

Snow-to-Liquid Ratio Prediction over the Contiguous United States Using Machine Learning

MICHAEL D. PLETCHER^a, PETER G. VEALS,^a RANDY J. CHASE,^b STEVEN D. HILBERG,^c NOAH NEWMAN,^c ANDREW A. ROSENOW,^d AND W. JAMES STEENBURGH^a

^a *Department of Atmospheric Sciences, The University of Utah, Salt Lake City, Utah*

^b *Tomorrow.io, Boston, Massachusetts*

^c *Community Collaborative Rain, Hail, and Snow Network (CoCoRaHS) snowfall observations, station geographic information, and fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) vertical profiles of atmospheric variables, we developed a random forest (RF)-based algorithm for predicting SLR across the CONUS. SLR forecasts and quantitative snowfall forecasts (QSFs) based on this algorithm were then derived from the NOAA High-Resolution Rapid Refresh (HRRR). The resulting RF SLR mean absolute error (MAE) was 2.92 across the CONUS compared to MAEs for current operational methods of 3.51–4.76. Improvements were also seen in QSF. Additionally, the RF outperformed operational SLR methods within individual regional snow climates. Despite these improvements, there were situations when the RF struggled, such as high (>20) and low (<5) SLR events that featured relatively cold and calm and warm and windy environments, respectively, in the eastern CONUS. The use of CoCoRaHS observations for training provides a template for improving snowfall forecasts at a CONUS-wide scale with the RF providing an algorithm for improving SLR prediction for operational applications.*

^d *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma and NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma*

(Manuscript received 20 June 2025, in final form 29 October 2025, accepted 19 December 2025)

ABSTRACT: Winter storms have widespread impacts across the contiguous United States (CONUS), resulting in travel and economic disruptions, loss of life, and damage to property and infrastructure. Snow-to-liquid ratio (SLR), the ratio of new (i.e., freshly fallen) snow to liquid precipitation equivalent, is used operationally to forecast snowfall amount during winter storms. However, no clear best prediction method for SLR has been established, and some operational SLR prediction methods were developed for specific regions, potentially limiting their utility for CONUS-wide applications. Using Community Collaborative Rain, Hail, and Snow Network (CoCoRaHS) snowfall observations, station geographic information, and fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) vertical profiles of atmospheric variables, we developed a random forest (RF)-based algorithm for predicting SLR across the CONUS. SLR forecasts and quantitative snowfall forecasts (QSFs) based on this algorithm were then derived from the NOAA High-Resolution Rapid Refresh (HRRR). The resulting RF SLR mean absolute error (MAE) was 2.92 across the CONUS compared to MAEs for current operational methods of 3.51–4.76. Improvements were also seen in QSF. Additionally, the RF outperformed operational SLR methods within individual regional snow climates. Despite these improvements, there were situations when the RF struggled, such as high (>20) and low (<5) SLR events that featured relatively cold and calm and warm and windy environments, respectively, in the eastern CONUS. The use of CoCoRaHS observations for training provides a template for improving snowfall forecasts at a CONUS-wide scale with the RF providing an algorithm for improving SLR prediction for operational applications.

SIGNIFICANCE STATEMENT: Despite improved weather forecasts in recent years, forecasting winter storms remains challenging. Using manual snowfall observations collected by trained observers across the contiguous United States (CONUS), we developed a machine learning method to predict snow-to-liquid ratio (SLR), an important variable used to forecast snowfall and related hazards during winter storms. This machine learning method outperforms existing methods used operationally by the National Weather Service (NWS) and is available for application with operational weather prediction models for improved winter storm forecasts.

KEYWORDS: Snowfall; Winter/cool season; Forecast verification/skill; Forecasting techniques; Machine learning

1. Introduction

Winter storms can have dramatic societal impacts including travel and economic disruptions, loss of life, and damage to property and infrastructure. On average in the contiguous United States (CONUS), ~900 fatalities are attributed to winter-precipitation-related vehicle and aviation accidents annually (Black and Mote 2015). Downwind of the Great Lakes and along the Interstate Highway 95 (I-95) corridor, populated cities can be gridlocked by multiday winter storms associated with lake-effect precipitation or extratropical cyclones, with damages exceeding \$100 million (U.S. dollars) in major events (Sanders and Bosart 1985; Kocin and Uccellini 1990; Kocin and Schumacher 2004; Kocin et al. 1995; Niziol

et al. 1995; Greybush et al. 2017; Lazo et al. 2020; Novak et al. 2023). Despite advancing skill for winter weather forecasts (Novak et al. 2023, see their Fig. 1), forecasting winter storms remains challenging, with precipitation type and amount, snowfall intensity and distribution, precipitation onset, and snow level critical for reducing loss of life and travel and economic disruptions (Steenburgh 2003; Byun et al. 2008; Alcott and Steenburgh 2010; Picca et al. 2014; Ganetis and Colle 2015; Seeherman and Liu 2015; Greybush et al. 2017; Gerber et al. 2018, 2019; Pletcher et al. 2024). Snowfall amount can be especially difficult to forecast as producing a quantitative snowfall forecast (QSF) typically involves precipitation-type identification, a quantitative precipitation forecast (QPF), and specification of a snow-to-liquid ratio (SLR), all three of which can vary in space and time and represent potential sources of QSF error.

SLR, or the ratio of new (i.e., freshly fallen) snowfall to liquid precipitation equivalent (LPE), historically was assumed

Corresponding author: Michael D. Pletcher, michael.pletcher@utah.edu

to follow the 10:1 rule [e.g., 10-mm QPF produces 10-cm QSF (Potter 1965; Judson and Doesken 2000)] but in reality varies over short distances and time scales due to variations in ice crystal habit, melting and freezing processes at the surface or aloft, precipitation-type changes, surface-snow compaction, and other processes, limiting effectiveness for many winter storm forecasting applications (Pomeroy and Brun 2001; Roebber et al. 2003; Baxter et al. 2005; Cobb and Waldstreicher 2005; Byun et al. 2008; Alcott and Steenburgh 2010; Pletcher et al. 2024; Wasserstein and Steenburgh 2024; Veals et al. 2025). Therefore, techniques have been developed to predict SLR using physical reasoning (e.g., Cobb and Waldstreicher 2005) or machine learning (e.g., Roebber et al. 2003; Hoopes et al. 2023; Pletcher et al. 2024). Often these techniques focus on specific locations or regions, which may introduce bias when applied to other areas.

Recently, machine learning (ML) applications in the atmospheric sciences have increased considerably due to their performance and computational benefits (McGovern et al. 2017; Chase et al. 2022, 2023; Roebber and Smith 2023). The earliest examples of ML in atmospheric sciences used linear regression to forecast surface weather variables such as temperature and precipitation, often based on training and application to NWP forecasts or analyses (e.g., Klein et al. 1959; Glahn and Lowry 1972; Kitzmiller et al. 1995; Billet et al. 1997; Hamill et al. 2004; Baars and Mass 2005). Deep learning (DL), a subset of ML, has also been used to forecast winter weather, with Roebber et al. (2003) among the earliest to apply DL for SLR prediction. Other winter weather ML applications have shown promise for improved snowfall and SLR forecasts over complex terrain (e.g., Sha et al. 2020; Hoopes et al. 2023; Pletcher et al. 2024; Veals et al. 2025), estimates of snowpack characteristics (e.g., Cui et al. 2023; Mayer et al. 2022), precipitation-type identification (e.g., McGovern et al. 2017; Filipiak et al. 2023), and orographic precipitation gradients (e.g., Wolvin et al. 2024).

Random forests (RFs; Breiman 2001) have been used extensively in the atmospheric sciences due to their computational speed, ability to extract nonlinear feature-label relationships, ease of use in open-source software packages, and overall performance (Pedregosa et al. 2011; Herman and Schumacher 2018a; Chase et al. 2022). For example, RFs can outperform more complex architectures like those based on neural networks when trained properly (Grinsztajn et al. 2022). RFs, a form of supervised learning, are ensembles of decision trees that make deterministic predictions by aggregating predictions from each decision tree in the ensemble (Breiman 2001). RFs have been used to forecast convective weather (e.g., Gagne et al. 2017; Hill et al. 2020, 2023; Mecikalski et al. 2021; Loken et al. 2022), precipitation (e.g., Gagne et al. 2014; Herman and Schumacher 2018a,b; Loken et al. 2019), aviation weather (e.g., Williams 2014; McGovern et al. 2017), and winter weather (e.g., Pletcher et al. 2024; Veals et al. 2025), but we are unaware of previous efforts to use RFs to forecast CONUS-scale SLR.

One challenge for CONUS-scale SLR prediction is the need for reliable SLR observations for ML training and validation across many regional snow climates. SLR requires the

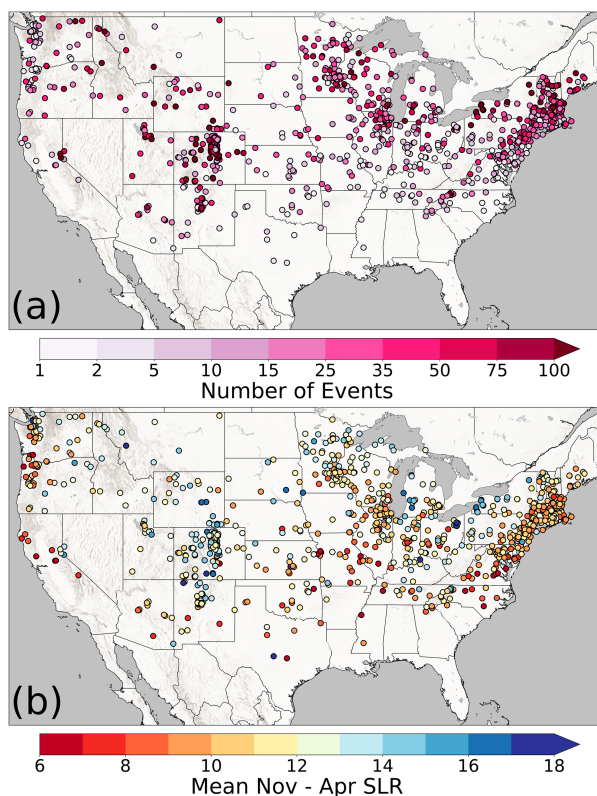


FIG. 1. (a) Number of SLR events and (b) mean SLR at each of the 921 CoCoRaHS sites used in this study.

measurement of new snow depth and water equivalent. Sometimes, the latter is determined using automated gauge measurements, which can suffer from undercatch, resulting in a high SLR bias (MacDonald and Pomeroy 2007; Rasmussen et al. 2012; Thériault et al. 2012). A method that can reduce SLR measurement errors is the use of skilled, manual measurements of the water equivalent of new snowfall using a snowboard and coring tube, an approach that is often used by avalanche forecasters. Recently, Pletcher et al. (2024) and Veals et al. (2025) utilized such manual snowfall measurements to train and test SLR forecasts produced by RFs at mountainous locations over the western CONUS.

In this paper, we demonstrate the value of the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS) as a suitable training, validation, and testing dataset for CONUS-scale SLR prediction. CoCoRaHS spans all 48 CONUS states and involves volunteers measuring LPE and snowfall daily during storms, typically in the morning (Cifelli et al. 2005; Reges et al. 2016). Using CoCoRaHS, we develop an RF SLR prediction algorithm based on training to station geographic information and fifth generation European Centre for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) atmospheric profiles. We then apply and evaluate the RF using the High-Resolution Rapid Refresh (HRRR; Dowell et al. 2022) forecasts.

The remainder of this paper is organized as follows. Section 2 describes the data and methods used, including the CoCoRaHS

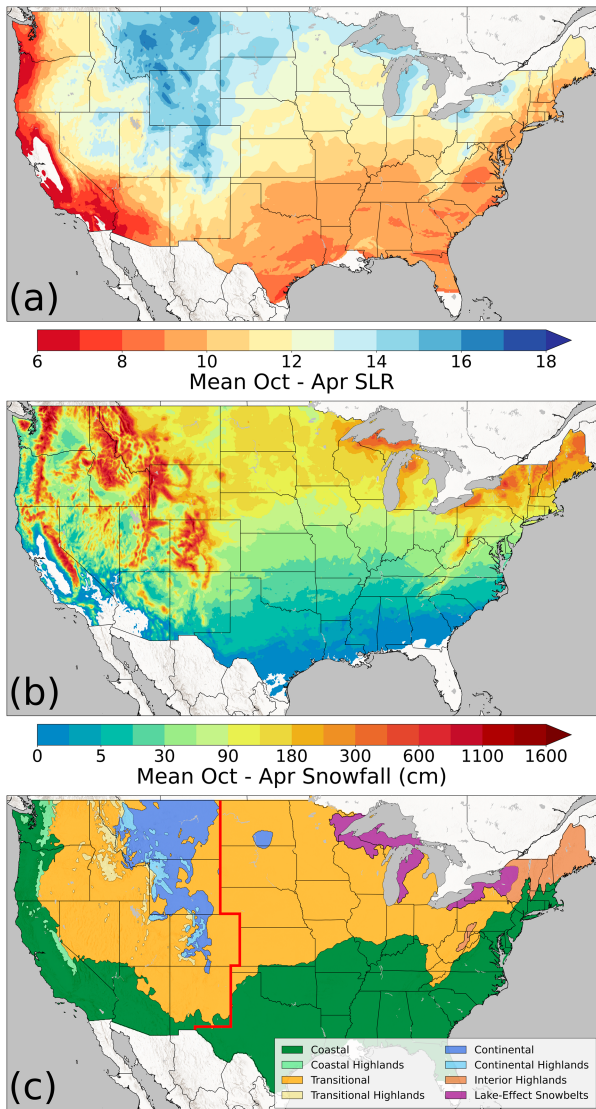


FIG. 2. CONUS maps of (a) mean October–April SLR from B2005, (b) mean 2007/08–2023/24 October–April NSA snowfall (cm), and (c) the resulting eight regional snow climates following thresholds in Table 1. The red line in (c) separates the western and eastern CONUS.

network and its regional SLR climatology, the RF methodology, current operational SLR methods, and the evaluation metrics. Section 3 examines the CONUS-wide performance of the RF and operational SLR methods for SLR and QSF. Sections 4 and 5 present the results of these algorithms for the eastern and western CONUS, respectively, and section 6 concludes with a summary of findings and ideas for future work.

2. Data and methods

a. CoCoRaHS SLR observations

We use 24-h snowfall and LPE measurements collected at ~0700 LST by CoCoRaHS volunteers to derive daily

TABLE 1. List of B2005 and NSA thresholds used to determine the eight snow climates.

Snow climate	Threshold	
	East	West
Coastal	Baxter SLR < 10.5 NSA < 175 cm	Baxter SLR < 10.5 NSA < 450 cm
Coastal highlands	—	Baxter SLR < 12 NSA ≥ 450 cm
Transitional	Baxter SLR ≥ 10.5 NSA < 175 cm	Baxter SLR ≥ 10.5 NSA < 450 cm
Transitional highlands	—	Baxter SLR ≥ 12 NSA ≥ 450 cm
Continental	Baxter SLR ≥ 14.5 NSA < 175 cm	Baxter SLR ≥ 14.5 NSA < 450 cm
Continental highlands	—	Baxter SLR ≥ 14.5 NSA ≥ 450 cm
Interior highlands	NSA ≥ 175 cm	—
Lake-effect snowbelts	Baxter SLR ≥ 12.5 NSA ≥ 175 cm LAT ≥ 40.5° LON ≥ −94°	—

SLR observations during the 2000/01–2023/24 cool seasons (November–April). These SLRs were then used as labels for ML training, validation, and testing. To reduce uncertainties from gauge undercatch, we only included measurements from observers who obtained snowfall using a snowboard and manually melted snowfall or weighed snowfall on a scale to obtain LPE. This was done by sending an online questionnaire survey to 29 519 CoCoRaHS volunteers, with 1182 sites responding to the survey. We focus on 921 sites that responded to the survey and confirmed the explicit use of a core sample to obtain LPE (Fig. 1).

SLR was based on the ratio of snowfall and LPE and limited to observations with at least 2.79 mm (0.11 in.) of LPE and 5.09 cm (2 in.) of snowfall. These thresholds reduce the influence of rounding and measurement errors on SLR calculations and have been used in previous studies (e.g., Judson and Doesken 2000; Roebber et al. 2003; Alcott and Steenburgh 2010; Pletcher et al. 2024; Veals et al. 2025). Additionally, observed SLRs of exactly 10, <2, or >50 were removed as 10:1 values were sometimes used as placeholders when LPE or snow depth observations were missing, and SLRs < 2 or >50 are prone to measurement and rounding errors (Veals et al. 2025).

In the eastern CONUS (defined for this paper as east of the eastern borders of Montana Wyoming, Colorado, and New Mexico), sites with the largest number of observed SLR events are near the Great Lakes and in northern New York and New England where there is relatively abundant lake effect and synoptic snowfall [Fig. 1a (Niziol et al. 1995; Novak et al. 2004)]. Sites with a relatively large number of observed SLR events are also found along the northeast seaboard and in portions of the upper Midwest. Consistent with Baxter et al. (2005, hereinafter B2005), whose 30-yr climatology of SLR for the CONUS used daily National Weather Service (NWS) Cooperative Observer Program (COOP) observations of

TABLE 2. Descriptions of atmospheric and geographic features used for RF training.

Feature	Levels	Group
Temperature (T , K)	300, 600, 900, 1200, 1500, 1800, 2100, 2400 m AGL	Atmospheric
Wind speed (SPD, m s^{-1})	300, 600, 900, 1200, 1500, 1800, 2100, 2400 m AGL	Atmospheric
RH (R , %)	300, 600, 900, 1200, 1500, 1800, 2100, 2400 m AGL	Atmospheric
Latitude (LAT, $^{\circ}$)	—	Geographic
Longitude (LON, $^{\circ}$)	—	Geographic
Site elevation (ELEV, m MSL)	—	Geographic

LPE and snowfall from more than 7000 stations, observed November–April mean SLRs increase from $\sim 10:1$ along the eastern seaboard to higher values over the interior, including $>16:1$ downstream of the Great Lakes (Fig. 1b). There is, however, large regional and local variability in both the number of events and mean SLR. While some of this variability reflects the regional snow climate, it is also influenced by the irregular sampling of winter storms by the volunteer CoCoRaHS observers.

In the western CONUS (defined for this paper as west of the eastern borders of Montana, Wyoming, Colorado, and New Mexico), sites with the highest number of observed events are near and in the Colorado Front Range, near Lake Tahoe, and in some intermountain areas. Mean SLRs are lowest at coastal sites and highest at interior intermountain and high elevation sites, consistent with prior studies (B2005; Veals et al. 2025). We also use observations from 14 snow-safety sites described in Veals et al. (2025) as an independent testing dataset to examine the capabilities of the RF in data-sparse areas. These site observations, however, which

are high quality and are more frequent due to abundant mountain snowfall, were not used for RF training due to our desire to evaluate CoCoRaHS as a dataset for developing CONUS-wide SLR algorithms.

b. Regional snow climates

Given the large regional variations in SLR and snowfall, we used the mean October–April SLR from B2005 (Fig. 2a) and mean 2007/08–2023/24 October–April snowfall from the National Operational Hydrologic Remote Sensing Center (NOHRSC) National Snow Analysis (NSA; Carroll et al. 2001; Fig. 2b) and to divide the CONUS into eight regional snow climates (Fig. 2c) for training, validation, and testing. Guided by a k -means clustering (Hartigan and Wong 1979) first guess and previously identified snow climates in the literature (e.g., Eichenlaub 1979; Niziol et al. 1995; Serreze et al. 1998; Mock and Birkeland 2000), we established thresholds for mean snowfall and SLR in the eastern and western CONUS (Table 1) to identify eight snow climates. These are coastal, which also includes low SLR and snowfall regions of

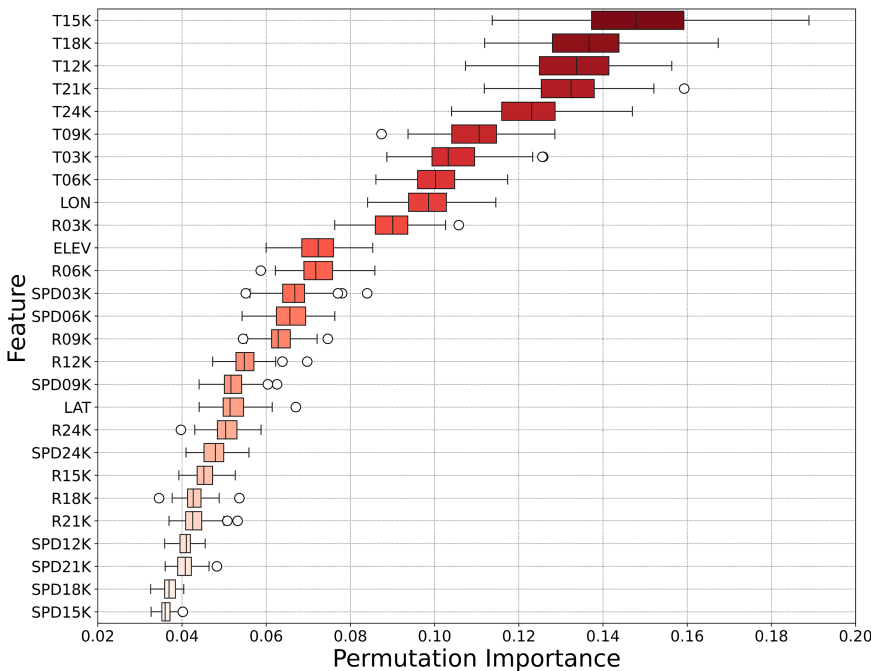


FIG. 3. Boxplots of aggregate permutation feature importance obtained from the 100 unique training datasets for each of the 27 features used to train the RF. The color of the boxplots corresponds to the feature importance value.

TABLE 3. Descriptions of each operational SLR method.

Method	Description
MaxTAloft	Maximum temperature between 2000 ft AGL and 400-hPa relationship based on Alaska snowfall cases (NBM v4.2)
Cobb	925–300-mb layer snow ratio, vertical velocity, and RH (NBM v4.2)
Thickness	850–700-mb thickness relationship based on 63 mid-Atlantic snowfall cases (NBM v4.2; evaluated in eastern CONUS only)

the interior southern CONUS, coastal highlands, transitional, transitional highlands, continental, continental highlands, northeast (NE) interior highlands, and lake-effect belts. Due to a low number of sites in the coastal highland, transitional highland, and continental highland snow climates, they were

combined with their lowland complementing snow climates (coastal, transitional, and continental) in the western CONUS. Thus, the regional training, validation, and testing were performed using the coastal, transitional, continental, interior highlands, and lake-effect snowbelt snow climates.

c. Atmospheric and geographic feature assembly

We used daily CoCoRaHS SLR observations, vertical profiles from the ECMWF ERA5 (Hersbach et al. 2020), and HRRR forecasts to train, validate, and test several ML-based SLR prediction algorithms regionally and CONUS wide, ultimately deciding to focus on an RF. Twenty-one cool seasons (2000/01–2021/22) were used for training and validation using station geographic information and ERA5 data, whereas two cool seasons (2022/23–2023/24) of HRRR forecasts were used for testing the application of the RF within an operational modeling system, including comparisons with operational

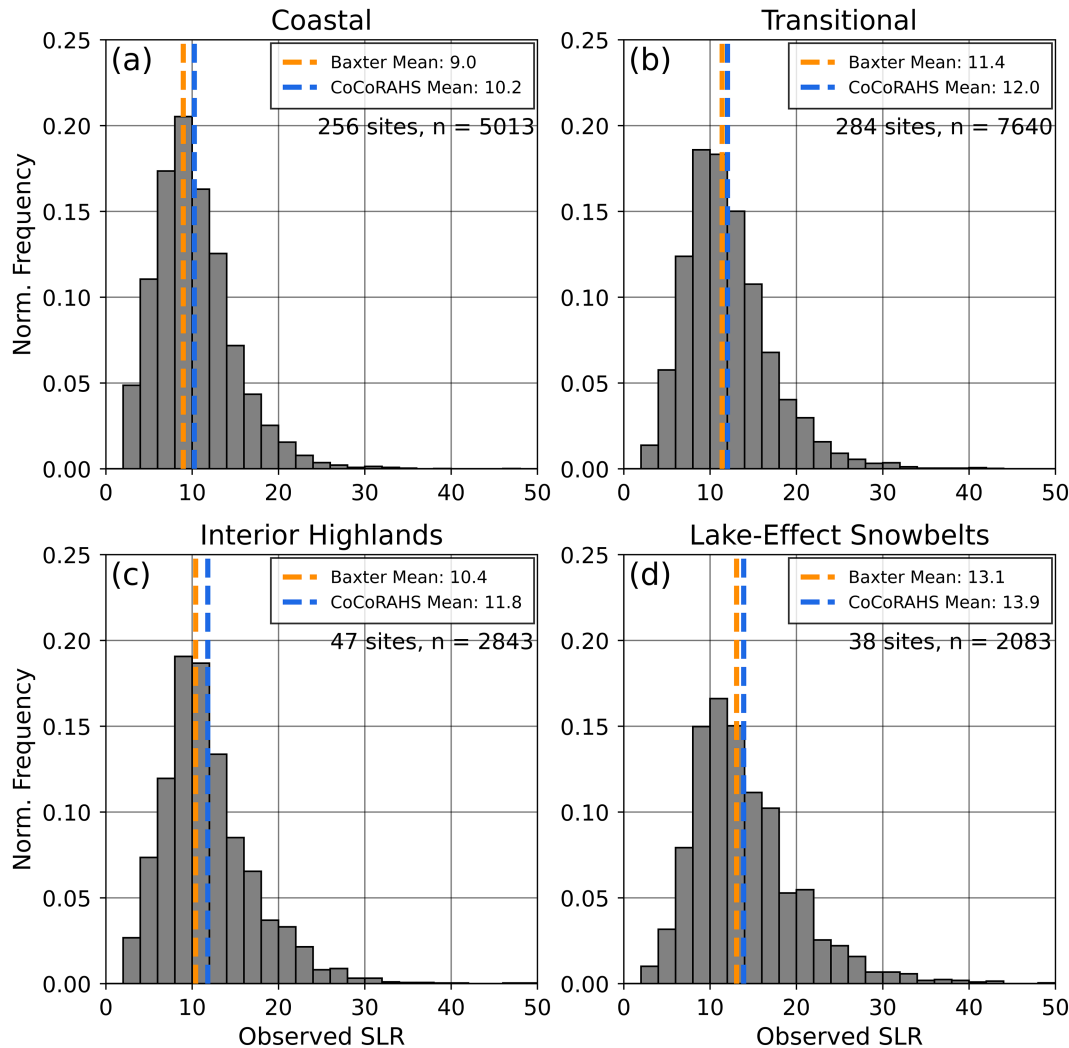


FIG. 4. 1D histograms of observed SLR for the (a) coastal, (b) transitional, (c) interior highlands, and (d) lake-effect snowbelt eastern CONUS snow climates. B2005 and CoCoRaHS means are annotated as yellow and blue vertical lines, respectively. Number of sites and events are annotated below top-right insets.

SLR techniques. The training, validation, and testing datasets contained 13 987, 9325, and 7863 events, respectively.

Based on prior studies (e.g., Roebber et al. 2003; Cobb and Waldstreicher 2005; Alcott and Steenburgh 2010; Herman and Schumacher 2018a; Pletcher et al. 2024), we chose temperature (T), wind speed (SPD), and relative humidity (R) at multiple levels above ground level (AGL) as atmospheric features and the site latitude (LAT), site longitude (LON), and site elevation (ELEV) as geographic features for training (Table 2). Although using latitude and longitude as input features has reduced RF performance for some applications (Milà et al. 2024), we found they slightly improved the performance metrics defined later in section 2f and include them in the RF training. Although LPE is an important SLR predictor, we did not use it here due to the differing resolution and biases of ERA5 precipitation analyses compared to the HRRR, which degraded RF performance. Other atmospheric features such as cloud-top temperature, lapse rate, 500-hPa geopotential height, specific humidity, vertical velocity, and maximum daily solar angle were considered but not included due to a limited or negative impact on skill.

The T , SPD, and R were obtained from the nearest ERA5 and HRRR profiles to each CoCoRaHS site. Due to the high density of CoCoRaHS sites in some regions (e.g., Colorado's Front Range, Fig. 1) and the ERA5's coarse horizontal resolution, some CoCoRaHS sites were matched with the same ERA5 profile. We did not perform any spatial interpolation of the ERA5 profiles to address this as it did not affect the performance metrics defined later in section 2f during the RF training. The T , SPD, and R were then linearly interpolated from pressure coordinates to 300–2400 m AGL (defined relative to the CoCoRaHS station elevation) at 300-m intervals. “SPD21K” is used to represent SPD at 2100 m AGL, “Q15K” is for Q at 1500 m AGL, “T09K” is for T at 900 m AGL, and so on. Increasing the vertical resolution or lowering the lowest level to 100 m provided no statistically significant increase in skill but added computational load. The use of AGL as a vertical coordinate enables the RF to be applied to any operational modeling system after data from that modeling system are interpolated to AGL. Internal testing showed that the 24-h mean (vs maximum or minimum) of the hourly atmospheric features yielded optimal performance for training.

d. Machine learning algorithms

We selected a CONUS-scale RF as the primary model and developed it using the scikit-learn Python package (Pedregosa et al. 2011). RFs are ensemble algorithms that can be used to make probabilistic or deterministic predictions. Here, we focus on the latter. To introduce randomness during training, the RF randomly selects subsets of features and labels [also referred to as predictions (or forecasts) and observations, respectively] for each tree. Deterministic predictions are then made by aggregating the predictions made by each tree in the ensemble, with the final output an average of each tree's predictions.

During model validation, we determined that the CONUS-scale RF outperformed regionally trained RFs based on the

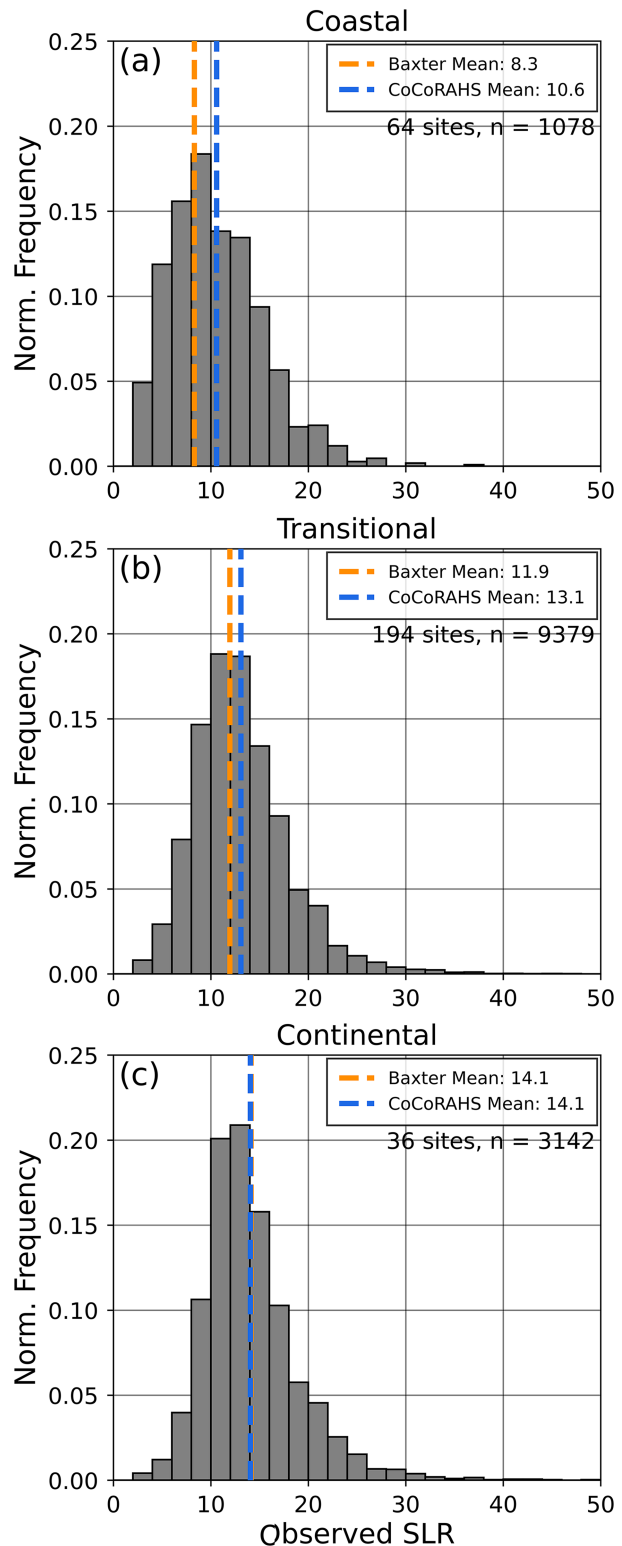


FIG. 5. As in Fig. 4, but for the (a) coastal, (b) transitional, and (c) continental western CONUS snow climates.

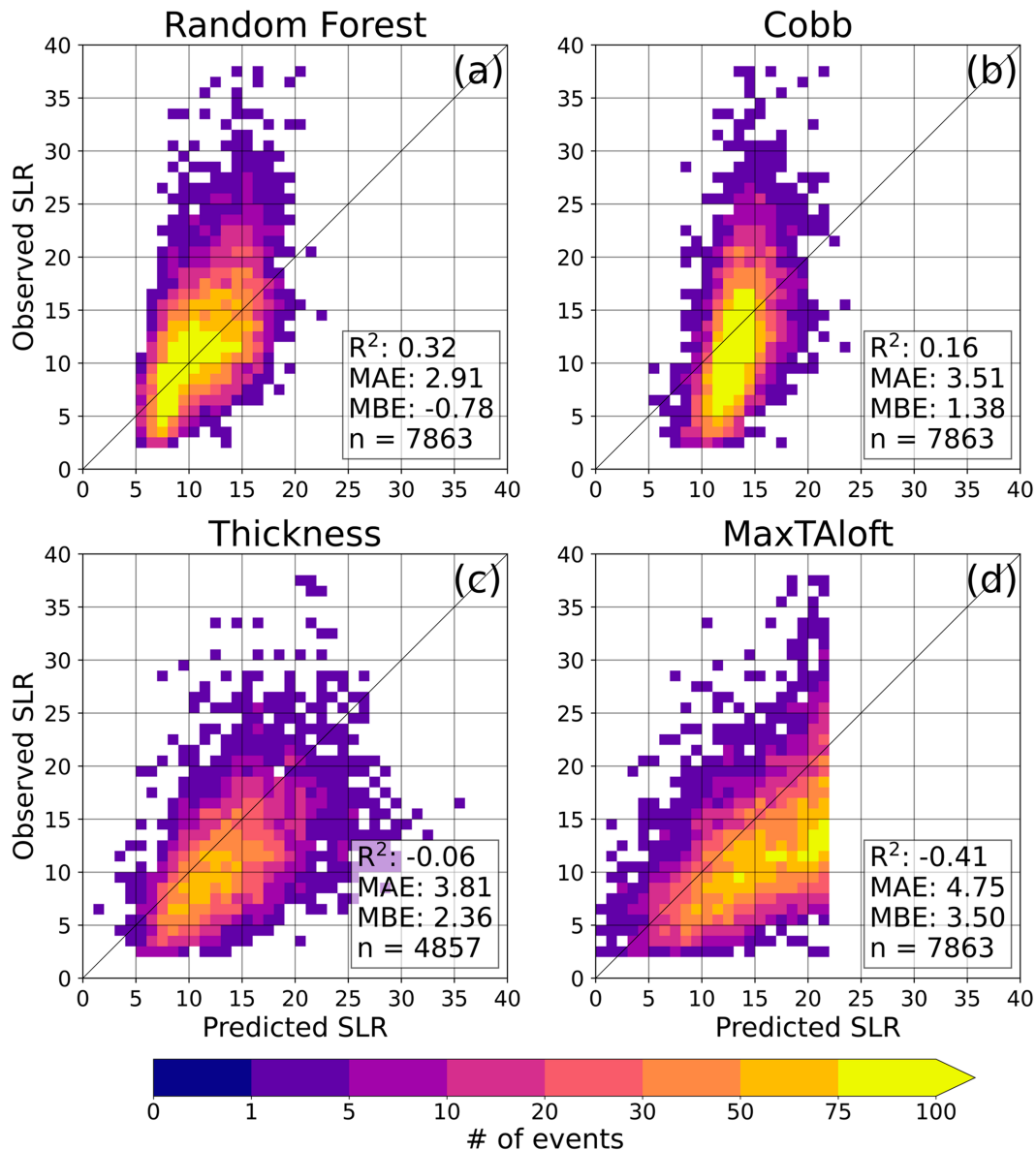


FIG. 6. Bivariate histograms of predicted and observed SLR for (a) RF, (b) Cobb, (c) Thickness, and (d) MaxTAloft for the 2022/23–2023/24 HRRR validation period. The R^2 , MAE, MBE, and number of events are annotated at bottom right. Bin width is set to 1.

five snow climates (coastal, transitional, continental, interior highlands, lake-effect snowbelts); thus, only the results from the CONUS-scale RF are included, which is named RF hereafter. We were also requested by NCEP to produce a faster, globally applicable, SLR model. This model, which is simpler and described in the [appendix](#), was based on linear regression (LR) and exhibited degraded performance compared to the RF.

To train and validate the RF, we performed k -fold cross validation ([Kohavi 2001](#)) using 100 randomized 60/40 train/validate splits during the 21 cool-season training period. The training and validation sets are separated temporally from the testing set to prevent overfitting that can result from training

and testing on the same events. We treat the RF's mean performance metrics from these 100 randomized splits as representative of its overall performance, with these mean performance metrics provided in the [appendix](#).

After training and validation, we used the 100 unique training datasets to establish the order of the RF's feature importances for making predictions. This was done using the `permutation_importance` function from `scikit-learn`, which determined that T15K, T18K, T12K, T21K, and T24K were the five most important features ([Fig. 3](#)). This is consistent with other studies that find a strong correlation between SLR and temperature and makes physical sense, given the importance of temperature for crystal habit, melting (if $>0^\circ\text{C}$), aggregation,

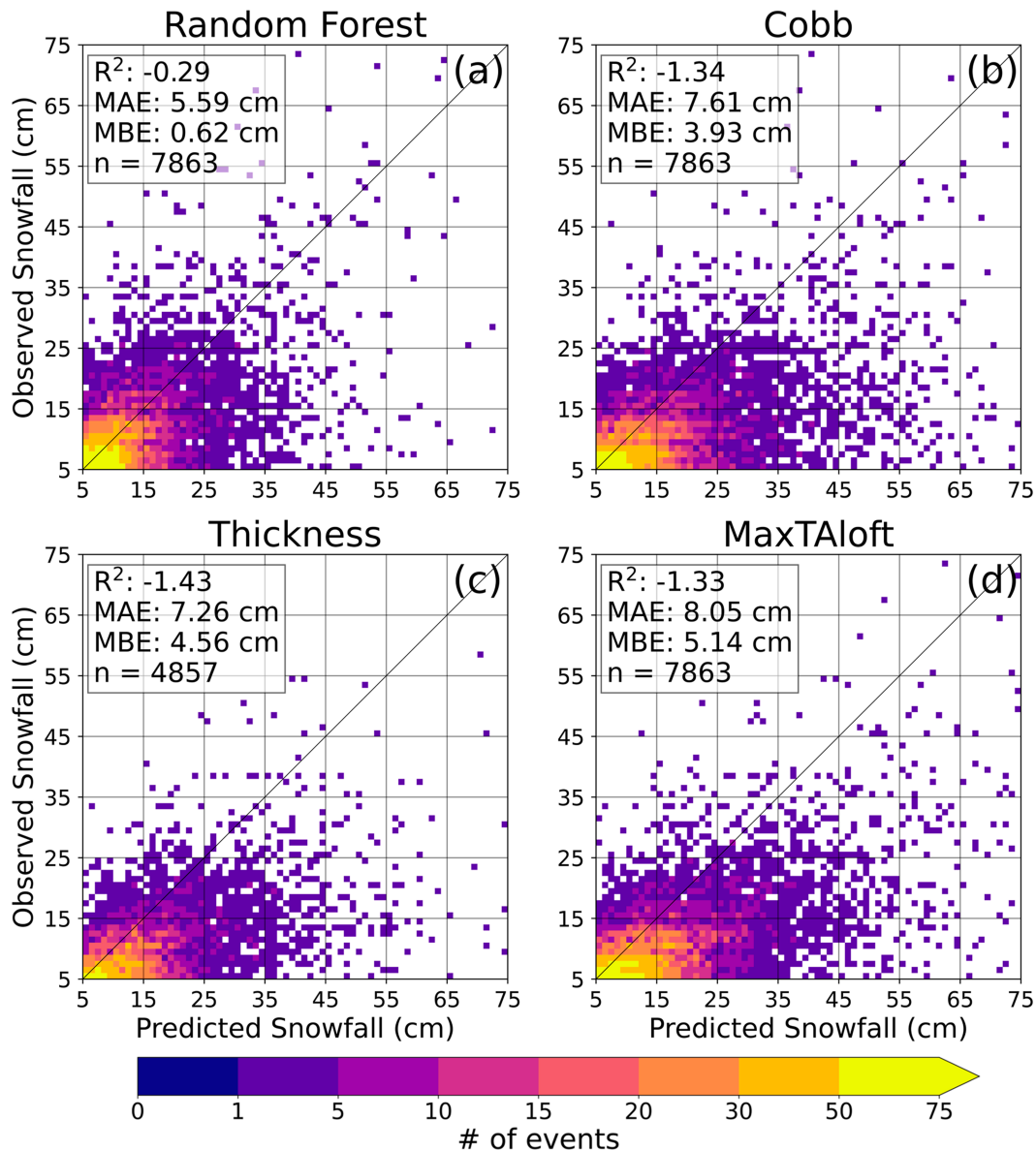


FIG. 7. As in Fig. 6, but for QSF (cm) derived from each SLR method.

and other processes (Nakaya 1954; Power et al. 1964; Roebber et al. 2003; Bailey and Hallett 2009; Alcott and Steenburgh 2010; Pletcher et al. 2024). Permutation importance is a model examination technique that is useful for evaluating the contribution of each feature to the RF's aggregate performance from the 100 training iterations. Permutations are invoked by randomly shuffling values of one feature at a time and observing the degradation of the RF's performance based on the training dataset (Breiman 2001).

We then tuned the RF's hyperparameters using a randomized search scoring method based on the 100 unique validation sets (scikit-learn; RandomizedSearchCV). While keeping input features constant, the optimal RF hyperparameters were identified as follows: 100 trees, a maximum depth of 100 nodes, a minimum of 2 samples for both node splitting and leaf nodes, the

square of the number of features for the best split search, and bootstrap sampling for tree building. These parameters were selected by RandomizedSearchCV as they maximized the coefficient of determination (R^2 ; Wilks 2011, p. 186) of the predictions. Other tested hyperparameters did not yield improved results when modified based on the aggregate results from the 100 iterations and were unchanged from their default settings.

e. Operational SLR methods

We compare the performance of the RF to existing operational algorithms used by the NWS in the National Blend of Models (NBM; Craven et al. 2020) version 4.2 (v4.2) (Table 3). MaxTAloft is based on observations from Alaska (see The COMET Program 2023) and predicts SLR using a fifth-degree polynomial of the form

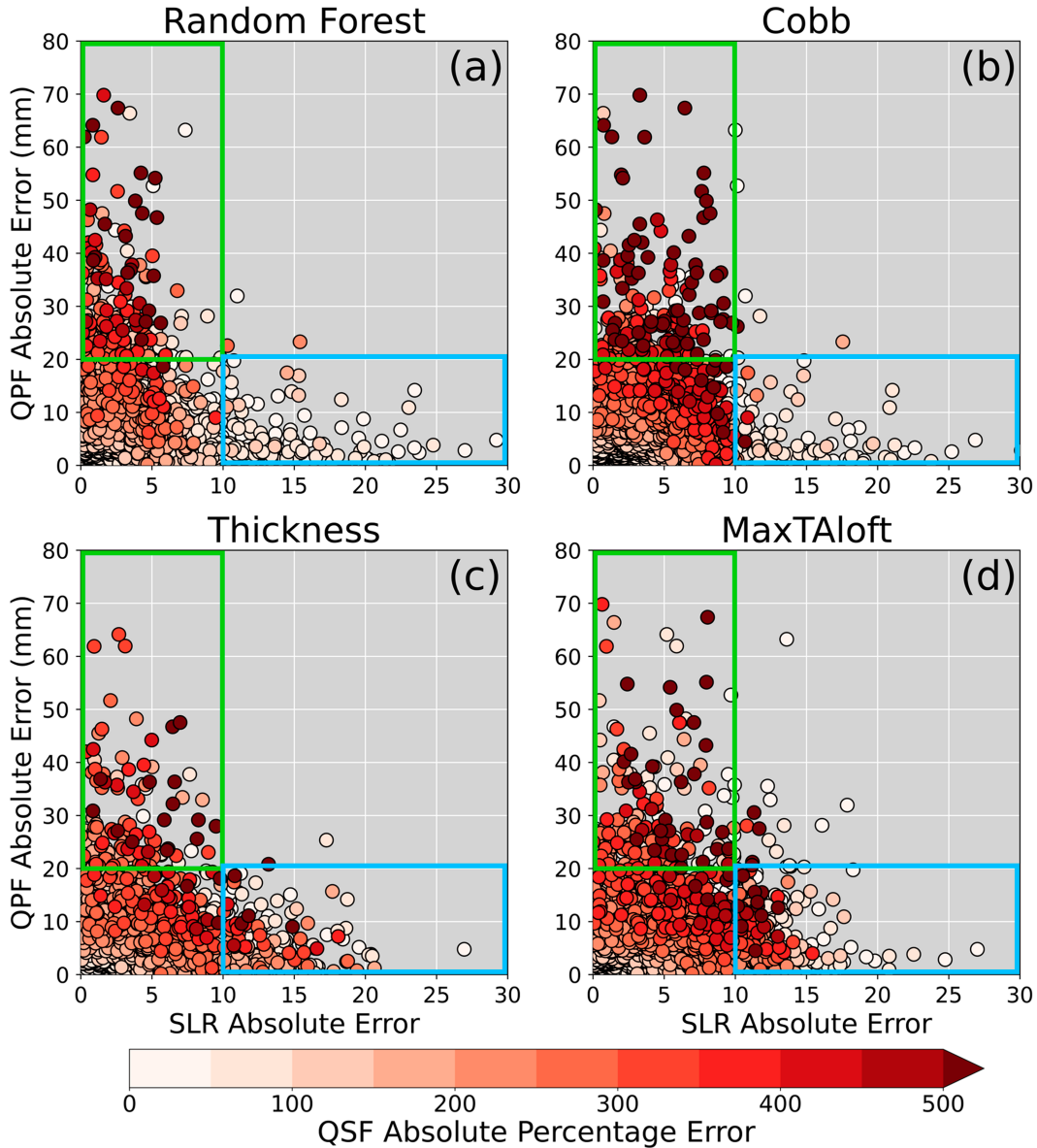


FIG. 8. Scatterplots of SLR AE vs QPF (mm) AE for (a) RF, (b) Cobb, (c) Thickness, and (d) MaxTAloft. Scatter-points are colored by QSF APE. The green and blue boxes correspond to QPF and SLR AE thresholds described in the text.

$$\begin{aligned} \text{SLR}_{\text{MaxTAloft}} = & 0.000\,004\,5 \times T_{\text{Max}}^5 + 0.000\,443\,2 \times T_{\text{Max}}^4 \\ & + 0.013\,090\,3 \times T_{\text{Max}}^3 + 0.058\,596\,8 \times T_{\text{Max}}^2 \\ & - 1.815\,080\,9 \times T_{\text{Max}} + 5.980\,572\,2, \end{aligned} \quad (1)$$

where T_{Max} is the maximum temperature ($^{\circ}\text{C}$) between 2000 ft (610 m) AGL and 400 hPa. Cobb is based on Cobb and Waldstreicher (2005) and predicts SLR using vertical velocity, temperature, and relative humidity (RH) in “cloudy” layers ($\geq 80\%$ RH). Cobb has been revised several times since its inception in 2005, and the most recent iteration (NBM v4.2) uses an improved polynomial curve to make SLR predictions.

Thickness, which was developed from mid-Atlantic snowfall cases (see The COMET Program 2023), predicts SLR using a simple linear regression equation of the form

$$\text{SLR}_{\text{Thickness}} = -(0.165\,59 \times \text{Thickness}_{850-700\text{mb}}) + 263.35, \quad (2)$$

where $\text{Thickness}_{850-700\text{mb}}$ is the thickness in meters between the 850- and 700-mb layers (1 mb = 1 hPa). In this study, Thickness is not used or evaluated at western CONUS sites where a portion of the 850–700-hPa layer is below site elevation, resulting in a smaller sample size than other techniques. We also evaluated the Roebber method, which is based on

Roebber et al. (2003), but it exhibited poor performance and will be removed from the NBM in the near future, so we have not included it. In part, this poor performance reflected the conversion of the probabilistic (Roebber et al. 2003) forecast to a deterministic forecast. NBM v4.2 adjusts these methods (MaxTAlot, Cobb, thickness, Roebber) in near-freezing environments using the surface wet-bulb temperature. However, the adjustments were excluded as they increased the overall SLR mean absolute error (MAE). This could be because we only include snowfall events > 2 in., which may reduce the number of events with ground melt.

f. SLR and QSF calculations and performance metrics

Based on forecast hours 12–36 from the 0000 to 1200 UTC HRRR initializations, the 24-h QSF for each observing period was calculated by multiplying 1-h PPFs by 1-h SLRs and integrating over the 24-h period. This approach produced better results compared to calculating the 24-h QSF by multiplying 24-h QPF with the 24-h mean SLR derived from 24-h mean predictor variables. The 24-h SLR, SLR_{24h} , was then calculated using

$$SLR_{24h} = \frac{QSF_{24h}}{QPF_{24h}}, \quad (3)$$

where QSF_{24h} and QPF_{24h} are the 24-h QSF and QPF, respectively. If the HRRR produced little to no precipitation (i.e., QPF_{24h} was ≤ 0.254 mm), a time-averaged SLR was calculated from the 1-h SLRs during the 24-h observing period to enable a comparison of forecast and observed SLRs during periods when there was no QPF but snowfall was observed. The 24-h subscripts are dropped hereafter for brevity.

Several performance metrics are used to evaluate SLR forecasts and QSFs from the held-out HRRR testing dataset, with absolute error (AE; Wilks 2011, p. 279) and MAE (Wilks 2011, p. 278) used as the primary metrics. The R^2 is also used as it measures how much observed variance is explained by the model and is often used in ML validation studies (e.g., Liu et al. 2022; Sweet et al. 2023). We calculate R^2 using the `r2_score` function from scikit-learn. A perfect R^2 value is 1, while an R^2 value of 0 indicates the model performs the same as the mean of the observations. The R^2 can also be negative, which indicates the model performs worse than the mean of the observations. The R^2 should not be confused with the square of the Pearson correlation coefficient, which can be used to assess the goodness of fit of a simple linear relationship between predicted and observed values. Root-mean-square error (RMSE; Wilks 2011, p. 308), mean bias error (MBE; Wilks 2011, p. 280), and absolute percentage error (APE; Jolliffe and Stephenson 2003, p. 194) are also used as performance metrics. We use multiple measures of performance to ensure reliable comparisons of the RF and NBM methods.

3. CoCoRaHS SLR climatology

To highlight SLR variability for different regions, we examine the SLR climatology for CoCoRaHS sites in each of the snow climates shown in Fig. 2c based on all available observations

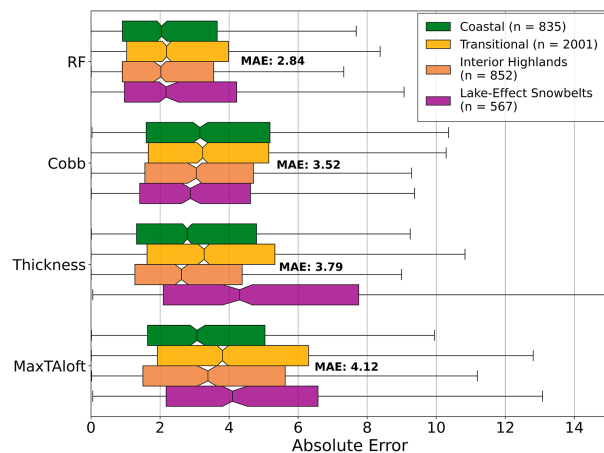


FIG. 9. Notched boxplots of AE grouped by SLR method for the eastern CONUS snow climates. The black vertical line indicates the median, the box indicates the 25th–75th percentiles, and the whiskers indicate the 5th and 95th percentiles.

from December 2000 to April 2024. For comparison, we also present the mean SLRs from the CoCoRaHS observations and the B2005 climatology. In the eastern CONUS, the coastal region is characterized by the lowest CoCoRaHS and B2005 means (10.2 and 9.0, respectively) and narrowest SLR distributions, whereas the lake-effect snowbelts exhibit the highest CoCoRaHS and B2005 means (13.9 and 13.1, respectively) and widest SLR distributions (Figs. 4a,d). High SLRs > 20 are more common in this region than any other in the eastern CONUS. The transitional and interior highlands have relatively similar means and distributions that are intermediate to the coastal and lake-effect snowbelt regions (Figs. 4b,c). Although snowfall is most plentiful in the lake-effect snowbelts and interior highlands, transitional sites accounted for the most ($\sim 44\%$) reports in the eastern CONUS due to the large geographic coverage of the region and number of CoCoRaHS stations.

In the western CONUS, the coastal snow climate features the lowest CoCoRaHS and B2005 means (8.3 and 10.6, respectively; Fig. 5a). A dramatic transition in snow climate occurs across the Cascade and Sierra Ranges (see also Fig. 2 of B2005), with a shift to higher mean SLRs and more frequent high SLRs > 20 in the transitional and continental snow climates of the western CONUS interior. The transitional snow climate features CoCoRaHS and B2005 means of 11.9 and 13.1, respectively (Fig. 5c). The continental snow climate also has a relatively high SLR of 14.1 for the CoCoRaHS B2005 means (Figs. 5e,f). The transitional and continental snow climates of the western CONUS interior also feature long tails to higher SLRs, including events > 30 or even > 40 .

The analysis above illustrates that the CoCoRaHS data generally capture the wide range of SLR climates across the CONUS. In the eastern United States, the coastal regions feature the lowest mean SLR, variability in SLR, and frequency of high SLR events, whereas the lake-effect snowbelts feature the highest mean SLR, variability in SLR, and frequency of

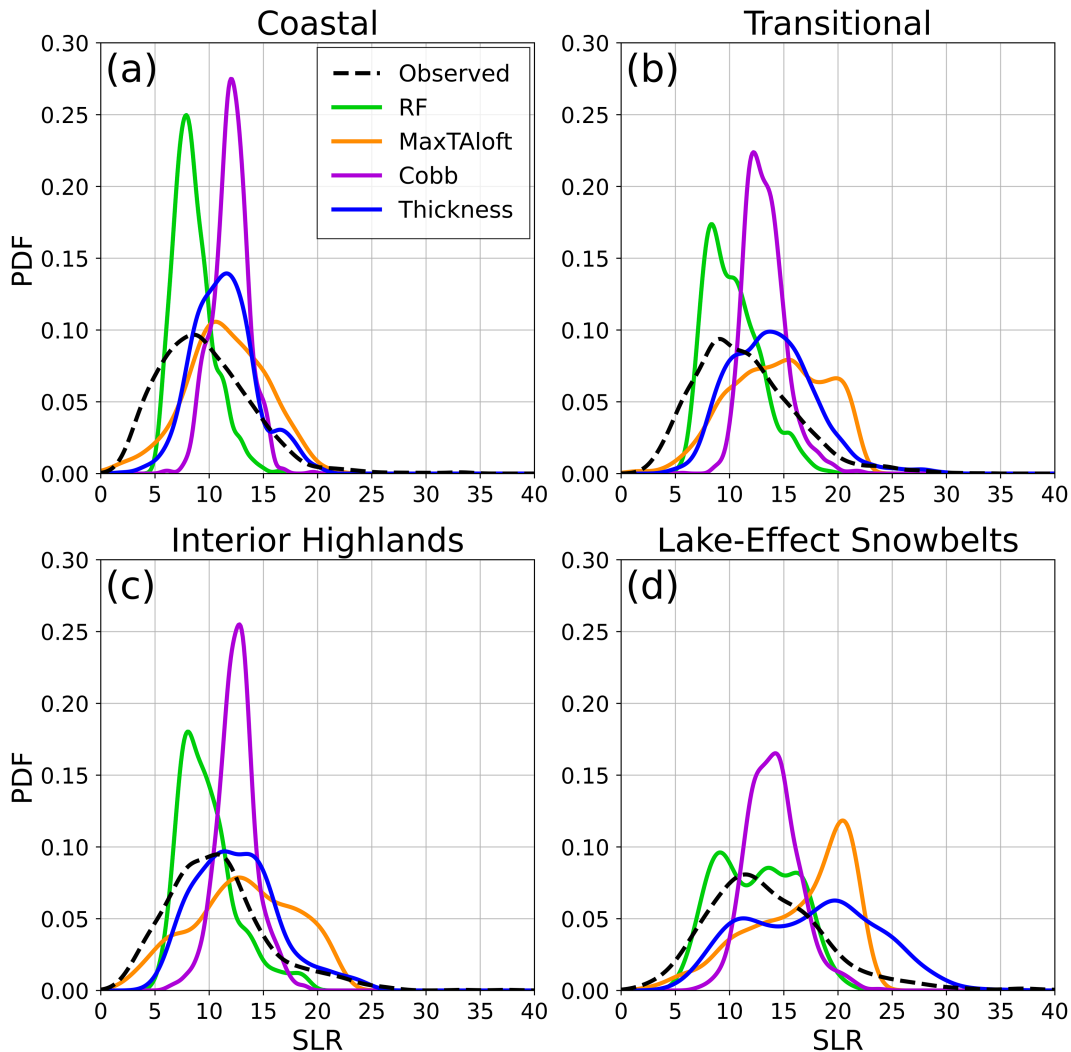


FIG. 10. PDFs of observed and forecast SLR for each method for the (a) coastal, (b), transitional, (c) interior highlands, and (d) lake-effect snowbelt eastern CONUS snow climates.

high SLR events. In the western United States, the coastal regions similarly feature the lowest mean SLR, variability in SLR, and frequency of low SLR events, whereas the western CONUS interior features higher mean SLR, variability in SLR, and frequency of high SLR events. These characteristics are generally consistent with B2005, although the B2005 mean SLRs are lower than the CoCoRaHS means (Figs. 3 and 4). This is surprising since some of the B2005 COOP observations likely used gauge data for LPE, which presumably would be subject to undercatch. The cause for this discrepancy is unclear.

4. CONUS-wide forecast performance

a. SLR forecast performance

We now compare the CONUS-wide, HRRR-derived RF and operational SLR methods forecast accuracy using bivariate histograms and performance metrics (Fig. 6). Overprediction

or underprediction is indicated by event pairs falling below or above the 1:1 line, respectively. The RF has relatively high accuracy for SLRs < 10; exhibits the highest R^2 (0.32), lowest MAE (2.91), and lowest MBE (−0.78); but underpredicts SLRs ≥ 17.5 , especially for SLRs > 20 (Fig. 6a). Cobb is less accurate with an R^2 value of 0.16, MAE of 3.51, and MBE of 1.38 and more event pairs falling farther below the 1:1 line for SLRs between 5 and 15, signaling overprediction of these events (Fig. 6b). Thickness clearly overpredicts SLR with an MBE of 2.36 and exhibits relatively high scatter, leading to a negative R^2 score (−0.06) and a higher MAE (3.81) (Fig. 6c; recall that this is only verified at lower-elevation sites). MaxTAloft exhibits a large positive SLR bias, resulting in the highest MBE (3.5) and the worst performing SLR forecasts (R^2 of −0.41 and MAE of 4.75) (Fig. 6d). Even though the MaxTAloft polynomial curve peaks at an SLR of ~ 22 (corresponding to a MaxTAloft of $\sim -12^\circ\text{C}$), it does not predict higher values (hence the cutoff evident in Fig. 6d).

TABLE 4. Descriptions of each eastern CONUS error subset.

Set	Description	Threshold	Mean observed SLR	Mean RF SLR	Minimum forecast error	Maximum forecast error
Least accurate AE	SLR events that exhibit the highest AE	≥ 7.8 (95th percentile eastern CONUS AE)	22.9	11.7	7.8	29.2
Least accurate APE	SLR events that exhibit the highest APE	$\geq 75.2\%$ (95th percentile eastern CONUS APE)	4.2	8.9	75.6%	339.7%
Most accurate AE	SLR events that exhibit the lowest AE	≤ 0.2 (5th percentile eastern CONUS AE)	9.5	9.5	0.002	0.2
Most accurate APE	SLR events that exhibit the lowest APE	$\leq 2\%$ (5th percentile eastern CONUS APE)	10.3	10.3	0.01%	2%

MaxTAloft’s poor performance is consistent with studies of SLR accuracy over the western CONUS (e.g., [Pletcher et al. 2024](#); [Veals et al. 2025](#)). Only the RF produces better forecasts than those based on the observed climatological CoCoRaHS

SLR mean at each site, which produced an R^2 of 0.15 and MAE of 3.41 (not shown). Thus, the RF is the only method that improved upon the climatological SLR mean for each CoCoRaHS site, a common standard for forecast skill ([Wilks 2011](#)).

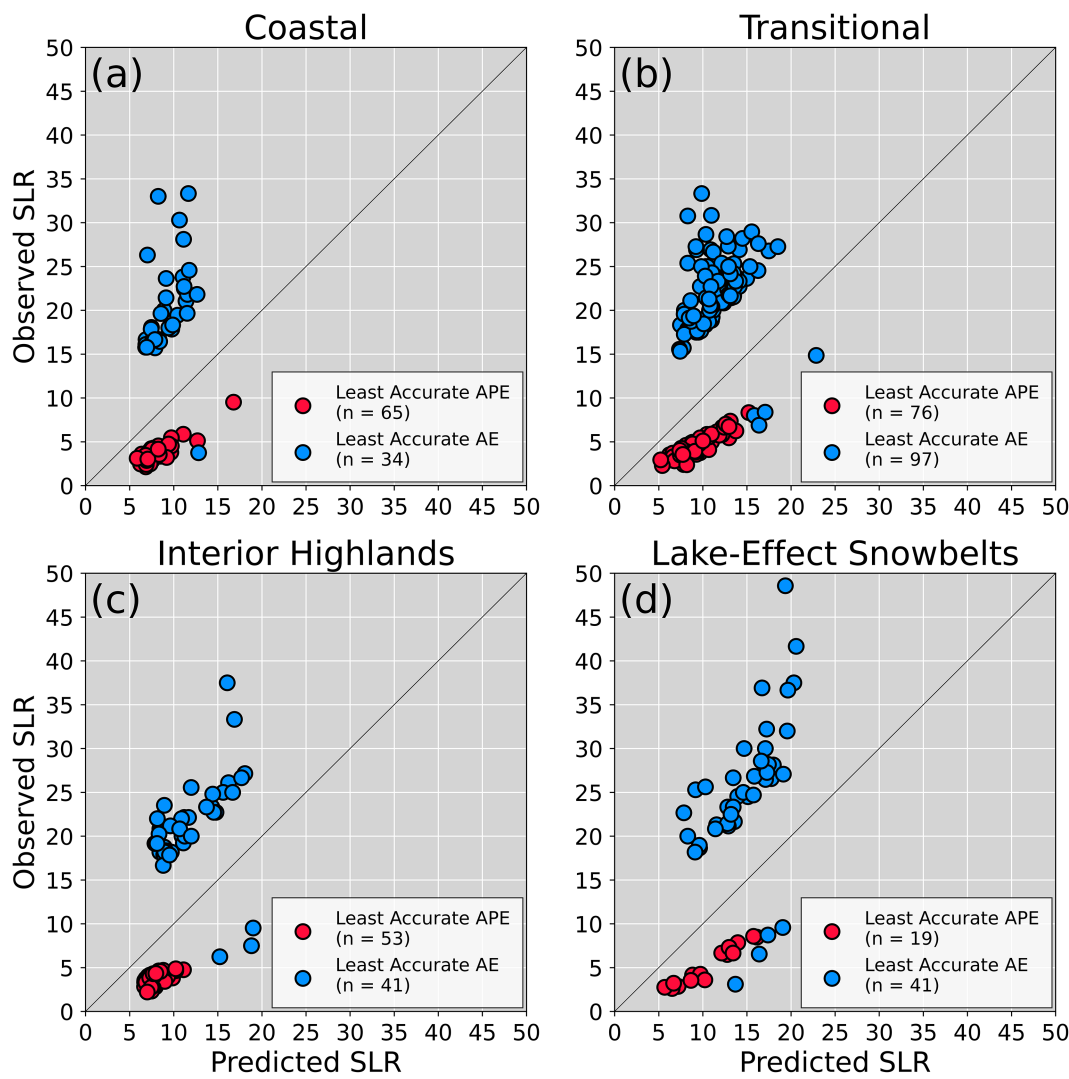


FIG. 11. Scatterplots of forecast and observed SLR for the least accurate AE (blue scatterpoints) and APE (red scatterpoints) subsets for the (a) coastal, (b) transitional, (c) interior highlands, and (d) lake-effect snowbelt eastern CONUS snow climates.

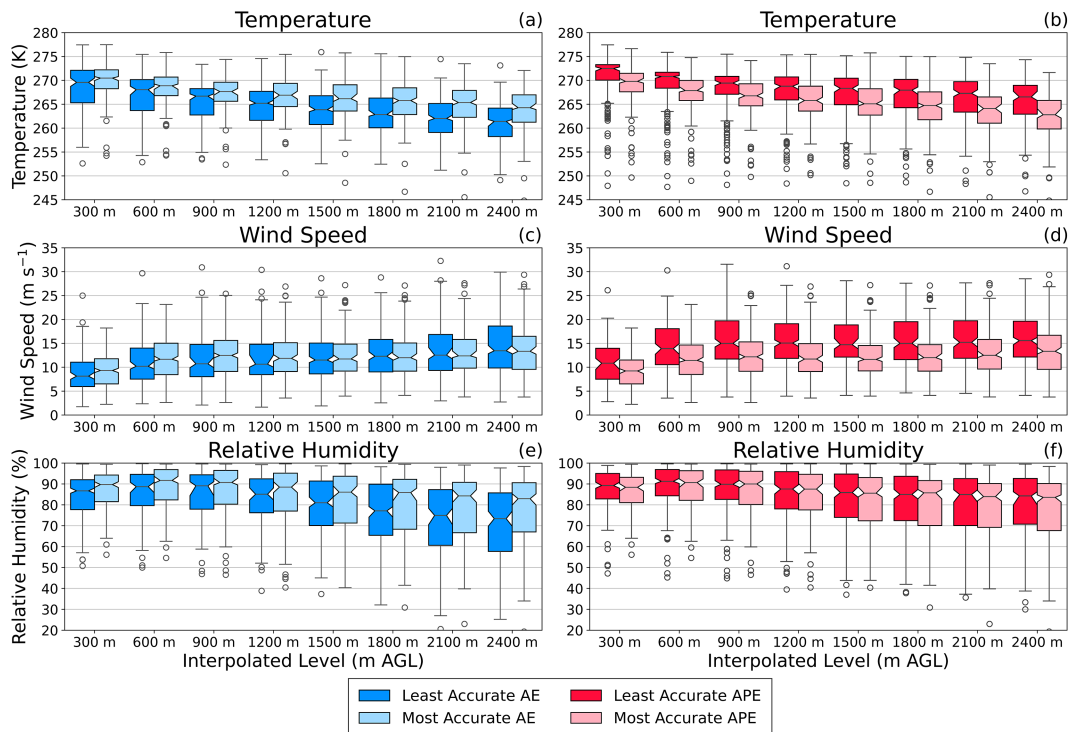


FIG. 12. Notched boxplots of 24-h mean temperature (K), wind speed (m s^{-1}), and RH (%) at each interpolated level (m AGL) used in the RF for the (a),(c),(e) AE and (b),(d),(f) APE subsets. The black horizontal line indicates the median, the box indicates the 25th–75th percentiles, the whiskers indicate the 5th and 95th percentiles, and the circles indicate outliers.

b. QSF performance

In this section, we verify QSFs to determine if the RF's improved SLRs lead to better snowfall forecasts. Although QSFs derived from the RF exhibit considerable scatter, they exhibit the highest R^2 , lowest MAE, and lowest MBE (-0.29 , 5.59 , and 0.62 cm, respectively) of the methods (Fig. 7a). The R^2 of -0.29 indicates worse performance than the mean, which we attribute to shortcomings in the HRRR QPF since validation of SLR alone for these events described in the previous section yielded an R^2 of 0.32 (Fig. 6a). We anticipate that the RF-derived QSF performance will improve further in the future as model QPF accuracy improves. QSFs derived from Cobb, MaxTAloft, and Thickness are less accurate, exhibiting increased scatter and an overprediction of snowfall at most thresholds, resulting in lower R^2 scores, higher MAEs, and higher MBEs (Figs. 7b–d). QSFs derived from Cobb are the most accurate of the NBM methods but exhibit a 25% higher MAE (7.61 cm) than the RF-derived QSFs.

A critical question for QSF performance concerns the relative contributions of SLR and QPF error. To examine this, we present scattergrams of 24-h SLR AE versus 24-h QSF APE (mm) with each scatter point colored by the 24-h QSF APE (Figs. 8a–d). Since SLR errors varied by method but HRRR QPF errors remained constant, a rightward shift in red and dark red scatterpoints demonstrates a larger SLR contribution to QSF error. For RF and Cobb, most of the large

($\geq 300\%$) QSF APE events occur when the SLR AE is ≤ 10 and the QPF AE is ≥ 20 mm, as shown in the green boxes (Figs. 8a,b). Within the green boxes, 75 and $120 \geq 300\%$ QSF APE events occur for RF and Cobb, respectively, compared with 51 and 66 events for Thickness and MaxTAloft, respectively. However, a shift is evident for MaxTAloft and Thickness-derived QSFs which feature more large ($>300\%$) QSF APE events when SLR AE is >10 and QPF AE is ≤ 7.5 mm (blue boxes, Figs. 8c,d). Thickness and MaxTAloft account for 23 and 37 such events, respectively, whereas RF and Cobb account for 0 and 2, respectively. This indicates that thickness and MaxTAloft produce more larger QSF errors from larger SLR errors. Conversely, the RF and Cobb-derived largest QSF APE events are a result of large QPF errors, as evidenced by the largest QSF APE confined to when SLR AE ≤ 5 and QPF AE is ≥ 30 mm. This suggests that QPF is the primary contributor to RF or Cobb-derived QSF errors, whereas SLRs contribute more strongly to MaxTAloft or Thickness-derived QSF errors.

5. Eastern CONUS forecast performance

a. Regional performance

We now examine the performance in the eastern CONUS snow climates using AE as a measure of performance. This is

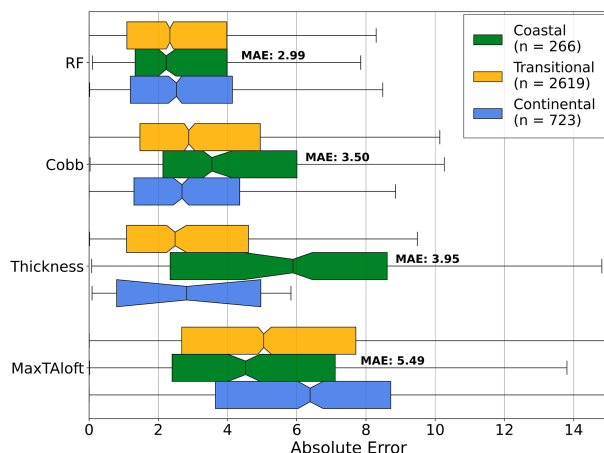


FIG. 13. As in Fig. 9, but for the western CONUS snow climates.

a region that experiences major regional or localized societal impacts from winter storms associated with nor'easters and lake-effect storms, resulting in loss of life and billions in damages (Sanders and Bosart 1985; Kocin and Uccellini 1990; Kocin and Schumacher 2004; Niziol et al. 1995; Greybush et al. 2017; Zaff 2022; Novak et al. 2023). The AE for each snow climate is grouped by the SLR method and displayed using boxplots, with statistically significant differences in median AE occurring when the boxplot notches do not overlap (Fig. 9). Here and in similar figures, statistical significance follows McGill et al. (1978) and is based on a 95% confidence level evaluated with bootstrap resampling with 1000 samples. MAE is displayed in boldface at right of each group. For the eastern CONUS, SLR forecasts derived from the RF exhibited the lowest MAE (2.84), whereas Cobb and Thickness exhibited somewhat higher MAEs (3.52 and 3.79, respectively). MaxTAloft produced the worst overall performance with an MAE of 4.12. The RF also produced the lowest median AE for each snow climate, with the difference in median compared to other methods statistically significant. Thus, the RF was consistently better across all of the eastern CONUS snow climates.

Probability density functions (PDFs) show the SLR forecast distributions compared to observed in the four snow climates (Figs. 10a–d). In addition to exhibiting the lowest overall MAE, the RF's mode (as indicated by the peak of the curves) is aligned closest to observed for the coastal, transitional, and interior highland snow climates (Figs. 10a–c). Its distribution, however, is narrower than observed. In the lake-effect snowbelts, the RF exhibits a bimodal distribution (Fig. 10d). Cobb's distribution is also narrow but shifted to higher values than RF due to an overprediction of SLRs ≥ 10 and SLRs ≤ 15 in the coastal, transitional, and interior highlands (Figs. 10a–c). Although the MaxTAloft and Thickness PDFs were broader than RF and Cobb, they were shifted to higher SLRs than observed, consistent with overprediction (Figs. 10a–d).

b. RF forecast environments

To investigate the factors affecting RF performance over the eastern CONUS, we subsetting the HRRR testing dataset

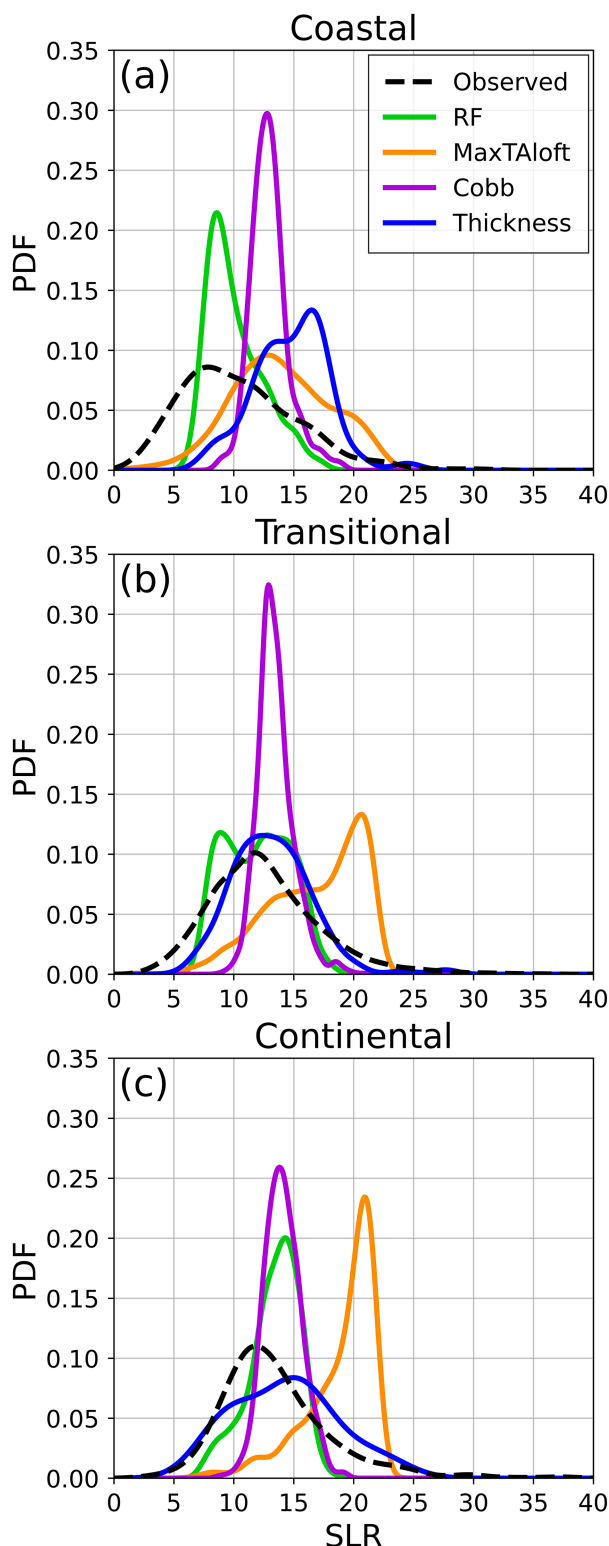


FIG. 14. As in Fig. 10, but for the (a) coastal, (b) transitional, and (c) continental western CONUS snow climates.

TABLE 5. Descriptions of each western CONUS error subset.

Set	Description	Threshold	Mean observed SLR	Mean RF SLR	Minimum forecast error	Maximum forecast error
Least accurate AE	SLR events that exhibit the highest AE	≥ 8 (95th percentile eastern CONUS AE)	23.3	11.7	8	27
Least accurate APE	SLR events that exhibit the highest APE	$\geq 69.4\%$ (95th percentile eastern CONUS APE)	5.9	8.9	69.4%	406.8%
Most accurate AE	SLR events that exhibit the lowest AE	≤ 0.2 (5th percentile eastern CONUS AE)	11.6	11.6	0.0002	0.2
Most accurate APE	SLR events that exhibit the lowest APE	$\leq 2\%$ (5th percentile eastern CONUS APE)	12.2	12.2	0.0015%	2%

based on the most and least accurate RF SLR forecasts. Accuracy was evaluated using AE and APE, with AE useful for identifying large absolute errors and APE useful for large absolute relative errors. We identified the most accurate absolute and relative forecasts based on RF SLRs with an AE or APE less than or equal to their eastern CONUS 5th percentile thresholds (0.2% and 2%, respectively) and the least accurate absolute and relative forecasts based on RF SLRs with an AE or APE greater than or equal to their eastern CONUS 95th percentile thresholds (7.8% and 75.2%, respectively). This resulted in four subsets containing 213 events each that we refer to as least accurate AE, least accurate APE, most accurate AE, and most accurate APE. Table 4 summarizes each of these subsets.

Although the mean observed SLRs were similar for the most accurate AE and APE subsets (9.5 and 10.3, respectively), the mean observed SLRs for the least accurate AE and APE subsets were 22 and 4.2, respectively. Thus, the largest absolute forecast errors typically occurred during high SLR events, whereas the largest relative forecast errors typically occurred during low SLR events. This is further illustrated by scatterplots of RF and observed SLR for the least accurate AE and APE subsets in each eastern CONUS snow climate which show that the largest absolute forecast errors typically occurred during high (>15) observed SLR events, whereas the largest relative forecast errors occurred during low (<10) observed SLR events (Fig. 11). There were, however, 11 least accurate AE events that featured low observed SLRs, so large absolute forecast errors are not confined exclusively to high observed SLR events. In the coastal region, least accurate APE events were more common than least accurate AE events (Fig. 11a), reflecting the higher frequency of low SLR events in that region. In contrast, in the lake-effect snowbelts, least accurate AE events were more common (Fig. 11b), reflecting the higher frequency of high SLR events in that region. Collectively, these results reflect the relatively narrow distribution of RF SLR forecasts compared to observed (e.g., Fig. 10), which yields a tendency to underforecast high SLR events and overforecast low SLR events.

We now compare the RF's atmospheric features (T , SPD, and R) during the most and least accurate AE and APE periods to examine if there is a useful signal for anticipating SLR forecast accuracy. These atmospheric features are grouped by each vertical level in notched boxplots and colored based on the four accuracy subsets (Fig. 12). To focus on least accurate

AE events associated with SLR underprediction, we excluded the 11 events with observed SLRs < 10 (5.2% of the population) from this subset. The remaining least accurate AE events featured lower median temperatures than the most accurate AE events, with the differences statistically significant at all levels except 300, 600, and 900 m AGL (Fig. 12a). There were also statistically significant differences in median wind speed and relative humidity below 1200 m AGL and above 1500 m AGL, respectively, with the least accurate AE wind speed and relative humidity values lower at these levels (Figs. 12c,e). The least accurate APE events featured higher median temperatures and wind speeds than the most accurate APE events at all levels, with statistically significant differences that increase with height (Figs. 12b,d). However, there were no statistically significant differences in median relative humidity values (Fig. 12f). These results suggest that the largest RF AEs, which are associated with underforecasting, occur during colder, less windy storms. We suggest that these are likely events where dendritic growth and light winds favor high SLR snowfall. Furthermore, statistically significant lower median RH values above 1500 m AGL could be a result of shallower events or events that featured a shorter precipitation time frame that lowered the 24-h mean RH values. In contrast, the largest RF APEs, which are associated with overforecasting, occur during relatively warm and windy storms, which have been shown to produce low SLRs (Roebber et al. 2003; Cobb and Waldstreicher 2005; Alcott and Steenburgh 2010). These results are consistent with Figs. 6 and 10 which show that the distribution of RF SLR forecasts is narrower than observed.

6. Western CONUS forecast performance

a. Regional performance

The western CONUS frequently experiences high-impact winter storms associated with diverse synoptic weather patterns and orographic forcings that can disrupt transportation and commerce (Ferber et al. 1993; Poulos et al. 2002; Rutz et al. 2014; Tessendorf et al. 2019; Wasserstein and Steenburgh 2024). For the western CONUS, the RF exhibited the lowest overall MAE (2.99), followed by Cobb (3.5) (Fig. 13). The RF also exhibited the lowest overall MAE for each western CONUS snow climate, although differences in median AE between the RF and Cobb were not statistically significant in the continental

snow climate. Thickness and MaxTAloft performed worse, with overall MAEs of 3.95 and 5.49, respectively, and poorer performance in each snow climate. For Thickness, however, there were a small number of samples for the continental snow climate ($n = 17$) due to the elevation of many western CONUS CoCoRaHS sites, limiting statistical confidence estimates as reflected in the wide notches around the median. Aside from Cobb, each method performed worse in the western CONUS compared to the eastern CONUS, perhaps due to the greater SLR variability (cf. Figs. 4 and 5).

The challenges of predicting the SLR variability in the western CONUS are further illustrated by a comparison of observed and forecast SLR PDFs (Figs. 14a–c). The RF's mode is centered near observed for each snow climate, but its distribution is too narrow for the coastal and continental snow climates, and it exhibits an inaccurate bimodal distribution for the transitional snow climate. The Cobb distribution is also too narrow for each snow climate, and its mode is shifted to higher SLRs, indicating frequent overforecasting. The Thickness distribution is also shifted to higher SLRs than observed but is broader, indicating that it forecasts a wider range of SLRs, although the sample size for this method is small. MaxTAloft's mode is shifted too far to the right of the observed mode for each snow climate, especially transitional and continental snow climates.

b. RF forecast environments

As done in section 5b for the eastern CONUS, we also identified the most and least accurate RF SLR forecasts based on the western CONUS 5th and 95th percentiles of AE and APE. The resulting thresholds for most accurate absolute (AE) and relative (APE) SLR forecasts were 0.2% and 2%, respectively. The least accurate absolute and relative SLR errors were 8% and 69.4%, respectively. Each subset contained 181 events and are summarized in Table 5.

The most accurate AE and APE subsets exhibited comparable mean observed SLRs (11.6 and 12.2, respectively), whereas the least accurate AE and APE subsets exhibited dissimilar mean observed SLRs (23.3 and 5.9, respectively). Thus, we again use the least accurate AE and APE subsets to highlight the RF's struggles to forecast high and low SLRs, respectively. Scatterplots of RF and observed SLR from each western CONUS snow climate subset further illustrate the RF's biases (Fig. 15). Least accurate AE events were less common than least accurate APE events at coastal sites, with the RF underforecasting SLR for the former (MBE of -7.7) and overforecasting SLR for the latter (MBE of 4.9) (Fig. 15a). Like the eastern CONUS, the greater frequency of least accurate APE events at coastal sites reflects the relatively common occurrence of low SLRs in the coastal snow climate (see Fig. 5). Least accurate AE events were more common than least accurate APE events at transitional and continental sites (MBEs of -10.3 and 5.6, respectively) because of a greater frequency of high SLRs for those snow climates (Figs. 15b,c). Similar to the eastern CONUS, these results indicate the RF's inability to forecast the full distribution of observed SLR.

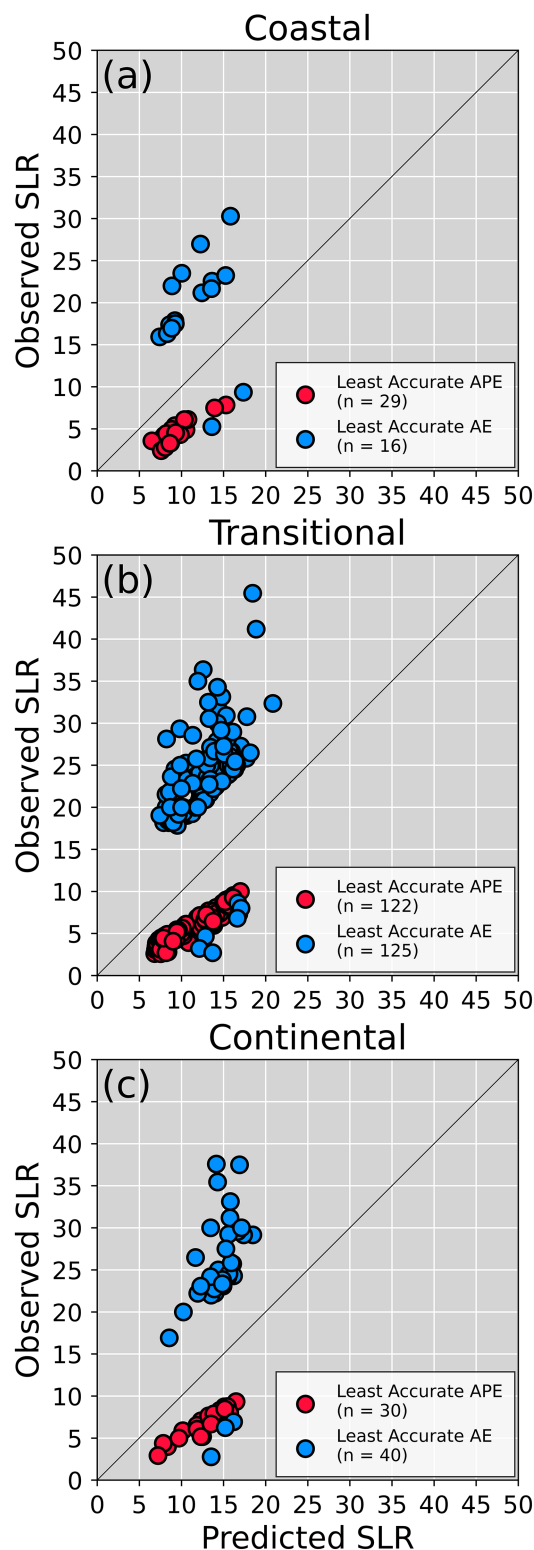


FIG. 15. As in Fig. 11, but for the (a) coastal, (b) transitional, and (c) continental western CONUS snow climates.

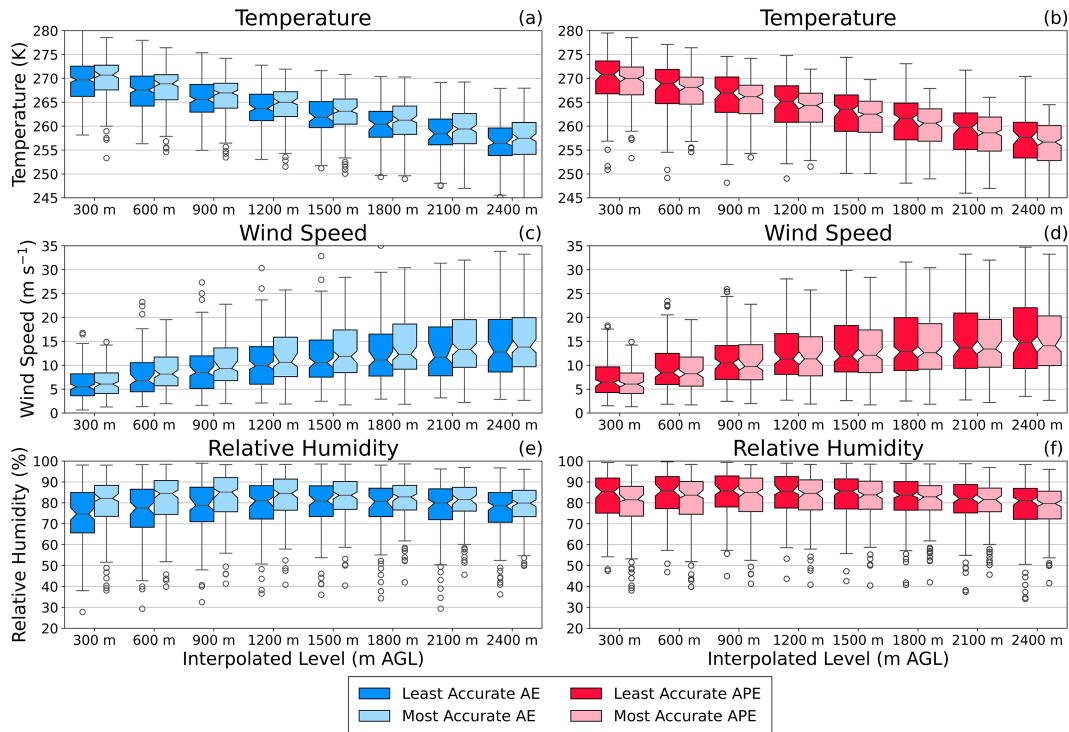


FIG. 16. As in Fig. 12, but for the western CONUS.

As in section 5b, we now compare the RF's atmospheric features (T , SPD, and R) during the most and least accurate AE and APE periods to examine if there is a useful signal for anticipating SLR forecast accuracy. As in the eastern CONUS, we excluded the 12 least accurate AE events with SLRs < 10 to allow a clearer comparison of atmospheric features between the least and most accurate AE subsets. Median temperatures and wind speeds were lower for the least accurate AE subset; however, none of the differences were statistically significant at any level (Figs. 16a,c). Median relative humidity values were also lower at all levels for the least accurate AE events, with statistically significant differences at 300 and 600 m AGL (Fig. 16e). The least accurate APE events featured higher median temperatures than the most accurate APE events at all levels, but the differences were not statistically significant (Fig. 16b). Differences in wind and relative humidity were also small and not statistically significant, except for median relative humidity at 300 and 600 m AGL, which was higher for the least accurate APE events (Figs. 16d,f). Although in the eastern CONUS there were statistically significant differences in the RF's features at more vertical levels, the contrasts in the western CONUS were more subtle, so there is not a clear distinction of which features affect the RF's western CONUS accuracy.

c. Veals et al. (2025) performance comparison

Finally, we examined the performance of the RF and the NBM methods at the 14 mountain sites described in Veals et al. (2025) (see their Fig. 1 for a map of the site locations),

which were not used for RF training. These sites provide snowfall and SLR observations collected by snow-safety teams and are mainly located at ski areas or along highways or railroad corridors. From November 2022 to April 2024 (i.e., the HRRR testing period), there were 1919 observations available, which reflects frequent snowfall at these mountain locations. Thickness was not included in this analysis since the 14 sites were located at elevations where a portion of the 850–700-hPa layer was below site elevation.

Consistent with earlier results, the RF exhibited the highest R^2 (0.25), lowest MAE (3.23), and lowest MBE (1.05) compared to the NBM methods at these sites (cf. Figs. 17a–c), although these values are slightly worse than the CONUS-wide, eastern CONUS, and western CONUS RF results (cf. Figs. 6a, 9, and 13). The positive MBE reflects a tendency to overpredict SLR events < 15 . These results are consistent with Veals et al. (2025), who found that an RF developed with data from the 14 sites outperformed the NBM methods. Additionally, these results also show that a CONUS-wide RF provides value and outperforms the NBM methods in mountainous areas of the west that are susceptible to avalanche hazards and road closures from high-impact winter storms, even at sites that are not included in the RF training.

7. Conclusions

This study used the CoCoRaHS observing network to train, validate, and test a new RF-based method for predicting SLR across the CONUS. We focused specifically on observations collected using a snowboard and coring tube to limit the

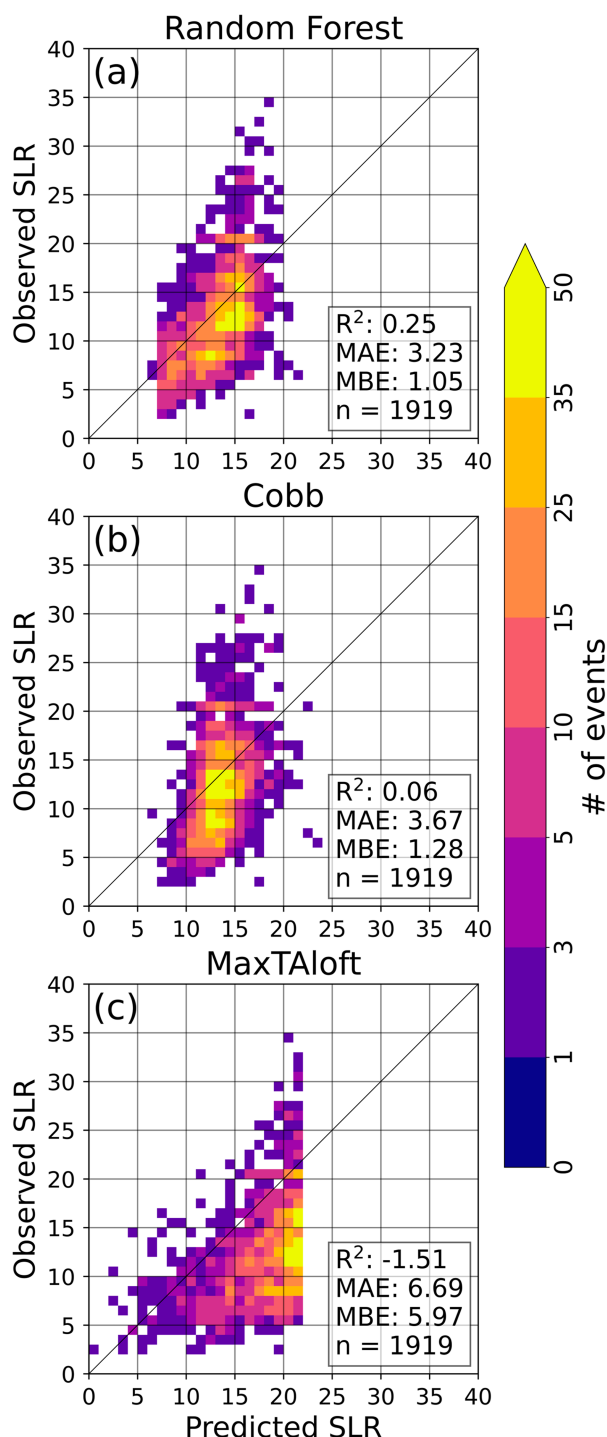


FIG. 17. As in Fig. 6, but for (a) RF, (b) Cobb, and (c) MaxTAloft at the 14 snow-safety sites described in Veals et al. (2025).

impacts of undercatch on LPE measurement. The RF trained using these SLR observations, ERA5 atmospheric profiles, and station geographic information produced the most accurate SLR forecasts across the CONUS when applied to the HRRR, outperforming existing NBM methods (i.e., Cobb,

Thickness, and MaxTAloft) and regionally trained and applied RFs. QSFs derived from the RF SLR forecasts were also more accurate than those derived from the NBM methods, with the largest contributor to error being the QPF rather than the SLR forecast. This was also true for Cobb-derived QSFs, whereas the SLR errors dominated Thickness and MaxTAloft QSF errors.

The RF also outperformed the NBM methods across the eastern CONUS and in the four eastern CONUS snow climates (coastal, transitional, interior highlands, and lake-effect snowbelts). Nevertheless, the distribution of RF SLR forecasts in these regions was narrower than observed, which reflects overforecasting SLR during low SLR (i.e., high density) events and underforecasting during high SLR (i.e., low density) events, the former typically warm and windy environments and the latter colder. Similarly, the RF outperformed the NBM methods across the western CONUS and in the three western CONUS snow climates (coastal, transitional, and continental), although the difference between RF and Cobb was not significant at a 95% confidence level in the continental snow climate. The RF also exhibited overforecasting of low SLR events and underforecasting of high SLR events in the western CONUS, although the atmospheric conditions contributing to these errors were less clear, with few statistically significant differences in the RF's atmospheric features during most and least accurate forecasts. Finally, the RF outperformed the NBM methods at high-elevation western CONUS sites not used in the training, demonstrating its potential to be applied in other regions and in mountain locations prone to avalanches and road closures during high-impact winter storms.

These results, along with those from Pletcher et al. (2024) and Veals et al. (2025), reinforce that machine learning can be applied to high-quality SLR observations to improve SLR forecasts and QSFs. This work extends that of Veals et al. (2025) by developing an RF that can be applied CONUS-wide while also still exhibiting forecast accuracy at mountain sites in the western CONUS. An issue for future work is how to address the RF's narrow SLR forecast distribution, which degrades forecast accuracy during high and low SLR events. It is possible that more accurate high and low SLR forecasts could be produced by training with HRRR forecasts rather than ERA5 analyses, including training with QPF (see Veals et al. 2025). However, the HRRR's data record is shorter, especially for the current HRRR version, which was updated in 2020. Training to the HRRR could also limit applicability to other modeling systems. It is also possible that the RF's architecture may not allow for accurate prediction of high and low SLRs, despite accounting for nonlinear feature-label relationships. Alternative approaches that might improve predictions are deep learning models such as a convolutional neural network (LeCun et al. 1989) or a vision transformer (ViT; Dosovitskiy et al. 2020), which can optimize performance by automatically extracting important features and encoding spatial information [i.e., the shape of vertical profiles (Chase et al. 2023)]. A potential issue with these alternative implementations is increased computational load, which could be a consideration for operational use.

Acknowledgments. We thank the CoCoRaHS observing network and snow-safety personnel for providing and collecting the snowfall observations used in this study as well as the University of Utah Center for High Performance Computing for computer support. We also thank Jimmy Correia Jr., Kirstin Harnos, Massey Bartolini, and others from the Weather Prediction Center for testing and evaluating preliminary versions of the RF during the Winter Weather Experiment. This article is based on research supported by NOAA Weather Program Office Grants NA19OAR4590137 and NA22OAR4590531 and NOAA Climate Program Office Grant NA23OAR4310279-T1-01 to the University of Utah. Funding was also provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA21OAR4320204, U.S. Department of Commerce. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect those of the National Oceanic and Atmospheric Administration.

Data availability statement. The CoCoRaHS observations, ERA5 training and validation dataset, HRRR testing dataset, trained models, and related scripts can be found on the associated GitHub repository at <https://github.com/mdpletcher/utahrfsr>. The HRRR forecast data are available from Amazon Web Services (<https://registry.opendata.aws/noaa-hrrr-pds/>). Observations from three of the fourteen sites used in Veals et al. (2025) are available on GitHub (https://github.com/pveals/Veals_et_al_2025) and online from the University of Utah Research Data Repository (<https://hive.utah.edu/concern/datasets/0r967383v>). The observations from the remaining 11 sites are available by request from the corresponding author of Veals et al. (2025) due to their proprietary nature. The ERA5 reanalysis data are available from the Copernicus Climate Change Service (Hersbach et al. 2018a,b). Terrain data used in Figs. 1 and 2 are available from ESRI (<https://services.arcgis.com/arcgis/rest/services/Elevation>). The NSA data are available from NOHRSC (<https://www.nohrsc.noaa.gov/snowfall/>).

APPENDIX

RF Versus LR Performance Comparison

Here, we compare the performance of the RF to the simpler, CONUS-wide LR model mentioned in section 2d. LR-based models are easier to interpret, computationally inexpensive, and can be used as a comparison baseline for more complex methods like the RF (Rudin 2019; Chase et al. 2022, 2023). The LR model is currently being evaluated for use in NWS operations, including global models. Thus, we excluded the input features LAT, LON, and ELEV from the LR training as they provide negligible improvements to MAE and R^2 and may introduce bias when applied outside the CONUS. Table A1 summarizes the aggregate performance metrics of the RF and LR based on the 100 train/validate splits, with the RF superior in all metrics except MBE, although both models had low scores

TABLE A1. Aggregate performance metrics from the 100 random train/validate splits for the RF and LR models. Boldface indicates the best metric of the two model configurations.

Model	MBE	MAE	R^2	RMSE
RF	0.06	3.05	0.35	4.16
LR	0.008	3.42	0.2	4.63

consistent with near-neutral bias. The LR model can be found on GitHub (<https://github.com/mdpletcher/utahrfsr/tree/main/models/lr>).

REFERENCES

- 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. *Wea. Forecasting*, **25**, 323–337, <https://doi.org/10.1175/2009WAF2222311.1>.
- Baars, J. A., and C. F. Mass, 2005: Performance of National Weather Service forecasts compared to operational, consensus, and weighted model output statistics. *Wea. Forecasting*, **20**, 1034–1047, <https://doi.org/10.1175/WAF896.1>.
- 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, <https://doi.org/10.1175/2009JAS2883.1>.
- 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>.
- Billet, J., M. DeLisi, B. G. Smith, and C. Gates, 1997: Use of regression techniques to predict hail size and the probability of large hail. *Wea. Forecasting*, **12**, 154–164, [https://doi.org/10.1175/1520-0434\(1997\)012<0154:UORTTP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0154:UORTTP>2.0.CO;2).
- Black, A. W., and T. L. Mote, 2015: Characteristics of winter-precipitation-related transportation fatalities in the United States. *Wea. Climate Soc.*, **7**, 133–145, <https://doi.org/10.1175/WCAS-D-14-00011.1>.
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Byun, K.-Y., J. Yang, and T.-Y. 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>.
- Carroll, T., D. Cline, G. Fall, A. Nilsson, L. Li, and A. Rost, 2001: NOHRSC operations and the simulation of snow cover properties for the coterminous U.S. *Proc. 69th Annual Meeting of the Western Snow Conf.*, Sun Valley, ID, Western Snow Conference, 1–14, <https://ftp.nohrsc.noaa.gov/technology/pdf/wsc2001.pdf>.
- Chase, R. J., D. R. Harrison, A. Burke, G. M. Lackmann, and A. McGovern, 2022: A machine learning tutorial for operational meteorology. Part I: Traditional machine learning. *Wea. Forecasting*, **37**, 1509–1529, <https://doi.org/10.1175/WAF-D-22-0070.1>.
- , —, G. M. Lackmann, and A. McGovern, 2023: A machine learning tutorial for operational meteorology. Part II: Neural networks and deep learning. *Wea. Forecasting*, **38**, 1271–1293, <https://doi.org/10.1175/WAF-D-22-0187.1>.
- Cifelli, R., N. Doesken, P. Kennedy, L. D. Carey, S. A. Rutledge, C. Gimmestad, and T. Depue, 2005: The Community Collaborative Rain, Hail, and Snow network: Informal education

- for scientists and citizens. *Bull. Amer. Meteor. Soc.*, **86**, 1069–1078, <https://doi.org/10.1175/BAMS-86-8-1069>.
- Cobb, D. K., Jr., 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, <https://ams.confex.com/ams/pdfpapers/94815.pdf>.
- Craven, J. P., D. E. Rudack, and P. E. Shafer, 2020: National blend of models: A statistically post-processed multi-model ensemble. *J. Oper. Meteor.*, **8** (1), 1–14, <https://doi.org/10.1519/nwajom.2020.0801>.
- Cui, G., M. Anderson, and R. Bales, 2023: Mapping of snow water equivalent by a deep-learning model assimilating snow observations. *J. Hydrol.*, **616**, 128835, <https://doi.org/10.1016/j.jhydrol.2022.128835>.
- Dosovitskiy, A., and Coauthors, 2020: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv, 2010.11929v2, <https://doi.org/10.48550/arXiv.2010.11929>.
- Dowell, D. C., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, <https://doi.org/10.1175/WAF-D-21-0151.1>.
- Eichenlaub, V. L., 1979: *Weather and Climate of the Great Lakes Region*. University of Notre Dame Press, 335 pp.
- Ferber, G. K., C. F. Mass, G. M. Lackmann, and M. W. Patnoe, 1993: Snowstorms over the Puget Sound lowlands. *Wea. Forecasting*, **8**, 481–504, [https://doi.org/10.1175/1520-0434\(1993\)008<0481:SOTPSL>2.0.CO;2](https://doi.org/10.1175/1520-0434(1993)008<0481:SOTPSL>2.0.CO;2).
- Filipiak, B. C., N. P. Bassill, K. L. Corbosiero, A. L. Lang, and R. A. Lazear, 2023: Probabilistic forecasting methods of winter mixed-precipitation events in New York State utilizing a random forest. *Artif. Intell. Earth Syst.*, **2**, e220080, <https://doi.org/10.1175/AIES-D-22-0080.1>.
- Gagne, D. J., A. McGovern, and M. Xue, 2014: Machine learning enhancement of storm-scale ensemble probabilistic quantitative precipitation forecasts. *Wea. Forecasting*, **29**, 1024–1043, <https://doi.org/10.1175/WAF-D-13-00108.1>.
- , —, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, <https://doi.org/10.1175/WAF-D-17-0010.1>.
- Ganetis, S. A., and B. A. Colle, 2015: The thermodynamic and microphysical evolution of an intense snowband during the northeast U.S. blizzard of 8–9 February 2013. *Mon. Wea. Rev.*, **143**, 4104–4125, <https://doi.org/10.1175/MWR-D-14-00407.1>.
- Gerber, F., and Coauthors, 2018: Spatial variability in snow precipitation and accumulation in COSMO-WRF simulations and radar estimations over complex terrain. *Cryosphere*, **12**, 3137–3160, <https://doi.org/10.5194/tc-12-3137-2018>.
- , R. Mott, and M. Lehning, 2019: The importance of near-surface winter precipitation processes in complex alpine terrain. *J. Hydrometeorol.*, **20**, 177–196, <https://doi.org/10.1175/JHM-D-18-0055.1>.
- Glahn, H. R., and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, [https://doi.org/10.1175/1520-0450\(1972\)011<1203:TUOMOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2).
- Greybush, S. J., S. Saslo, and R. Grumm, 2017: Assessing the ensemble predictability of precipitation forecasts for the January 2015 and 2016 East Coast winter storms. *Wea. Forecasting*, **32**, 1057–1078, <https://doi.org/10.1175/WAF-D-16-0153.1>.
- Grinsztajn, L., E. Oyallon, and G. Varoquaux, 2022: Why do tree-based models still outperform deep learning on tabular data? arXiv, 2207.08815v1, <https://doi.org/10.48550/arXiv.2207.08815>.
- Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, [https://doi.org/10.1175/1520-0493\(2004\)132<1434:ERIMFS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2).
- Hartigan, J. A., and M. A. Wong, 1979: Algorithm AS 136: A K-means clustering algorithm. *J. Roy. Stat. Soc.*, **28C**, 100–108, <https://doi.org/10.2307/2346830>.
- Herman, G. R., and R. S. Schumacher, 2018a: Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests. *Mon. Wea. Rev.*, **146**, 1571–1600, <https://doi.org/10.1175/MWR-D-17-0250.1>.
- , and —, 2018b: “Dendrology” in numerical weather prediction: What random forests and logistic regression tell us about forecasting extreme precipitation. *Mon. Wea. Rev.*, **146**, 1785–1812, <https://doi.org/10.1175/MWR-D-17-0307.1>.
- Hersbach, H., and Coauthors, 2018a: ERA5 hourly data on pressure levels from 1959 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 3 March 2025, <https://doi.org/10.24381/cds.bd0915c6>.
- , and Coauthors, 2018b: ERA5 hourly data on single levels from 1959 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed 3 March 2025, <https://doi.org/10.24381/cds.adbb2d47>.
- , and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hill, A. J., G. R. Herman, and R. S. Schumacher, 2020: Forecasting severe weather with random forests. *Mon. Wea. Rev.*, **148**, 2135–2161, <https://doi.org/10.1175/MWR-D-19-0344.1>.
- , R. S. Schumacher, and I. L. Jirak, 2023: A new paradigm for medium-range severe weather forecasts: Probabilistic random forest-based predictions. *Wea. Forecasting*, **38**, 251–272, <https://doi.org/10.1175/WAF-D-22-0143.1>.
- Hoopes, C. A., C. L. Castro, A. Behrangi, M. R. Ehsani, and 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>.
- Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 254 pp.
- Judson, A., and N. Doesken, 2000: Density of freshly fallen snow in the central Rocky Mountains. *Bull. Amer. Meteor. Soc.*, **81**, 1577–1588, [https://doi.org/10.1175/1520-0477\(2000\)081<1577:DOFFSI>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<1577:DOFFSI>2.3.CO;2).
- Kitzmler, D. H., W. E. McGovern, and R. E. Saffie, 1995: The WSR-88D severe weather potential algorithm. *Wea. Forecasting*, **10**, 141–159, [https://doi.org/10.1175/1520-0434\(1995\)010<0141:TWSWPA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010<0141:TWSWPA>2.0.CO;2).
- Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. *J. Meteor.*, **16**, 672–682, [https://doi.org/10.1175/1520-0469\(1959\)016<0672:OPOFDM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1959)016<0672:OPOFDM>2.0.CO;2).
- Kocin, P. J., and L. W. Uccellini, 1990: Snowstorms along the northeastern coast of the United States: 1955 to 1985. *Meteor. Monogr.*, No. 44, Amer. Meteor. Soc., 280 pp., <https://doi.org/10.1007/978-1-940033-93-8>.

- , and —, 2004: A snowfall impact scale derived from north-east storm snowfall distributions. *Bull. Amer. Meteor. Soc.*, **85**, 177–194, <https://doi.org/10.1175/BAMS-85-2-177>.
- , P. N. Schumacher, R. F. Morales, and L. W. Uccellini, 1995: Overview of the 12–14 March 1993 Superstorm. *Bull. Amer. Meteor. Soc.*, **76**, 165–182, [https://doi.org/10.1175/1520-0477\(1995\)076<0165:OOTMS>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<0165:OOTMS>2.0.CO;2).
- Kohavi, R., 2001: A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proc. 14th Int. Joint Conf. on Artificial Intelligence*, Montreal, Quebec, Canada, Morgan Kaufmann, 1137–1143, <https://dl.acm.org/doi/10.5555/1643031.1643047>.
- Lazo, J. K., H. R. Hosterman, J. M. Sprague-Hilderbrand, and J. E. Adkins, 2020: Impact-based decision support services and the socioeconomic impacts of winter storms. *Bull. Amer. Meteor. Soc.*, **101**, E626–E639, <https://doi.org/10.1175/BAMS-D-18-0153.1>.
- LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. Jackel, 1989: Handwritten digit recognition with a back-propagation network. *Proc. Third Int. Conf. on Neural Information Processing Systems*, Denver, CO, MIT Press, 396–404, <https://dl.acm.org/doi/10.5555/2969830.2969879>.
- Liu, Q., M. Yang, K. Mohammadi, D. Song, J. Bi, and G. Wang, 2022: Machine learning crop yield models based on meteorological features and comparison with a process-based model. *Artif. Intell. Earth Syst.*, **1**, e220002, <https://doi.org/10.1175/AIES-D-22-0002.1>.
- Loken, E. D., A. J. Clark, A. McGovern, M. Flora, and K. Knopfmeier, 2019: Postprocessing next-day ensemble probabilistic precipitation forecasts using random forests. *Wea. Forecasting*, **34**, 2017–2044, <https://doi.org/10.1175/WAF-D-19-0109.1>.
- , —, and K. Knopfmeier, 2022: Comparing and interpreting differently designed random forests for next-day severe weather hazard prediction. *Wea. Forecasting*, **37**, 871–899, <https://doi.org/10.1175/WAF-D-21-0138.1>.
- MacDonald, J., and P. W. Pomeroy, 2007: Gauge undercatch of two common snowfall gauges in a prairie environment. *Proc. 64th Eastern Snow Conf.*, St. John's, NL, Canada, Eastern Snow Conference Publisher, 119–124, www.easternsnow.org/esc-2007.
- Mayer, S., A. van Herwijnen, F. Techel, and J. Schweizer, 2022: A random forest model to assess snow instability from simulated snow stratigraphy. *Cryosphere*, **16**, 4593–4615, <https://doi.org/10.5194/tc-16-4593-2022>.
- McGill, R., J. W. Tukey, and W. A. Larsen, 1978: Variations of box plots. *Amer. Stat.*, **32**, 12–16, <https://doi.org/10.1080/00031305.1978.10479236>.
- McGovern, A., K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, <https://doi.org/10.1175/BAMS-D-16-0123.1>.
- Mecikalski, J. R., T. N. Sandmæl, E. M. Murillo, C. R. Homeyer, K. M. Bedka, J. M. Apke, and C. P. Jewett, 2021: Random forest model to assess predictor importance and nowcast severe storms using high-resolution radar–GOES satellite–lightning observations. *Mon. Wea. Rev.*, **149**, 1725–1746, <https://doi.org/10.1175/MWR-D-19-0274.1>.
- Milà, C., M. Ludwig, E. Pedesma, C. Tonne, and H. Meyer, 2024: Random forests with spatial proxies for environmental modelling: Opportunities and pitfalls. *Geosci. Model Dev.*, **17**, 6007–6033, <https://doi.org/10.5194/gmd-17-6007-2024>.
- Mock, C. J., and K. W. Birkeland, 2000: Snow avalanche climatology of the western United States Mountain ranges. *Bull. Amer. Meteor. Soc.*, **81**, 2367–2392, [https://doi.org/10.1175/1520-0477\(2000\)081<2367:SACOTW>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2367:SACOTW>2.3.CO;2).
- Nakaya, U., 1954: *Snow Crystals, Natural and Artificial*. Harvard University Press, 510 pp.
- Niziol, T. A., W. R. Snyder, and J. S. Waldstreicher, 1995: Winter weather forecasting throughout the eastern United States. Part IV: Lake effect snow. *Wea. Forecasting*, **10**, 61–77, [https://doi.org/10.1175/1520-0434\(1995\)010<0061:WWFTE>2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010<0061:WWFTE>2.0.CO;2).
- Novak, D. R., L. F. Bosart, D. Keyser, and J. S. Waldstreicher, 2004: An observational study of cold season–banded precipitation in northeast U.S. cyclones. *Wea. Forecasting*, **19**, 993–1010, <https://doi.org/10.1175/815.1>.
- , and Coauthors, 2023: Innovations in winter storm forecasting and decision support services. *Bull. Amer. Meteor. Soc.*, **104**, E715–E735, <https://doi.org/10.1175/BAMS-D-22-0065.1>.
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Picca, J. C., D. M. Schultz, B. A. Colle, S. Ganetis, D. R. Novak, and M. J. Sienkiewicz, 2014: The value of dual-polarization radar in diagnosing the complex microphysical evolution of an intense snowband. *Bull. Amer. Meteor. Soc.*, **95**, 1825–1834, <https://doi.org/10.1175/BAMS-D-13-00258.1>.
- Pletcher, M. D., P. G. Veals, M. E. Wessler, D. Church, K. Harnos, J. Correia, R. J. Chase, and W. J. Steenburgh, 2024: Validation of cool-season snowfall forecasts at a high-elevation site in Utah's Little Cottonwood Canyon. *Wea. Forecasting*, **39**, 1261–1277, <https://doi.org/10.1175/WAF-D-23-0176.1>.
- Pomeroy, J. W., and E. Brun, 2001: Physical properties of snow. *Snow Ecology*, H. G. Jones et al., Eds., Cambridge University Press, 45–126.
- Potter, J. G., 1965: Water content of freshly fallen snow. *CIR-4232*, TEC-569, 12 pp.
- Poulos, G. S., D. A. Wesley, J. S. Snook, and M. P. Meyers, 2002: A Rocky Mountain storm. Part I: The blizzard—Kinematic evolution and the potential for high-resolution numerical forecasting of snowfall. *Wea. Forecasting*, **17**, 955–970, [https://doi.org/10.1175/1520-0434\(2002\)017<0955:ARMSPI>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0955:ARMSPI>2.0.CO;2).
- Power, B. A., P. W. Summers, and J. D'Avignon, 1964: Snow crystal forms and riming effects as related to snowfall density and general storm conditions. *J. Atmos. Sci.*, **21**, 300–305, [https://doi.org/10.1175/1520-0469\(1964\)021<0300:SCFARE>2.0.CO;2](https://doi.org/10.1175/1520-0469(1964)021<0300:SCFARE>2.0.CO;2).
- 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>.
- Reges, H. W., N. Doesken, J. Turner, N. Newman, A. Bergantino, and Z. Schwalbe, 2016: CoCoRaHS: The evolution and accomplishments of a volunteer rain gauge network. *Bull. Amer. Meteor. Soc.*, **97**, 1831–1846, <https://doi.org/10.1175/BAMS-D-14-00213.1>.
- Roebber, P. J., and S. Smith, 2023: Prospects for machine learning activity within the United States National Weather Service. *Bull. Amer. Meteor. Soc.*, **104**, E1333–E1344, <https://doi.org/10.1175/BAMS-D-22-0181.1>.
- , 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](https://doi.org/10.1175/1520-0434(2003)018<0264:ISFBDS>2.0.CO;2).
- Rudin, C., 2019: Stop explaining black box machine learning models for high stakes decisions and use interpretable models

- instead. *Nat. Mach. Intell.*, **1**, 206–215, <https://doi.org/10.1038/s42256-019-0048-x>.
- Rutz, J. J., W. J. Steenburgh, and F. M. Ralph, 2014: Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Mon. Wea. Rev.*, **142**, 905–921, <https://doi.org/10.1175/MWR-D-13-00168.1>.
- Sanders, F., and L. F. Bosart, 1985: Mesoscale structure in the megalopolitan snowstorm of 11–12 February 1983. Part I: Frontogenetical forcing and symmetric instability. *J. Atmos. Sci.*, **42**, 1050–1061, [https://doi.org/10.1175/1520-0469\(1985\)042<1050:MSITMS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<1050:MSITMS>2.0.CO;2).
- Seeherman, J., and Yi Liu, 2015: Effects of extraordinary snowfall on traffic safety. *Accid. Anal. Prev.*, **81**, 194–203, <https://doi.org/10.1016/j.aap.2015.04.029>.
- Serreze, M. C., M. P. Clark, D. L. McGinnis, and D. A. Robinson, 1998: Characteristics of snowfall over the eastern half of the United States and relationships with principal modes of low-frequency atmospheric variability. *J. Climate*, **11**, 234–250, [https://doi.org/10.1175/1520-0442\(1998\)011<0234:COSOTE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<0234:COSOTE>2.0.CO;2).
- Sha, Y., D. J. Gagne II, G. West, and R. Stull, 2020: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: Daily precipitation. *J. Appl. Meteor. Climatol.*, **59**, 2075–2092, <https://doi.org/10.1175/JAMC-D-20-0058.1>.
- Steenburgh, W. J., 2003: One hundred inches in one hundred hours: Evolution of a Wasatch Mountain winter storm cycle. *Wea. Forecasting*, **18**, 1018–1036, [https://doi.org/10.1175/1520-0434\(2003\)018<1018:OHIOH>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<1018:OHIOH>2.0.CO;2).
- Sweet, L.-B., C. Müller, M. Anand, and J. Zscheischler, 2023: Cross-validation strategy impacts the performance and interpretation of machine learning models. *Artif. Intell. Earth Syst.*, **2**, e230026, <https://doi.org/10.1175/AIES-D-23-0026.1>.
- Tessendorf, S. A., and Coauthors, 2019: A transformational approach to winter orographic weather modification research: The SNOWIE project. *Bull. Amer. Meteor. Soc.*, **100**, 71–92, <https://doi.org/10.1175/BAMS-D-17-0152.1>.
- The COMET Program, 2023: Understanding NBM v4.0 snowfall products. The University Corporation for Atmospheric Research, accessed 29 October 2025, https://www.meted.ucar.edu/education_training/lesson/10166.
- Thériault, J. M., 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>.
- Veals, P. G., M. Pletcher, A. J. Schwartz, R. J. Chase, K. Harnos, J. Correia, M. E. Wessler, and W. J. Steenburgh, 2025: Predicting snow-to-liquid ratio in the mountains of the western United States. *Wea. Forecasting*, **40**, 1825–1838, <https://doi.org/10.1175/WAF-D-24-0233.1>.
- Wasserstein, M. L., and W. J. Steenburgh, 2024: Diverse characteristics of extreme orographic snowfall events in Little Cottonwood Canyon, Utah. *Mon. Wea. Rev.*, **152**, 945–966, <https://doi.org/10.1175/MWR-D-23-0206.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 704 pp.
- Williams, J. K., 2014: Using random forests to diagnose aviation turbulence. *Mach. Learn.*, **95**, 51–70, <https://doi.org/10.1007/s10994-013-5346-7>.
- Wolvin, S., C. Strong, S. Brewer, and W. J. Steenburgh, 2024: Evaluation of a convolutional neural network to predict wintertime orographic precipitation gradients of the western CONUS. *23rd Conf. on Artificial Intelligence for Environmental Science*, Baltimore, MD, Amer. Meteor. Soc., 439397, <https://ui.adsabs.harvard.edu/abs/2024AMS...10439397W/abstract>.
- Zaff, D. A., 2022: The once-in-a-generation Buffalo blizzard and lake effect event of December 23–26 2022. *32nd Conf. on Weather Analysis and Forecasting*, Madison, WI, Amer. Meteor. Soc., J3.3, <https://ams.confex.com/ams/WAFNWPM/meetingapp.cgi/Paper/425172>.