Evaluation of Ensemble Snowfall Forecasts Using Operationally Used Snow-to-Liquid Ratios in Five Winter Storms

ANDREW A. ROSENOW,[a,b] HEATHER D. REEVES,[a,b] AND DANIEL D. TRIPP[a,b]

[a] Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma
[b] NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 17 November 2022, in final form 3 August 2023, accepted 16 August 2023)

ABSTRACT: The prediction of snow accumulation remains a forecasting challenge. While the adoption of ensemble numerical weather prediction has enabled the development of probabilistic guidance, the challenges associated with snow accumulation, particularly snow-to-liquid ratio (SLR), still remain when building snow-accumulation tools. In operations, SLR is generally assumed to either fit a simple mathematical relationship or conform to a historic average. In this paper, the impacts of the choice of SLR on ensemble snow forecasts are tested. Ensemble forecasts from the nine-member High-Resolution Rapid Refresh Ensemble (HRRRE) were used to create 24-h snowfall forecasts for five snowfall events associated with winter cyclones. These snowfall forecasts were derived from model liquid precipitation forecasts using five SLR relationships. These forecasts were evaluated against daily new snowfall observations from the Community Collaborative Rain Hail and Snow network. The results of this analysis show that the forecast error associated with individual members is similar to the error associated with choice of SLR. The SLR with the lowest forecast error showed regional agreement across nearby observations. This suggests that, while there is no one SLR that works best everywhere, it may be possible to improve ensemble snow forecasts if regions where SLRs perform best can be determined ahead of time. The implications of these findings for future ensemble snowfall tools will be discussed.

SIGNIFICANCE STATEMENT: Snowfall prediction remains a challenge. Computer models are used to address the inherent uncertainty in forecasts. This uncertainty includes aspects like the location and rate of snowfall. Meteorologists run multiple similar computer models to understand the range of possible weather outcomes. One aspect of uncertainty is the snow-to-liquid ratio, or the ratio of snow depth to the amount of liquid water it melts into. This study tests how common predictions of snow-to-liquid ratio impact snowfall forecasts. The results show that snow-to-liquid ratio choices are as impactful as the models' differing snow rate or snow location forecasts, and that no particular snow-to-liquid ratio is most accurate. These results underscore the importance of better snow-to-liquid ratio prediction to improve snowfall forecasts.

KEYWORDS: Snow; Snowfall; Ensembles; Forecast verification/skill; Forecasting techniques

1. Introduction

Accumulating snow represents a significant threat to life and property. Snow, by its nature, is typically a long-duration threat, lasting for hours to days. Even with phenomena that happen over short time scales (e.g., snow squalls), mitigation efforts such as pretreating or plowing roads and runways take a substantial amount of time. The effectiveness of these mitigation efforts depends on how quickly snow accumulates. Thus, accurate snowfall accumulation forecasts are important for stakeholders to operate with optimal effectiveness and safety.

There are two critical components to a successful forecast of snow accumulation. Both are problematic. The first is an accurate prediction of the liquid-equivalent precipitation, known as the quantitative precipitation forecast (QPF), improvement of which has long been a point of emphasis for the cold season (Ralph et al. 2005). Ensemble NWP has been shown to improve the performance of QPF for winter storms (e.g., Greybush et al. 2017), though these ensembles have known issues with biases and under-dispersiveness (Romine et al. 2014).

The second component, and the topic of this research, is the conversion from QPF to the actual snow accumulation. The snow-to-liquid water ratio (SLR) depends on multiple factors including QPF itself, crystal habit/size, and morphological changes to the snowpack after the snow has fallen, such as temperature, wind, melting, compaction, and saltation (see review in Roebber et al. 2003). These factors can all have significant variation in both time and space. However, NWP models are incapable of representing most of the factors that contribute to SLR. Thus, the uncertainty in snow accumulation forecasts can exceed the uncertainty due to QPF alone (Roebber et al. 2003 and references therein).

SLR prediction methods vary in how much information they require as inputs to determine SLR. The simplest is a constant SLR democratically applied everywhere. An SLR of 10:1 was used for many years in operational forecasting, and there are still some references to its use in media today [see Roebber et al. (2003) for a review of the origins of the 10:1 relationship]. This method fails to account for known temporal and spatial variations in SLR as demonstrated in Baxter et al.

Corresponding author: Andrew A. Rosenow, andrew.rosenow@noaa.gov

DOI: 10.1175/WAF-D-22-0202.1

(2005), and the limited use that the 10:1 method sees is likely due to its mathematical simplicity rather than any inherent skill in the method. The Baxter et al. (2005) climatological assessment of SLR shows clear spatial trends across the CONUS as well as monthly trends. The Baxter et al. (2005) climatology has been updated and is currently used as the SLR for first guess field of the national snowfall analysis produced by the National Operational Hydrological Remote Sensing Center (NOHRSC 2022). This method is also simple, as it relies on predesignated SLRs for each point in the CONUS. While the climatological method allows for a prescribed degree of spatial and temporal variability, it does not account for rapid changes in SLR at a set location during a single event. More sophisticated methods are required to obtain this degree of SLR variability.

There are multiple techniques that allow for a dynamic SLR (i.e., one that allows for variation at short temporal and spatial scales). 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 (E. Kuchera 2021, personal communication). Such a methodology exploits the fact that the snow habit, and consequently, SLR, are functionally dependent on the environmental temperature (Roebber et al. 2003; Alcott and Steenburgh 2010). While not as simple and intuitive as the 10:1 or climatological methods, the Kuchera method requires only the temperature profile to run. A technique that accounts for more forcings on the SLR is presented in Cobb and Waldstreicher (2005; Cobb 2011, the "Cobb method"). This method accounts for the temperature, vertical velocity, and relative humidity within the cloud above each grid point. This method implicitly considers ascent in the dendritic-growth zone. If the column's ascent is primarily between $-10°$ and $-20°$C, the algorithm produces larger SLR values for that grid point. The Kuchera and Cobb methods are relatively simple, rules-based algorithms. Techniques that exploit machine learning (Roebber et al. 2003; Ware et al. 2006) and efforts to directly derive SLR from microphysical parameterization schemes (Smirnova et al. 2016) have also been presented in the literature.

This study uses snowfall observations to assess the relative contribution of the choice of ensemble member, with the primary impact of member choice being the associated QPF, and choice of SLR to the final snowfall forecast error. For this study, we will be focusing on the error in forecasts of snowfall, rather than SLR. This is being done for two reasons. First, SLR is not measured directly, and is thus derived from two measurements, snow depth and melted liquid depth. This makes SLR more uncertain than just the snowfall measurement alone. Also, the constituent observations can be biased toward numbers that yield "nice" SLRs, especially where the measurements would yield a value of 10:1 (Baxter et al. 2005). Second, we are approaching this problem from the perspective of improving the ultimate snowfall forecast, and its role in developing decision support tools. Snowfall and/or snow rate are the important parameter for decision support; this work is aimed at understanding a potential source of uncertainty in those fields in SLR choice.

The next section will describe the data sources and methodology used herein. The overall results of the analysis are presented in section 3. Detailed analysis from two cases within the dataset will be presented in section 4. Finally, the implications of the findings of this study are discussed in section 5.

2. Data and methodology

a. *Ensemble NWP*

The ensemble NWP data used in this study is from the High-Resolution Rapid Refresh Ensemble (HRRRE; Kalina et al. 2021), which is an experimental ensemble consisting of perturbed versions of the operational High Resolution Rapid Refresh model. This ensemble has nine members run out to 36 hours, with hourly output. The HRRRE was chosen for this study due to its relatively small horizontal grid spacing of 3 km and hourly output, both of which are necessary to represent phenomena with short time scales and small geographic footprints (e.g., lake effect snow or snow squalls), as well as the hourly output enabling SLR updates every hour where changes in the model fields cause the SLR calculation to change. The HRRRE uses a stochastic parameter perturbation to create ensemble spread in a method that is still consistent with atmospheric physics beyond the traditional perturbations to initial and boundary conditions.

For each member of the HRRRE, an hourly snowfall field is computed from the hourly QPF using the four SLRs discussed in detail in section 1 (i.e., 10:1, climatology, Kuchera, and Cobb). Snow is assumed to occur in any grid box in a particular member if that member has a nonzero snow accumulation in the grid box. If no snow accumulation is produced, snow accumulation at that grid point is zero, even if other members produced snow there. Assuming precipitation falls as all snow where any snow falls in the model results in an overestimate in transition zones, as some of the precipitation actually falls as other precipitation types in reality, which could include particles like graupel in otherwise all-snow environments. However, this methodology is chosen over using the fraction of mass in a grid point that fell as snow in order to minimize the impact of the HRRRE microphysics scheme on the results of this study. Reeves (2016) demonstrated that these mixed or transition precipitation type observations represent 2%–4% of precipitation observations in near-0°C environments. Of these 2%–4%, roughly half are freezing drizzle or freezing rain, with no snow. Thus, the impact of these transition zones on the final results should be limited.

b. *Snow-to-liquid algorithms*

The snow algorithms are chosen to represent SLRs that National Weather Service forecasters are trained to use (WDTD 2022). The four algorithms used in this study are the 10:1 relationship, the climatological SLR, the Kuchera method, and the Cobb method (Cobb 2011). These four algorithms represent a spectrum of data requirements to run, from no environmental data to both thermodynamic and kinematic profiles as discussed in section 1 and summarized below.

1) 10:1: This SLR relationship assumes 10 in. of snowfall for every 1 in. of liquid-equivalent precipitation.
2) Climatological SLR: This relationship is used for the NOHRSC analysis and is based on the climatology presented in Baxter et al. (2005). The NOHRSC climatology uses a singular value for an individual month, representing the median, as opposed to the Baxter et al. (2005) climatology, which includes multiple SLRs for various percentiles. Values with this SLR generally ranged from 6–7:1 across the southern United States to 20–21:1 in the northern Rocky Mountains during the months used in this study. The NOHRSC dataset is used here both in deference to its operational use and to remove the requirement to choose a percentile from the climatology to use in the forecast. Each event in this study uses the SLR from the month during which the event occurred.
3) The Kuchera method: This method calculates the SLR using the following relationship:

$$\mathrm{SLR} = \begin{cases} 12 + 2 \times (271.16 - T_{\max}) & \text{for } T_{\max} > 271.16\ \mathrm{K} \\ 12 + (271.16 - T_{\max}) & \text{for } T_{\max} \le 271.16\ \mathrm{K} \end{cases}.$$

4) The Cobb method: This method requires the most input data to run of the methods considered here. It considers the vertical profiles of temperature, vertical velocity, and relative humidity with respect to ice at each grid point. First, the column maximum vertical velocity ($w_{\max}$) within a cloud is determined. A cloud is defined as any altitude where $\mathrm{RH}_i \ge 90\%$. Then, each vertical level in the cloud is assigned an SLR to contribute to the column's SLR based on the temperature of the layer. For this work, the updated temperature/SLR relationships specified in Cobb (2011) were used. The temperature/SLR relationship is zero when above freezing, 3:1 at 0°C, increases to 26:1 for a layer average temperature of −16°C, and then decreases to 7.2:1 below −30°C. A weighting factor for this layer, W_F, is then calculated:

$$W_{F\_\mathrm{layer}} = w\left(\frac{w}{w_{\max}}\right)^2 \Phi_2 - \Phi_1,$$

where w is the layer vertical velocity, and Φ_2 and Φ_1 are the altitudes at layer top and layer bottom, respectively. Finally, the SLR value for that grid point is calculated by summing over each layer's SLR weighted value:

$$\mathrm{SLR} = \sum \frac{W_{F\_\mathrm{layer}} \times \mathrm{SLR}_{\mathrm{layer}}}{W_{F\_\mathrm{total}}}.$$

5) HRRR method: The HRRR postprocessed data files include a snowfall field, previously mentioned for the masking performed in this study. This field is calculated using the same framework used here, by multiplying the liquid equivalent precipitation by the method's SLR. The HRRR uses an SLR relationship based solely on the lowest model level temperature (Benjamin et al. 2021). This method varies SLR from less than 5:1 at 0°C to 17:1 at −15°C and below. Most of the analysis in this paper will focus on the other SLRs, as the HRRR SLR method is not used widely outside the production of the HRRR model output to the authors' best knowledge.

c. *Observational data*

Model snowfall forecasts are verified using observations from the Community Collaborative Rain Hail and Snow (CoCoRAHS; Cifelli et al. 2005) network. CoCoRAHS observers receive training and materials to support their data collection efforts. Observers report precipitation daily at 0700 local time (LT), including the 24-h-accumulated new snow depth, which will be used in this study. While CoCoRAHS observers report at 0700 LT, they are trained to take their snow observations as near as possible to the cessation of snowfall, though the snow board used for measurement is not cleared until the reporting time every day (CoCoRAHS 2014a). This should act to minimize losses to melting and compaction in the reported accumulation. CoCoRAHS also undergoes automated and manual quality control (CoCoRAHS 2014b). Additionally, as with most volunteer networks, the location and number of observation sites are weighted by population density, meaning that more densely populated locations will be overrepresented in the dataset. To limit the impact of imprecise measurements of small snow depths, only observations with at least 50.8 mm (2 in.) of new snow are considered in this study following Baxter et al. (2005) and Roebber et al. (2003).

d. *Snow cases*

For this study, five 24-h periods during snow events from the 2020/21 winter season are analyzed (Table 1). These events are geographically large snowfall events associated with midlatitude cyclones to maximize the variety of snow-to-liquid ratios in the dataset. One pair of days, 31 January/1 February are snow events from the same cyclone. For 1 February, the only available HRRRE run where data were available for the whole 24-h study period was initialized at 0600 UTC instead of 1200 UTC, so snow accumulating between forecast hours 6 and 30 is used for this case. For all other cases, snow accumulating between forecast hours 0 and 24 are used. The model data and CoCoRAHS snowfall reports are only processed in a limited area for each individual event to limit the analysis to the cyclone of interest. The defining corners of each subdomain's rectangle are provided in Table 1. Table 1 also shows the number of CoCoRAHS observations that fall within the analysis domain for each case.

3. Results—Overall

A statistical summary of the performance of the SLR algorithms across all events is presented in Fig. 1. Here, the algorithms' performance for each event have been combined into an event total, with each CoCoRAHS observation represented nine times for each event, one verification per

TABLE 1. Snow cases used in this study. The number of observations refers to those observations that meet the 5.08-cm snowfall minimum that are also within the study region.

Case	Model initialization time	Region (SW lon, NE lon, SW lat, NE lat)	No. of CoCoRAHS observations
Start: 1200 UTC 31 Jan 2021 End: 1200 UTC 1 Feb 2021	1200 UTC 31 Jan 2021	−96°, −68°, 32°, 45°	788
Start: 1200 UTC 1 Feb 2021 End: 1200 UTC 2 Feb 2021	0600 UTC 1 Feb 2021	−96°, −68°, 32°, 45°	1054
Start: 1200 UTC 15 Feb 2021 End: 1200 UTC 16 Feb 2021	1200 UTC 15 Feb 2021	−100°, −68°, 29°, 50°	1297
Start: 1200 UTC 17 Feb 2021 End: 1200 UTC 18 Feb 2021	1200 UTC 17 Feb 2021	−100°, −68°, 27°, 45°	365
Start: 1200 UTC 13 Mar 2021 End: 1200 UTC 14 Mar 2021	1200 UTC 13 Mar 2021	−116°, −95°, 30°, 45°	692

ensemble member. In performing this verification, snowfall errors are considered in an absolute sense; a positive and negative departure are identical in these statistics. The absolute error for each station/member/SLR combination is compared to the errors for the other SLRs for that station/member combination. The SLR that produces the smallest absolute error is considered to be the best algorithm for that station/member combination; the largest absolute error causes the SLR to be considered the worst.

Across the events, there is no clear "best" or "worst" SLR algorithm. The HRRR algorithm is the one with the smallest error for the largest fraction of sites in four of the five events (Fig. 1a). In three of these events, more than one-third of the sites are best predicted using this algorithm. But even the algorithm that has the fewest fraction of best forecasts, 10:1, still manages to produce the best forecast for at least 19% of the sites' observations for three of the cases. The climatological SLR provides the best forecast at less than 15% of the sites in four of the five cases. It only produces the most best forecasts for one event, 15 February, with the best forecast at 53.6% of the sites. The Cobb algorithm never has the highest proportion of best forecasts, and has the lowest individual event performance of any algorithm, with only 5.6% of the best-forecast locations on 15 February.

The 10:1 and climatological SLRs have the most sites with the worst snowfall forecast, accounting for more than 40% of the worst-forecast sites when combined for all but one of the events (Fig. 1b). On 13 March, climatology has the worst

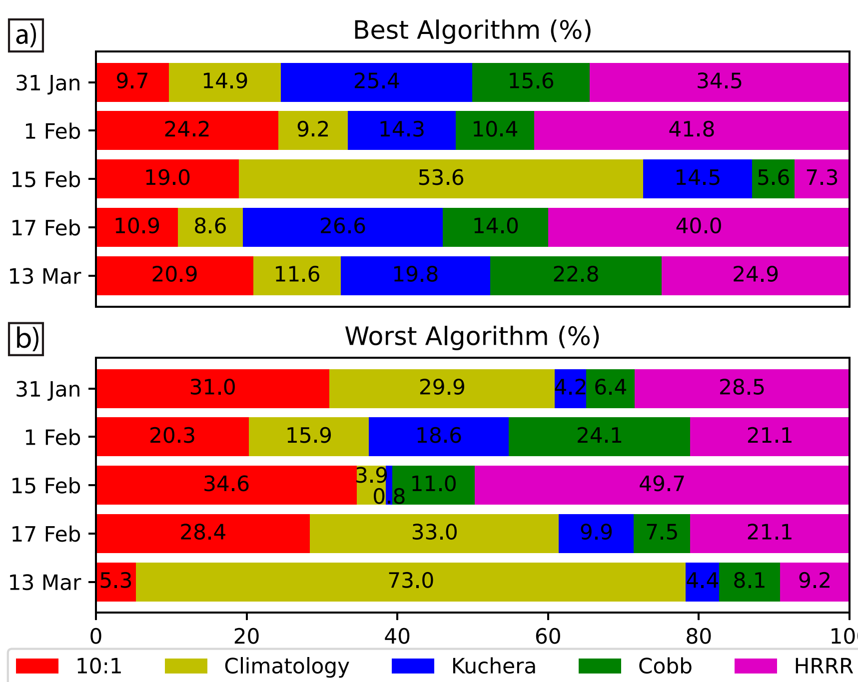

FIG. 1. Percentage of CoCoRAHS observations for each snow event where each SLR relationship (color) is the (a) best and (b) worst performing. Percentages may not add up to exactly 100% due to rounding.

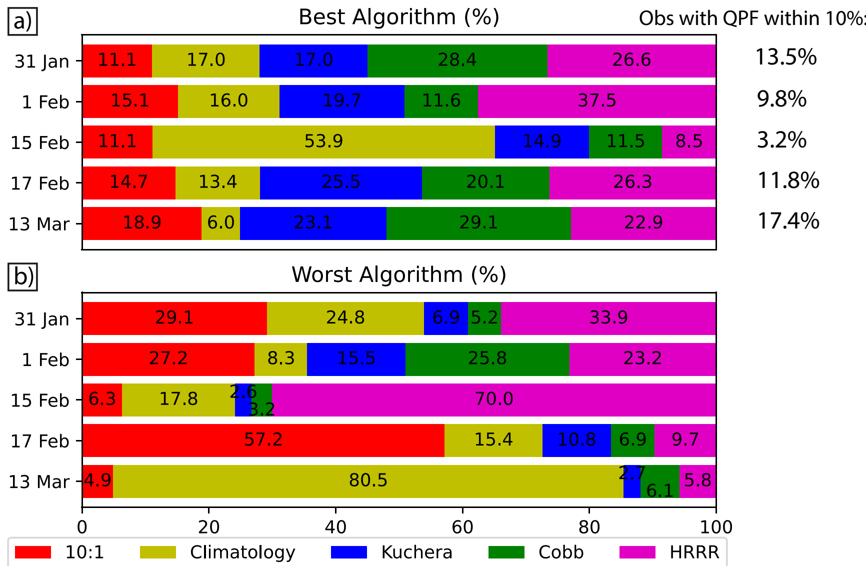

FIG. 2. As in Fig. 1, but for snow observations where the HRRRE QPF is within 10% of the observed CoCoRAHS liquid precipitation. Percentage of CoCoRAHS dataset meeting this additional criterion is indicated at the right side of (a).

forecast at 73.0% of the sites. On 1 February, the algorithm with the most sites having the highest error is the Cobb algorithm at 24.1%, but all algorithms are the worst at least 15% of the sites in this event. Performance can vary widely within the same event; on 31 January, despite being the best algorithm at 34.5% of the sites, the HRRR algorithm is also the worst at 28.5% of the sites. Interestingly, the Cobb algorithm is neither commonly the best algorithm, nor is it commonly the worst algorithm, never going over 25% in either the best or worst categories.

This analysis was also performed without the HRRR algorithm included, since that algorithm sees little use outside of the HRRR model itself. The results (not shown) were very similar to the five algorithms: there was no dominant SLR that clearly produces the lowest error across events. The major difference is that the Kuchera SLR was now typically the SLR with the most forecasts with the lowest error, replacing the HRRR. This is likely because of the two algorithms' substantial similarity as max temperature algorithms. In addition, the analysis was also repeated with a 15.24 cm (6 in.) minimum snowfall threshold. The heavier snowfall dataset still had the same general distribution, where there was not a clear best SLR (not shown).

To assess the sensitivity of these results to population bias, all CoCoRAHS observations are placed into bins on the HRRR grid based upon what HRRR grid point the observation was taken within. For HRRR grid points with multiple snowfall observations, subsequent observations are only retained if they differ by at least 5.00 cm from all other observations associated with that grid point. To simplify the analysis, this experiment was performed on the non-HRRR SLRs. This has the effect of removing observations that agree with others, while retaining those that are substantially different and may yield a different best or worst SLR. This method implicitly assumes that the substantially different subgrid measurements are correct and the difference is due to subgrid variability. Some of these measurements may be erroneous outliers, so this methodology increases their impact on the results. Removing the similar observations within the same grid point from each case reduces the number of snow-depth observations in each case by an average of 7.7% (in the four-SLR dataset). These removals do not appreciably change the results in Fig. 1, with a median difference of 0.3%, and a maximum difference of 3.9%. As the difference is small, the rest of this paper will use the full CoCoRAHS dataset.

The final experiment performed was to test the sensitivity of the results to QPF error. In this test, pairs of observations and member forecasts for that station are only kept if the difference between the observed liquid precipitation and the member's QPF is 10% or less. This test is meant to be restrictive, and test the SLRs only if the underlying QPF is close to correct. The results of this experiment are presented in Fig. 2. With these restrictive criteria, less than 20% of the original observation/member pairs are included in the analysis for every event. As with the other tests, the overall performance of the SLRs is similar to Fig. 1. There are some significant changes, such as the 10:1 SLR with a percentage difference of almost 30% higher on 17 February, or the HRRR having a 20% percentage difference on 15 February, though that case only has 3.2% of its 1297 observations meeting the QPF criteria. Despite those changes, only one case (15 February) has an SLR above 50% (climatology); only one other case, 1 February, has an SLR above 30%. Even if a certain SLR would become clearly the highest performing algorithm in this dataset, there is no practical way to know ahead of time

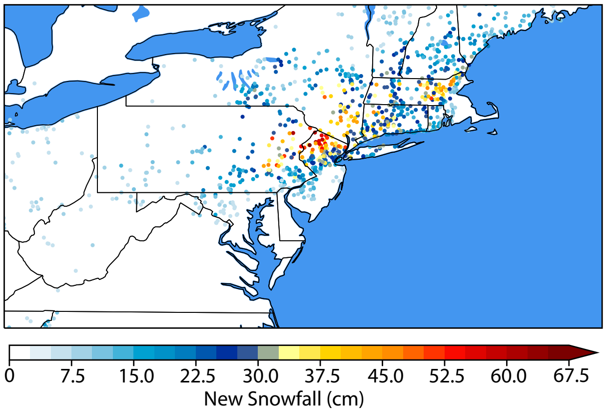

FIG. 3. CoCoRAHS 24-h new snow accumulations observed on 2 Feb 2021.

what locations in the model forecast will turn out to accurately predict QPF.

As with the algorithms with the lowest error, there are several cases for which the QPF restriction substantially changes the worst forecasts in Fig. 2b, with the HRRR algorithm on 15 February and the 10:1 relationship on 17 February had a percentage difference that was more than 20% higher than in Fig. 1b. This makes three cases where over 50% of the worst forecasts were a single SLR, but those SLRs were all different (HRRR, 10:1, and climatology).

4. Results—Case studies

An algorithm being best or worst based on the percentage of observations at which the algorithm performed best does not on its own tell the whole story. A closer look at the results from individual events is necessary to more fully understand how the choice of SLR impacts the ultimate forecast. Two of the five events in the analysis are examined in more detail below. This analysis will start with the four non-HRRR SLRs, as those are more widely available in output.

a. 1 February 2021

This event was the second day of a large, cyclone-driven precipitation event across the eastern United States. On 1 February, a new low pressure center developed off the New Jersey coast and slowly deepened (not shown). This slow-moving low produced copious amounts of snow over the northeastern United States (Fig. 3). A large swath of 24-h accumulations over 30 cm is present from south-central Pennsylvania through far southern Maine, with maximum snowfalls over 61 cm (2 ft) in northern New Jersey.

Before discussing snowfall errors, it is important to consider the role of liquid melted depth in the snowfall errors in the case study. Figure 4a contains a plot of the 24-h liquid precipitation observed by CoCoRAHS observers for this event. Because some volunteers may submit partial observations, not all stations with snowfall observations are represented here due to missing total precipitation. The HRRRE quantitative precipitation forecast covering the same timeframe for the 1 February event is presented in Fig. 4b. There is a consistent high bias to the HRRRE QPFs when compared with CoCoRAHS observations, with a mean difference of 5.3 mm more QPF in the HRRRE compared to the observed precipitation across all CoCoRAHS sites. There are regions where this bias is larger, such as eastern Massachusetts, and other regions where there was more precipitation observed than predicted by the HRRRE ensemble mean, such as northern and central New Jersey. It is worth noting that some of the difference between forecast and observations can be due to observation error, as measuring solid precipitation comes with additional uncertainty due to challenges such as gauge undercatch (e.g., Rasmussen et al. 2012).

Next, the spatial distribution of the SLR method with the lowest snowfall error is examined. Figure 5a shows how the best SLR varied across the CoCoRAHS observations based on HRRRE member 5, chosen for illustrative purposes. The best SLR for a given observation site tended to agree with other nearby snow observations' SLRs, with several large regions of agreement for the SLRs. The 10:1 SLR produced the

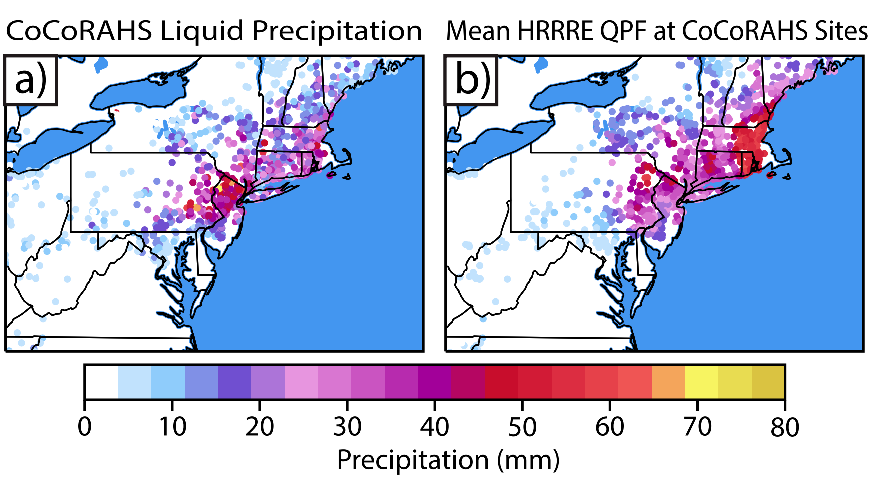

FIG. 4. (a) Observed liquid precipitation at CoCoRAHS stations used in this study for the 1 Feb case. (b) Ensemble mean QPF for each of the CoCoRAHS stations in (a).

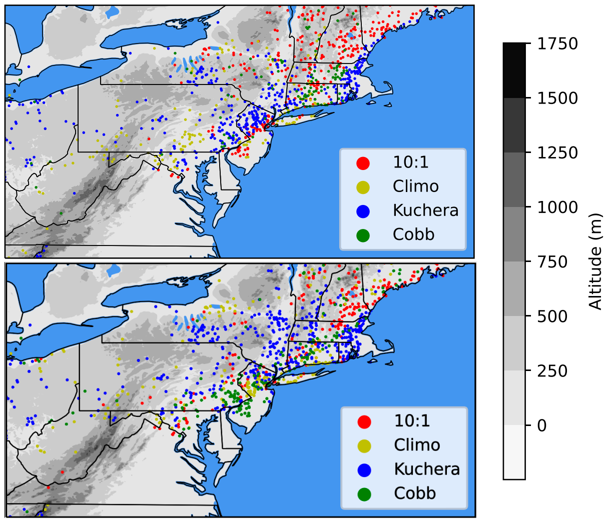

FIG. 5. (a) CoCoRAHS observations colored by the SLR that produced the lowest snowfall error for HRRRE member 5 on 1 Feb 2021. Terrain elevation shaded in grayscale. (b) As in (a), but for HRRRE member 2.

best forecasts in this member across northern New York, Vermont, New Hampshire, and southern Maine. The Kuchera ratio performed well in this member across eastern Pennsylvania and eastern Maine, as well as Ohio. The climatological ratio did well across south-central Pennsylvania, as well as southwestern Pennsylvania into West Virginia. The Cobb ratio was best from central Connecticut into central Massachusetts. In short, the best SLR varied regionally.

To illustrate the member-to-member variability, the same plot for another member, member 2, is displayed in Fig. 5b. In several regions, the best SLR method changes consistently between members; that is, where nearby observations share an SLR in Fig. 5a, the entire region consistently changes to another SLR in Fig. 5b. For instance, the region from southern Vermont and western Massachusetts changed from mostly having 10:1 as the best algorithm in Fig. 5a to the Kuchera being the best algorithm in Fig. 5b. Additionally, north-central New Jersey changed between members from either 10:1 or Kuchera in Fig. 5a to mostly Climatology in Fig. 5b as the best SLR, while southeast Pennsylvania and southern New Jersey changed to the Cobb algorithm as the best in Fig. 5b. For both members presented, the maximum swath in snowfall from Fig. 3 generally has the Kuchera algorithm as the best forecast for many sites.

In addition to declaring a best or worst forecast from the SLR algorithms, it is important to examine how absolute snowfall error varies across the algorithms. Figure 6 presents the error distribution from two perspectives: the absolute error (|forecast − observation|) and bias error (forecast − observation; Wilks 2019). These results are shown in Fig. 6a, where all forecast absolute errors from all CoCoRAHS observations and all members are binned by SLR. The distribution of errors by SLR method is very similar across all four SLRs, with the 10:1 having the lowest median error at 7.4 cm, both the Kuchera and Cobb algorithms having median errors of 7.8 cm, and the climatology method having a median error of 8.3 cm.

To test the statistical veracity of the median differences, bootstrap resampling analysis was performed. (Hamill 1999) Each median was calculated with 9999 resamples to calculate 95th percentile confidence intervals. For the four SLRs investigated, the 95th confidence intervals draw similar conclusions as the medians themselves, with the 10:1 SLR ([7.16, 7.54] cm) having the lowest range, the Kuchera SLR ([7.602, 8.006] cm) and the Cobb SLR ([7.561, 7.969] cm) having mostly overlapping confidence ranges, and the climatology SLR ([7.994, 8.493] cm) having the highest median absolute error.

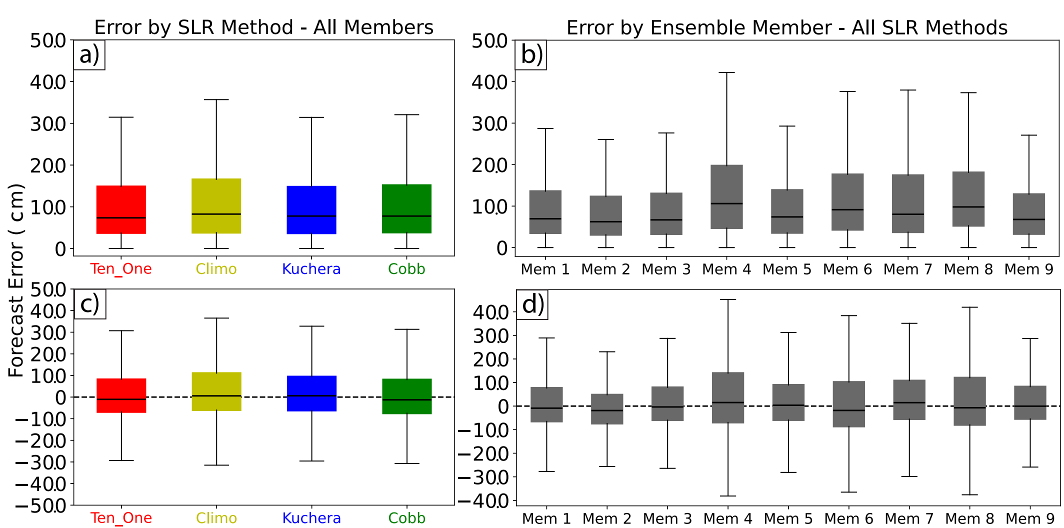

FIG. 6. Snow accumulation absolute forecast errors for all observations on 1 Feb 2021, sorted by (a) SLR method and (b) ensemble member. Whiskers represent 1.5 × the interquartile range. (c),(d) As in (a) and (b), but the errors are the bias errors.

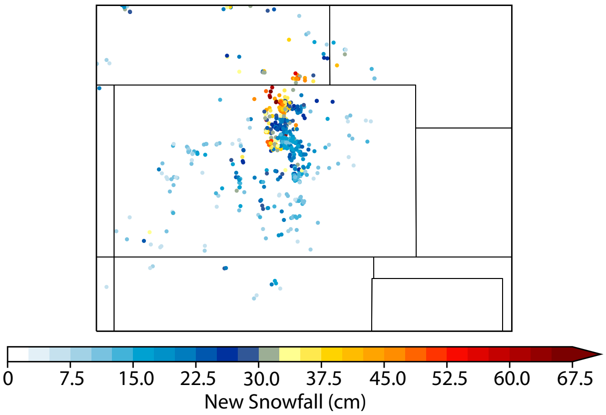

FIG. 7. As in Fig. 3, but for new snowfall observed on 14 Mar 2021.

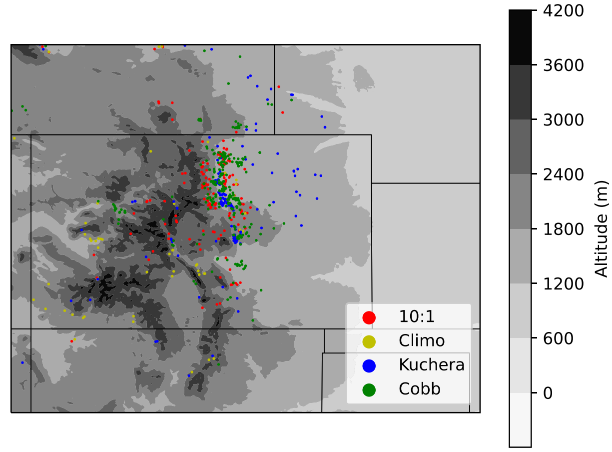

FIG. 8. As in Fig. 5, but for HRRRE member 5 for 13 Mar 2021.

In Fig. 6b, all forecasts and all SLRs are binned by the individual ensemble member. Median absolute snowfall errors across members varied from member 2 at the low end with 6.3 cm to member 4 at the high end with a median error of 10.6 cm. The best case in terms of member, member 2, had a lower median error than the 10:1 ratio, the SLR with the lowest median error, and the worst member has a higher median error than the SLR with the highest median error, the climatology SLR. Thus, the choice of member here was slightly more impactful than the choice of SLR.

Bootstrapping analysis was performed for the per-member diagnostics as well. One benefit of this analysis on the individual members is that it reduces issues associated with the sampling disparity between having nine members per SLR in Fig. 6a and four SLRs per member in Fig. 6b. As with the per-SLR medians, the 95th percentile confidence intervals for the individual members range from member 2's ([6.002, 6.478] cm) to member 4's ([10.154, 10.998] cm). In this case, any SLR choice gives you a higher median error across the ensemble than specifically choosing member 2, and a lower median error than choosing member 4.

When examining the bias error distribution in Fig. 6c, all SLRs have a near-zero median. This indicates that for this case, the individual algorithms have minimal bias, with the climatology and Kuchera having a slightly positive median error, and the 10:1 and Cobb having a slightly negative median error. The 10:1 median bias error being negative is consistent with the median SLR being higher than 10:1 (e.g., Roebber et al. 2003). Similarly, in the member-by-member plots in Fig. 6d, the individual members have small positive or negative biases in the median error of similar magnitude to those of the individual SLR ratios. For this case, the choice of SLR does not significantly cause bias in the median snowfall forecast.

In addition, comparing the median errors to the percentages where the SLRs are the best forecast in Fig. 1 yields interesting results. In Fig. 1a, the Kuchera algorithm had the most best forecasts by quantity, yielding the lowest error 42.1% of the time. However, its median error is in the middle of the SLRs in Fig. 7a, and the 10:1 ratio, while being best for 29.2% of forecasts, has the lowest median error. The Cobb algorithm had the same median error as the Kuchera SLR, but was only the algorithm with the lowest error 13.8% of the time, the smallest fraction of all algorithms for this case. These results indicate that determining a best method is not a straightforward endeavor. Rather, the definition of best is dependent on what aspect of the distribution the statistical analysis is optimized toward (lowest median error, largest number of sites with the lowest error, etc.). This is analogous to the multiple ways a forecast can more generally be considered best (Murphy 1993), where the use of the resultant forecast drives the way in which it scored.

b. 13 March 2021

The 13 March event was the result of sustained lee cyclogenesis in eastern Colorado, which created a long-duration precipitation event. Temperatures are initially above freezing for the most part in the plains east of the mountains, and these locations undergo a rain-to-snow transition as the event goes on (not shown). Figure 7 shows the CoCoRAHS snowfall report for this event, with the largest accumulations (>30.0 cm) reported from southeastern Wyoming to central Colorado, with a tight gradient between these heavier snows and lesser amounts across nearby plains.

A map of the SLR with the best snowfall forecast for each CoCoRAHS observation in this event from HRRRE member 5 forecast is shown in Fig. 8 as an example. As with the 1 February case, SLR methods tend to be the best in clusters. Here, the Kuchera algorithm is the best for most of the observations across the eastern plains of Colorado and Wyoming. For the high density of observations in and around the urban corridor east of the Front Range mountains in northern Colorado, the Cobb algorithm was commonly the SLR with the lowest absolute error, alongside the Kuchera algorithm. The 10:1 relationship performed best near and to the west of the leading mountains in northern Colorado, including many observations in the highest snowfall regions in Fig. 7. The sites where the climatological SLR performs best are generally limited to sites in the mountains, particularly southwestern Colorado. The member-to-member variability of best forecast by the various

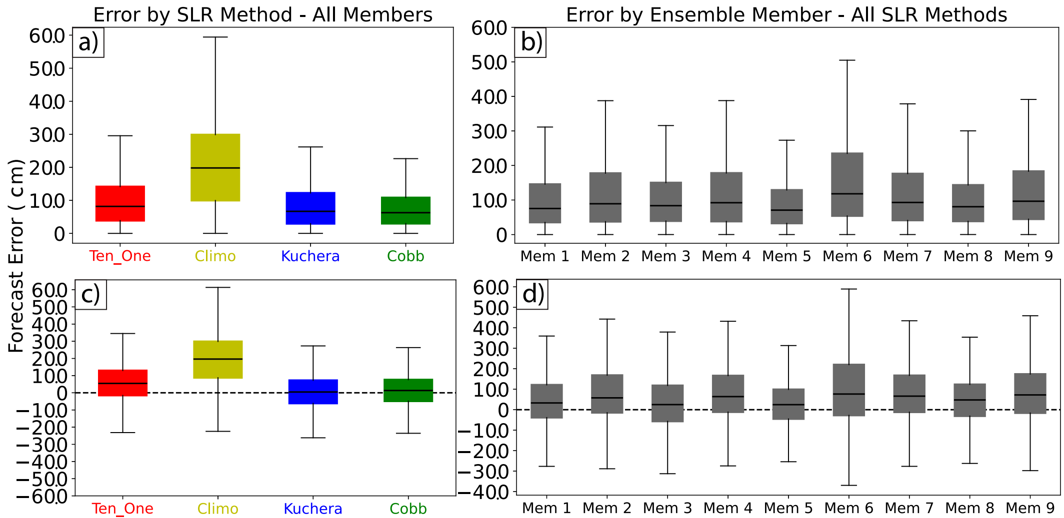

FIG. 9. As in Fig. 6, but for 13 Mar 2021.

SLRs has a similar consistency as the previous case. In some areas, the SLR method producing the best snowfall forecast are clustered. In others, there are variations, even among nearby CoCoRAHS sites. Where some nearby observations—here, mainly those in the eastern plains—consistently have the same best SLR, but others change between members, often changing simultaneously with nearby observations (not shown).

The distribution of snowfall forecast errors for the 13 March event is shown in Fig. 9. When separated by SLR (Fig. 9a), three SLR methods have similar interquartile ranges: the 10:1, Kuchera, and Cobb SLRs. The latter two have slightly lower errors, with the Kuchera SLR having a median error of 6.7 cm, and the Cobb SLR having a median error of 6.3 cm. The 10:1 ratio has a median error of 8.2 cm, about 25% larger than the NWP-derived SLRs. The remaining SLR, climatology, does poorly on this event, with a median error of 19.8 cm. This large error for the climatology SLR suggests that this storm is atypical when compared to historic events in the region in March. The better performance of the 10:1 ratio, which is lower than the climatological SLR this time of year in these regions, and the Kuchera algorithm, which would produce smaller SLRs with the near-freezing profiles in the eastern plains, suggest that the climatology SLR is too high for this event. Denver, Colorado, actually sees its highest monthly snowfall in March (NWS 2022). As snowfall is common this time of year, Colorado's March SLR climatology is more robust than other parts of the CONUS, for which March snow may be uncommon. This highlights the importance of understanding the SLR algorithm(s) being applied to a forecast. In the case of the 13 March storm, for instance, an experienced forecaster evaluating NWP forecasts could notice that the storm is not consistent with typical March snow events in Colorado, and disregard any snow accumulation forecasts using the climatological SLR.

Similar to the previous case, the errors in individual ensemble members are similar to the error distribution across the SLR methodologies. The exception is the climatological SLR, due to its consistently worse performance than other SLRs. The upper bound of the interquartile ranges for the individual members are higher than those of the three more accurate SLRs, with many above 15 cm, but this is influenced by the high errors associated with the climatological SLR (Fig. 9). Using only the three other SLRs (10:1, Kuchera, Cobb) yields lower errors on a member-by-member basis (not shown). With the exception of this case, the rest of the cases in this study were similar to the 1 February case in that there is not an SLR that is consistently worse for almost all observations. A potential follow-on to this study would be to gather a much larger ensemble NWP dataset and determine the types of events for which the individual SLRs may be disregarded in a forecast.

The larger absolute error of the climatological SLR is likewise represented in the median bias error, with a median error of nearly 20.0 cm in Fig. 9c. This imparts a strong positive bias across the climatology snowfall forecasts. The 10:1 has a smaller, but still substantial positive median error near 5.0 cm. The Kuchera algorithm has a near-zero median error in Fig. 9c, and the Cobb algorithm has a small positive median error. The combination of the climatology SLR and the 10:1 SLR errors strongly influence the individual member errors in Fig. 9d, where all nine HRRRE members have a positive bias to their median snowfall forecast.

Performing bootstrapping on the calculation of the median shows the Cobb SLR ([61.25, 64.51] cm) has the lowest 95% confidence interval for the median, though there is slight overlap with the Kuchera SLR ([64.81, 68.79] cm) confidence interval. The 10:1 SLR ([79.65, 84.44] cm) has a higher median range, and the climatology SLR ([193.02, 202.29] cm) lower bound is more than double the upper bound of the third-place SLR's confidence interval. The median absolute error across the ensemble is lower for both the Kuchera and Cobb algorithms than the best-performing member, member 5 ([67.30, 73.87] cm). The worst-performing member, member 6 ([110.72, 127.00] cm) performed substantially worse than the

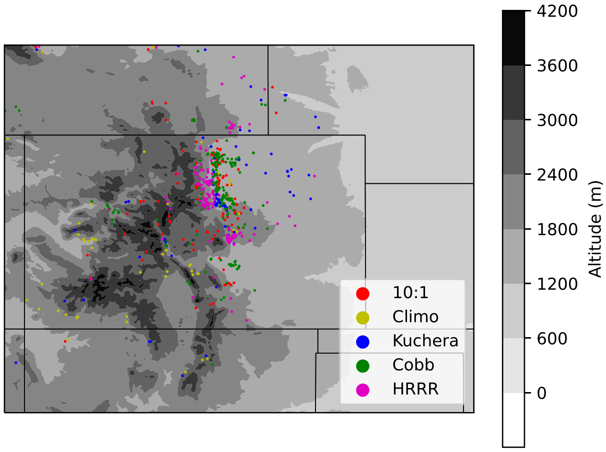

FIG. 10. As in Fig. 8, but including HRRR snowfall forecasts.

10:1 SLR. Here, the choice of SLR is more important than choice of member, as using the ensemble of SLRs is worse than two SLRs for all members, and worse than three SLRs for some members.

To illustrate the sensitivity of the geographic dependency on the number of SLRs included, the analysis in Fig. 8 was rerun, including the HRRR SLR with the other four SLRs. The results including the HRRR are presented in Fig. 10. Consistent with the similarities between the algorithms, a substantial number of the best forecast changes from Fig. 8 to the HRRR are from the Kuchera algorithm, particularly near the Front Range in northern Colorado. However, not all locations that were best-forecast by the Kuchera SLR are supplanted by the HRRR forecast. Locations across northeast Colorado in particular still have the smallest error when the Kuchera algorithm is used with this member, even when the HRRR snowfall is added in. While a few other stations switch to the HRRR, most other regions remain with the same best algorithm as in Fig. 8. Consistent with the overall statistics in Fig. 1, the addition of the HRRR does not make any method an obvious best algorithm, regardless of member chosen.

There are five different forecasts per member, per location. To this point, it has been demonstrated that changing SLR relationships with the same underlying NWP forecasts yields no strong signal as to the most performant choice of SLR method. It is also instructive to consider the practical impacts of the choice of SLR. (The hourly progression of snow rate during the 13 March case for Cheyenne, Wyoming, is shown in Fig. 12.) Each of the SLRs' ensemble mean snow rate is shown in Fig. 11. These ensemble means vary by more than that 2.54 cm h^{-1}, due to climatology's poor performance for this event. Even ignoring the climatology SLR as an outlier, the mean snow rates differed by over 1.0 cm h^{-1}, with the difference maximizing at forecast hour 17.

The differences between SLRs is significant enough to change decision-making substantially. Included on Fig. 11 is a line at 2.54 cm h^{-1}, representing a hypothetical decision point for a stakeholder. This is done for illustrative purposes to show the impact of SLRs on exceedance of a threshold; the general interpretation of this figure would be similar with other snow-rate decision points. These decision points would, for example, be part of a decision support tool requested by a partner. Blindly trusting the climatological SLR, for instance, would only drop below the decision point from hours 14 to 16; all other mean snowfall rates drop below the decision point from hours 13 to 17. The more complex SLRs vary with time (Kuchera, Cobb, HRRR), and these variations too can impact decision-making. The Kuchera SLR snow rate, for instance, is in the middle of these three SLRs through hour 10, but the snow rate jumps up to higher than the 10:1 SLR snow rate at hour 12. This makes the Kuchera algorithm the only algorithm above the decision point at hour 22, besides

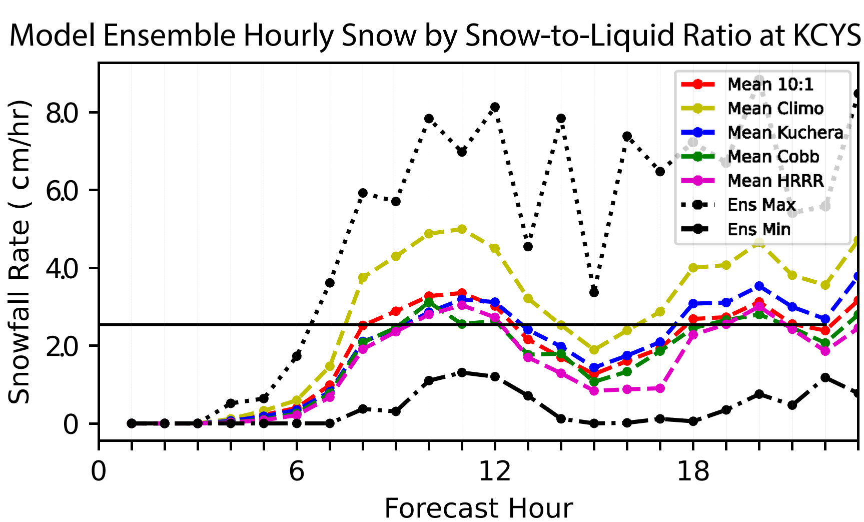

FIG. 11. Hourly snowfall rate across the HRRRE ensemble and all SLRs for Cheyenne, WY, for the 13 Mar case. Colored lines represent ensemble mean snowfall rates for each snowfall forecast. Black lines represent the extrema (minimum and maximum) at each time; these forecasts may come from different members and SLRs at different hours.

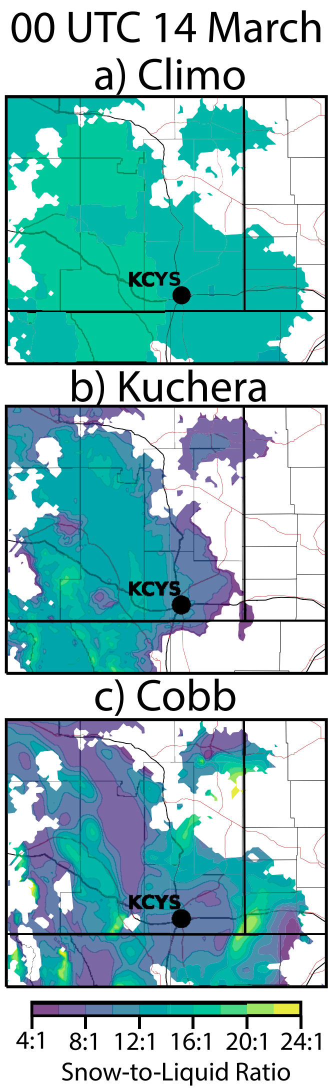

FIG. 12. Derived SLR for HRRRE member 6 at 0000 UTC 14 Mar 2021 for (a) climatology SLR, (b) Kuchera SLR, and (c) Cobb cSLR. The location of Cheyenne, WY (KCYS), is indicated by the dot.

climatology. If one considers the ensemble maximum and minimum snow rates, a wide spectrum of solutions are possible. In the worst case, the ensemble exceeds the decision point at hour 7 and never goes below for the rest of the period. In the best case, the ensemble never reaches above 2.54 cm h^{-1}. These results suggest that a probabilistic approach to SLR might be required to capture the true range of solutions.

To illustrate the distribution of the underlying SLRs for the snow rates in Fig. 11, the spatial variability of SLRs (other than the constant 10:1) near Cheyenne, Wyoming (KCYS), from one ensemble member is shown in Fig. 12. These SLR analyses are valid at forecast hour 12 in Fig. 11, a time with substantial spread between the mean climatological SLR snow rate and the other methods. The climatology SLR does not vary much spatially, with the highest value of 14.9:1 at KCYS, consistent with the largest snow rate at forecast hour 12 in Fig. 11. The Kuchera SLR has higher values across the colder high terrain to the west, and lower values over the plains to the east, with a value of 10.2:1 at KCYS. The Cobb SLR method produces the most variable field, owing to its reliance on the model's vertical velocity. The Cobb SLR has a value of 9.7:1 at KCYS. The Cobb method's spatial variability shows up in Fig. 11 in the Cobb method's temporal variability (Cobb 2011).

These plots also show potential limits of the methods. For instance, in Fig. 12b), the Kuchera method produces no output southeast of KCYS, as max column temperatures are too warm. West of KCYS, there is a hole in the Cobb analysis near the Medicine Bow Mountains, where orographic descent would be expected on the downwind of the mountains. As SLR itself is not a field that is typically plotted, these limitations are only potentially detectible in downstream fields (e.g., snowfall).

5. Discussion and conclusions

Here, the results of calculating the ensemble error statistics for 24-h snowfalls across five events are presented. The overarching result of this work is that snowfall depth errors due to the choice of SLR are of similar magnitude to the snowfall errors due to the choice of ensemble member. This is despite the SLRs in this study ranging from simplistic (everywhere is 10:1) to complex (SLR depends on temperature, relative humidity, and vertical velocity in a column). This implies that for the purposes of producing snowfall forecasts, there is not a clear choice of the best SLR to produce the most accurate forecast. This remains true when model-calculated snowfall from the HRRR is included as an option; the HRRR's performance was within the envelope of performance of the SLR methods in this study. Additionally, constraining the analysis to observations with low QPF errors did not appreciably change the results and did not yield a best SLR despite the accurate liquid precipitation input. The small portion of observations that met the restrictive criteria for QPF from this study, less than 20%, illustrates the inherent challenge of snow forecasting, as the errors in the forecast and observations of liquid equivalent precipitation tend to be relatively large compared to the actual quantity.

Upon closer examination of the SLR errors in this study, there appear to be potential ways to overcome the inability to improve snowfall forecasts by selecting a single SLR. When the best SLR for individual observations are examined, there is often a regionality to them; that is, where a given SLR is best for one observation, there are often several nearby observations that have the same SLR as giving the lowest error. Thus, if there is a way to determine where an SLR is going to perform better than the others, errors could be lowered, regardless of the underlying reasons for this regionality (actual variation in SLR, NWP errors, variations in atmospheric parameters that feed SLR algorithms, etc.). One avenue of future work would be to determine the meteorological conditions under which these SLR relationships (and others) have higher and lower errors. This would follow the conclusions of Ware et al. (2006), which suggested that meteorological conditions, principally thermodynamic profiles, can constrain SLRs to improve forecasts.

For example, in the 13 March case presented here, the Kuchera SLR was consistently the best performing in eastern Colorado, a region where temperatures were marginal for snow. As the only algorithm that can respond to marginal near-surface temperatures due to its use of maximum-column temperature, it could be expected that the Kuchera algorithm would do well for that region when forecasting. Even where it is clear a certain SLR will perform better in one region, other SLRs will typically outperform that algorithm in other regions. Thus, a "best" SLR field would likely have to consist of multiple SLR methods, at least for the cases in this study. One possibility for handling this would be to handle SLR in a probabilistic framework, where SLR is treated as another axis of uncertainty in the ultimate snowfall forecast. An in-depth look at the underlying physical mechanisms for the SLR algorithm's regionality may be useful to these efforts.

What constitutes the best SLR also depends on how the best algorithm is defined. Sometimes, a given SLR would have the smallest error across the most stations by number, but a different SLR would have a lower median error. This illustrates that finding regions where individual SLRs perform poorly would likewise be important to producing a best guess SLR, as this would help eliminate big misses that bring down an otherwise performant SLR. Discussions with stakeholders will be necessary to determine what type of "goodness" they prefer to maximize in their forecasts (Murphy 1993).

Another potential method to improve SLRs used to forecast snowfall is to eliminate from consideration any that are known to be inapplicable to a given event. For instance, the 13 March case was not a storm consistent with the climatological SLRs for that time of year, so being able to exclude the climatological SLR from use improves the forecast. This would be true for any algorithm; as another example, the 10:1 relationship would be a poor choice for lake-effect snow, where SLRs are typically higher (Baxter et al. 2005).

These potential improvements to SLR could lead to improved decision support. Variation in SLR alone can yield a spread of rates over 1 cm h^{-1} across the ensemble, which in turn can change the amount of time a decision threshold is reached. Any ability to narrow the range of SLRs possible for a particular event would reduce uncertainty in decision-support tools.

This study represents an initial look at this problem. Future studies are needed to determine the applicability of these results to ensembles beyond the HRRRE, and to ensure the results presented here are generalizable over a larger seasonal dataset. Likewise, the selection of SLRs could be expanded beyond those examined here to determine if there are certain meteorological regimes not well-handled by the SLRs in this study.

This study raises some additional questions about the role of SLR algorithms. The first of these being whether substantial value can be had from input-intensive SLR relationships, or if the uncertainty in the underlying fields is too great, even at relatively short forecast times. Indeed, it is not obvious from these results that one should always expect a great improvement in forecasts from running the input-intensive Cobb algorithm over just using 10:1, or even the Kuchera algorithm, which only has one input. However, as suggested above, if constraining the expense of the NWP-based algorithms to regions where they perform well is possible, they can be used to improve the SLR forecast. Additionally, the interaction in spatial variability of SLR estimates and QPF features across the ensemble should be examined, particularly in the context of mesoscale structures of interest (e.g., frontogenetical bands).

Another area of future investigation with SLR algorithms is whether they can represent another axis of uncertainty in ensemble design. If each member/SLR combination is considered to be a separate forecast, does this effectively increase ensemble size, and thus the ability to represent snowfall probabilities better? An ensemble approach to SLR ratios would allow SLRs that do not work well for a given event to be outliers, enabling the forecaster to eliminate those from consideration if the solution does not seem realistic.

Finally, this study focused on four SLRs, but more exist; how these other SLR relationships would fit in this framework would also be a useful follow-on study. As NWP microphysical packages continue to advance and better represent the particle size distribution, it is possible that information from these, up to and potentially including an explicitly modeled SLR, may be useful. It may also be worthwhile to revisit artificial intelligence as a solution (e.g., Roebber et al. 2003; Ware et al. 2006) as either a new approach to SLR or as a way of choosing SLR relationships.

Acknowledgments. This work was supported by NOAA under Grant NA19OAR4590230. Additional funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA21OAR4320204, U.S. Department of Commerce. HRRRE data were provided by NOAA's Global Systems Laboratory. The authors thank Evan Kuchera for his discussion of the SLR relationship he developed. The authors acknowledge CoCoRAHS for the collection and storage of snow accumulation measurements. We thank Michael Baldwin and three anonymous reviewers for suggestions to improve this manuscript.

Data availability statement. HRRRE data were obtained directly from NOAA's Global Systems Division. CoCoRAHS data are available from https://api.cocorahs.org/. Climatological SLR datasets are available online at https://www.nohrsc.noaa.gov/snowfall/data/resources/SLR_climatology/. Because the authors were not authorized to share the experimental data from the HRRRE, the derived snow accumulation data created from the HRRRE as a part of this research cannot be made openly available.

REFERENCES

Alcott, T. I., and J. W. 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.

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., E. P. James, J. M. Brown, E. J. Szoke, J. S. Kenyon, R. Ahmadov, and D. D. Turner, 2021: Diagnostic fields developed for hourly updated NOAA weather models. NOAA Tech. Memo. OAR GSL-66, 55 pp., https://repository.library.noaa.gov/view/noaa/32904.

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., 2011: Snow/liquid ratio. 2 pp., https://www.weather.gov/media/mdl/SLR.pdf.

——, 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.

CoCoRAHS, 2014a: "In depth" snow measuring. CoCoRAHS, 73 pp., https://media.cocorahs.org/docs/MeasuringSnow2.1.pdf.

——, 2014b: CoCoRaHS data quality assurance and quality control. CoCoRAHS, 6 pp., https://media.cocorahs.org/docs/CoCoRaHS_QA_QC_April_2019.pdf.

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.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

Kalina, E. A., I. Jankov, T. Alcott, J. Olson, J. Beck, J. Berner, D. Dowell, and C. Alexander, 2021: A progress report on the development of the High-Resolution Rapid Refresh ensemble. *Wea. Forecasting*, **36**, 791–804, https://doi.org/10.1175/WAF-D-20-0098.1.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

NOHRSC, 2022: Index of /snowfall/data/resources/SLR_climatology. NOAA, accessed 17 November 2022, https://www.nohrsc.noaa.gov/snowfall/data/resources/SLR_climatology/.

NWS, 2022: Denver seasonal snowfall. NWS, accessed 17 November 2022, https://www.weather.gov/bou/SeasonalSnowfall.

Ralph, F. M., and Coauthors, 2005: Improving short-term (0–48 h) cool-season quantitative precipitation forecasts: Recommendations from a USWRP workshop. *Bull. Amer. Meteor. Soc.*, **86**, 1619–1632, https://doi.org/10.1175/BAMS-86-11-1619.

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.

Reeves, H. D., 2016: The uncertainty of precipitation-type observations and its effect on the validation of forecast precipitation type. *Wea. Forecasting*, **31**, 1961–1971, https://doi.org/10.1175/WAF-D-16-0068.1.

Roebber, P. J., S. L. Bruening, D. M. Schultz, and J. V. Cortinas Jr., 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.

Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.*, **142**, 4519–4541, https://doi.org/10.1175/MWR-D-14-00100.1.

Smirnova, T. G., J. M. Brown, S. G. Benjamin, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle Land Surface Model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, https://doi.org/10.1175/MWR-D-15-0198.1.

Ware, E. C., D. M. Schultz, H. E. Brooks, P. J. Roebber, and S. L. Bruening, 2006: Improving snowfall forecasting by accounting for the climatological variability of snow density. *Wea. Forecasting*, **21**, 94–103, https://doi.org/10.1175/WAF903.1.

WDTD, 2022: Snowfall forecasting. Accessed 17 November 2022, https://training.weather.gov/wdtd/courses/woc/winter/fcst-hzds/snow-forecast/presentation_html5.html.

Wilks, D., 2019: *Statistical Methods in the Atmospheric Sciences*. 4th ed. Elsevier, 840 pp.