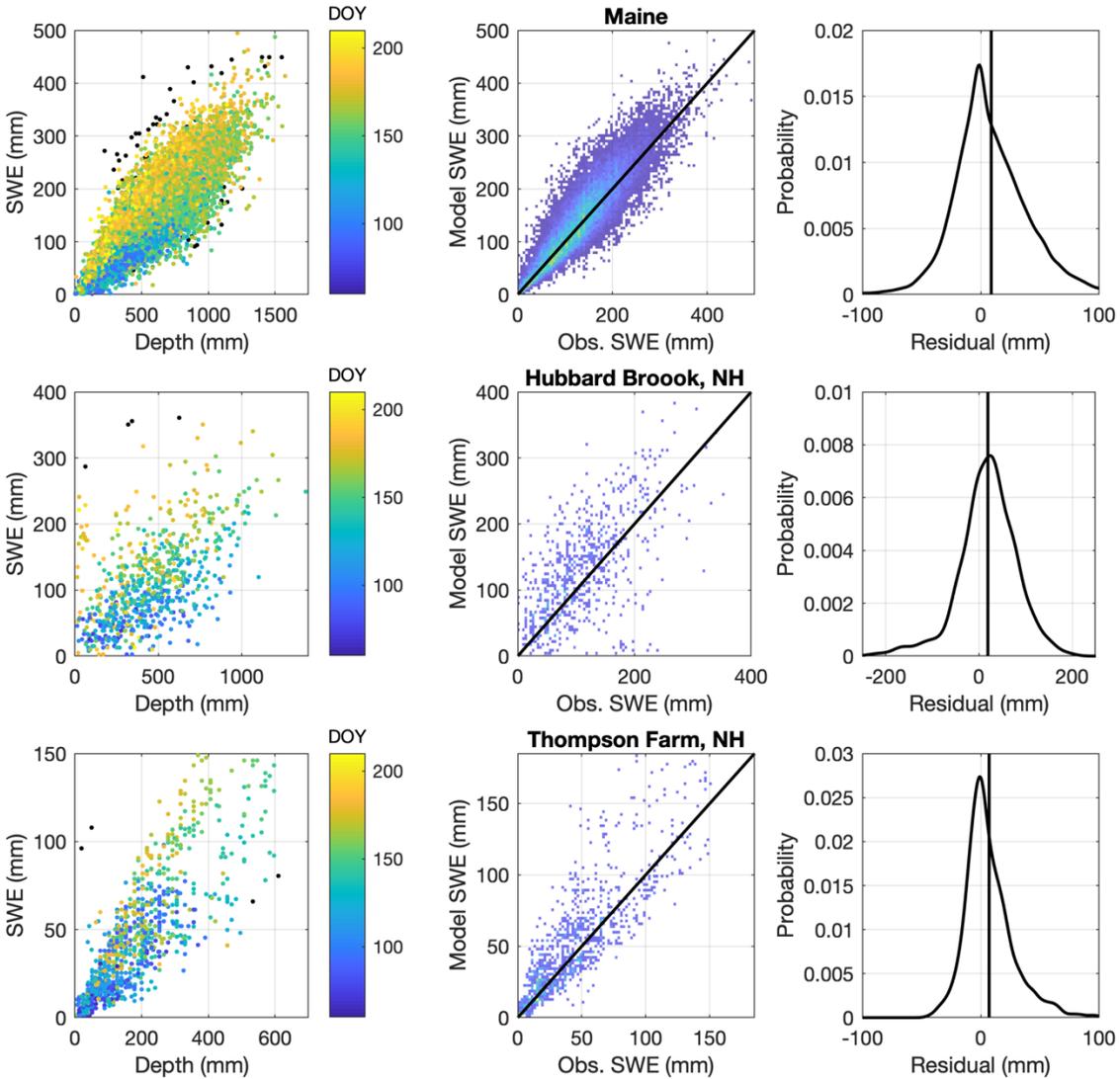
765

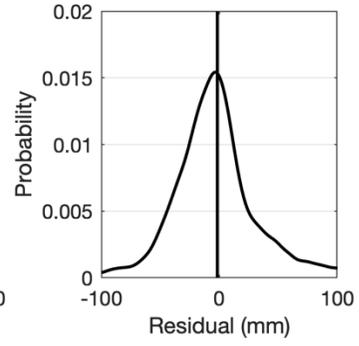
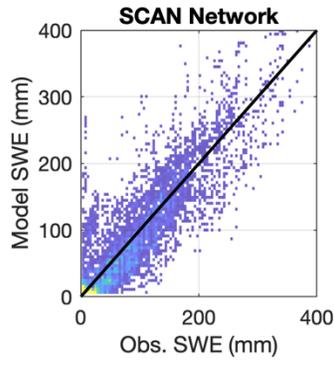
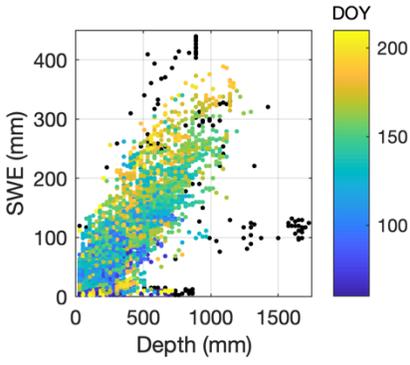
766 Figure 11: Comparison of the multi-variable, two-equation model of the present study with the model of Sturm et al.
767 (2010). The subpanels show probability density functions of the residuals of the model fits for all of the data binned
768 together, as well as for the data broken out by the snow classes identified by Sturm et al. (1995). The gray lines
769 show the Sturm result and the colored lines show the current result. The vertical lines show the mean error, or the
770 model bias, for both the Sturm and the current result. The models are being applied to the validation data set (50% of
771 the aggregated snow pillow data for CONUS, AK, and BC).



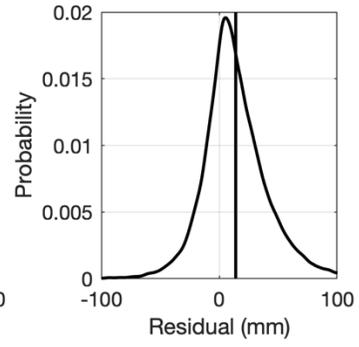
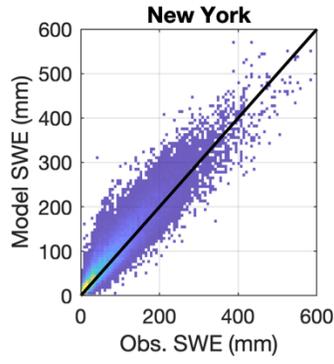
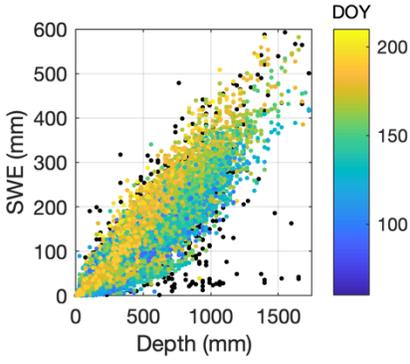
772

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

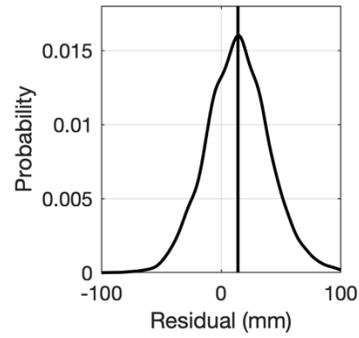
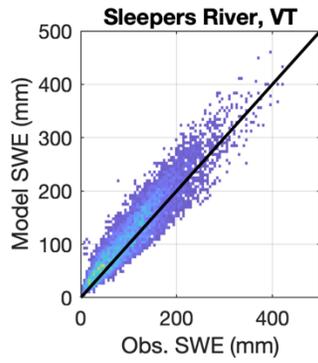
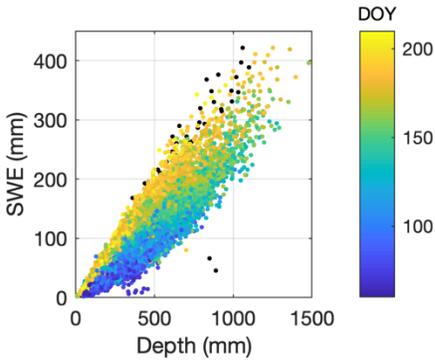




782



783



784

785

786

1 Converting Snow Depth to Snow Water Equivalent Using 2 Climatological Variables

3
4 David F. Hill¹, Elizabeth A. Burakowski², Ryan L. Crumley³, Julia Keon⁴, J. Michelle Hu⁵,
5 Anthony A. Arendt⁶, Katreen Wikstrom Jones⁷, Gabriel J. Wolken⁸

6
7 ¹Civil and Construction Engineering, Oregon State University, OR, USA

8 ²Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, NH, USA

9 ³Water Resources Graduate Program, Oregon State University, OR, USA

10 ⁴Civil and Construction Engineering, Oregon State University, OR, USA

11 ⁵Civil and Environmental Engineering, University of Washington

12 ⁶Applied Physics Laboratory, University of Washington

13 ⁷Alaska Division of Geological & Geophysical Surveys, Fairbanks, AK, USA

14 ⁸Alaska Division of Geological & Geophysical Surveys, Fairbanks, AK, USA; International Arctic Research Center,
15 University of Alaska Fairbanks, Fairbanks, AK, USA

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

17
18
19 **Abstract.** We present a simple method that allows snow depth measurements to be converted to snow water
20 equivalent (SWE) estimates. These estimates are useful to individuals interested in water resources, ecological
21 function, and avalanche forecasting. They can also be assimilated into models to help improve predictions of total
22 water volumes over large regions. The conversion of depth to SWE is particularly valuable since snow depth
23 measurements are far more numerous than costlier and more complex SWE measurements. Our model regresses
24 SWE against snow depth and climatological (30-year normal) values for mean annual precipitation (*MAP*) and mean
25 February temperature (\bar{T}_{Fmean}), producing a power-law relationship. Relying on climatological normals rather than
26 weather data for a given year allows our model to be applied at measurement sites lacking a weather station.
27 Separate equations are obtained for the accumulation and the ablation phases of the snowpack, which introduces
28 ‘day of water year’ (*DOY*) as an additional variable. The model is validated against a large database of snow pillow
29 measurements and yields a bias in SWE of less than 0.5 mm and a root-mean-squared-error (RMSE) in SWE of
30 approximately 65 mm. When the errors are investigated on a station-by-station basis, the average RMSE is about 5%
31 of the *MAP* at each station. The model is additionally validated against a completely independent set of data from
32 the northeast United States. Finally, the results are compared with other models for bulk density that have varying
33 degrees of complexity and that were built in multiple geographic regions. The results show that the model described
34 in this paper has the best performance for the validation data set.

Formatted: Font:Italic

35 **1 Introduction**

36 In many parts of the world, snow plays a leading-order role in the hydrological cycle (USACE, 1956; Mote et al.,
37 2018). Accurate information about the spatial and temporal distribution of snow water equivalent (SWE) is useful to
38 many stakeholders (water resource planners, avalanche forecasters, aquatic ecologists, etc.), but can be time
39 consuming and expensive to obtain.

40
41 Snow pillows (Beaumont, 1965) are a well-established tool for measuring SWE at fixed locations. Figure 1 provides
42 a conceptual sketch of the variation of SWE with time over a typical water year. A comparatively long accumulation
43 phase is followed by a short ablation phase. While simple in operation, snow pillows are relatively large in size and
44 they need to be installed prior to the onset of the season’s snowfall. This limits their ability to be rapidly or
45 opportunistically deployed. Additionally, snow pillow installations tend to require vehicular access, limiting their
46 locations to relatively simple topography. ~~Finally, snow pillow sites are not representative of the lowest or highest~~
47 ~~elevation bands within mountainous regions (Molotch and Bales, 2005). In the western United States (USA), the~~
48 ~~Natural Resources Conservation Service (NRCS) operates a large network of Snow Telemetry (SNOTEL) sites,~~
49 ~~featuring snow pillows. The NRCS also operates the smaller Soil Climate Analysis Network (SCAN) which~~
50 ~~provides the only, and very limited, snow pillow SWE measurements in the eastern USA.~~

51
52 SWE can also be measured manually, using a snow coring device that measures the weight of a known volume of
53 snow to determine snow density (Church, 1933). These measurements are often one-off measurements, or in the
54 case of ‘snow courses’ they are repeated weekly or monthly at a given location. The simplicity and portability of
55 coring devices expand the range over which measurements can be collected, but it can be challenging to apply these
56 methods to deep snowpacks due to the length of standard coring devices. Note that there are numerous different
57 styles of coring devices, including the Adirondack sampler and the Mt. Rose / Federal sampler (Church and Marr,
58 1937).

59
60 There are a number of issues that affect the accuracy of both snow pillow and snow coring measurements. With
61 coring measurements, if the coring device is not carefully extracted, a portion of the core may fall out of the device.
62 Or, snow may become compressed in the coring device during insertion. These effects have led to varying
63 conclusions, with some studies (e.g., Sturm et al., 2010) showing a low SWE bias and other studies (e.g., Goodison,
64 1978) showing a high SWE bias. As noted by Johnson et al. (2015) a good rule of thumb is that coring devices are
65 accurate to around ± 10%. Also, studies comparing different styles of snow samplers report statistically different
66 results, suggesting that SWE measurements are sensitive to the design of the specific coring device, such as the
67 presence of holes or slots, the device material, etc. (Beaumont and Work, 1963; Dixon and Boon, 2012). With snow
68 pillows, some studies (e.g., Goodison et al., 1981) note that ice bridging can lead to low biases in measured SWE,
69 with the snow surrounding the pillow partly supporting the snow over the pillow. Other studies (Johnson and Marks,
70 2004; Dressler et al., 2006; Johnson et al., 2015) note a more complex situation with SWE under-reported at times,

Deleted: ,
Deleted: and

Moved (insertion) [1]
Deleted: 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; Johnson et al., 2015) note a more complex situation with SWE under-reported at times, but over-reported at other times.
Deleted: these
Deleted: and/or the need to dig very deep snowpits

Moved down [2]: Studies comparing different styles of snow samplers report statistically different results, suggesting that SWE measurements are sensitive to the design of the coring device, such as the presence of holes or slots, the device material, etc. (Beaumont and Work, 1963; Dixon and Boon, 2012).

Moved (insertion) [2]
Deleted: S

89 [but over-reported at other times. Note that when snow pillow data are evaluated, they are most commonly compared](#)
90 [to coring measurements at the same location.](#)

91 ~~.....~~
92 All methods of measuring SWE are challenged by the fact that SWE is a depth-integrated property of a snowpack.
93 This is why the snowpack must be weighed, in the case of a snow pillow, or a core must be extracted from the
94 surface to the ground. This measurement complexity makes it difficult to obtain SWE information with the spatial
95 and temporal resolution desired for watershed-scale studies. Other snowpack properties, such as the depth h , are
96 much easier to measure. ~~For example, using a graduated device such as a meterstick or an avalanche probe to~~
97 ~~measure the depth takes only seconds. Automating depth measurements at a fixed location can easily be done using~~
98 low-cost ultrasonic devices (Goodison et al., 1984; Ryan et al., 2008). High-spatial-resolution measurements of
99 snowpack depth are commonly made with Light Detection and Ranging (LIDAR). One example of this is the
100 Airborne Snow Observatory program (ASO; Painter et al., 2016). The comparatively high expense of airborne
101 LIDAR surveys typical limits measurements geographically (to a few basins) and temporally (weekly to monthly
102 interval).

103
104 Given the relative ease in obtaining depth measurements, it is common to use h as a proxy for SWE. Figure 1 shows
105 a conceptual sketch of the variation of SWE with h over a typical water year. Noting the arrows on the curve, we see
106 that SWE is multi-valued for each h . This is due to the fact that the snowpack increases in density throughout the
107 water year, producing a hysteresis loop in the curve. A large body of literature exists on the topic of how to convert
108 h to SWE. It is beyond the scope of this paper to provide a full review of these ‘bulk density equations,’ where the
109 density is given by $\rho_b = SWE/h$. Instead, we refer readers to the useful comparative review by Avanzi et al. (2015).
110 Here, we prefer to discuss a limited number of previous studies that illustrate the spectrum of methodologies and
111 complexities that can be used to determine ρ_b or SWE.

112
113 Many studies express ρ_b as an increasing function (often linear) of h . In some cases (e.g., Lundberg et al., 2006) a
114 second equation is added where ρ_b attains a constant value when a threshold h is exceeded. A single linear equation
115 captures the process of densification of the snowpack during the accumulation phase, but performs poorly during the
116 ablation phase, where depths are decreasing but densities continue to increase or approach a constant value.

117 Other approaches choose to parameterize ρ_b in terms of time, rather than h . Pistocchi (2016) provides a single
118 equation while Mizukami and Perica (2008) provide two sets of equations, one set each for early and late season.
119 Each set contains four equations, each of which is applicable to a particular ‘cluster’ of stations. This clustering was
120 driven by observed densification characteristics and the resulting clusters are relatively spatially discontinuous.

121 Jonas et al. (2009) take the idea of region- (or cluster-) specific equations and extend it further to provide
122 coefficients that depend on time and elevation as well. They use a simple linear equation for ρ_b in terms of h and the
123 slope and intercept of the equation are given as monthly values, with three elevation bins for each month (36 pairs of
124 coefficients). There is an additional contribution to the intercept (or ‘offset’) which is region-specific (one of 7
125 regions).

Deleted: [1]

Deleted: U
Deleted: , and
Deleted: are
Deleted: automated

Deleted: and Magnusson

133

134 These classifications, whether based on region, elevation, or season, are valuable since they acknowledge that all
135 snow is not equal. [McKay and Findlay \(1971\)](#) discuss the controls that climate and vegetation exert on snow density,
136 [and](#) Sturm et al. (2010) address this directly by developing a snow density equation where the coefficients depend
137 upon the 'snow class' (5 classes). Sturm et al. (1995) explain the decision tree, based on temperature, precipitation,
138 and wind speed, that leads to the classification. The temperature metric is the 'cooling degree month' calculated
139 during winter months only. Similarly, only precipitation falling during winter months was used in the classification.
140 Finally, given the challenges in obtaining high quality, high-spatial-resolution wind information, vegetation
141 classification was used as a proxy. Using climatological values (rather than values for a given year), Sturm et al.
142 (1995) were able to develop a global map of snow classification.

143

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

155

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

162

163 The present paper seeks to generalize the ideas of Mizukami and Perica (2008), Jonas [et al.](#) (2009), and Sturm et al.,
164 (2010). Specifically, our goal is to regress physical and environmental variables directly into the equations. In this
165 way, environmental variability is handled in a continuous fashion rather than in a discrete way (model coefficients
166 based on classes). The main motivation for this comes from evidence (e.g., Fig. 3 of Alford, 1967) that density can
167 vary significantly over short distances on a given day. Bulk density equations that rely solely on time completely
168 miss this variability and equations that have coarse (model coefficients varying over either vertical bins or horizontal
169 grids) spatial resolution may not fully capture it either.

Deleted: and Magnusson

Deleted: (and correspondingly, accuracy)

Deleted: .

Deleted: Also, many of these models resolve vertical variations in snow density which are wholly absent from the bulk (vertically averaged) density approaches reviewed above.

Deleted: f

Deleted: and Magnusson

179

180 Our approach is most similar to Mizukami and Perica (2008), Jonas [et al. \(2009\)](#), and Sturm et al., (2010) in that a
181 minimum of information is needed for the calculations; we intentionally avoid approaches like Meloysund et al.
182 (2007) and McCreight and Small (2014). This is because our interests are in converting h measurements to SWE
183 estimates in areas lacking weather instrumentation. The following sections introduce the numerous data sets that
184 were used in this study, outline the regression model adopted, and assess the performance of the model.

Deleted: and Magnusson

185 2 Methods

186

187 2.1 Data

188

189 2.1.1 Snow Depth and Snow Water Equivalent

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

192

193 2.1.1.1 USA NRCS Snow Telemetry and Soil Climate Analysis Networks

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

200

201 2.1.1.2 Canada (British Columbia) Snow Survey Data

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

209

210 2.1.1.3 Northeast USA Data

211 [In addition to the data from the SCAN sites, snow data for this project from the northeast US come from two](#)
212 networks and three research sites (Figure 2b). The Maine Cooperative Snow Survey (MCSS, 2018) network
213 includes h and SWE data collected by the Maine Geological Survey, the United States Geological Survey, and
214 numerous private contributors and contractors. MCSS snow data are collected using the Standard Federal or

Deleted: S

217 Adirondack snow sampling tubes typically on a weekly to bi-weekly schedule throughout the winter and spring,
218 1951-present. The New York Snow Survey network data were obtained from the National Oceanic and Atmospheric
219 Administration's Northeast Regional Climate Center at Cornell University (NYSS, 2018). Similar to the MCSS,
220 NYSS data are collected using Standard Federal or Adirondack snow sampling tubes on weekly to bi-weekly
221 schedules, 1938-present.

222
223 The Sleepers River, Vermont Research Watershed in Danville, Vermont (Shanley and Chalmers, 1999) is a USGS
224 site that includes 15 stations with long-term weekly records of h and SWE collected using Adirondack snow tubes.
225 Most of the periods of record are 1981-present, with a few stations going back to the 1960s. The sites include
226 topographically flat openings in conifer stands, old fields with shrub and grass, a hayfield, a pasture, and openings in
227 mixed softwood-hardwood forests. The Hubbard Brook Experiment Forest (Campbell et al., 2010) has collected
228 weekly snow observations at the Station 2 rain gauge site, 1959-present. Measurement protocol collects ten samples
229 2 m apart along a 20 m transect in a hardwood forest opening about $\frac{1}{4}$ hectare in size. At each sample location along
230 the transect, h and SWE are measured using a Mt. Rose snow tube and the ten samples are averaged for each
231 transect. Finally, the Thompson Farm Research site includes a mixed hardwood forest site and an open pasture site
232 (Burakowski et al. 2013; Burakowski et al. 2015). Daily (from 2011-2018), at each site, a snow core is extracted
233 with an aluminum tube and weighed (tube + snow) using a digital hanging scale. The net weight of the snow is
234 combined with the depth and the tube diameter to determine ρ_b , similar to a Federal or Adirondack sampler.

235

236 2.1.1.4 Chugach Mountains (Alaska) Data

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

244

245 2.1.1.5 Data Pre-Processing

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

- Deleted: *SWE*
- Formatted: Font:Italic
- Deleted: using a Federal sampler
- Deleted: of those 71 sites
- Deleted: took 3 SWE and h measurements within $1 m^2$ and averaged the result. Additionally, we used an avalanche probe to
- Deleted: in snow densities
- Deleted: Outlier Detection and Removal

262

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

272

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

279

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

289

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

Dataset Name	Dataset Type	Number of retained stations	Number <u>and percentage</u> of retained data points	Precision (<i>h</i> / SWE)
NRCS SNOTEL	Snow pillow (SWE), ultrasonic (<i>h</i>)	791	1,900,000 (99.3%)	(0.5 in / 0.1 in)
NRCS SCAN	Snow pillow (SWE), ultrasonic (<i>h</i>)	5	7094 (97.8%)	(0.5 in / 0.1 in)

Deleted: uncleaned

Deleted: cleaning process

Formatted: Font:Italic

Formatted: Font:Italic

Formatted Table

Deleted: NRCS SCAN

Deleted: (0.5 in / 0.1 in)

Formatted: Font:Bold

Deleted: 7094

British Columbia Snow Survey	Snow pillow (SWE), ultrasonic (<i>h</i>)	31	61,000 (97.5%)	(1 cm / 1 mm)
Maine Geological Survey	Adirondack or Federal sampler (SWE and <i>h</i>)	431	28,000 (99.3%)	(0.5 in / 0.5 in)
Hubbard Brook (Station 2), NH	Mount Rose sampler (SWE and <i>h</i>)	1	704 (99.4%)	(0.1 in / 0.1 in)
Thompson Farm, NH	Snow core (SWE and <i>h</i>)	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 <i>h</i>)	523	44,614 (98.2%)	(0.5 in / 0.5 in)
Chugach Mountains, AK	Avalanche probe (<i>h</i>)	71	71 (100%)	(1 cm)

Formatted: Font:Bold

Formatted: Font:Bold

Formatted Table

Deleted: Federal sampler (SWE and *h*) and

Deleted: a

Deleted: 0.5 in / 0.5 in) for sampler; 1 cm for probe

297

298

2.1.2 Climatological Variables

299

30-year climate normals at 800 m (nominal) resolution for CONUS and for the period 1981-2010 were obtained from the 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.

307

308

2.2 Regression Model

309

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. Additional validation was done with the northeast USA datasets ([SCAN snow pillow and various snow coring datasets](#)) which were completely left out of the model building process.

315

316

2.2.1 One-Equation Model

317

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.,

320

321

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

322

323

This equation can be log-transformed into

327

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

329

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

333

334 2.2.2 Two-Equation Model

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

340

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

342

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

344

345 where DOY is the number of days from the start of the water-year (October 1 is the origin), and DOY^* is the critical
346 or dividing day-of-water-year separating the two phases. Put another way, DOY^* is the day of peak SWE.

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

356

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

359

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

361

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

362

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

365

366 2.2.3 Two-Equation Model with Climate Parameters

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

373

$$374 (6) \quad SWE = f(h, z, MAP, \bar{T}_{Jmin}, \bar{T}_{Jmean}, \bar{T}_{Jmax}, \bar{T}_{Fmin}, \bar{T}_{Fmean}, \bar{T}_{Fmax}, \bar{T}_{Mmin}, \bar{T}_{Mmean}, \bar{T}_{Mmax}, \bar{T}_{Amin}, \bar{T}_{Amean}, \bar{T}_{Amax})$$

375

376 where z is the elevation (m), MAP is the mean annual precipitation (mm) and the temperatures ($^{\circ}C$) represent the
 377 mean of minimum, mean, and maximum daily values for the months January through April (J, F, M, A). For
 378 example, \bar{T}_{Jmin} is the climatological normal of the average of the daily minimum temperatures observed in January.

379 In the stepwise regression, explanatory variables were accepted if they improved the adjusted R^2 value by 0.001.

380 The result of the regression yielded

381

$$382 (7) \quad SWE_{acc} = Ah^{a_1} MAP^{a_2} (\bar{T}_{Fmean} + 30)^{a_3}; \quad DOY < DOY^*$$

383

$$384 (8) \quad SWE_{abl} = Bh^{b_1} MAP^{b_2} (\bar{T}_{Fmean} + 30)^{b_3}; \quad DOY \geq DOY^*$$

385

386 or, in log-transformed format,

387

$$388 (9) \quad \log_{10}(SWE_{acc}) = \log_{10}(A) + a_1 \log_{10}(h) +$$

$$389 \quad a_2 \log_{10}(MAP) + a_3 \log_{10}(\bar{T}_{Fmean} + 30); \quad DOY < DOY^*$$

390

$$391 (10) \quad \log_{10}(SWE_{abl}) = \log_{10}(B) + b_1 \log_{10}(h) +$$

$$392 \quad b_2 \log_{10}(MAP) + b_3 \log_{10}(\bar{T}_{Fmean} + 30); \quad DOY \geq DOY^*$$

393

394 indicating that only snow depth, mean annual precipitation and mean February temperature were relevant. Manual
 395 tests of model construction with other variables included confirmed that Eqns. (7-8) yielded the best results. In the
 396 above equations, note that an offset is added to the temperature in order to avoid taking the log of a negative
 397 number. These two SWE estimates for the individual (*acc* and *abl*) phases of the snowpack are then blended with
 398 Eqn. (5) to produce a single equation for SWE spanning the entire water year. The obtained regression coefficients

400 were $(A, a_1, a_2, a_3) = (0.0128, 1.070, 0.132, 0.506)$ and $(B, b_1, b_2, b_3) = (0.0271, 1.038, 0.201, 0.310)$. The
 401 physical interpretation of these coefficients is straightforward. *If a_1 and b_1 were equal to unity, then the density,*
 402 *given by (SWE/h) , would be a constant at a given location. Since they are greater than unity, they capture the effect*
 403 *that snow density increases as depth increases. Turning to the coefficients on the climate variables, both a_2 and b_2*
 404 *are greater than zero. So, for two locations with equal depth, equal temperature characteristics, but different*
 405 *precipitation characteristics, the regression model predicts that the wetter location (larger MAP) will have a greater*
 406 *density. Finally, regarding temperature, both a_3 and b_3 are greater than zero. Therefore, for two locations with equal*
 407 *depth, equal precipitation characteristics, but different temperature characteristics, the regression model predicts that*
 408 *the warmer location (larger \bar{T}_{mean}) will have a greater density. These trends are similar in concept to Sturm et al.*
 (2010), whose snow classes (based on climate classes) indicate which snow will densify more rapidly.

Formatted: Font:Italic

Formatted: Font:Italic

Deleted: The fact that the coefficients on depth are greater than unity indicates that the density (SWE/h) increases as the snowpack depth increases. The positive coefficients associated with MAP and \bar{T}_{mean} indicate that snow densities should be higher in warmer, wetter locations than in colder, drier locations.
 Formatted: Font:Italic

409 **3 Results**

410 A comparison of the three regression models (one-equation model, Eq. (2); two-equation model, Eqs. (3-5); multi-
 411 variable two-equation model, Eqs. (5, 7-8)) is provided in Figure 6. The left column shows scatter plots of modeled
 412 SWE to observed SWE for the validation data set with the 1:1 line shown in black. The right column shows
 413 histograms of the model residuals. The vertical lines in the right column show the mean error, or model bias.

414 Visually, it is clear that the one-equation model performs relatively poorly with a large negative bias. This large
 415 negative bias is partially overcome by the two-equation model (middle row, Figure 6). The cloud of points is closer
 416 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
 417 see that the multi-variable two-equation model yields the best result by far. The residuals are now evenly distributed
 418 with a negligible bias. Several metrics of performance for the three models, including R^2 (Pearson coefficient), bias,
 419 and root-mean-square-error (RMSE), are provided in Table 2. *Figure 7 shows the distribution of model residuals for*
 420 *the multi-variable two-equation model as a function of DOY.*

Deleted: This is easily explained. The SNOTEL measurements are uniformly spaced in time (daily). Given that the accumulation season is much longer than the ablation season (Figure 1), there are many more data points that are representative of the accumulation season. The model fit is weighted towards these points, which leads to large negative residuals in the ablation season.

422 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	0.972	-0.5	67

423 *It is useful to also consider the model errors in a non-dimensional way.* Therefore, an RMSE was computed at each
 424 station location and normalized by the *mean annual maximum SWE (SWE_{max})* at that location. Figure 8 shows the
 425 probability density function of these normalized errors. The average RMSE is approximately *11%* of SWE_{max} , with
 426 most falling into the range of *5-25%*. The spatial distribution of these normalized errors is shown in Figure 9. For
 427 the SNOTEL stations, *it appears there is a slight regional trend, in terms of stations in continental climates (northern*
 428 *Rockies) having smaller relative errors than stations in maritime climates (Cascades).* The British Columbia stations
 429 also show higher *relative errors.*

Deleted: Model errors will have varying impact on the local snow regime depending on the total precipitation in a specific region.

Deleted: PRISM estimate of MAP

Deleted: 7

Deleted: 5

Deleted: MAP

Deleted: 2-8

Deleted: 8

Deleted: there are no clear regional patterns governing the normalized errors, with the possible exception of the Sierra Nevada, where the errors are consistently higher than elsewhere

Deleted: overall

459 **3.1 Results for Snow Classes**

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

464 (11)
$$\rho_b = (\rho_{max} - \rho_0)[1 - e^{(-k_1h - k_2DOY)}] + \rho_0$$

466 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,
 467 and DOY in this case begins on January 1. This means that their DOY for October 1 is -92. The coefficients vary
 468 with snow class and the values determined by Sturm et al. (2010) are shown in Table 3.

470 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

471 To make a comparison, the snow class for each [SNOTEL and British Columbia snow survey \(Rows 1 and 3 of Table](#)
 472 [1\)](#) site was determined using a 1-km snow class grid (Sturm et al., 2010). [The aggregated dataset from these stations](#)
 473 [was made up of 27% Alpine, 14% Maritime, 10% Prairie, 11% Tundra, and 38% Taiga data points.](#) Equation (11)
 474 was [then](#) used to estimate snow density (and then SWE) for every point in the validation dataset described in Section
 475 2.2. Figure 10 compares the SWE estimates from the Sturm model and from the present multi-variable, two-equation
 476 model (Equations 5, 7-8). The upper left panel of Figure 10 shows all of the data, and the remaining panels show the
 477 results for each snow class. In all cases, the current model provides better estimates. Plots of the residuals by snow
 478 class are provided in Figure 11, giving an indication of the bias of each model for each snow class. Summaries of the
 479 model performance, broken out by snow class, are given in Table 4.

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

Model	Sturm et al. (2010)			Multi-variable two-equation model		
	R ²	Bias (mm)	RMSE (mm)	R ²	Bias (mm)	RMSE (mm)
All Data	0.928	-29.2	111	0.972	-0.5	67
Alpine	0.973	10.1	55	0.971	-0.3	55
Maritime	0.968	-16.8	109	0.970	-4.5	105
Prairie	0.967	18.7	56	0.965	-0.2	51
Tundra	0.956	-10.5	82	0.969	-6.1	67
Taiga	0.943	-80.0	151	0.971	2.4	62

483
 484 **3.2 Comparison to Pistocchi (2016)**

Deleted: SNOTEL

Deleted: (including CONUS, AK, and BC)

Deleted: and

Deleted: 9

Deleted: 9

Deleted: 0

Formatted: Font:Bold

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

493
494 (12) $\rho_b = \rho_0 + K(DOY + 61)$,
495

496 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
497 November 1. Application of this model to the validation dataset yields a bias of 55 mm and an RMSE of 94 mm.
498 These results are comparable to the Sturm et al. (2010) model, with a larger bias but smaller RMSE.

499
500 **3.3 Results for Northeast USA**

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

507
508 Figure 12 graphically summarizes the datasets and the performance of the multi-variable two-equation model of the
509 current study. The RMSE values are comparable to those found for the western stations, but, given the
510 comparatively thinner snowpacks in the northeast, represent a larger relative error (Table 5). The bias of the model
511 is consistently positive, in contrast to the western stations where the bias was negligible.

512
513 Table 5: Performance metrics for the multi-variable two-equation model applied to various northeastern USA
514 datasets.

Dataset Name	R ²	Bias (mm)	RMSE (mm)
Maine Geological Survey, ME	0.91	8.9	33.3
Hubbard Brook (Station 2), NH	0.63	18.9	64.2
Thompson Farm, NH	0.85	7.1	21.6
NRCS SCAN	0.87	-1.8	38.7
Sleepers River, VT	0.93	14.0	29.7
New York Snow Survey	0.93	13.8	31.2

515
516 **4 Discussion**

517 The results presented in this study show that the regression equation described by equations (5, 7-8) is an
518 improvement (lower bias and RMSE) over other widely used bulk density equations. The key advantage is that the
519 present method regresses in relevant physical parameters directly, rather than using discrete bins (for snow class,
520 elevation, month of year, etc.), each with its own set of model coefficients. The comparison (Figs. 10-11; Table 4) to
521 the model of Sturm et al. (2010) reveals a peculiar behavior of that model for the Taiga snow class, with a large

- Formatted: Font:Not Bold
- Formatted: Font:Not Bold
- Formatted: Font:Not Bold, Superscript
- Formatted: Font:Not Bold
- Formatted: Font:Not Bold, Italic
- Formatted: Font:Not Bold
- Formatted: Font:Not Bold, Superscript
- Formatted: Font:Not Bold
- Formatted: Font:12 pt, Not Bold, Font color: Auto
- Formatted: Normal, Line spacing: single
- Deleted: 2

Deleted: 1

Deleted: 3.3 Results for Chugach Mountains ... [2]

Deleted: 9-10

527 negative bias in the Sturm estimates. Inspection of the coefficients provided for that class (Table 3) shows that the
528 model simply predicts that $\rho_b = \rho_{max} = 0.217$ for all conditions.

529
530 When our multi-variable two-equation model, developed solely from western North American data, is applied to
531 northeast USA locations, it produces SWE estimates with smaller RSME values and larger biases than the western
532 stations. When comparing the SWE- h curves of the SNOTEL data (Figure 4b) to those of the east coast data sets
533 (left column; Figure 12), it is clear that the northeast data generally have more scatter. This is confirmed by
534 computing the correlation coefficients between SWE and h for each dataset. It is unclear if this disparity in
535 correlation is related to measurement methodology or is instead a 'signal to noise' issue. Comparing Figures 4 and
536 12 shows the considerable difference in snowpack depth between the western and northeastern data sets. When the
537 western dataset is filtered to include only measurement pairs where $h < 1.5$ m, the correlation coefficient is reduced
538 to a value consistent with the northeast datasets. This suggests that the performance of the current (or other)
539 regression model is not as good at shallow snowpack depths. This is also suggested upon examination of the time
540 series of observed $\rho_b = SWE/h$ for a given season at a snow pillow site. Very early in the season, when the depths
541 are small, the density curve has a lot of variability. Later in the season, when depths are greater, the density curve
542 becomes much smoother. Very late in the season, when depths are low again, the density curve becomes highly
543 variable again.

544
545 Measurement precision and accuracy affect the construction and use of a regression model. Upon inspection of the
546 snow pillow data, it was observed that the precision of the depth measurements was approximately 25 mm and that
547 of the SWE measurements was approximately 2.5 mm. To test the sensitivity of the model coefficients to the
548 measurement precision, the depth values in the training dataset were randomly perturbed by +/- 25 mm and the SWE
549 values were randomly perturbed by +/- 2.5 mm and the regression coefficients were recomputed. This process was
550 repeated numerous times and the mean values of the perturbed coefficients were obtained. These adjusted
551 coefficients were then used to recompute the SWE values for the validation data set and the bias and RMSE were
552 found to be -10.5 mm and 72.7 mm. This represents a roughly 10% increase in RMSE, but a considerable increase in
553 bias magnitude (see Table 4 for the original values). This sensitivity of the regression analysis to measurement
554 precision underscores the need to have high-precision measurements for the training data set. Regarding accuracy,
555 random and systematic errors in the paired SWE - h data used to construct the regression model will lead to
556 uncertainties in SWE values predicted by the model. As noted in the introduction, snow pillow errors in SWE
557 estimates do not follow a simple pattern. Additionally, they are complicated by the fact that the errors are often
558 computed by comparing snow pillow data to coring data, which itself is subject to error. Lacking quantitative
559 information on the distribution of snow pillow errors, we are unable to quantify the uncertainty in the SWE
560 estimates.

561
562 Another important consideration has to do with the uncertainty of depth measurements that the model is applied to.
563 For context, one application of this study is to crowd-sourced, opportunistic snow depth measurements from

Deleted: 1

Deleted: 1

Deleted: is very noisy

Deleted: *

Deleted: When applied to the Chugach coring measurements, the model appears to perform well. The higher values of bias and RMSE (when compared to Tables 4 and 5) are higher in part since the measurements (and model estimates) of SWE are only at times of larger snow depth. 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%. Propagating this uncertainty through the regression equations yields a slightly higher (16%) uncertainty in the SWE estimates. Clearly, this is a function of surface roughness of the underlying terrain. Backcountry areas characterized by fields and meadows are likely to have smaller coefficients of variation for ensembles of depth measurements over a small radius. As a result, SWE estimated from a single depth measurement should be comparatively accurate. Areas of high surface roughness, characterized by crags, rocks and fallen logs will have large coefficients of variation and larger numbers of depth measurements should be collected and averaged to provide the best possible SWE estimate. *

Deleted: also affects

Deleted: found to be $(A, a_1, a_2, a_3) = (0.0188, 0.9737, 0.2034, 0.4301)$ and $(B, b_1, b_2, b_3) = (0.0386, 0.9535, 0.2745, 0.2184)$

Formatted: Font:Not Italic

Moved up [1]: 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; Johnson et al., 2015) note a more complex situation with SWE under-reported at times, but over-reported at other times.

Deleted: , however

Deleted: It also raises the interesting question of whether or not future resources should be directed towards expanding networks (greater spatial coverage) of current technologies or towards refining instrumentation (better accuracy) at currently instrumented stations.

608 programs like the Community Snow Observations (CSO; Hill et al., 2018) project. In the CSO program,
609 backcountry recreational users submit depth measurements, typically taken with an avalanche probe, using a
610 smartphone in the field. The measurements are then converted to SWE estimates which are assimilated into
611 snowpack models. These depth measurements are ‘any time, any place’ in contrast to repeated measurements from
612 the same location, like snow pillows or snow courses. Most avalanche probes have cm-scale graduated markings, so
613 measurement precision is not a major issue. A larger problem is the considerable variability in snowpack depth that
614 can exist over short (meter scale) distances. The variability of the Chugach avalanche probe measurements was
615 assessed by taking the standard deviation of 8 h measurements per site. The average of this standard deviation over
616 the sites was 22 cm and the average coefficient of variation (standard deviation normalized by the mean) over the
617 sites was 15%. This variability is a function of the surface roughness of the underlying terrain, and also a function of
618 wind redistribution of snow. Propagating this uncertainty through the regression equations yields a slightly higher
619 (16%) uncertainty in the SWE estimates. CSO participants can do three things to ensure that their recorded depth
620 measurements are as representative as possible. First, avoid measurements in areas of significant wind scour or
621 deposition. Second, avoid measurements in terrain likely to have significant surface roughness (rocks, fallen logs,
622 etc.). Third, take several measurements and average them.

623
624 Expansion of CSO measurements in areas lacking SWE measurements can increase our understanding of the
625 extreme spatial variability in snow distribution and the inherent uncertainties associated with modeling SWE in
626 these regions. It could also prove useful for estimating watershed-scale SWE in regions like the northeastern USA,
627 which is currently limited to five automated SCAN sites with historical SWE measurements for only the past two
628 decades. Additionally, historical snow depth measurements are more widely available in the Global Historical
629 Climatology Network (GHCN-Daily; Menne et al. 2012), with several records extending back to the late 1800s.
630 While many of the GHCN stations are confined to lower elevations with shallower snow depths, the broader
631 network of quality-controlled snow depth data paired with daily GHCN temperature and precipitation measurements
632 could potentially be used to reconstruct SWE in the eastern US given additional model development and refinement.

633 5 Conclusions

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

639
640 We developed this model with data from paired SWE-*h* measurements from the western United States and British
641 Columbia. The model was tested against entirely independent data (primarily snow course; some snow pillow) from
642 the northeastern United States and was found to perform well, albeit with larger biases and root-mean-squared-
643 errors. The model was tested against other well-known regression equations and was found to perform better.
644

Deleted: Recalling the Chugach discussion above, even in flat areas, with a smooth snow surface (away from major drifting or wind scour), terrain features such as rocks, logs, and vegetation can produce large variations in probe measurements.

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

654 **6 Acknowledgements**

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

658 **7 Data Access**

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

- 660 1. NRCS Snow Telemetry: <https://www.wcc.nrcs.usda.gov/snow/SNOTEL-wedata.html>
- 661 2. NRCS Soil Climate Analysis Network: <https://www.wcc.nrcs.usda.gov/scan/>
- 662 3. British Columbia Automated Snow Weather Stations:
663 [https://www2.gov.bc.ca/gov/content/environment/air-land-water/water/water-science-data/water-data-
665 tools/snow-survey-data/automated-snow-weather-station-data](https://www2.gov.bc.ca/gov/content/environment/air-land-water/water/water-science-data/water-data-
664 tools/snow-survey-data/automated-snow-weather-station-data)
- 666 4. Maine Cooperative Snow Survey: <https://mgs-maine.opendata.arcgis.com/datasets/maine-snow-survey-data>
- 667 5. New York Snow Survey: <http://www.nrcc.cornell.edu/regional/snowsurvey/snowsury.html>
- 668 6. Sleepers River Research Watershed. Snow data not available online; request data from contact at:
669 <https://nh.water.usgs.gov/project/sleepers/index.htm>
- 670 7. Hubbard Brook Experimental Forest: <https://hubbardbrook.org/d/hubbard-brook-data-catalog>
- 671 8. CONUS PRISM Data: <http://www.prism.oregonstate.edu/>
- 672 9. British Columbia PRISM Data: <http://climatebcdata.climatewna.com/>
- 673 10. Alaska PRISM Data: <https://irma.nps.gov/Portal/>

674 A Matlab function for calculating SWE based on the results in this paper has been made publicly available at Github
675 (URL provided upon paper acceptance).

676 **References**

677

678 Alford, D.: Density variations in alpine snow, *J. Glaciol.*, 6(46), 495-503,

679 <https://doi.org/10.3189/S0022143000019717>, 1967.

680

681 Avanzi, F., De Michele, C., and Ghezzi, A.: On the performances of empirical regressions for the estimation of bulk

682 snow density, *Geogr. Fis. Dinam. Quat.*, 38, 105-112, doi:10.4461/GFDQ.2015.38.10, 2015.

683

684 Beaumont, R.: Mt. Hood pressure pillow snow gage, *J. Appl. Meteorol.*, 4, 626-631, <https://doi.org/10.1175/1520->

685 0450(1965)004<0626:MHPPSG>2.0.CO;2 1965.

686

687 Beaumont, R., and Work, R.: Snow sampling results from three samplers, *Hydrol. Sci. J.*, 8(4), 74-78,

688 <https://doi.org/10.1080/02626666309493359>, 1963.

689

690 Burakowski, E.A., Ollinger, S., Lepine, L., Schaaf, C.B., Wang, Z., Dobb, J.E., Hollinger, D.Y., Kim, J.-H., Erb, A.,

691 and Martin, M.E.: Spatial scaling of reflectance and surface albedo over a mixed-use, temperate forest landscape

692 during snow-covered periods, *Remote Sens. Environ.*, 158, 465-477, <https://doi.org/10.1016/j.rse.2014.11.023>,

693 2015.

694

695 Burakowski, E.A., Wake, C.P., Stampone, M., and Dobb, J.: Putting the Capital 'A' in CoCoRAHS: An

696 Experimental Program to Measure Albedo using the Community Collaborative Rain Hail and Snow (CoCoRaHS)

697 Network, *Hydrol. Process.*, 27(21), 3024-3034, <https://doi.org/10.1002/hyp.9825>, 2013.

698

699 Campbell, J., Ollinger, S., Flerchinger, G., Wicklein, H., Hayhoe, K., and Bailey, A.: Past and projected future

700 changes in snowpack and soil frost at the Hubbard Brook Experimental Forest, New Hampshire, USA, *Hydrol.*

701 *Process.*, 24, 2465-2480, <https://doi.org/10.1002/hyp.7666>, 2010.

702

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

704 1933.

705

706 [Church, J.E., and Marr, J.C.: Further improvement of snow-survey apparatus, *Transactions of the American*](#)

707 [*Geophysical Union*, 18\(2\), 607-617, 10.1029/TR018i002p00607, 1937.](#)

708

709 Daly, C., Neilson, R., and Phillips, D.: A statistical-topographic model for mapping climatological precipitation over

710 mountainous terrain, *J. Appl. Meteorol.*, 33, 140-158, <https://doi.org/10.1175/1520->

711 0450(1994)033<0140:ASTMFM>2.0.CO;2, 1994.

712

713 Deeb, E., Marshall, H.-P., Forster, R., Jones, C., Hiemstra, C., and Siqueira, P.: Supporting NASA SnowEx remote
 714 sensing strategies and requirements for L-band interferometric snow depth and snow water equivalent estimation,
 715 Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, 1395-
 716 1396, DOI: 10.1109/IGARSS.2017.8127224, 2017.

717

718 Dixon, D., and Boon, S.: Comparison of the SnowHydro sampler with existing snow tube designs, Hydrol. Process.,
 719 26(17), 2555-2562, <https://doi.org/10.1002/hyp.9317>, 2012.

720

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

723

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

726

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

729

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

732

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

735

736 Goodison, B., Glynn, J., Harvey, K., and Slater, J.: Snow Surveying in Canada: A Perspective, Can. Water Resour.
 737 J., 12(2), 27-42, <https://doi.org/10.4296/cwrij1202027>, 1987.

738

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

741

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

744

745 [Johnson, J., and Marks, D.: The detection and correction of snow water equivalent pressure sensor errors, Hydrol.
 746 Proc., <https://doi.org/10.1002/hyp.5795>, 2004.](#)

747

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

Deleted: ”

Formatted: Superscript

Deleted:

Formatted: Superscript

Formatted: Default Paragraph Font, Font:12 pt, Not Bold

752
753 Jonas, T., [Marty, C.](#), and Magnusson, M.: Estimating the snow water equivalent from snow depth measurements, *J.*
754 *Hydrol.*, 378, 161-167, <https://doi.org/10.1016/j.jhydrol.2009.09.021>, 2009.
755
756 Kang, D.H., and Barros, A.P.: Full-system testing in laboratory conditions of an L-band snow sensor system for in
757 situ monitoring of snow-water content, *IEEE Trans. Geosci. Remote Sens.*, 49(3), 908-919,
758 10.1109/TGRS.2010.2072786, 2011.
759
760 Leinss, S., Wiesmann, A., Lemmetyinen, J., and Hajnsek, I.: Snow water equivalent measured by differential
761 interferometry, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 8(8), 3773-3790, 10.1109/JSTARS.2015.2432031,
762 2015.
763
764 Leys, C., Klein, O., Dominicy, Y., and Ley, C.: Detecting multivariate outliers: use a robust variant of the
765 Mahalanobis distance, *J. Exp. Soc. Psychol.*, 74, 150-156, <https://doi.org/10.1016/j.jesp.2017.09.011>, 2018.
766
767 Liang, X., Lettermaier, D., Wood, E., and Burges, S.: A simple hydrologically based model of land surface water
768 and energy fluxes for general circulation models, *J. Geophys. Res. Atmos.*, 99(D7), 14,415-14,428,
769 <https://doi.org/10.1029/94JD00483>, 1994.
770
771 Liston, G., and Elder, K.: A distributed snow evolution modeling system (SnowModel), *J. Hydrometeorol.*, 7, 1259-
772 1276, <https://doi.org/10.1175/JHM548.1>, 2006.
773
774 Lundberg, A., Richardson-Naslund, C., and Andersson, C.: Snow density variations: consequences for ground
775 penetrating radar, *Hydrol. Process.*, 20, 1483-1495, <https://doi.org/10.1002/hyp.5944>, 2006.
776
777 Maine Geological Survey: Maine Cooperative Snow Survey Dataset,
778 https://www.maine.gov/dacf/mgs/hazards/snow_survey/, 2018.
779
780 [McKay, G., and Findlay, B., 1971: Variation of snow resources with climate and vegetation in Canada, Proceedings](#)
781 [of the 39th Western Snow Conference, Billings, MT, 17-26, 1971.](#)
782
783 De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D.: The Mahalanobis distance, *Chemometr. Intell. Lab. Syst.*,
784 50(1), 1-18, [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7), 2000.
785
786 McCreight, J., and Small, E.: Modeling bulk density and snow water equivalent using daily snow depth
787 observations, *The Cryosphere*, 8, 521-536, <https://doi.org/10.5194/tc-8-521-2014>, 2014.
788

Formatted: Superscript

789 Meloyund, V., Leira, B., Hoiseth, K., and Liso, K.: Predicting snow density using meteorological data, *Meteorol.*
790 *Appl.*, 14, 413-423, <https://doi.org/10.1002/met.40>, 2007.

791

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

795

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

799

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

802

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

805

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

807

808 Pagano, T., Garen, D., Perkins, T., and Pasteris, P.: Daily updating of operational statistical seasonal water supply
809 forecasts for the western U.S., *J. Am. Water Resour. Assoc.*, 45(3), 767-778, [https://doi.org/10.1111/j.1752-](https://doi.org/10.1111/j.1752-1688.2009.00321.x)
810 [1688.2009.00321.x](https://doi.org/10.1111/j.1752-1688.2009.00321.x), 2009.

811

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

817

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

820

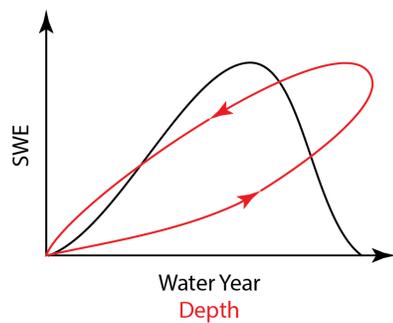
821 Rousseeuw, P.: Least Median of Squares Regression, *J. Am. Stat. Assoc.*, 79, 871-880, DOI:
822 10.1080/01621459.1984.10477105, 1984.

823

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

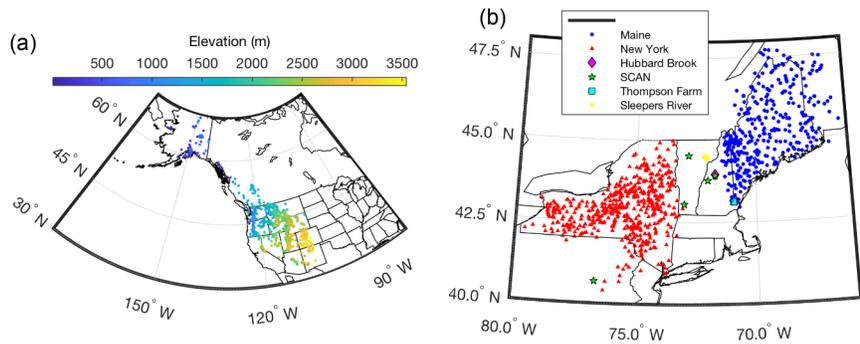
826
827 Schaefer, G., Cosh, M., and Jackson, T.: The USDA Natural Resources Conservation Service Soil Climate Analysis
828 Network (SCAN), *J. Atmos. Ocean. Technol.*, 24, 2073-2077, <https://doi.org/10.1175/2007JTECHA930.1>, 2007.
829
830 Serreze, M., Clark, M., Armstrong, R., McGinnis, D., and Pulwarty, R.: Characteristics of the western United States
831 snowpack from snowpack telemetry (SNOTEL) data, *Water Resour. Res.*, 35(7), 2145-2160,
832 <https://doi.org/10.1029/1999WR900090>, 1999.
833
834 Shanley, J., and Chalmers, A.: The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont, *Hydrol.*
835 *Process.*, 13(12-13), 1843-1857, [https://doi.org/10.1002/\(SICI\)1099-1085\(199909\)13:12/13<1843::AID-](https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1843::AID-)
836 [HYP879>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1843::AID-HYP879>3.0.CO;2-G), 1999.
837
838 Sokol, J., Pultz, T., and Walker, A.: Passive and active airborne microwave remote sensing of snow cover," *Int.*
839 *J. Remote. Sens.*, 24, 5327-5344, <https://doi.org/10.1080/0143116031000115076>, 2003.
840
841 Sturm, M., Holmgren, J., and Liston, G.: A seasonal snow cover classification system for local to global
842 applications, *J. Clim.*, 8, 1261-1283, [https://doi.org/10.1175/1520-0442\(1995\)008<1261:ASSCCS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1261:ASSCCS>2.0.CO;2), 1995.
843
844 Sturm, M., Taras, B., Liston, G.E., Derksen, C., Jonas, T., and Lea, J.: Estimating snow water equivalent using snow
845 depth data and climate classes, *J. Hydrometeorol.*, 11, 1380-1394, <https://doi.org/10.1175/2010JHM1202.1>, 2010.
846
847 [U.S. Army Corps of Engineers: Snow hydrology: Summary report of the snow investigations of the North Pacific](#)
848 [Division, 437pp., 1956.](#)
849
850 Vuyovich, C., Jacobs, J., and Daly, S.: Comparison of passive microwave and modeled estimates of total watershed
851 SWE in the continental United States, *Water Resour. Res.*, 50(11), 9088-9012,
852 <https://doi.org/10.1002/2013WR014734>, 2014.
853
854 Wang, T., Hamann, A., Spittlehouse, D.L., and Murdock, T.: ClimateWNA - High-Resolution Spatial Climate Data
855 for Western North America, *J. Appl. Meteorol. Climatol.*, 51, 16-29, <https://doi.org/10.1175/JAMC-D-11-043.1>,
856 2012.
857
858 Wigmosta, M.S., Vail, L., and Lettenmaier, D.: A distributed hydrology-vegetation model for complex
859 terrain, *Water Resour. Res.*, 30, 1665-1679, <https://doi.org/10.1029/94WR00436>, 1994

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

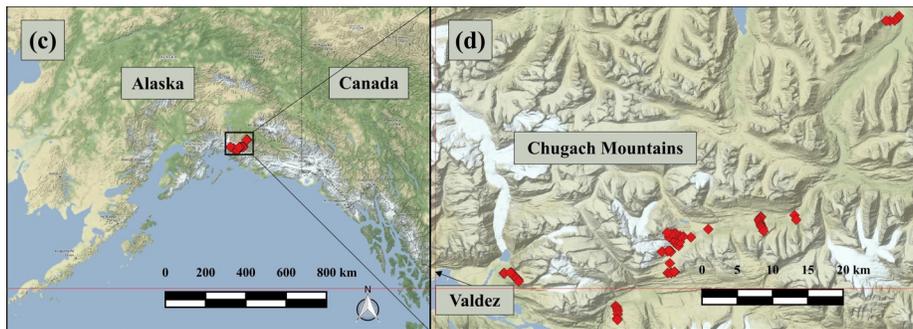


863

864 Figure 2: Distribution of measurement locations used in this study. (a) Western USA and Canada station locations,
865 with colors indicating station elevation in meters. (b) Northeast USA locations, with stations colored according to
866 data source. (c, d) Measurement sites in the Chugach Mountains, southcentral Alaska.
867



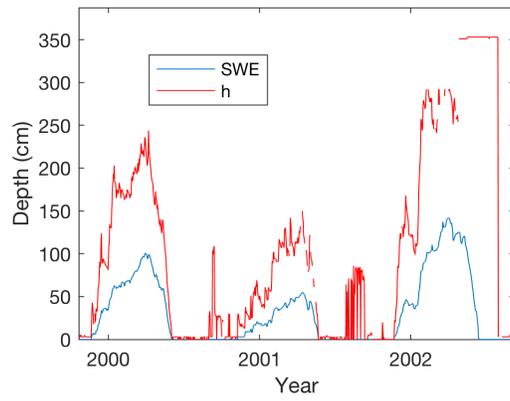
868



869

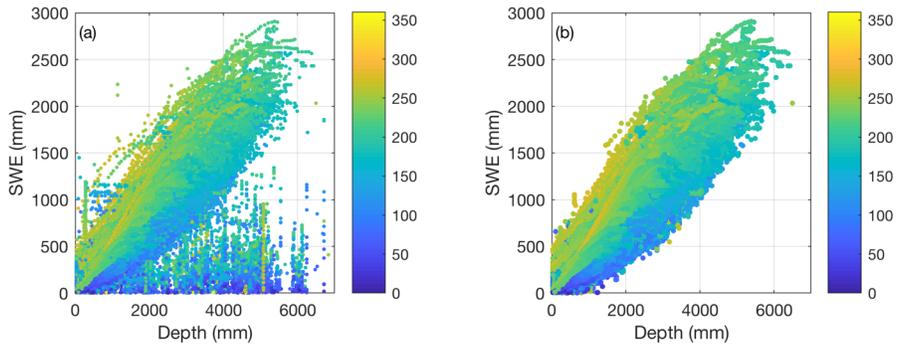
870

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



873

874 Figure 4: Scatter plot of SWE vs. h for the complete SNOTEL dataset before (a) and after (b) removing outliers.
875 Symbols are colored by 'day of water year' (DOY ; October 1 is the origin).
876

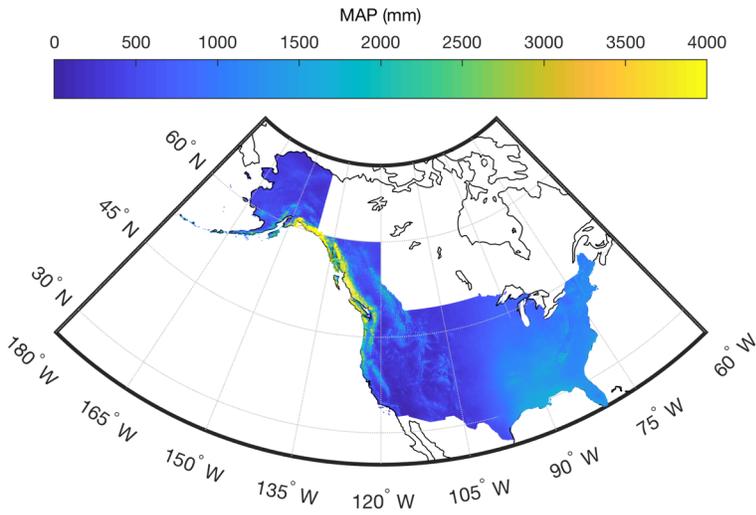


877

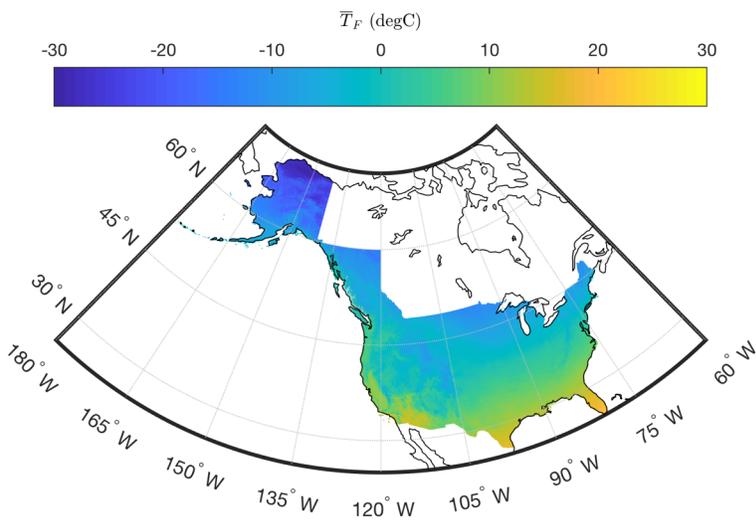
878 Figure 5: Gridded maps of mean annual precipitation (MAP) and mean February temperature (\bar{T}_{Fmean}) for the study
879 regions. Climate normals are from the PRISM data set (1981-2010 for CONUS and British Columbia; 1971-2000
880 for Alaska).

Formatted: Font:Italic

Deleted: (\bar{T}_F)



881



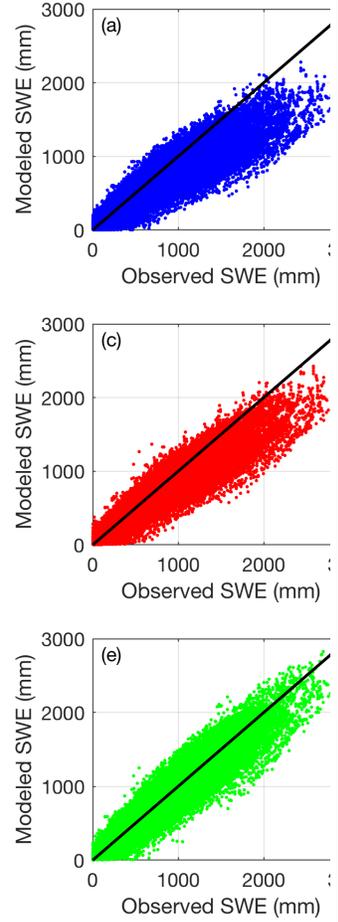
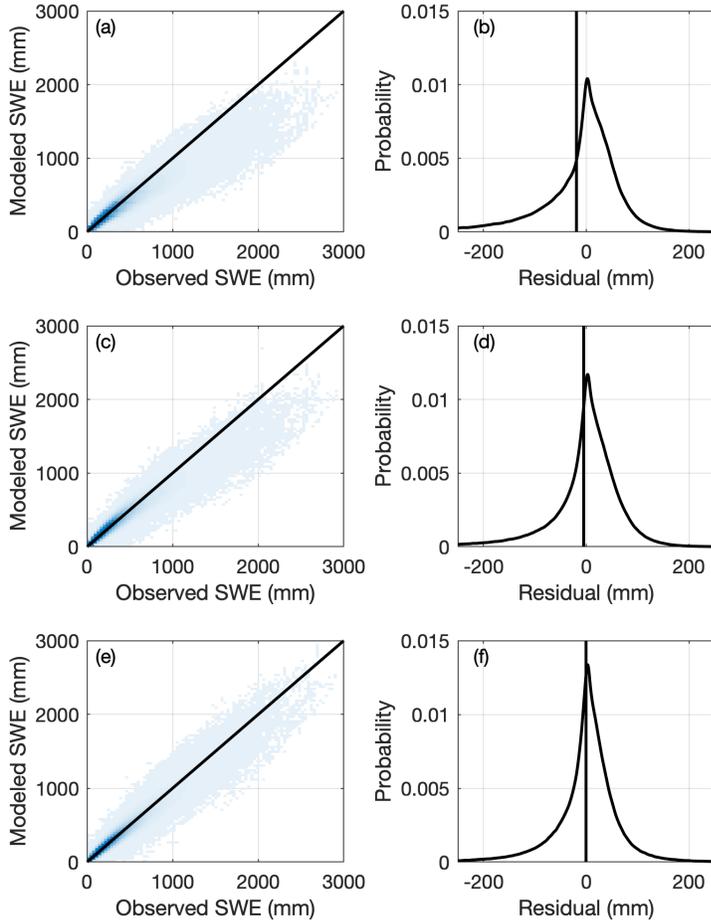
882
883

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

Deleted: Scatter plots

Deleted: Top

890



Deleted:

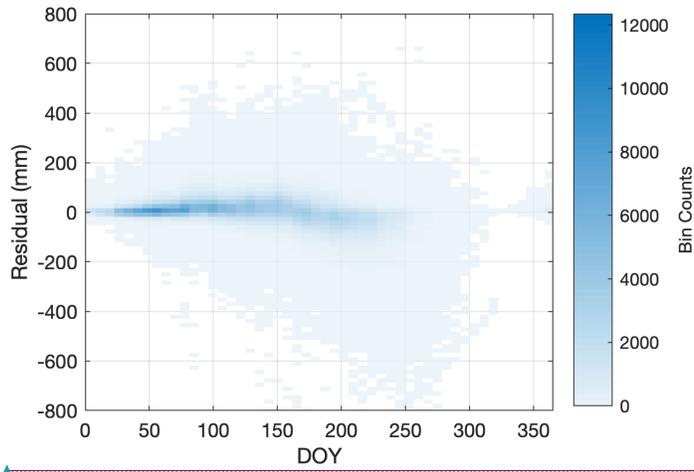
891

895
896

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

Formatted: Font:Italic

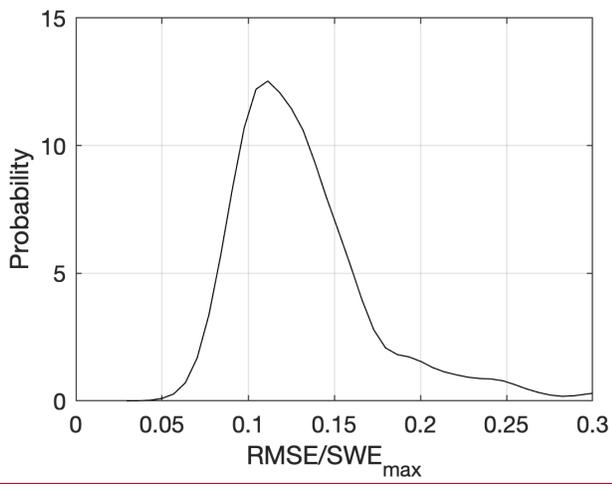
Formatted: Font:10 pt



897

898
899

Figure 8: Probability density function of snow pillow station root-mean-square error (RMSE) normalized by station mean annual maximum SWE.

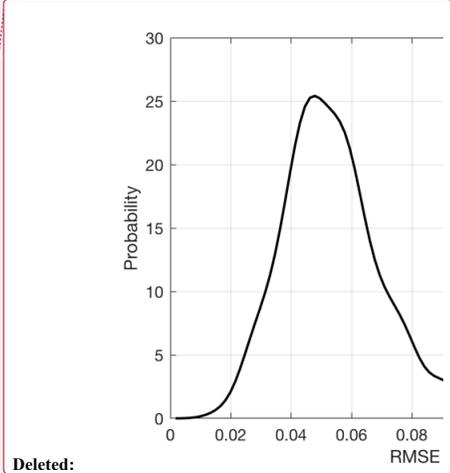


900

Deleted: 7

Deleted: precipitation

Deleted: (MAP)

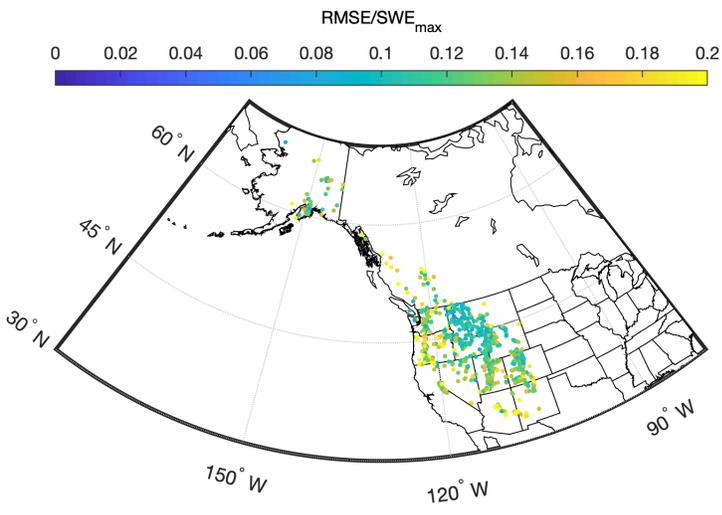


Deleted:

905
906
907

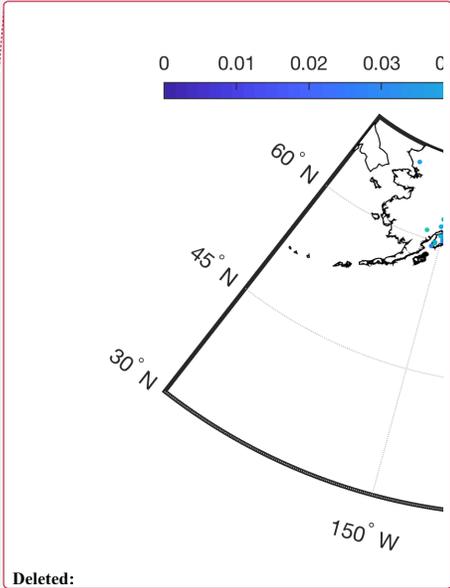
908

Figure 9. Spatial distribution of snow pillow station root-mean-square error (RMSE) normalized by station mean annual maximum SWE.



Deleted: 8

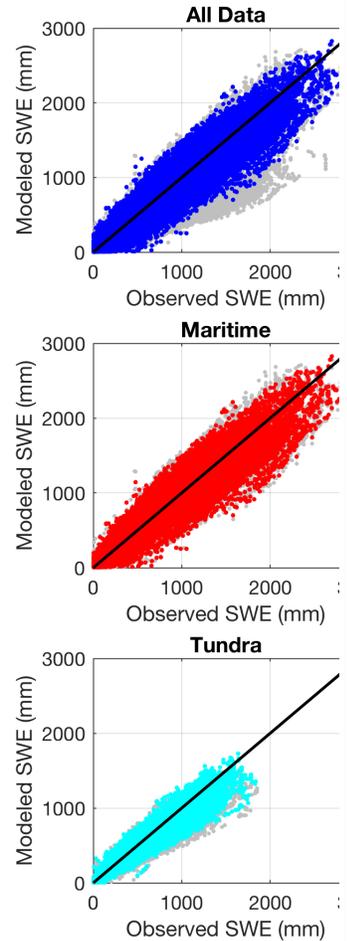
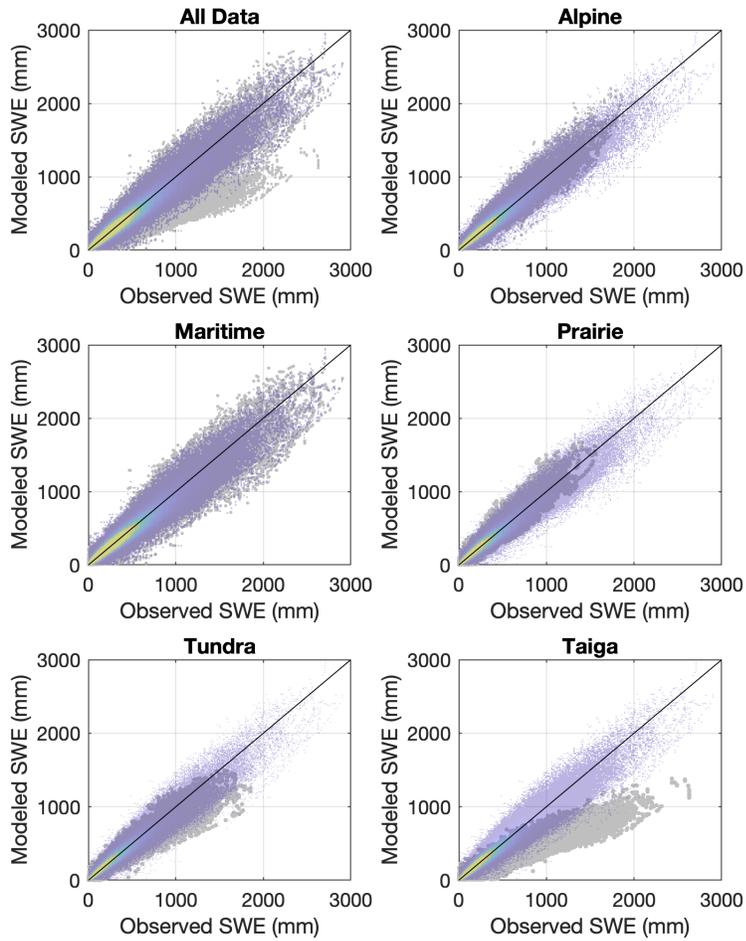
Deleted: of RMSE/MAP at snow pillow stations.



p12 Figure 10: Comparison of the multi-variable, two-equation model of the present study with the model of Sturm et al.
913 (2010). The subpanels show modeled SWE vs. observed SWE for all of the data binned together, as well as for the
914 data broken out by the snow classes identified by Sturm et al. (1995). The gray symbols show the Sturm result and
p15 the transparent heat maps (colored to show relative density of points) show the current result. The models are being
916 applied to the validation data set (50% of the aggregated snow pillow data for CONUS, AK, and BC).

Deleted: 9

Deleted: colored symbols (draped on top)

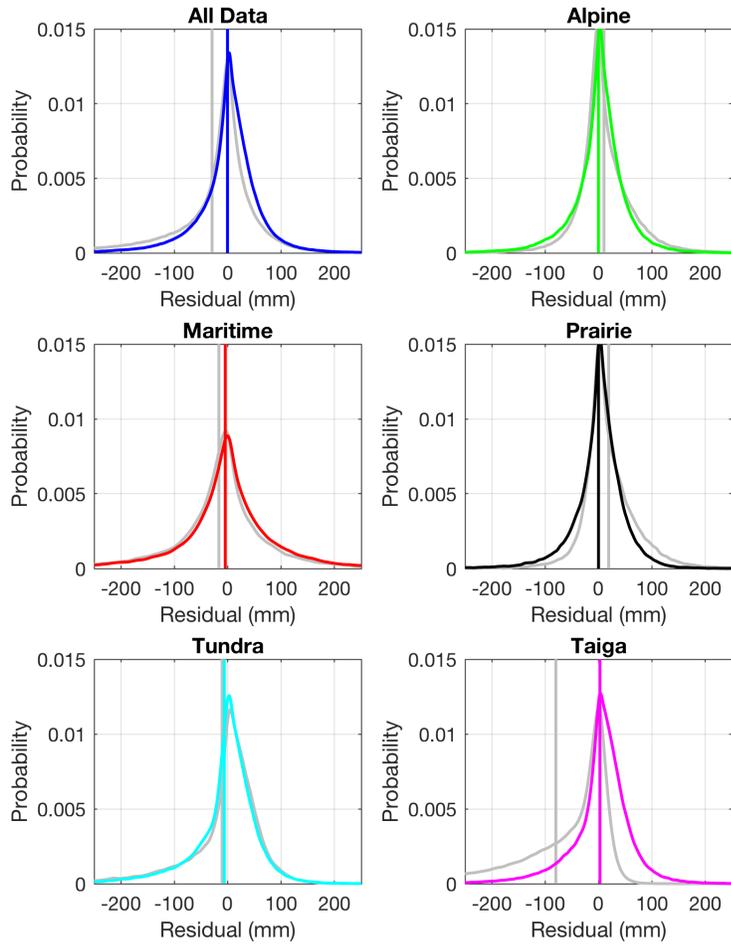


Deleted:

p17

921 Figure 1: Comparison of the multi-variable, two-equation model of the present study with the model of Sturm et al.
922 (2010). The subpanels show probability density functions of the residuals of the model fits for all of the data binned
923 together, as well as for the data broken out by the snow classes identified by Sturm et al. (1995). The gray lines
924 show the Sturm result and the colored lines show the current result. The vertical lines show the mean error, or the
925 model bias, for both the Sturm and the current result. The models are being applied to the validation data set (50% of
926 the aggregated snow pillow data for CONUS, AK, and BC).

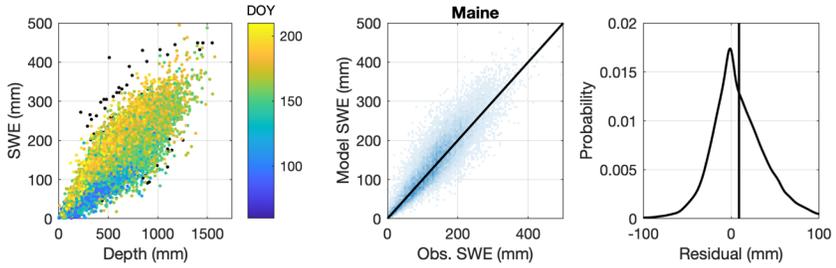
Deleted: 0



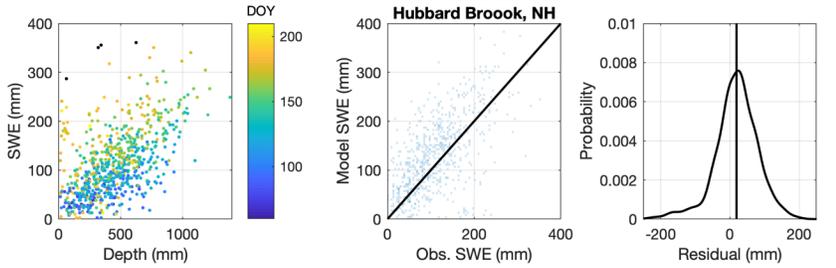
927

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

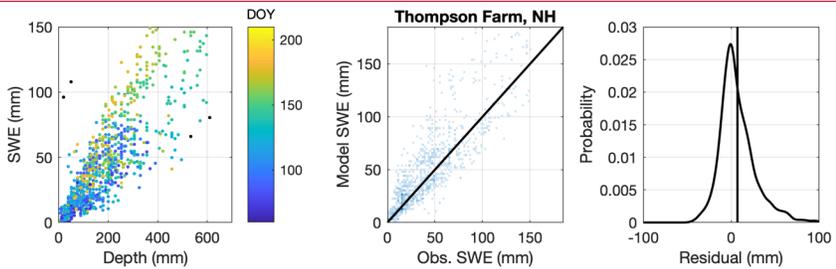
935



936

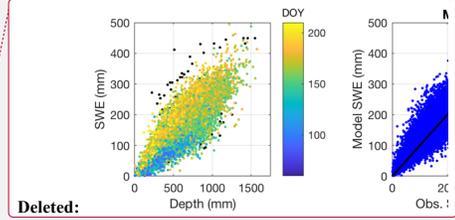


937

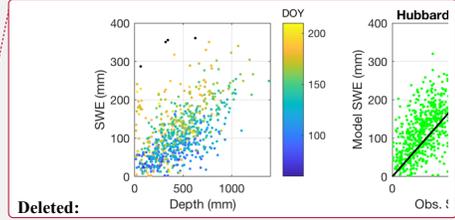


938

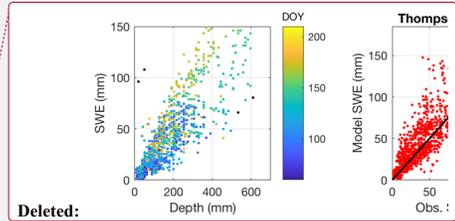
Deleted: 1



Deleted:

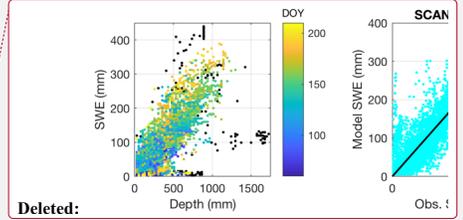
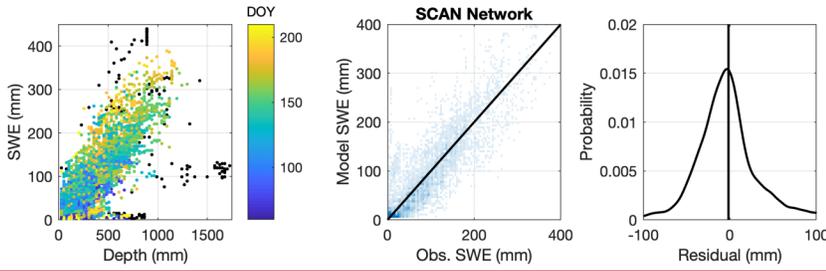


Deleted:

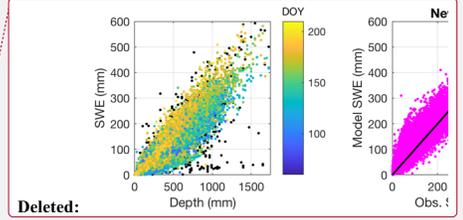
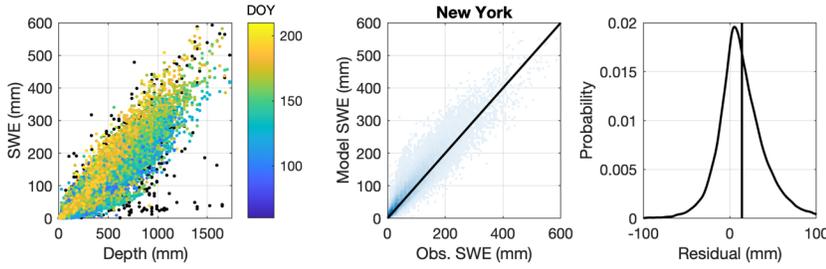


Deleted:

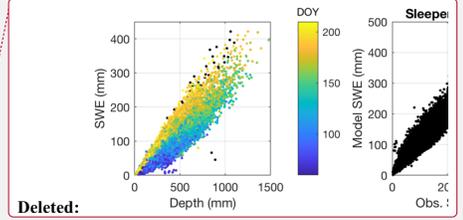
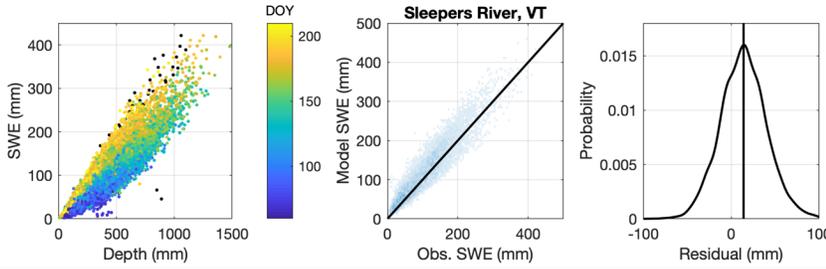
943



944



945



946

Deleted: Page Break

... [3]

Finally, SWE can be estimated with remote sensing methods, including satellite, airborne, and fixed platforms (e.g., Sokol et al., 2003; Vuyovich et al., 2014). Microwave frequencies are commonly used, but these frequencies do not work well in the presence of liquid water in the snowpack (Leinss et al., 2015). Recent attention has focused on the superior ability of L-band frequencies to measure SWE in wet snowpacks. Kang and Barros (2011) developed and tested an L-band snow sensor system in laboratory conditions and Deeb et al. (2017) discuss the application of L-band measurements to field-scale snow depth and SWE estimates for the SnowEx project.

3.3 Results for Chugach Mountains

The results for the Federal sampler core measurements in the Chugach Mountains are shown in Figure 12, using a format consistent with Figure 11. The three different measurement campaigns (March, April, and May) can be seen by the different symbol colors in the left panel. One notable difference between Figures 11 and 12 is that the Chugach dataset only spans spring months and not the full water year. So, the cluster of data points does not start at the origin. The R², bias (mm) and RMSE (mm) are 0.89, -50.0 and 118.0, respectively.

Figure 12: Results from application of the multi-variable, two-equation model to the Chugach Mountains, AK. The left column shows the measured SWE- h data. The symbols are colored by DOY. The middle panel plots the model estimates of SWE against the observations of SWE with the 1:1 line included. The right panel shows the model residuals, with the vertical line indicating the mean error, or bias.

