Predicting Snow-to-Liquid Ratio in the Mountains of the Western United States

Peter G. Veals[1], Michael Pletcher[1], Andrew J. Schwartz[2], Randy J. Chase[4], Kirstin Harnos[3], Jimmy Correia[3], Michael E. Wessler[5], and W. James Steenburgh[1]

[1]Department of Atmospheric Sciences, University of Utah, Salt Lake City, UT

[2]Central Sierra Snow Laboratory, University of California Berkeley, Soda Springs, CA

[3]NOAA/NWS/Weather Prediction Center, College Park, MD

[4]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO

[5]National Weather Service Western Region Headquarters, Salt Lake City, UT

Submitted to *Weather and Forecasting*

5 December 2024

Revised 6 April 2025

Corresponding author address: Peter G. Veals, Department of Atmospheric Sciences, University of Utah, 135 South 1460 East Room 819, Salt Lake City, UT, 84112. E-mail: peter.veals@utah.edu

Abstract

The snow-to-liquid-ratio (SLR) and its inverse, snow density, are crucial for forecasting snowfall in numerical weather prediction models and for estimating snow-water-equivalent (SWE) on the ground using remote sensing. SLR also varies widely in space and time, making it challenging to forecast accurately, particularly in the heterogenous terrain and climate of the mountains of the western United States. This study utilizes high quality, manually-collected measurements of new snowfall and new SWE from 14 mountainous sites across the region to build multiple linear regression (MLR) and random forecast (RF) algorithms to predict SLR as a function of atmospheric variables.

When an MLR algorithm is trained on a simple combination of wind speed and temperature from either the ERA5 reanalysis, the GFS, or the HRRR, it predicts SLR with considerably more skill than existing SLR prediction methods. When a more extensive set of variables is considered, the skill improves further.

The variables used to achieve the most skillful prediction of SLR are temperature, wind speed, relative humidity, specific humidity, maximum solar altitude angle during the observing period, CAPE, and HRRR QPF. When an RF algorithm is trained using these variables, it can predict SLR with R^2=0.43 and MAE=2.94. For the existing SLR prediction techniques currently used in operations, R^2 ranges from 0.04 to 0.23 and MAE ranges from to 4.01 to 9.45. Therefore the algorithms built in this paper can drastically improve SLR prediction over the mountains of the western US.

1. Introduction

Snowstorms in the western United States are essential for providing the water that sustains life, agriculture, industry, and hydropower in the region (Diffenbaugh et al. 2015; Li et al. 2017; Hagenstad et al. 2018). They can also pose a serious hazard to life, property, and commerce (Blattenberger and Fowles 1995; Spencer 2009; Black and Mote 2015; Seeherman and Liu 2015). Therefore, the accurate prediction and estimation of snowfall is essential, yet it remains difficult and error-prone.

Freshly-fallen snow is mostly air, with a snow-to-liquid ratio (SLR) ranging from 2:1 or lower to as high as 100:1 (e.g. Judson and Doesken 2000; Roebber et al. 2003). Consequently, for a given amount of liquid precipitation, the corresponding amount of snow can vary widely. This is a problem because contemporary snowfall prediction typically involves converting the liquid precipitation equivalent forecast that is directly predicted by numerical forecast systems (referred to as a quantitative precipitation forecast or QPF) to snow amount using an SLR (Alcott and Steenburgh 2000; Roebber et al. 2003; Byun et al. 2008; Pletcher et al. 2024). Conversely, for a given amount of snow, the corresponding amount of liquid can vary widely, posing a challenge for snow surveys and analyses that ingest the depth of new snowfall, yet seek to estimate the amount of liquid that has fallen (e.g., Raleigh and Small 2017). Therefore, the accurate prediction of SLR is essential to properly forecast snowfall and to measure snow-water-equivalent (SWE) from snow depth.

There are many factors that influence SLR. For example, the ice crystal habit can affect how densely crystals will pack together, increased riming of crystals removes some air space, high winds and the resulting ice crystal collisions can remove crystal branches causing tighter packing of crystals, natural compaction of snow under its own weight can densify it, and melting or rainfall can fill the spaces between crystals with water (Pomeroy and Brun 2001; Roebber et al. 2003; Baxter et al. 2005; Byun et al. 2008; Alcott and Steenburgh 2010; Steenburgh 2023). All of these processes are not explicitly or reliably accounted for in operational numerical forecast systems, motivating the need for other approaches to predict SLR.

Although SLR is known to vary, the simplest approach is to assume a fixed SLR. In the past, and even for convenience today, an SLR of 10:1 was/is sometimes assumed. The 10:1 rule was based on the findings of a single study conducted in eastern Canada that found a median SLR of 10 (Potter 1965; Roebber et al. 2003). However, not only is an SLR of 13 more appropriate for most of the US (Baxter et al. 2005), using a fixed SLR can be problematic over regions like the western US where significant intra- and inter-storm SLR variability occurs (Judson and Doesken 2000; Alcott and Steenburgh 2010; Pletcher et al. 2024).

To enable variable SLR prediction, the National Weather Service (NWS) National Blend of Models (NBM; Craven et al. 2020) uses four SLR methods referred to as 850–700-mb Thickness, Cobb, MaxTaloft, and Roebber in NWS training resources (Craven et al. 2020; The COMET Program 2023). The methods used and their weighting varies depending on the numerical modeling systems. The 850–700-mb Thickness method is used only for global ensembles and the data it requires would be subterranean at many western US sites, so it is not considered here. The Cobb method derives from Cobb and Waldstreicher (2005), although it has been revised several times. The Cobb method in NBM version 4.1 first identifies the maximum upward vertical velocity (UVV) contained in a cloudy layer, then calculates a weighting factor based on UVV and layer thickness, then applies a temperature-SLR relation to each model layer, and finally computes a weighted sum of the SLR from all model layers (The COMET program 2023). Most recently in NBM v4.2, a melting factor was added to adjust SLR based on the surface wet-bulb temperature and 1-h precipitation rate in marginal snow environments (Rudack et al. 2024).

The MaxTaloft SLR method is based on data from Alaska and uses a 5th degree polynomial to calculate SLR based on temperature:

$$\mathrm{SLR}_{\mathrm{MaxTAloft}} = 0.0000045 * \mathrm{T_{Max}}^5 + 0.0004432 * \mathrm{T_{Max}}^4 + 0.0130903 * \mathrm{T_{Max}}^3 + 0.0585968 * \mathrm{T_{Max}}^2 - 1.8150809 * \mathrm{T_{Max}} + 5.9805722, \quad (1)$$

where $\mathrm{T_{Max}}$ is the maximum temperature (℃) between 610 m AGL and 400 hPa (The COMET Program 2023; Pletcher et al. 2024).

The Roebber method is derived from Roebber et al. (2003) who trained an artificial neural network using snowfall observations from NWS sounding sites and input variables that include monthly solar radiation, temperature and relative humidity at multiple levels, wind speed, and 6-h SWE. The training data consisted primarily of data from the eastern two-thirds of the US (the only western US sites were Great Falls, Lander, Salt Lake City, and Denver, all in in non-mountain areas of the western interior). The artificial neural network predicted SLR in three classes [heavy (1:1 < ratio < 9:1), average (9:1 ≤ ratio ≤ 15:1), and light (ratio > 15:1)], but was modified to produce a deterministic SLR for the NBM (The COMET Program 2023).

The Kuchera method, although not used in the NBM, is often used by forecasters and meteorological websites. As stated in Rosenow et al. (2023): “The so-called Kuchera method has become commonplace in operational meteorology, including the NWS, despite having not been formally published. This technique was created by performing a linear regression on snow depth and liquid equivalent observations using the maximum temperature in a column below 500 hPa, $T_{\max}$, as the sole predictor of SLR”. The Kuchera algorithm is defined as:

$$\mathrm{SLR}_{\mathrm{Kuchera}} = \begin{cases} 12 + 2\ x\ (271.16 - T_{max}),\ T_{max} > 271.16 \\ 12 + (271.16 - T_{max}),\ T_{max} \leq 271.16 \end{cases} \quad (2)$$

Other SLR methods, based on a single air temperature variable, abound. Van Cleave (2013) relies on 700 hPa temperature, and a method recently implemented in the HRRR model (Benjamin et al. 2021) uses the temperature in the lowest model layer. Byun et al. (2008) use the 2 m temperature.

There are also a number of subregional SLR methods, such as that developed by Hoopes et al. (2023) for the mountain ranges of southern Arizona using gridded SLR analyses derived from Broxton et al. (2019). SLR in this case, however, is based on the 24-h change in the total depth of the snowpack divided by the 24-h change in snowpack SWE, which is a problematic due to compaction of the pre-existing snowpack.

These legacy SLR methods have not been carefully evaluated over the western US. However, Pletcher et al. (2024) compared the performance of the NBM SLR methods to a random forest algorithm trained on local data at one western US site: Alta Ski Area in the Wasatch Range of northern Utah. They found that the random-forest SLR algorithm produced substantially better SLR forecasts than the NBM methods, suggesting that an algorithm based on high quality regional observations might yield substantial forecast improvements for SLR over the western US.

One issue that affects nearly all of the SLR algorithms created to date is the paucity of high-quality snowfall and SWE (therefore SLR) observations from conventional meteorological networks like the NWS Cooperative Observer Program (COOP; Mehta 2023) and Automated Surface Observing Station (ASOS; NOAA 1998). Most of the datasets used to build these SLR algorithms use SWE from precipitation gauges, and when precipitation is falling as snow, precipitation gauges suffer from a problem known as undercatch, whereby wind flowing up and over the gauge orifice prevents many hydrometeors from falling into the gauge (Rasmussen et al. 2012). Undercatch is nonexistent during calm conditions, but grows with increasing wind speed. MacDonald and Pomeroy (2007) showed that for just an 8 ms^{-1} wind speed, an unshielded gauge will only capture ~30% of the SWE that falls, and a gauge with an Alter shield will only capture ~60% of the SWE that falls. Thériault et al. (2012) suggest that undercatch is likely even greater that this, especially when the snowflakes have little riming. Compounding this issue, gauge SWE amounts from the National Weather Service Cooperative Summary of the Day (COOP) do not distinguish between SWE produced by snow or other precipitation types (e.g., ice pellets or liquid precipitation), adding additional uncertainty to snowfall measurements.

Another issue affecting legacy SLR algorithms is the fact that many of them were trained using observations mostly or completely from non-mountainous regions. Hydrometeor growth over mountainous regions is heavily influenced by the regions of ascent over the windward slopes of the terrain, with the majority of hydrometeor growth often happening <2 km above ground level (AGL) on these slopes. Storms in the interior ranges of the Western US also generally feature a temperature profile that decreases monotonically with height (e.g., Geerts et al. 2015; Aikins et al. 2016; Friedrich et al. 2021). However, the dynamics of mountain waves, precipitation spillover, and flow blocking can often make for complex and difficult-to-predict accumulation patterns that can be far removed from the windward slopes (e.g., Neiman et al. 2002; Yuter et al. 2011; Geerts et al. 2015; Veals et al. 2020). Over flatter regions like the eastern half of the US, midlatitude cyclones generally produce the majority of cool-season precipitation. The hydrometeors in these cyclones can see much of their growth occur >50 km away from their eventual accumulation location, can be lofted along the way, and often originate 7 km or more AGL. They also often feature complex temperature profiles, with warm layers and inversions (e.g., Lackmann and Thompson 2019; Janiszeski et al. 2024). Yet these are general differences, with a broad spectrum of atmospheric conditions possible for both flatland and mountain regions.

In this study, we develop algorithms to predict SLR over the western US using high-quality, manually measured snowfall and SWE observations from 14 geographically and climatologically diverse mountain observing sites. In Section 2, we describe the characteristics of these sites and observations, the techniques used to generate SLR algorithms, and the methods employed for verification. Section 3 then examines the fidelity of these algorithms, which exhibit significant improvement relative to NBM SLR methods based on randomized testing, and examines the role of the atmospheric variables in predicting SLR. Section 4 summarizes the main conclusions, which illustrate the value of SLR algorithms based on high-quality regional data.

2. Data and Methods

a. Snowfall Observations

Data were obtained from 14 sites across the western US (Fig. 1, Table 1) where an observer takes manual observations of new snowfall and new SWE once or twice daily from a board that is wiped clean after each observation and placed atop the snowpack. SWE is based on a sample collected on the board by a coring tube and scale. This reduces errors due to undercatch, although high-wind situations can still create representativeness errors in some circumstances. Twelve of the observations come from snow-safety (i.e., avalanche mitigation) teams working for departments of transportation on avalanche-prone highways or at ski resorts where avalanche mitigation is frequently conducted. The remaining datasets are HLY, operated by an avalanche forecaster at the Sawtooth Avalanche Center, and CSSL, the Central Sierra Snow Laboratory (cssl.berkeley.edu). To mitigate the influence of rounding and measurement errors on SLR calculations, we only used observations from periods with snowfall $>$ 5.08 cm and SWE $>$ 0.28 cm. These thresholds are consistent with prior studies (Judson and Doesken 2000; Roebber et al. 2003; Alcott and Steenburgh 2010; Pletcher et al. 2024). We omitted observations (186 total) with SLR=10.0, as there were some cases in which the observer likely used 10.0 as a placeholder due to a missing depth or SWE observation. We tried omitting observations with SLR=20.0, as they appeared much more frequently than observations with SLR=19.0 or SLR=21.0, but omitting them did not significantly affect our results, so we chose to keep them. Careful investigation suggested that the increased frequency of SLR=20.0 reflects a tendency for observers, when measuring snowfall amounts $\leq$ 4 in (10.2 cm; the raw observations are taken in inches at most sites), to round the SWE amount to either 0.15 in (.38 cm) or 0.2 in (0.51 cm). Observations with SLR≤2 and SLR>50 were also omitted, as values in these ranges are more prone to rounding and/or measurement error.

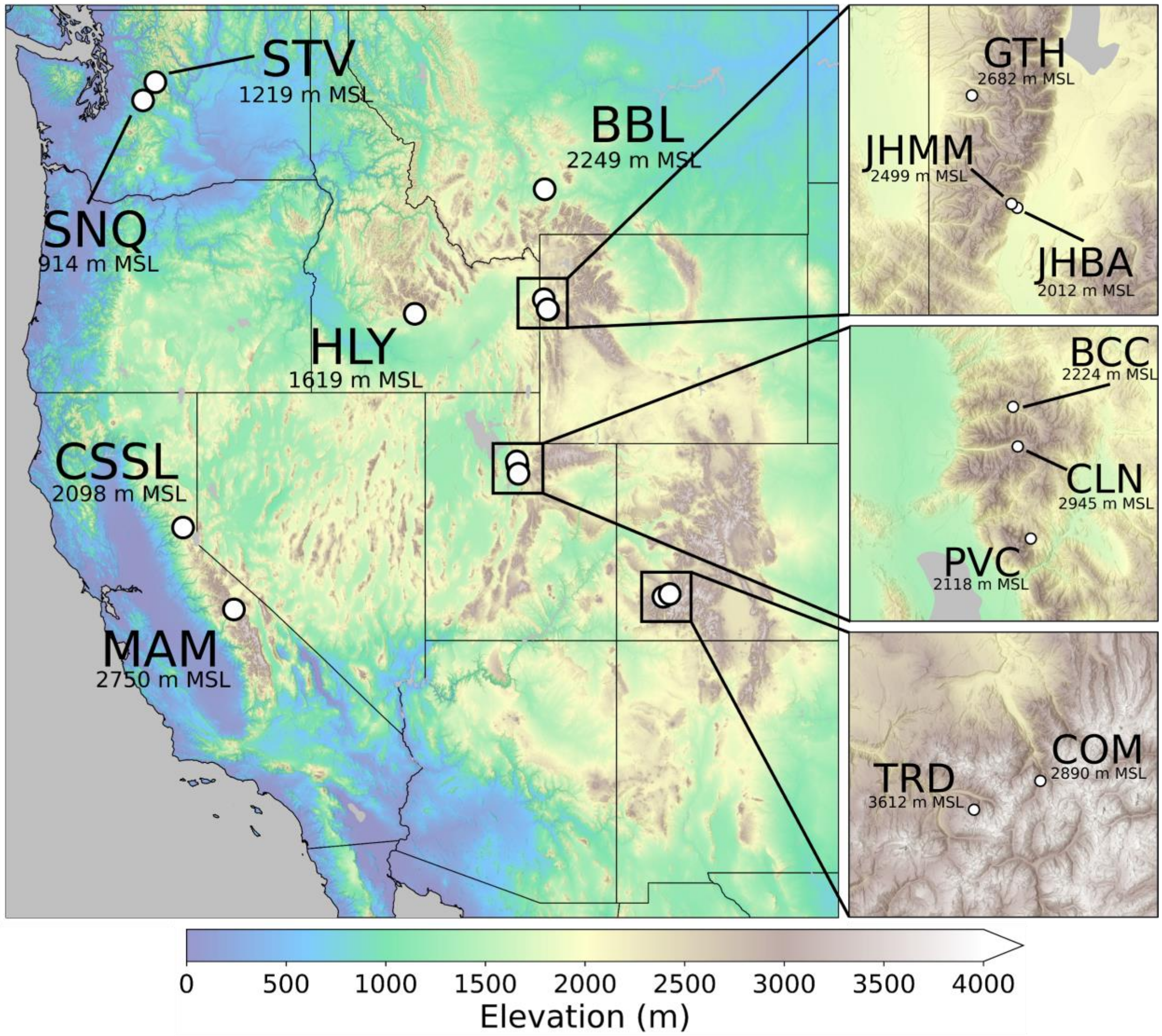

Figure 1. Topography (m MSL following scale at bottom) of the Western US, with the locations of each of the 14 observing sites used in this study, and their elevations indicated.

Site	Data Available	Elevation (m)	Approx. Observation Frequency	*N* Observations
BBL	2018–2024	2249	24 h	177
BCC	2018–2024	2224	10 h and 14 h	234
CLN	2018–2024	2945	12 h	444
COM	2018–2024	2890	24 h	170
CSSL	2021–2024	2098	8 h and 16 h	170

GTH	2018–2020, 2022–2024	2682	24 h	155
HLY	2019–2024	1619	24 h	48
JHBA	2018–2024	2012	24 h	108
JHMM	2018–2024	2499	24 h	232
MAM	2018–2024	2750	6–18 or 24 h*	177
PVC	2018–2024	2118	11 h and 13 h	144
SNQ	2018–2024	914	24 h	271
STV	2020–2024	1219	24 h	176
TRD	2018–2024	3612	24 h	194

Table 1. For each of the 14 sites in this study, the name, period that data is available, elevation, approximate observation frequency, and number of observations used. *MAM observations are most often ~24 h, but during intense storms, 2 observations per day are taken, with intervals ranging from ~6–18 h.

The period of record varies widely among the sites, but we limit the data to the study period 2 October 2018 – 30 April 2024 for all sites. The beginning date was selected because it marks the beginning of HRRR data availability beyond forecast hour 18, which is required for our analysis. Therefore, data for all sites comes from the full study period, except for CSSL, GTH, HLY, and STV, which had some missing seasons in the study period (Table 1). Within the study period, we only consider the heart of the cool season, which we define as the months of November–April. This leaves 2700 total observations that are used in this study.

b. Atmospheric Variables

We build and evaluate one algorithm that uses atmospheric variables from the ERA5 reanalysis (Hersbach et al. 2020), another algorithm that uses the High-Resolution Rapid Refresh model (HRRR; Benjamin et al. 2016), and another algorithm that uses the Global Forecast System model (NOAA EMC, 2024). The selection of atmospheric variables to consider in our algorithm was influenced by prior studies (e.g., Roebber et al. 2003; Alcott and Steenburgh 2010), the physics of ice crystal growth and metamorphism, and the available variables from the datasets used. These include:

(1) Temperature (T), specific humidity (Q), relative humidity (RH), and wind speed (SPD) linearly-interpolated from pressure coordinates to height AGL in increments of 400 m, spanning the surface to 4800 m AGL. The use of AGL coordinates makes the algorithm applicable at all elevations and grid points (pressure levels like 850 hPa are underground at many high elevation grid points, and 700 hPa may be near the surface at high elevations but 3 km above the flatlands and valleys). "T04" is used to denote T at 400 m AGL, "Q12" for Q at 1200 m AGL, "SPD24" for SPD at 2400 m AGL, and so on.

(2) Cloud-top temperature (CTT). The cloud top was defined as either (A) the location of the maximum RH lapse rate, or (B) the first location, moving upward from the 2nd pressure level above ground, where first RH drops below 80% for 2 consecutive levels.

(3) Solar altitude angle (Solar) during the observation period. We considered the mean and maximum, with the maximum having the strongest effect on SLR. Hereafter, we use the variable name "Solar" to denote the maximum solar altitude angle during the observing period.

(4) SWE amount – both the model-forecasted (hereafter QPF) and observed amounts.

(5) Temperature lapse rate – 1-2 km, 2-3 km, 3-4 km, 1-3 km, and 3-5 km AGL.

(6) Convective Available Potential Energy (CAPE) – as calculated and output by the modelling/reanalysis system

The process for attributing the atmospheric variable to the corresponding period of the SLR observation was as follows. To attribute the ERA5 reanalysis data to a 12-hour SLR observation, the mean of the ERA5 data within that 12-hour observation period is used. For a 24-hour observation, the mean of the ERA5 data within that 24-hour observation period is used, and so on for 8-hour observations, 16-hour observations, etc.

For the GFS and HRRR, which are forecasts, the process was slightly different. We began with the assumption that any forecast hour before Forecast Hour 10 (Fhr10) was of limited utility in a real-world setting, as it provides very little lead time for consumers of the forecast to make plans or decisions. We then tested two different time matching schemes:

(1) Only use data from the model initialization closest to the observation period, for which the observation period can fit entirely between Fhr10 and Fhr42.

(2) Assemble a timeseries of only Fhr12-Fhr24 from only 00Z and 12 model initializations, which creates one continuous timeline of atmospheric data.

Both method 1 and method 2 performed similarly for SLR prediction, so we opted for method 2 because it was much easier to deal with for coding, data management, and understanding any issues when debugging was required. For all datasets, we experimented with the mean, maximum, and minimum value of the atmospheric variables during the observation period from the gridpoint nearest the observing site, and found the mean to be best for predicting SLR.

c. SLR Algorithm

We create four different SLR algorithm versions in this study, referred to as V1–V4. For V1 and V2, the atmospheric variables described above were fed into a multiple linear regression (MLR) calculator from the *scikit-learn* Python package (Pedregosa et al. 2011), along with the SLR observations from our 14 sites, to produce a predictive algorithm for SLR. For V3 and V4, a random forest (RF; Breiman 2001) regressor from *scikit-learn* is used instead to produce the predictive algorithm. To select the optimal hyperparameters for the RF, we began with the default values, including: 100 trees, no constraint on the maximum depth of the tree, a minimum of 2 samples required to split a node, a minimum of 1 sample per node. We then experimented with a broad range of values for these hyperparameters, and none of the other values achieved a better predictive skill, so the default values are used in this study.

We chose the MLR technique because it is computationally inexpensive, easy to share, and easy to implement in any coding language. We chose the RF because, for input consisting of >6 atmospheric variables, the resulting algorithm exhibits greater skill than the corresponding MLR algorithm in predicting SLR, can learn nonlinear relationships, and has been useful in other meteorological tasks (e.g., Pletcher et al. 2024; Chase et al. 2023 and references therein). We also experimented with other machine learning techniques, including Support Vector Regression (SVR; Vapnik 1995), and a type of neural network known as a Multilayer Perceptron (MLP; Gardner and Dorling 1998). We did not include the SVR in this work because it was an order of magnitude slower than the RF, making it unsuitable for forecasting applications, and we did not include the MLP because it was less skillful than the RF for our application.

All four algorithms (V1–V4) were built using a 60/40 train/test split, where 60% of the data (randomly selected) are used to train the algorithm, and the remaining 40% that have been withheld are used to test the performance of the algorithm. We evaluate the algorithm using the 40% of observations that have been withheld for testing, and the evaluation is done 1000 times (i.e. k-fold cross validation with k=1000). This sampling procedure is done because the split into the 60/40 train/test observations is random, and the resulting algorithm and its performance can depend to some extent on which observations are dealt into testing or training. By doing 1000 permutations of the train/test split process, we can account for this variability when evaluating algorithm performance. We consider the mean performance of the 1000 splits as a good estimate of model skill. Because observations from the same site that are adjacent to each other in time (for example the 00Z and 12Z observations from the same day) may occur under similar atmospheric conditions, it is possible that including an observation in the testing dataset and its temporally adjacent counterpart in the training dataset would artificially inflate the skill of the algorithm. We tried omitting any observations from testing that had a temporally adjacent counterpart in training, and it did not substantially affect our results, so we opt to permit temporally adjacent observations to keep a larger sample size.

d. NDFD data

We obtained SLR forecasts from the NWS's National Digital Forecast Database (NDFD) for all cool seasons (November through April) in the entire period of record, which yielded the cool season from November 2020 through April 2024. The NDFD is a gridded dataset consisting of the forecasts sent out by each NWS forecast office. To compute SLR from the NDFD, we divide the forecast hour 6-12 snowfall by the forecast hour 6-12 SWE accumulation, and then make a continuous timeseries of SLR from these 6-12 hour forecasts. The data for each site come from the nearest NDFD gridpoint to the site.

3. Results

a. Climatology of the 14 sites

The distribution of SLR from each site over its period of record is shown in Fig. 2, with observations limited to those with snowfall $>$ 5.08 cm, SWE $>$ 0.28 cm, SLR >2.0, SLR<50, and SLR $\neq 10.0$. The sites in the Sierra Nevada and Cascade mountains (CSSL, MAM, SNQ, and STV), with their maritime snow climates (Trujillo and Molotch 2014), exhibit the lowest mean SLRs and narrowest SLR distributions (Fig. 2e,j,l,m). Although their SLR distributions are heavily skewed toward the lower values, they do still occasionally see events with SLR>20. Farther inland, COM and TRD have greater mean and median SLRs, and tails extending more to the right (Fig. 2d,n). Moving farther north and/or farther inland to BBL, BCC, CLN, GTH, JHBA, JHMM, and PVC, where colder storms are more common, the distributions and tails move even farther to the right, with SLRs >20 being quite common (Fig 2 a,b,c,f,h,i,k). However, even at these cold continental locations, there are still a significant number of dense snow events with SLR <6, highlighting the broad variability from storm to storm at these locations. The effects of the interaction between synoptic climatology and terrain orientation can also be seen when comparing 2 sites in the Wasatch Range of Utah (BCC and PVC). BCC is <26 km from PVC, and both sites are at nearly the same elevation, yet PVC has a much lower median and mean SLR than BCC and CLN. The primary difference is that PVC receives a much greater fraction of its cool season SWE from southwesterly flow events, which tend to be warmer and windier than other flow directions (Steenburgh 2023). The effects of elevation on snow climate are also apparent when comparing SNQ and STV, which are <45 km from one another, yet STV has significantly higher mean and median SLRs. STV is 300 m higher in elevation, which appears to have a strong impact on SLRs in the relatively warm maritime climate of the Cascades.

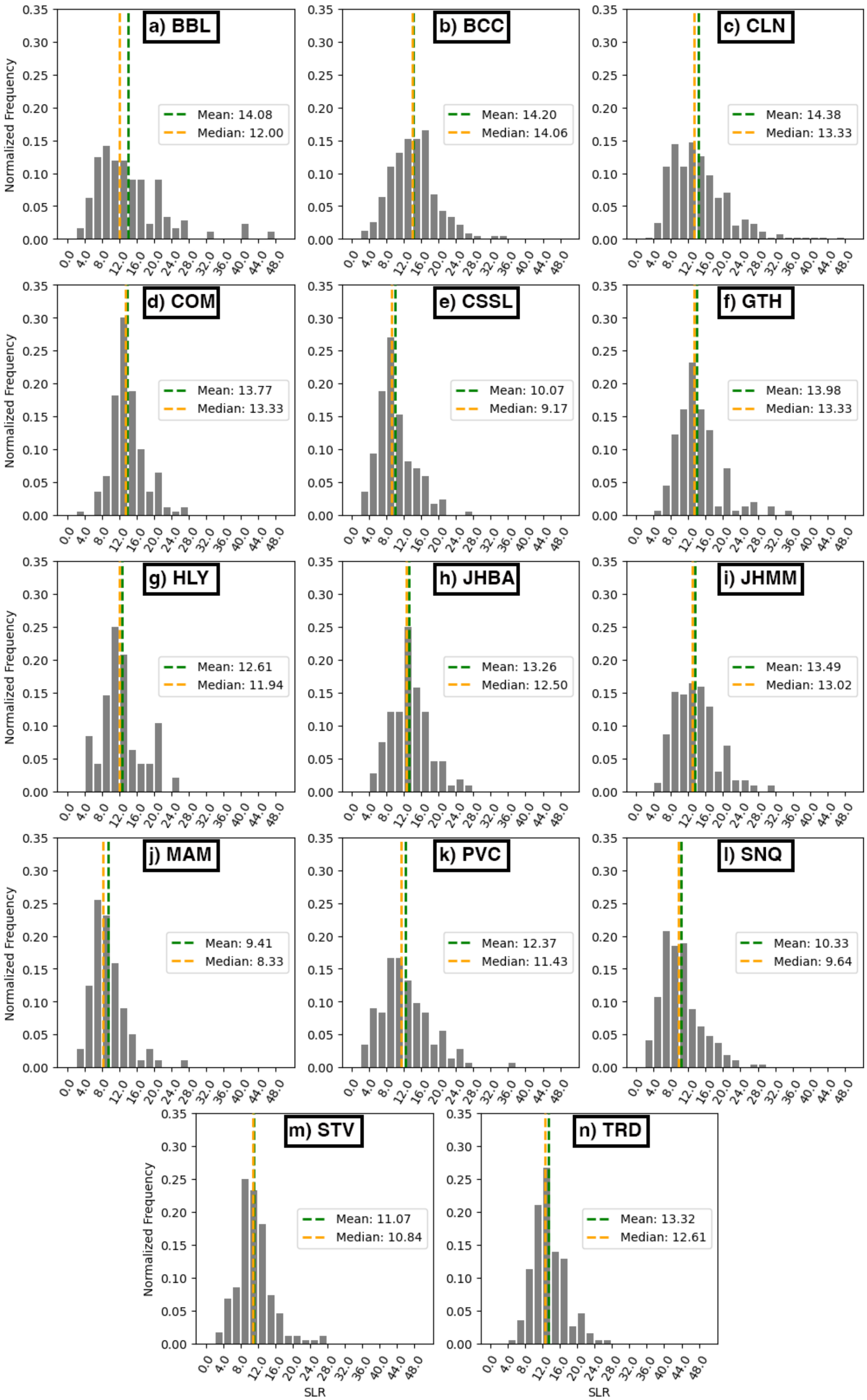
a) BBL
Mean: 14.08
Median: 12.00
b) BCC
Mean: 14.20
Median: 14.06
c) CLN
Mean: 14.38
Median: 13.33
d) COM
Mean: 13.77
Median: 13.33
e) CSSL
Mean: 10.07
Median: 9.17
f) GTH
Mean: 13.98
Median: 13.33
g) HLY
Mean: 12.61
Median: 11.94
h) JHBA
Mean: 13.26
Median: 12.50
i) JHMM
Mean: 13.49
Median: 13.02
j) MAM
Mean: 9.41
Median: 8.33
k) PVC
Mean: 12.37
Median: 11.43
l) SNQ
Mean: 10.33
Median: 9.64
m) STV
Mean: 11.07
Median: 10.84
n) TRD
Mean: 13.32
Median: 12.61
Normalized Frequency
SLR

Figure 2. Distribution of observed SLR for the full period of record at each of the 14 sites used in this study. Note that at each site, observations with SLR=10.0, SLR≤2.0, SLR≥50, SWE≤2.8 mm, and snowfall ≤50.8 mm have been removed.

b. V1 Algorithm

We began by creating a simple version of the algorithm (V1) that yielded the best possible skill using only temperature and wind speed to train an MLR. It only requires 4 variables (T04, SPD04, T24, and SPD24), with the addition of extra levels yielding negligible additional skill. We evaluate skill using R^2 and mean absolute error (MAE), with R being defined here as the Pearson correlation coefficient (Wilks 2019).

For our V1 algorithm, using data from ERA5, the mean R^2 value for predicted SLR relative to observed SLR is 0.31, and the mean MAE value is 3.27, with the standard deviation (σ) of the MAE at 0.07. Figure 3a shows a run of the algorithm that produced R^2 and MAE values equal to the mean of the 1000 train/test iterations. The predicted SLR most closely matches observed SLR for low and moderate SLR values, but for events when observed SLR is >20, the spread increases (Fig. 3a).

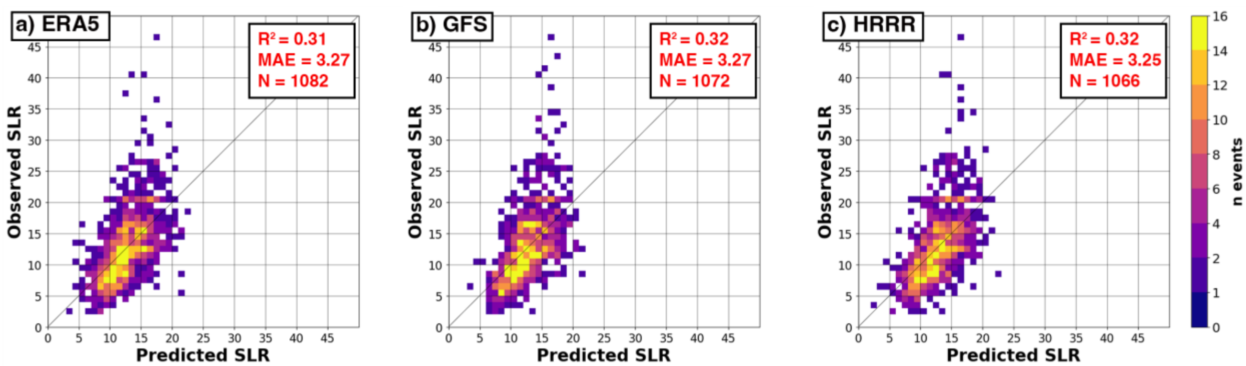

Figure 3. Observed SLR vs SLR predicted by the V1 algorithm, evaluated against the 40% of observations that are withheld from training to be used for testing, using (a) ERA5 reanalysis, (b) GFS, and (c) HRRR as the source of atmospheric input variables, shown here for one of the 1000 permutations for which R^2 and MAE were equal to the mean R^2 and mean MAE of the 1000 permutations.

When the V1 algorithm is trained and tested the same way, but instead using GFS (Fig. 3b) and HRRR (Fig. 3c) data, the performance is nearly identical. The mean R^2 and MAE using GFS data are 0.32 and 3.27, respectively. The mean R^2 and MAE using HRRR data are 0.32 and 3.25, respectively. The similar performance between the ERA5, GFS, and HRRR suggests the differences in their depictions of temperature and wind speed values are not large enough to appreciably affect the skill of SLR prediction for the V1 algorithm.

c. Legacy SLR methods

We also computed SLR with 5 commonly used methods for comparison: (1) MaxTaloft, (2) Cobb, (3) Roebber, [all used operationally in NBM (Craven et al. 2020) v4.2], (4) the Kuchera method (described above), and (5) a fixed SLR. A sixth source of SLR, the forecast from the NDFD, is included, as it reflects the final SLR that goes out to users of NWS forecasts.

When the Cobb method is applied to the HRRR data for each of the cases in our 2018-2024 study period, and compared to the high-quality manual observations from our dataset, R^2=0.04 and MAE=4.29 (Fig. 4a). The SLR values predicted by Cobb are mainly clustered in the 5-15 range, with large prediction errors and very little correlation with reality. It is also biased a bit low, tending to underpredict SLR compared observed. Cobb's poor performance at these sites may be a result of the different distribution of vertical velocities, both in reality and in the HRRR, over complex terrain compared to flatter terrain. There may also be a different relationship between the location of hydrometeor growth and where the resulting hydrometeors reach the ground, over complex terrain compared to flatter terrain (see discussion of these difference near the end of Section 1). In other words, an algorithm that relies on vertical velocity locations and values observed over flat land may not perform well over complex terrain.

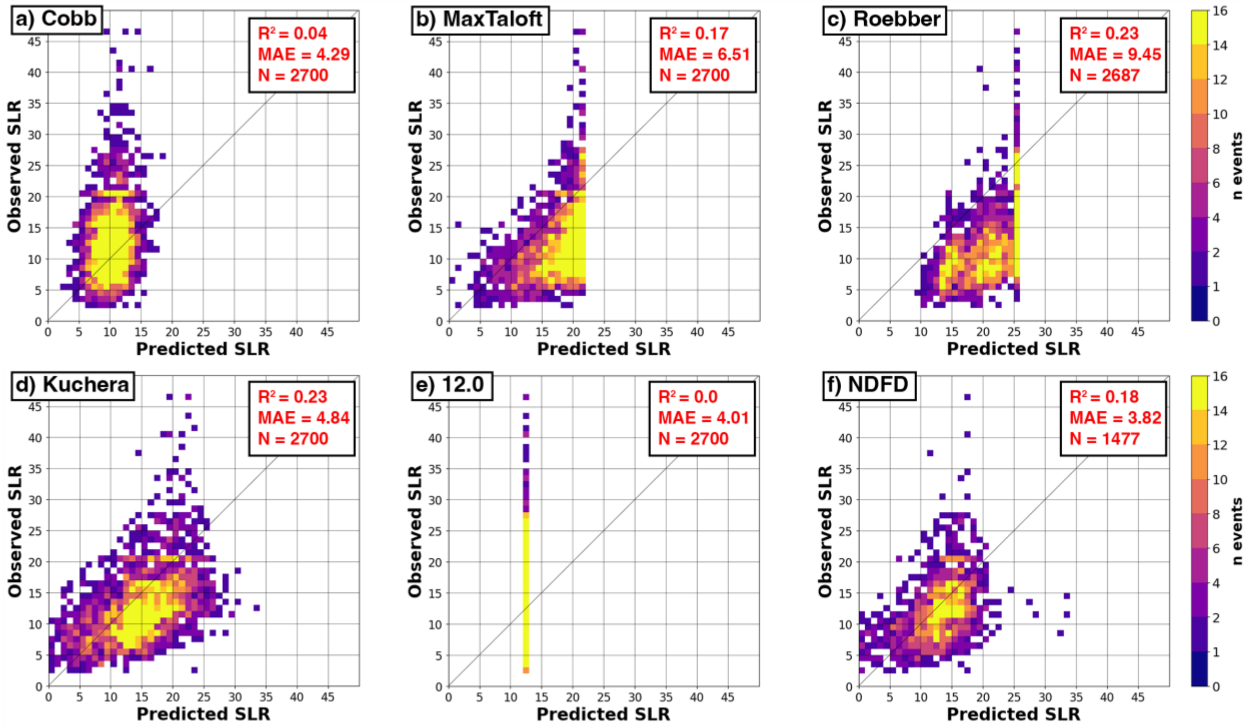

Figure 4. Observed SLR vs SLR from common sources of SLR prediction: (a) Cobb, (b) MaxTaloft, (c) Roebber, (d) Kuchera, (e) fixed 12.0 SLR, and (f) the NWS NDFD. No random test/training split needed for these datasets, so the total number of available observations are used.

When the MaxTaloft method is applied to the HRRR data for each of the cases in our 2018-2024 study period, and compared to the manual observations from our dataset, the R^2=0.17 and MAE=6.51 (Fig 4b). MaxTaloft appears to greatly overpredict SLR (a high bias), with an even larger MAE than Cobb, and it rarely predicts SLR<10.0. The abrupt cutoff at ~22 reflects the boundary of MaxTaloft's polynomial formula.

When the Roebber method is applied to the HRRR data for each of the cases in our 2018-2024 study period, and compared to the manual observations from our dataset, the R^2=0.23 and MAE=9.45 (Fig 4c). The predictions are biased quite high, with large prediction errors. The cluster of predictions at 25 reflects the fact that the NBM code caps the Roebber SLR prediction, with any prediction >25 set to 25.

When the Kuchera algorithm is applied to the HRRR data for each of the cases in our 2018-2024 study period, and compared to the observations from our dataset, the R^2=0.23 and MAE=4.84 (Fig 4d). The predicted SLR does exhibit some vague correlation with observed SLR, but the prediction errors are quite large and biased high.

Using a fixed SLR of 12.0 yields R^2=0.0 and MAE=4.01 (Fig. 4e). We experimented with fixed SLR values from 10.0 to 13.0, in increments of 0.1, and 12.0 yielded the best performance in our dataset. This differs of course from the 10.0 that is commonly used for a fixed SLR, likely due to the fact that many of the sites in our dataset have mean and median SLRs around 12, 13, or even 14 (Fig 2).

The final SLR that we evaluate is the forecast SLR from the NDFD. When compared to our observations, the R^2=0.18 and MAE=3.82 (Fig 6f). This is the lowest MAE and best performance of the 6 legacy SLR techniques. The NDFD is a gridded aggregate of the forecasts issued by each NWS office, so the methods used to produce the final SLR values varies with time and by office, but the NDFD SLR outperforms all of the NBM SLR techniques. It does not, however, outperform the V1 algorithm.

d. V2 and V3 Algorithms

To build the V2 algorithm, we began with the variables described in Section 2b and used recursive elimination, stepwise screening regression, and lasso regression (Wilks 2019) to identify the optimal set of input variables from the HRRR, selecting for the lowest MAE and highest R^2 value relative to observed SLR. The combination of variables that added skill to the algorithm includes: T04, SPD04, SPD24, RH04, RH24, Q04, Q24, model QPF, CAPE, and Solar. When these variables are input into an MLR, the performance of the V2 algorithm improves relative to V1, with R^2=0.39 and MAE=3.05 (Fig. 5a). This includes some improvement in predicting SLRs >20, although these events remain a challenge. In practical terms, the algorithm does not have the ability to discriminate between the conditions associated with an observed SLR of 15 and those associated with an SLR of 25.

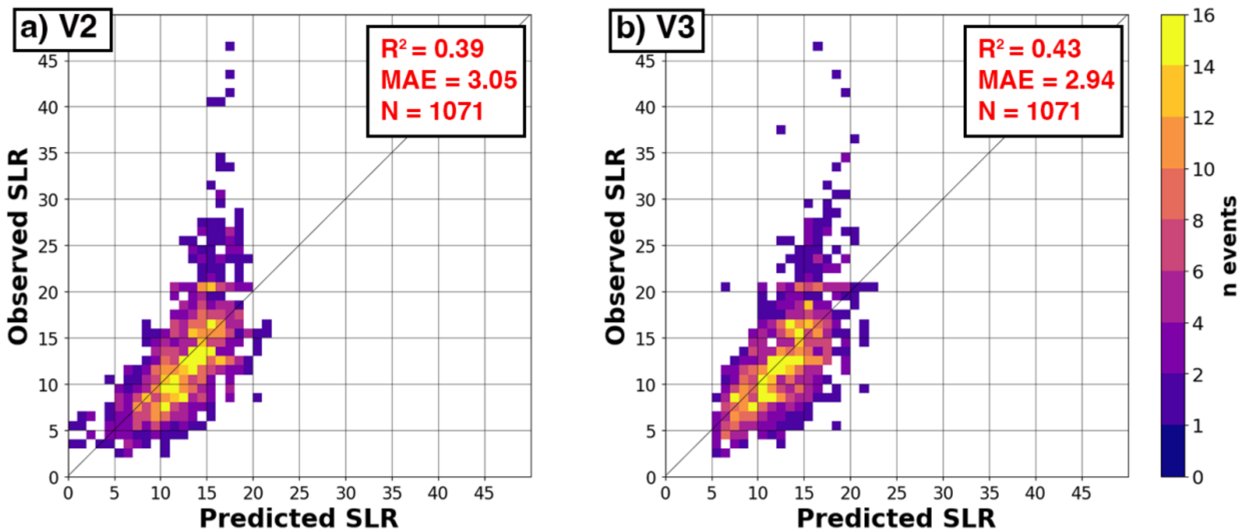

Figure 5. Observed SLR vs SLR predicted by (a) the V2 and (b) V3 algorithms, evaluated against the 40% of observations that are withheld from training to be used for testing, shown here for one of the 1000 permutations for which R^2 and MAE were equal to the mean R^2 and mean MAE of the 1000 permutations.

The V3 algorithm uses a RF regressor instead of MLR to increase skill. This V3 of the algorithm, when compared to observed SLR, yields R^2=0.43 and MAE=2.94 (Fig 5b). The increased skill of V3 relative to V2, given the same set of input variables, is a result of the RF's ability to detect and replicate nonlinear relationships between the input variables and SLR. The relationships predicted by the MLR are, by definition, linear. Even though V3 is more skillful than V2 on the whole, there is little or no improvement in anticipating the high SLR events (Fig 5b).

e. Importance of each variable

The logical next step in this study is to explore the importance of each variable to the fit of the algorithm. For V2, which is an MLR, the algorithm must be re-run using the standardized anomaly of each variable, rather than its actual value. This does not change the fit or skill of the algorithm; it simply makes the coefficient of each term in the regression equation represent the relative contribution of its corresponding variable to the fit of the equation. The train/test split and the MLR fit are performed 1000 times and the coefficients recorded for each of these 1000 permutations. This process determines that Q04 (400 m AGL specific humidity) to have the greatest contribution to the MLR equation with a median coefficient of -2.3 (Fig 6a). A negative coefficient indicates SLR decreases with increasing Q04. The next most important variable is 400 m AGL wind speed (SPD04), with a median coefficient of -0.8, followed by RH24 with a median coefficient of 0.7, and HRRR QPF with a median coefficient of 0.65 (Fig 6a).

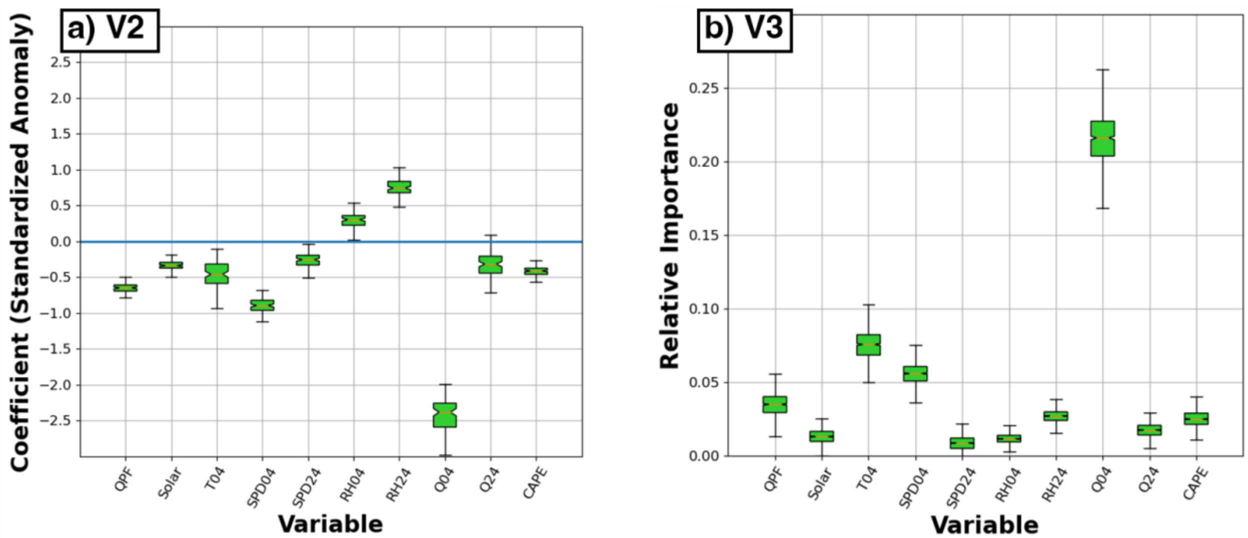

Figure 6. (a) For the V2 algorithm, with all atmospheric input variables from the HRRR converted to standardized anomalies, the coefficients for each term in the MLR equation. The magnitude of a coefficient is proportional to its importance in predicting SLR in the equation. The box-and-whisker plots represent the 1000 different values of the coefficient for the 1000 permutations. (b) For the V3 algorithm, the importance of each variable in the Random Forest regression. The box-and-whisker plots represent the different values of the coefficient for the 1000 permutations. The differences between boxes are statistically significant at the 95% confidence interval if the notches area around the medians do not overlap.

The importance of each variable in the RF version of algorithm V3, as again determined by 1000 permutations, looks somewhat different than for the MLR. The scoring of relative importance is also determined differently for the RF, with the nonlinear fits for each variable making it impractical to assign a single sign to the contribution of the variable. So feature importance for the RF is determined by the contribution to model fit when the values in a column are randomly reshuffled (Breiman 2001). As in the MLR algorithm, Q04 is again the greatest contributor to the fit by a wide margin, with a median relative importance of 0.24 (Fig 6b). Then in a distant 2^{nd}, 3^{rd}, and 4^{th} place are T04, SPD04, and HRRR QPF, respectively. The main contrast with the MLR relative importances is that T04 has a much more prominent role (2^{nd} most important variable) in the RF than it does in the MLR. This suggests that when temperature is permitted to have a nonlinear fit, it more accurately predicts SLR.

f. Relationships amongst variables

For the most important atmospheric variables identified in the previous section, the key relationships amongst them and with observed SLR are explored next. The observed SLR exhibits an increasing trend with decreasing T04 up until values of around -15 C, when the relationship changes sign and colder values actually tend to yield increasing SLR (Fig 7a). Something resembling this can be seen in a number of prior studies (Byun et al. 2008; Alcott and Steenburgh 2010). It suggests that when T04 is colder than about -15 C, the temperature in the zone where most crystals are growing (likely at or above the 400 m AGL level where T04 is defined) is cold enough that dendritic crystals are no longer favored, and crystal habits that pack more densely, like plates and columns, are beginning to be favored. See Nakaya (1954) and Bailey and Hallett (2009) for more information on favored crystal habit as a function of temperature.

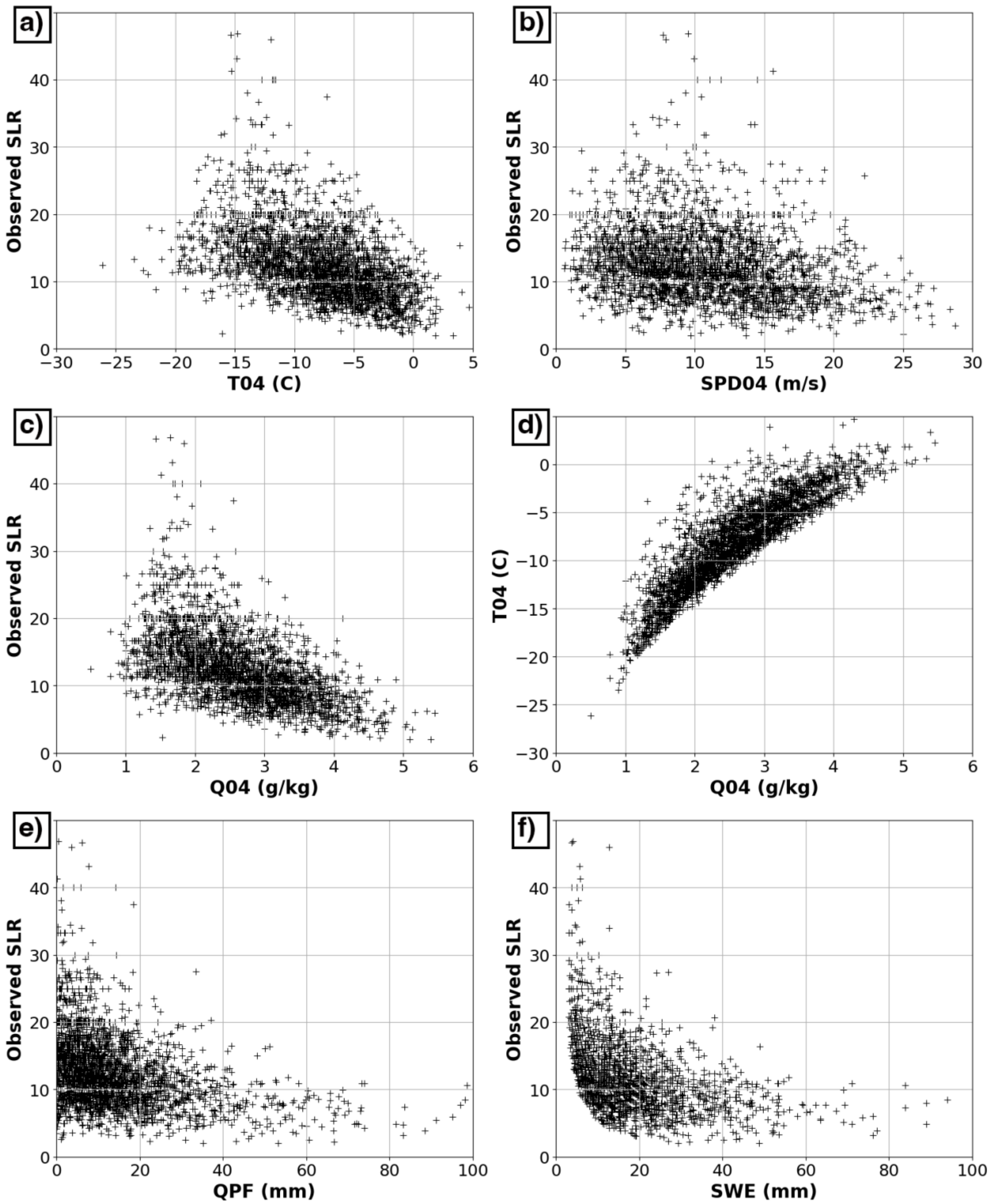

Figure 7. For all observations, (a) observed SLR vs T04, (b) observed SLR vs SPD04, (c) observed SLR vs Q04, (d) T04 vs Q04, (e) observed SLR vs QPF, and (f) observed SLR vs observed SWE.

The relationship between SPD04 and observed SLR is weaker, with a decreasing SLR trend with increasing SPD04 (Fig. 7a). This is due to the tendency for increasing wind speeds to fracture ice crystals, with the resulting fragments likely packing more densely than intact crystals (e.g., Steenburgh 2023). The relationship appears most useful for forecasting when values of SPD04 are $\geq$20 ms^{-1}. Under these conditions, only one instance of SLR >20 occurs in our dataset.

The relationship between Q04 and observed SLR is stronger than that of T04 and observed SLR (Fig. 7c). This is a bit surprising, as T04 and Q04 are very closely related (Fig 7d), with the Clausius-Clapeyron equation governing their relationship when the air is at saturation (Wallace and Hobbs 2006). The relationship between Q04 and observed SLR is also roughly the same shape (though the opposite sign) as that of T04 and observed SLR, with a trend reversal around Q04 $\cong$ 1.5 g kg^{-1}, below which observed SLR decreases. Yet an important factor appears to be that the trend reversal is not nearly as pronounced for Q04 as it is for T04. For moderate and high values of Q04, there is also much less variability in SLR for a given value of Q04 than there is for a moderate or high value of T04 (cf. 7a, 7c).

As for potential physical explanations of the Q04-SLR relationship, it is well established that at saturation in subfreezing clouds, increasing temperature (and therefore specific humidity) generally leads to increasing supercooled liquid water content (SLWC; e.g., Gultepe and Isaac 1997). Increasing SLWC, in turn, leads to increased riming (Waitz et al. 2022). Yet this would suggest temperature is just as important, at least near saturation, as specific humidity. This result is also at odds with previous studies, which conclude that temperature is more important. Therefore future work is needed to examine these physical processes, in particular the differences in snow growth over lowland regions compared to mountain regions.

HRRR QPF is negatively correlated with observed SLR (Fig 7e). This is likely due to the fact that an increasing amount of SWE on the interval board leads to increasing compaction of the entire column under its own weight. Supporting this assertion, the relationship between observed SWE and observed SLR is much stronger (Fig 7f). This means that for modelling systems with QPF forecasts that tend to be closer to reality, more skillful SLR prediction is possible. It also means that for non-forecasting applications, where observed SWE is available as a variable for SLR prediction, even greater forecasting skill than that of the V3 algorithm is possible.

g. V4 Algorithm

The V4 algorithm is identical to the V3 algorithm, except that it uses observed SWE instead of HRRR QPF. As mentioned above, this is only useful for applications that are run after a snowfall event is over, like snowfall analyses, in which observed SWE for the event is available.

The V4 algorithm, when compared to observed SLR, predicts SLR with R^2=0.54 and MAE=2.64. (Fig 8a). When the feature importance is calculated for the algorithm, using 1000 permutations, observed SWE is the greatest contributor to the fit by a wide margin, with a median relative importance of 0.28 (Fig 8b). Q04 is relegated to a distant 2nd place with a median importance of 0.13, followed by T04 in 3rd place at 0.07. The most dramatic illustration of V4's increased skill is how well it handles high SLR (>20) events (Fig 8a). It is clear that a major factor contributing to the difficulty of V1–V3 in forecasting high SLR events is that QPF forecasts have large errors. If SWE could be forecasted perfectly, high SLR events would not be nearly as challenging to predict. SWE amount is the single most important variable in predicting SLR accurately.

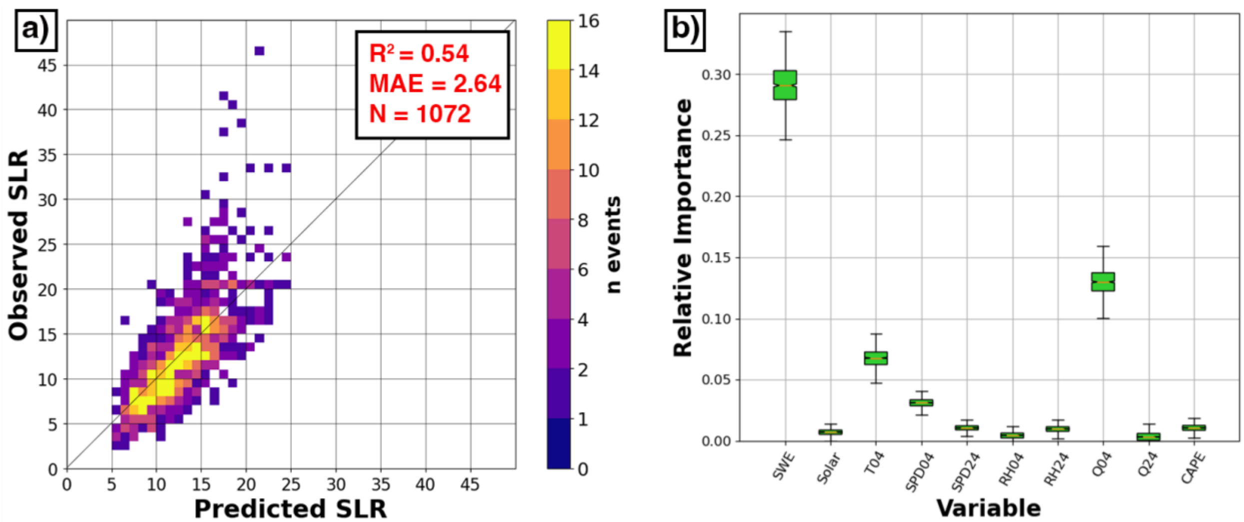

Figure 8. (a) Observed SLR vs SLR predicted by the V4 algorithm, evaluated against the 40% of observations that are withheld from training to be used for testing, using HRRR as the source of atmospheric input variables, shown here for one of the 1000 permutations for which R^2 and MAE were equal to the mean R^2 and mean MAE of the 1000 permutations. (b) For the V4 algorithm, the importance of each variable in the Random Forest regression. The box-and-whisker plots represent the different values of the coefficient for the 1000 permutations. The differences between boxes are statistically significant at the 95% confidence interval if the notches area around the medians do not overlap.

Because the V4 algorithm is trained with observed SWE, it is only skillful when it is applied to observed SWE. If it is applied to model QPF, it performs poorly (not shown). This is because model QPF often deviates substantially from, and is biased compared to, observed SWE. For model QPF to add skill, the algorithm must be trained on model QPF (this is what V2 and V3 are).

4. Summary and Conclusions

This study utilizes a novel dataset of manually-collected snowfall and SWE measurements taken every ~24-h or less from an interval board placed atop the snowpack by snow safety workers and scientists at 14 sites across the western US. We then use these observations and atmospheric variables from either the ERA5, the GFS, or the HRRR to train a MLR or RF algorithm to predict SLR. A simple model utilizing only temperature and wind speed at 2 levels as the predictive variables (known as V1) performs well when compared to observed SLR events that were withheld from training the algorithm, with R^2=0.32 and MAE=3.27 when using the GFS forecasts. When the training source was instead ERA5 reanalysis or HRRR forecasts, the R^2 and MAE values were within 1% of the above values.

These R^2 and MAE values from the simple V1 algorithm substantially outperform the 3 SLR algorithms (Cobb, MaxTaloft, and Roebber) that are used operationally in the NBM version 4.2. The respective R^2 values for those algorithms, when compared to observed SLR values, are 0.04, 0.17, and 0.23, with respective MAE values of 4.29, 6.51, and 9.45. The V1 algorithm also substantially outperforms a commonly used but unpublished algorithm known as the Kuchera method. When SLR predicted by Kuchera is compared to observed SLR, R^2=0.23 and MAE=4.84.

There are several potential explanations for the poor performance of the 4 legacy SLR algorithms relative to observations. The first is that these 4 algorithms were trained partly or entirely on SLR observations for which the liquid equivalent came from a precipitation gauge, and precipitation gauges, even with an Alter shield, can underatch falling snow by 40% or more (MacDonald and Pomeroy 2007; Thériault et al. 2012). Such undercatch would strongly affect the resulting SLR values. The observations used in our dataset are not immune to the effects of high winds, but using a weighed core from an interval board is substantially less error-prone than a precipitation gauge. Another issue is that algorithms like Cobb and MaxTAloft are mostly trained with variables from NWP over relatively flat terrain. In the complex terrain of the western US, steep slopes can dramatically increase or decrease vertical velocity, for example. Complex terrain may also cause most crystal growth and riming to happen at different heights above ground level than they do over flat terrain. Furthermore, the synoptic climatology of the sites used in this study may differ from that of the cyclone-dominated central and eastern US.

The V1 algorithm also substantially outperforms two other sources of SLR forecasts. If a fixed 12.0 SLR is used (12.0 yielded the best performance for our dataset) and compared to observed SLR, R^2=0.0 and MAE=4.01. The SLR forecasts from the NDFD are obtained for the 4 cool seasons that they are available from the archive (2020–2024) and compared to observed SLR. R^2=0.18 and MAE=3.82. The NDFD is a gridded aggregate of the forecasts issued by each NWS office.

We trained the V2 and V3 algorithms on the optimal combination of atmospheric variables from the HRRR to achieve the maximum possible skill, as defined by R^2 and MAE. The optimal combination of variables was T04, SPD04, SPD24, RH04, RH24, Q04, Q24, model QPF, CAPE, and Solar. The V2 algorithm is a MLR trained with these variables, and when its forecasted SLR is compared to observed SLR from a set of data points withheld for testing, the resulting R^2 and MAE are 0.39 and 3.05, respectively. The V3 algorithm is a RF model trained with the same variables. When V3 algorithm, which contains complex nonlinear relationships to predictive variables, is used to build an algorithm, the resulting R^2 and MAE are 0.43 and 2.94, respectively. For both the V2 and V3 versions of our algorithm, the most important variable is Q04 by a wide margin. In a distant 2nd, 3rd, and 4th place for importance in the V2 algorithm are SPD04, RH24, and model QPF, respectively. In a distant 2nd, 3rd, and 4th place for importance in the V3 algorithm are T04, SPD04, and model QPF, respectively. High SLR (>20) events are the most challenging for the algorithms, though these events account for 7% of the cases in the dataset.

The V4 algorithm is identical to V3, except that it uses observed SWE instead of HRRR QPF. This means it is only useful for applications that are run after a snowfall event is over, like snowfall analyses, in which observed SWE for the event is available. When forecasted SLR from V4 is compared to observed SLR from a set of data points withheld for testing, the resulting R^2=0.54 and MAE=2.64. The use of observed SWE as a predictor, instead of QPF, is responsible for the dramatic improvement in algorithm performance. The high SLR (>20) events do not pose nearly the forecasting challenge that they do for the V1, V2, and V3 algorithms. SWE becomes the most important variable by a wide margin, followed by Q04, T04, and SPD04 in a distant 2nd, 3rd, and 4th place, respectively. This shows that much of the difficulty in forecasting high SLR events is due to the uncertainty in QPF amounts. If QPF were a perfect forecast of SWE, high SLR events would not be nearly as difficult to predict.

In summary, the V1, V2, and V3 algorithms represent a large increase in skill for SLR prediction compared to the current algorithms and techniques used by forecasters, and the V4 algorithm is quite skillful for non-forecasting applications. But these algorithms were only trained to mountain sites with colder temperatures, located closer to cloud base, and different synoptic and mesoscale weather conditions than much of the lowlands and flatlands of the US. Therefore, future work will include developing an algorithm trained on high-quality manual snowfall observations from the entire US. A final important caveat is that because most of the legacy SLR algorithms are trained using precipitation gauge observations that often suffer from undercatch, these algorithms may verify more favorably when evaluated with such gauge observations. Conversely, the algorithms we've developed may not perform as well when these gauge observations are treated as the "truth".

Acknowledgements: This research was supported by NOAA grants NA19OAR4590137, NA22OAR4590531, and NA23OAR4310278. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding organization. Analyses were performed using Python software packages. We thank MesoWest, NOAA, NCAR, Unidata, and the University of Utah Center for High Performance Computing for the provision of datasets, software, and computing resources. We also thank Ned Bair, Aaron Barnett, Frank Carus, Steven Clark, Mike Cooperstein, Ella Darham, Daniel Nagy, Carol Peck, Jon Preuss, Mike Rheam, Dave Richards, Harlan Sheppard, John Stimberis, Jon Tukman, and Alan Willard for all their hard work gathering and sharing data. Jesse Meng was a huge help in getting the NBM algorithms coded up.

Data availability: The HRRR data were obtained from Amazon Web Services (https://registry.opendata.aws/noaa-hrrr-pds/). GFS data were obtained from the National Center for Atmospheric Research's Research Data Archive (https://rda.ucar.edu/datasets/d084001/). The ERA5 data were obtained from the Copernicus Climate Change Service (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels?tab=overview). The

NDFD data were obtained from Amazon Web Services (https://noaa-ndfd-pds.s3.amazonaws.com/index.html). The V1–V4 SLR algorithms are available on GitHib (https://github.com/pveals/Veals_etal_2025). The snowfall and SWE observations from CLN are available online from the University of Utah Research Data Repository (https://hive.utah.edu/concern/datasets/0r967383v). The CSSL and HLY observations are available on the GitHub (https://github.com/pveals/Veals_etal_2025), but due to their proprietary nature, the remainder of the snowfall and SWE observations are limited to access by verified researchers, and are available through the corresponding author.

References

Aikins, J., K. Friedrich, B. Geerts, and B. Pokharel, 2016: Role of a Cross-Barrier Jet and Turbulence on Winter Orographic Snowfall. *Mon. Wea. Rev.*, **144**, 3277–3300.

Alcott, T. I., and W. J. Steenburgh, 2010: Snow-to-liquid ratio variability and prediction at a high-elevation site in Utah's Wasatch Mountains. *Weather and Forecasting*, **25**, 323–337. https://doi.org/10.1175/2009waf2222311.1.

Bailey, M.P, and J. Hallett, 2009: A Comprehensive Habit Diagram for Atmospheric Ice Crystals: Confirmation from the Laboratory, AIRS II, and other Field Studies. *J. Atmos. Sci.*, **66**, 2888–2899.

Baxter, M. A., C. E. Graves, and J. T. Moore, 2005: A Climatology of Snow-to-Liquid Ratio for the Contiguous United States. *Wea. Forecasting*, **20**, 729–744, https://doi.org/10.1175/WAF856.1.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, doi:10.1175/MWR-D-15-

0242.1.

___, and coauthors, 2021: Diagnostic fields developed for hourly updated NOAA weather models. NOAA technical memo OAR GSL-66, 54pp, https://rapidrefresh.noaa.gov/Diag-vars-NOAA-TechMemo.pdf

Black, A.W., and T.L. Mote, 2015: Characteristics of Winter-Precipitation-Related Transportation Fatalities in the United States. *Wea. Clim. Society,* **7**, 133-145.

Blattenberger, G., and R. Fowles, 1995: Road closure to mitigate avalanche danger: a case study for Little Cottonwood Canyon. *Int. J. Forecast*., **11**, 159–174.

Breiman, L., 2001: Random Forests. *Machine Learning,* **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Broxton, P., X. Zeng, and N. Dawson, 2019: Daily 4 km gridded SWE and snow depth from assimilated in-situ and modeled data over the conterminous US, version 1. NASA National Snow and Ice Data Center Distributed Active Archive Center. **https://doi.org/10.5067/0GGPB220EX6A**

Byun, K., J. Yang, and T. Lee, 2008: A Snow-Ratio Equation and Its Application to Numerical Snowfall Prediction. *Wea. Forecasting*, **23**, 644–658, https://doi.org/10.1175/2007WAF2006080.1.

Chase, R.J., D.R. Harrison, A. Burke, G.M. Lackmann, and A. McGovern, 2023: A Machine Learning Tutorial for Operational Meteorology. Part I: Traditional Machine Learning. *Wea. Forecasting*, **37**, 1509–1529.

Cobb, D. K., and J. S. Waldstreicher, 2005: A simple physically based snowfall algorithm. *21st Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction*, Washington, DC, Amer. Meteor. Soc., 2A.2. http://ams.confex.com/ams/pdfpapers/94815.pdf.

Craven, J. P., Rudack, D. E., and Shafer, P. E., 2020: National blend of models: A statistically post-processed multi-model ensemble. *J. Operational Meteor.,* **8**, 1–14, https://doi.org/10.15191/nwajom.2020.0801.

Diffenbaugh, N.S, D.A. Swain, and D. Touma, 2015: Anthropogenic warming has increased drought risk in California. *Proc. Nat. Acad. Sci*, **112**, 3931–3936.

Friedrich, K., and coauthors, 2021: Microphysical Characteristics and Evolution of Seeded Orographic Clouds. *J. Appl. Meteorol. Climatol.*, **60**, 909–934.

Gardner, M.W., and S.R. Dorling, 1998: Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. *Atmos. Environment*, **32**, 14–15.

Geerts, B., Y. Yang, R. Rasmussen, S. Haimov, and B. Pokharel, 2015: Snow Growth and Transport Patterns in Orographic Storms as Estimated from Airborne Vertical-Plane Dual-Doppler Radar Data. *Mon. Wea. Rev.*, **143**, 644–665.

Gultepe, I., and G.A. Isaac, 1997: Liquid Water Content and Temperature Relationship from Aircraft Observations and Its Applicability to GCMs. *J. Climate*., **10**, 446–452.

Hagenstad, M., E. Burakowski, and R. Hill, 2018: The economic contributions of winter sports in a changing climate. Report published by Protect our Winters. https://protectourwinters.org/wp-content/uploads/2019/12/POW_EconReport_v22.pdf

Hersbach, H., and coauthors, 2020: The ERA5 global reanalysis. *Q. J. Roy. Meteor. Soc.*, **146**, 1999–2049.

Hoopes, C. A., C. L. Castro, A. Behrangi, M. R. Ehsani, P. Broxton, 2023: Improving prediction of mountain snowfall in the southwestern United States using machine learning methods. *Meteor. Appl.*, **30**, e2153, https://doi.org/10.1002/met.2153.

Janiszeski, A., R.M. Rauber, B.F. Jewett, and T. J. Zaremba, 2024: Reorganization of Snowfall

Beneath Cloud-top within the Comma Head Region of Two Extreme US East Coast Winter Cyclones. *Wea. Forecasting*, **39**, 1181–1202.

Judson, A., and N. Doesken, 2000: Density of freshly fallen snow in the central Rocky Mountains. *Bull. Amer. Meteor. Soc.*, **81**, 1577–1587.

Li, D., M.L. Wrzesien, M. Durand, J. Adam, and D. Lettenmaier, 2017: How much runoff originates as snow in the western United States, and how will that change in the future? *Geophys. Res. Lett.*, **44**, 6163–6172.

MacDonald JP, JW Pomeroy, 2007: Gauge Undercatch of Two Common Snowfall Gauges in a Prairie Environment. In *Proceedings of the 64th Eastern Snow Conference*, St. John's, Newfoundland, Canada: 119–126.

Mehta, A., 2023: Requirement and Standards for NWS Climate Observations. https://www.weather.gov/media/directives/010_pdfs/pd01013002curr.pdf

Nakaya, U., 1954: Snow Crystals: Natural and Artificial. Harvard University Press, 510 pp.

Neiman, P. J., F. M. Ralph, A. B. White, D. E. Kingsmill, and P. O. G. Persson, 2002: The Statistical Relationship between Upslope Flow and Rainfall in California’s Coastal Mountains: Observations during CALJET. *Mon. Weather Rev.*, **130**, 1468–1492.

NOAA, 1998: Automated Surface Observing System (ASOS) User’s Guide. https://www.weather.gov/media/asos/aum-toc.pdf.

NOAA EMC, 2024: The Global Forecast System. Accessed 31 January 2024, https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php

Pedregosa, F., and Coauthors, 2011: Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.*, **12**, 2825-2830. https://doi.org/10.48550/arXiv.1201.0490.

Pletcher, M.D, and coauthors, 2024: Validation of cool-season snowfall forecasts at a high-elevation site in Utah's Little Cottonwood Canyon. Early Online Release, *Wea and Forec*.

Pomeroy, J. W., and Brun, E., 2001: Physical properties of snow. *Snow Ecology*, H. G. Jones, Pomeroy, J. W., Walker, D. A., Hoham, R. W., Eds., Cambridge University Press, 45–126.

Potter, J. G., 1965: Water content of freshly fallen snow. Meteorology Branch, Dept. of Transport, CIR-4232, TEC-569, Toronto, ON, Canada, 12 pp. [Available from National Snow and Ice Data Center User Services, University of Colorado, Campus Box 449, Boulder, CO 80309-0449.]

Raleigh, M.S., and E.E. Small, 2017: Snowpack density modeling is the primary source of uncertainty when mapping basin-wide SWE with lidar. *Geophys. Res. Lett.*, **44**, 3700–3709.

Rasmussen, R., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. *Bull. Amer. Meteor. Soc.*, **93**, 811–829, https://doi.org/10.1175/BAMS-D-11-00052.1.

Roebber, P.J., S. L. Bruening, D. M. Schultz, and J. V. Cortinas, 2003: Improving Snowfall Forecasting by Diagnosing Snow Density. *Wea. Forecasting*, **18**, 264–287, https://doi.org/10.1175/1520-0434(2003)018<0264:ISFBDS>2.0.CO;2.

Rosenow, A.A., H.D. Reeves, and D. Tripp, 2023: Evaluation of Ensemble Snowfall Forecasts Using Operationally Used Snow-to-Liquid Ratios in Five Winter Storms. *Wea. and Forec.*, **38**, 2135–2147.

Rudack, D., R. James, S. Perfater, G. Manikin, and A. Just, 2024: National Blend of Models (NBM) v4.2 Upgrades and Improvements, 40 pp. https://vlab.noaa.gov/documents/6609493/6665561/NBM_v4.2_Eval_SlideDeck.pdf

Seeherman, J., and Liu, Y., 2015: Effects of extraordinary snowfall on traffic safety. Accid. Anal. Prev., **81**, 194-203, https://doi.org/10.1016/j.aap.2015.04.029.

Spencer, J.M., 2009: Winter Weather Related Fatalities in the Conterminous United Sates: An Analysis of Three Winter Fatality Databases, M.S. Thesis, Dept of Geography, Northern Illinois University, 91 pp. https://www.proquest.com/openview/4dfd6aff74007b68087582a4ffd811ea/1?pq-origsite=gscholar&cbl=18750

Steenburgh, W.J., 2023: *Secrets of the Greatest Snow on Earth*. 2nd Ed. Utah State University Press, 215 pp.

The COMET Program, 2023: Understanding NBM v4.0 snowfall products. The University Corporation for Atmospheric Research, 22 May 2024, https://www.meted.ucar.edu/education_training/lesson/10166.

Thériault, J., R. Rasmussen, K. Ikeda, and S. Landolt, 2012: Dependence of snow gauge collection efficiency on snowflake characteristics. *J. Appl. Meteor. Climatol.*, **51**, 745–762, https://doi.org/10.1175/JAMC-D-11-0116.1.

Thompson, G., and G.M. Lackmann, 2019: Hydrometeor Lofting and Mesoscale Snowbands. *Mon. Wea. Rev.*, **147**, 3879–3899.

Trujillo, E., and N.P. Molotch, 2014: Snowpack regimes of the Western United States. *Water Resour. Res.*, **50**, 5611–5623.

Vapnik, V., 1995: *The Nature of Statistical Learning*. Springer, 314pp.

Veals, P.G., W.J. Steenburgh, S. Nakai, and S. Yamaguchi, 2020: Intrastorm Variability of the Inland and Orographic Enhancement of a Sea-Effect Snowstorm in the Hokuriku Region of Japan. *Mon. Wea. Rev.*, **148**, 2527–2548.

Waitz, F., M. Schnaiter, T. Leisner, and E. Järvinen, 2022: In situ observations of riming in mixed-phase clouds using the PHIPS probe. *Atmos. Chem. Phys.*, **22**, 2087–7103.

Wilks, D.S., 2019: *Statistical Methods in the Atmospheric Sciences*. Fourth Edition. Elsevier, 840 pp.

Yuter, S. E., D. A. Stark, J. A. Crouch, M. J. Payne, and B. A. Colle, 2011: The impact of varying environmental conditions on the spatial and temporal patterns of orographic precipitation over the Pacific Northwest near Portland, Oregon. *J. Hydrometeor.*, **12**, 329–351.