Evaluation of Wintertime Precipitation Estimates and Forecasts in the Mountains of Colorado

JANICE L. BYTHEWAY,[a] WILLIAM R. CURRIER,[a] MIMI R. ABEL,[a] KELLY MAHONEY,[a] AND ROB CIFELLI[a]

[a] *NOAA/Physical Sciences Laboratory, Boulder, Colorado*

(Manuscript received 15 September 2023, in final form 2 January 2024, accepted 8 February 2024)

ABSTRACT: Wintertime precipitation poses many observational and forecasting challenges, especially in the complex topography of the western United States where radar beam blockage and difficulty siting in situ observations yields more sparse observations than in the eastern United States. Uncertainty in western U.S. winter precipitation is known to be high, so much so that some studies have found model simulated precipitation to produce similar or better large-scale estimates of annual precipitation than gridded observational products during climatologically anomalous years. This study evaluates high-resolution gridded precipitation estimates from Multi-Radar Multi-Sensor (MRMS) and Stage IV as well as forecasts from NOAA's High-Resolution Rapid Refresh (HRRR) model in the Colorado Rocky Mountains. Gridded precipitation estimates and forecasts are compared with in situ SNOTEL measurements for two seasons of wintertime precipitation. The influence of forecast length, lead time, and model elevation on seasonal precipitation predictions from the HRRR are investigated. Additional comparisons are made with the relatively dense network of observations deployed in Colorado's East River Watershed during the Study of Precipitation, the Lower Atmosphere and Surface for Hydrometeorology (SPLASH) campaign. Gridded products and forecasts are found to underestimate cold-season precipitation by 25%–65% relative to in situ and aircraft measurements, with longer forecast periods and lead times (6–24 h) having smaller biases (25%–30%) than shorter forecast periods and lead times (55%–65%). The assessment of multiple years of observations indicates that these biases are related more to the data and methods used to create the gridded products and forecasts than to precipitation characteristics.

SIGNIFICANCE STATEMENT: In the mountainous western United States, it is very challenging to both observe and forecast wintertime precipitation, yet snowfall plays an important role in providing the region's annual water supply. This study aims to increase our understanding of the biases in observations and forecasts of snowfall in the Colorado Rocky Mountains, which can in turn impact forecasts of water availability for the ensuing warm season. In this study we find high-resolution gridded precipitation estimates and forecasts to underestimate cold-season precipitation when compared with in situ observing stations, with longer-range forecasts (e.g., daily) being the least biased. These findings were consistent over two years of study and have broad implications for the hydrologic modeling and water management communities.

KEYWORDS: Complex terrain; Precipitation; Snow; Forecast verification/skill

1. Introduction

Annual snowfall produces approximately 70% of the runoff in the western United States (Li et al. 2017) making snow an important hydrologic resource necessary to fill the reservoirs that provide water supply through the drier summer months. In the Colorado River basin, which provides water to 40 million people and millions of acres of agricultural land across seven U.S. states, approximately 85% of the streamflow is a result of snowmelt in the basin's headwaters (Ikeda et al. 2010). Long-term drought and excessive demand in the basin have led to record low water levels, triggering use restrictions at the local level and complex interstate discussions on how to reduce usage and fairly allocate water among users (Ge et al. 2023).

Hydrologic models are used to predict the fate of the wintertime precipitation that falls in the mountains. Streamflow predictions can help water managers anticipate how much water will potentially be available to fill reservoirs for later use. Precipitation, along with radiation and antecedent soil moisture conditions, has been found to largely influence errors in modeled streamflow (Tang et al. 2023; Raleigh et al. 2015). Both the quantity and phase of the precipitation are important to accurately predict streamflow, particularly at relatively small spatial and temporal scales (Wayand et al. 2013; Currier et al. 2017; Li et al. 2019). While hydrologic models are being developed with increasing spatial and temporal resolution and more sophisticated process representations, the benefits of these advances cannot be fully exploited without accurate estimates of the initial meteorological state, including precipitation and snow liquid water equivalent (SWE) estimates (Li et al. 2019).

Despite the importance of wintertime precipitation, it is inadequately monitored due to challenges in directly observing

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-23-0158.s1.

Corresponding author: Janice Bytheway, janice.bytheway@noaa.gov

DOI: 10.1175/JHM-D-23-0158.1

snowfall in complex terrain, including uncertainties in in situ observations (e.g., gauge undercatch) or the logistical constraints of measuring precipitation in remote and protected areas. Additionally, the uncertainties introduced when interpolating between these point-based observations are large (Rasmussen et al. 2012; Goodison et al. 1998; Martinaitis et al. 2015; Henn et al. 2018; Kochendorfer et al. 2022). Remotely sensed estimates of precipitation in mountainous areas suffer from radar beam blockage (Maddox et al. 2002; Westrick et al. 1999), large uncertainty in the reflectivity–precipitation rate relationship (Kirstetter et al. 2015; Villarini and Krajewski 2009) and large uncertainties in satellite-based observations, particularly of frozen precipitation and precipitation falling over snow-covered surfaces (Bartsotas et al. 2018; Derin et al. 2016; Ebert et al. 2007; Tian and Peters-lidard 2010; Sun et al. 2018; Cao et al. 2018). The uncertainties associated with observationally based precipitation datasets in complex terrain are large enough that Lundquist et al. (2019) argued that model-simulated estimates of rangewide annual precipitation may be as good or better than products derived from interpolated in situ observations.

The assertion of Lundquist et al. (2019) that modeled precipitation estimates can be more skillful than interpolated observational datasets is limited to the mid- to northern latitudes and rangewide total annual precipitation, that is, over a long accumulation period and large spatial average. While some of the studies cited by Lundquist et al. (2019) examined precipitation on somewhat shorter seasonal time scales, all used well configured model simulations run over long periods of time (Gutmann et al. 2012; Wayand et al. 2013; Currier et al. 2017). This context (long duration simulations over a large spatial area) is often omitted in studies that cite model superiority over observations in estimating mountain precipitation, with some studies applying it to shorter-term numerical weather prediction (NWP) model forecasts (e.g., Jennings et al. 2023; English et al. 2021; Bytheway et al. 2022; Li et al. 2023; James et al. 2022; Gouttevin et al. 2023).

Many studies (e.g., Anderson et al. 2002; Westrick et al. 2002; Kunstmann and Stadler 2005; Smiatek et al. 2012; Förster et al. 2014, 2018) have found that hydrologic models forced with gauge observations provided better streamflow than those forced with NWP forecast precipitation [i.e., quantitative precipitation forecasts (QPF)]. Despite these findings, operational hydrologic models are often forced with short-range, frequently updated forecast models. For example, the operational short-term National Water Model (NWM) currently receives precipitation information from the operational High-Resolution Rapid Refresh (HRRR) model in conjunction with quantitative precipitation estimates (QPE) from the Multi-Radar Multi-Sensor (MRMS) system (Zhang et al. 2016; Martinaitis et al. 2020). A previous version of the HRRR was found to underestimate seasonal precipitation in a high-elevation watershed in southwest Colorado by 20% (Cao and Barros 2023). Thus, it is important to understand how a rapidly updating, high-resolution forecast model such as the HRRR fits into the idea that models may be equally or more skillful at representing wintertime precipitation in complex terrain than observationally based gridded products.

In this study, we compare the performance of wintertime precipitation forecasts from HRRR at a number of forecast lengths and lead times with the performance of low-latency high-resolution gridded precipitation estimates from MRMS and Stage IV. These products are selected for their comparable resolution to HRRR (1 h; <5 km) and status as operational products. While MRMS and Stage IV are multisensor products over much of the eastern United States, in western Colorado a lack of radar coverage causes to these products to become more reliant on in situ and other data sources (see section 2a). Reference precipitation estimates include observations of SWE and snow-adjusted precipitation at ground-based Snowpack Telemetry (SNOTEL) sites across the state of Colorado, which includes much of the upper Colorado River Watershed. Additionally, the dense observation network deployed for the Study of Precipitation, the Lower Atmosphere and Surface for Hydrometeorology (SPLASH) field campaign (de Boer et al. 2023) as well as airborne SWE measurements from the Airborne Snow Observatory (ASO) over the SPLASH domain are used for a more in-depth comparison at a variety of elevations. We focus on water years (WY) 2022 and 2023 both for the availability of the SPLASH observations, and for comparison of QPE and QPF performance in two very different water years.

We seek to answer the following science questions:

1) Is the operational HRRR model as good as, if not better than, observationally based, low-latency QPE products with high spatiotemporal resolution at estimating wintertime precipitation in Colorado?
2) Are there ranges of forecast hours, lead times, or forecast lengths that provide precipitation forecasts that may be better suited for forcing hydrologic models?
3) How do storm characteristics impact the results from 1 and 2 above?

Section 2 outlines the various datasets used in this study, including any quality control, regridding, or interpolation implemented to perform comparisons. In section 3, cold-season precipitation (1 October–30 April) from the high-resolution gridded QPE and QPF products is evaluated for both water years. Discussion and conclusions are presented in sections 4 and 5, respectively.

2. Data and methods

Figure 1 shows the elevation of the western Colorado domain of interest, including SNOTEL locations and highlighting the East River Watershed (ERW) where the SPLASH field campaign took place. Because the eastern portion of the state is relatively flat and does not have any SNOTEL sites, we focus only on the portion of the state west of 105°W. Unless otherwise noted, all datasets are evaluated at their native spatial resolution and highest available temporal resolution, and evaluations are performed over the period from 1 October through 30 April. Point-based datasets are compared with gridded datasets at the nearest gridbox location.

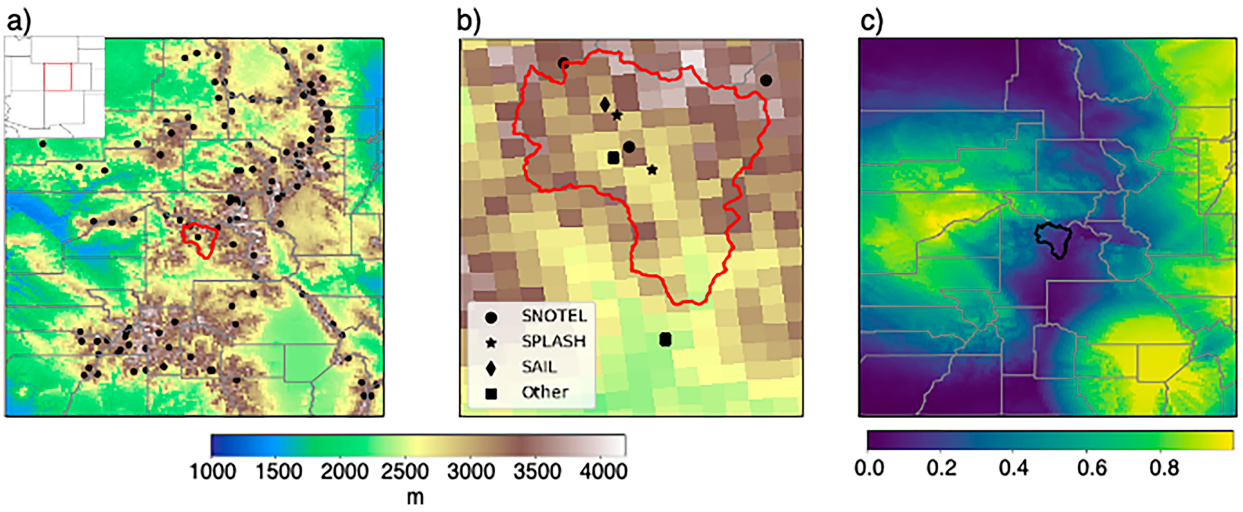

FIG. 1. (a) Study domain, with 3-km elevation from the HRRR model (m) indicated by shading. Gray outlines indicate Colorado counties, black dots the locations of SNOTEL sites, and the ERW is outlined in red. The inset shows the study domain within the southwestern United States. (b) Elevation and reference station locations within the ERW. (c) Seasonal average radar quality index used by MRMS. Values closer to 1 indicate higher-quality radar data.

a. Gridded QPE

1) MRMS MULTISENSOR PRECIPITATION

MRMS, version 12.1.0 (V12.1.0), gridded precipitation products provide estimates of hourly precipitation over the CONUS on a 1-km grid. The multisensor QPE product includes synthetic dual-polarization radar-based precipitation estimates with local gauge correction (Zhang et al. 2016, 2020). A radar quality index (RQI; Zhang et al. 2012) is used to estimate the reliability of the radar observations, and accounts for beam blockage and the height of the radar beam (Fig. 1c; Martinaitis et al. 2020). In areas with good radar coverage, dual-polarization variables (where available) are used to estimate precipitation rate, otherwise reflectivity–precipitation rate relationships are used. In areas where the RQI is unacceptable, surface wet-bulb temperature information from HRRR is used to determine whether the precipitation is likely to be liquid or solid. Liquid precipitation estimates in complex terrain are made by applying a climatological adjustment to gauge observations (i.e., mountain mapper; Daly et al. 1994, 2008; Schaake et al. 2004; Zhang et al. 2016). In regions with a low RQI where the precipitation is determined to be 1) convective, 2) stratiform in nonsteep topography, or 3) snow, hourly forecasts of precipitation from the HRRR are mapped to the 1-km MRMS grid and used in place of observations (Martinaitis et al. 2020).

MRMS V12.1.0 multisensor QPE is provided in two "passes" with varying latency. Pass 1 ingests approximately 50% of the ~20 000 gauges used in the product and is available just 15–20 min after observation time. Pass 2 has a latency of approximately 1 h and ingests 85%–90% of the gauge observations (Martinaitis et al. 2021; Zhang et al. 2023). This study includes both pass-1 and pass-2 multisensor QPE in our evaluation. A radar-only hourly QPE product is also available but is not used in this study because of the consistently low RQI values in the study area resulting from the complex terrain (Fig. 1c).

Previous evaluations of MRMS snowfall estimates have shown large uncertainty in the reflectivity–precipitation rate relationship for snow (Kirstetter et al. 2015), underestimation of snowfall rate because of the increase in the beam height and volume with distance from the radar, along with larger disagreement between MRMS and surface observations with temperatures approaching 0°C (Wen et al. 2017).

2) STAGE IV HOURLY

The NCEP Stage IV hourly QPE (Lin and Mitchell 2005; Nelson et al. 2016; Du 2011) is a multisensor product produced regionally by the 12 CONUS River Forecast Centers (RFCs) and mosaicked at the National Centers for Environmental Prediction (NCEP) onto a grid with variable spacing based on latitude that ranges from 4 to 4.2 km in the Colorado domain shown in Fig. 1a (NWS 2001). Each RFC determines the appropriate datasets to include in their QPE, which may include radar, gauge, and other datasets or interpolation methods. Some manual quality control is also performed at the RFCs, however, due to the latency at which RFCs send their hourly QPE maps to NCEP, the manually quality controlled fields may not be included in the mosaicked CONUS-wide grid. In most of the western United States, including the Colorado River basin RFC (CBRFC), Stage IV is a gauge-based product that includes interpolation and adjustment for terrain. While most of the western Colorado domain is part of the CBRFC, the eastern half of the domain contains portions of the Missouri basin, Arkansas Red basin, and West Gulf RFCs, who may employ radar or multisensor techniques in these areas.

Although Stage IV has been frequently used as a validation dataset for other QPE products and QPFs, it has known

biases, including discontinuities at RFC boundaries due in part to the different techniques used by each RFC in producing its QPE. These discontinuities are particularly evident when examining the dataset over longer accumulation periods (Nelson et al. 2016). Additionally, Nelson et al. (2016) caution against the use of hourly data from the western RFCs (including the CBRFC) due to the potential for bad gauges passing through the first round of automated quality control.

b. *HRRR precipitation forecasts*

The HRRR model is the current operational high-resolution, rapidly updating forecast model in the United States. Forecasts on a 3-km grid are produced hourly out to 18 h, with 48-h forecasts produced every 6 h starting at 0000 UTC (Benjamin et al. 2016). The current version of the HRRR, HRRRv4, is used in this study. HRRRv4 has been shown to have a reduced high bias in the frequency of light precipitation in the winter relative to previous versions 1 and 2 while also having variable performance depending on region, season, time of day, and availability of observations for data assimilation (DA; James et al. 2022). HRRR initial conditions are provided through a 36-member ensemble analysis system (Dowell et al. 2022). This includes a radar-based latent heating adjustment that uses a 3D radar reflectivity mosaic provided by MRMS to produce precipitation in the first few hours of the forecast (Weygandt et al. 2022).

The radar-based latent heating technique is only performed when the reflectivity provided by MRMS exceeds 28 dBZ, so that it is only applied to areas with ongoing convection or moderate to active condensation processes. In regions with no radar coverage or where the observed reflectivity is less than 28 dBZ, it is not included in the assimilation process and a model-computed temperature tendency is used instead. Given the complex terrain of Colorado, large parts of the study area likely do not include radar data in the assimilation (Fig. 1; Weygandt et al. 2022).

In this study, HRRR QPFs of liquid equivalent precipitation are evaluated at various forecast lengths and lead times. Over the course of each water year, we evaluate concatenated HRRR forecasts from 0 to 1 h, 0 to 6 h, and 0 to 24 h from successive model runs initialized every 1, 6, and 24 h, respectively, starting at 0000 UTC. We refer to these as hourly, 6-hourly, and daily or 24-h forecasts (FH1, FH1–6, FH1–24). We also concatenate forecast hours 5–6 from forecasts initialized hourly to evaluate 1-h forecasts with a 6-h lead time (FH5–6). Last, we concatenate hours 7–12 of forecasts initialized every 6 h to evaluate 6-h forecasts with a lead time of 12 h (FH7–12).

c. *Reference datasets*

1) SNOTEL

SNOTEL is a network of observation sites operated by the National Resource Conservation Service (NRCS) to predict spring and summer water supply based on the snowpack (Fleming et al. 2023). There are over 900 SNOTEL sites in the western United States, including Alaska, with most sites located in mountainous terrain sited in wind-protected clearings (Caron and Steenburgh 2020). SWE is measured using a bladder (also called a snow pillow) filled with antifreeze solution. The weight of accumulating snow forces the solution into a manometer column, with the change in height equal to the increase or decrease in water (SWE + any incorporated liquid) on the pillow. All SNOTEL sites include precipitation gauges that are also charged with an antifreeze solution and use the same manometer/pressure transducer principle as the pillow. While the snow pillows are able to measure SWE only, the gauge observations include all phases of precipitation. Raw SNOTEL data from the previous day are transmitted daily just after midnight. Additional network details are given in Serreze et al. (1999).

This study uses daily SNOTEL measurements of SWE and snow-adjusted gauge precipitation (Meyer et al. 2012; Kirk and Schmidlin 2018). Snow pillow measurements are set to zeros at the start of the water year, and the ensuing measurements are cumulative. Snow pillows tell only how much water is resting on the bladder and can rise or fall as a result of accumulation or melting. Snow-adjusted precipitation measurements from the gauges use changes in SWE from the snow pillow to adjust for possible gauge undercatch of snow. Note that SNOTEL measurements may be used by RFCs in the creation of the Stage IV gridded product, and therefore the two datasets are likely not entirely independent.

Limitations to the SNOTEL network include missed events due to light precipitation below the sensitivity of the pillow and gauge [2.54 mm (0.01 in.); Meyer et al. 2012; NRCS 2023] and the representativeness of the point scale measurements to the gridded products. Molotch and Bales (2005) and Meromy et al. (2013) found that SNOTEL observations of depth and SWE were often not representative of the mean of the surrounding area at a variety of grid scales, owing to the characteristics of the area surrounding the SNOTEL site (i.e., canopy density, slope and aspect of the terrain, etc.). These studies showed larger subgrid variability in low accumulation years and during the ablation period, as well as better agreement between the point and grid for smaller grids. Given the relatively high spatial resolution of the gridded data used in this study, the consideration of only the accumulation phase of SWE, and the average to above average nature of the water years studied, representativeness errors are expected to be ameliorated somewhat for the SNOTEL SWE estimates and have little to no impact on the precipitation gauge measurements except with respect to the SWE-based adjustment.

Additional uncertainties in the SWE measurements include snow or ice bridges above the pillow surface, reductions in SWE due to drifting or sublimation, or increases in SWE due to blowing snow or foreign objects on the pillow. With regard to the precipitation gauges, snow sticking to the cylinder that is not recorded until hours or days later leads to timing issues in the observations. For this study, several quality control steps are performed on the SNOTEL data following Serreze et al. (1999), including checks for zero SWE on 1 October, checks for extraneous missing values, and checks for truncated time series. Because most bias in the SNOTEL observations is expected to be a result of undercatch, biases in datasets relative to the SNOTEL should appear erroneously high (Caron and Steenburgh 2020).

2) AIRBORNE SNOW OBSERVATORY

SNOTEL measurements provide estimates of snow water equivalent only at a point and are limited to relatively flat clearings in otherwise forested areas of complex terrain. Measurements from Airborne Snow Observatories, Inc., provide snow depth, SWE, and albedo measurements from aircraft over a number of mountain watersheds in the western United States. Snow depth is calculated using the difference between snow-free and snow-covered surface elevation data. Using the snow-depth measurements, SWE is calculated using a snow density estimate made using the iSnobal model (Marks et al. 1999), which accounts for snow aging, compaction, and deposition of new snow. Uncertainties in SWE estimates from the ASO result from a combination of the uncertainties in snow-depth measurements (<2 cm at 50-m spatial resolution) and modeled density errors (3%–8%). Given these relatively small estimated uncertainties in the measured snow-depth and density calculations, uncertainties in SWE are also believed to be relatively small (Painter et al. 2016).

ASO flights are typically targeted to capture the peak SWE in a given basin, and the instantaneous measurements from each flight are available at 50 m spatial resolution. Not every basin is observed every year, though the ERW was observed in both 2022 and 2023.

To compare SWE from the ASO flights with the gridded QPE data, ASO SWE was regridded using the Python interface to the Earth System Modeling Framework (ESMF) regridding utility (ESMPy) to match the resolution of MRMS, HRRR, and Stage IV.

3) SPLASH/SAIL INSTRUMENTATION

The SPLASH field campaign was designed to improve our understanding of the physical processes that drive the cycle of water in the mountains from precipitation to streamflow. The project included a 2-yr (2021–23) deployment of a variety of instruments at five sites in the ERW of Colorado (Fig. 1b). Sites were selected to cover a range of locations within the East River Valley and included observations of precipitation, radiation, and snowpack. SPLASH occurred concurrently with the U.S. Department of Energy (DOE)–sponsored Surface-Atmosphere Integrated field Laboratory (SAIL; Feldman et al. 2023), which also included precipitation measurements, and the NSF-sponsored Sublimation of Snow (SOS) experiment. For a full description of the SPLASH campaign, see de Boer et al. (2023).

The SPLASH field campaign had two core instrument sites at Brush Creek (BCK) and Kettle Ponds (KPS). BCK was located on the south side of Crested Butte Mountain at 2722 m MSL, while KPS was located in a meadow approximately 75 m above the valley floor at 2861 m MSL. Both sites included both a disdrometer and heated gauge for measuring liquid precipitation, snow-depth stakes for snow-depth measurement, and a variety of other sensors for surface meteorology and atmospheric measurements (de Boer et al. 2023).

The SAIL campaign included a Pluvio-2 Weighing Bucket precipitation gauge situated just north of KPS at an elevation of 2886 m MSL. Both the SPLASH and SAIL campaigns included gap-filling X-band radars in the ERW, with the SPLASH radar situated at the Roaring Judy Fish Hatchery and the SAIL radar sited on the Mount Crested Butte ski slope, approximately 22 km north of the Roaring Judy site. The data from these radars are still being processed as of this writing.

In addition to the SPLASH and SAIL gauges, two additional gauges were included in this study—CBUC2 and GUN01, located in Crested Butte at 2700 m and Gunnison at 2406 m, respectively. CBUC2 is part of the HADS gauge network (Kim et al. 2009), while GUN01 is part of the Colorado Mesonet. While many other gauges belonging to various networks are available in the ERW, only CUBC2 and GUN01 had data of sufficient quality for inclusion in this study.

3. Results

Over most of the domain, WY2023 cool-season (1 October–30 April) precipitation totals are much larger than those from WY2022, particularly in southwestern Colorado (Fig. 2), and the differences are larger in the areas of highest terrain (Fig. 1), which typically received >250 mm more precipitation in WY2023 than WY2022. Differences are generally smaller east of the Continental Divide, and the southeastern part of the domain experienced more cool-season precipitation in WY2022 than in WY2023, as did a few locations along the Continental Divide. Although the two water years featured different precipitation characteristics, for brevity we will show here only the results from WY2022, with corresponding figures for WY2023 shown in the online supplemental material.

Figure 3 shows the total WY2022 cool-season precipitation from each of the gridded products. Overlaid is the ratio of precipitation from each product to snow-adjusted precipitation from SNOTEL gauges. All of the gridded products exhibit low biases at almost all SNOTEL sites. Of the QPE products, Stage IV is the least biased, owing to its use of the SNOTEL data in its production with potential contributions from the manual quality control used in creation of the product. Both MRMS products and HRRR 1-h forecasts consistently underestimate cool-season precipitation by 55%–65%, regardless of RQI in a given location (Fig. 1c). This is much larger than the underestimation found by Cao and Barros (2023) in the Senator Beck basin of southwest Colorado, though their study used a previous version of the HRRR and there are some SNOTEL sites where HRRR 1-h forecasts are in better agreement with the SNOTEL gauges.

The ability to include radar information in the MRMS precipitation estimates and HRRR data assimilation seems to have little impact on the bias in total cool-season precipitation when compared with the SNOTEL snow-adjusted precipitation. The similarity in bias between the MRMS pass 2 and HRRR 1-h forecasts can be explained by the relationship between the two products, specifically the use of HRRR 1-h forecasts in MRMS when RQI is low, and the lack of radar data assimilation in HRRR under the conditions described in section 2b. As the HRRR forecast length increases to 6 and 24 h, or the lead time increases to 6 or 12 h, the underestimation of precipitation decreases, which could potentially be

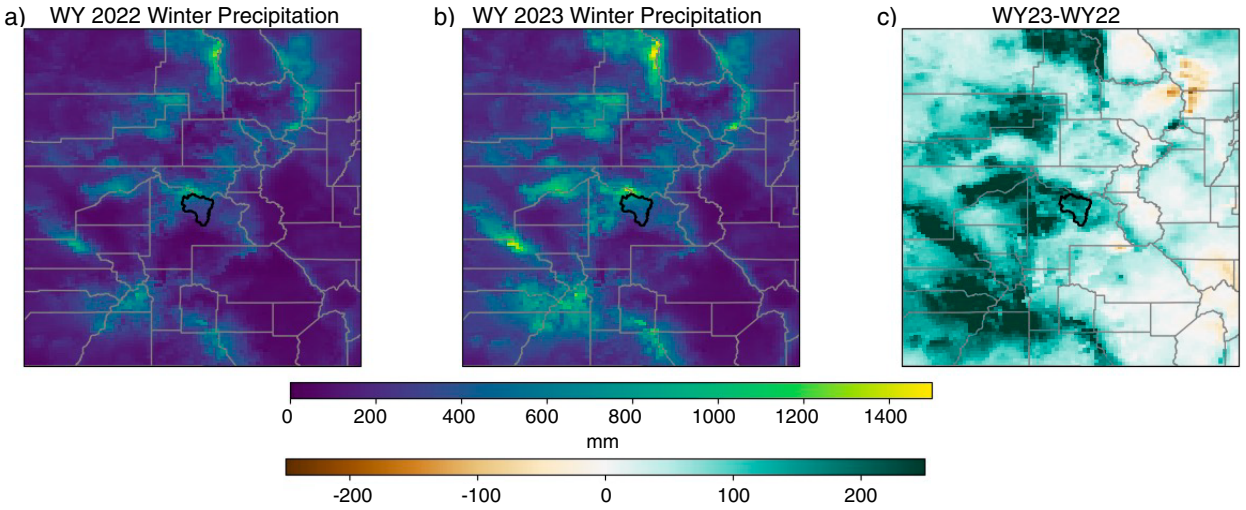

FIG. 2. (a),(b) Cool-season (1 Oct–30 Apr) precipitation totals from Stage IV in the study domain, and (c) WY2023 − WY2022.

related to reduced influence of DA at later forecast hours (Figs. 2e–h). This is further illustrated in Fig. 4, which shows scatterplots of WY2022 cold-season precipitation from the gridded QPE products versus snow-adjusted precipitation from the SNOTEL, with RMSE, fractional bias, and correlation coefficients included for reference. As HRRR forecast length and lead time increases from hourly to 12–24 h, RMSE and bias magnitudes for the domain are reduced by nearly one-half [RMSE reduced from 402.41 to 204.53 for 24-h forecasts (Figs. 4d,f), and bias reduced from −60.37% to −25% to −30% (Figs. 4d,f–h)].

Because the gridded products represent a much larger spatial area than point-based SNOTEL measurements and the aforementioned concerns regarding the representativeness of

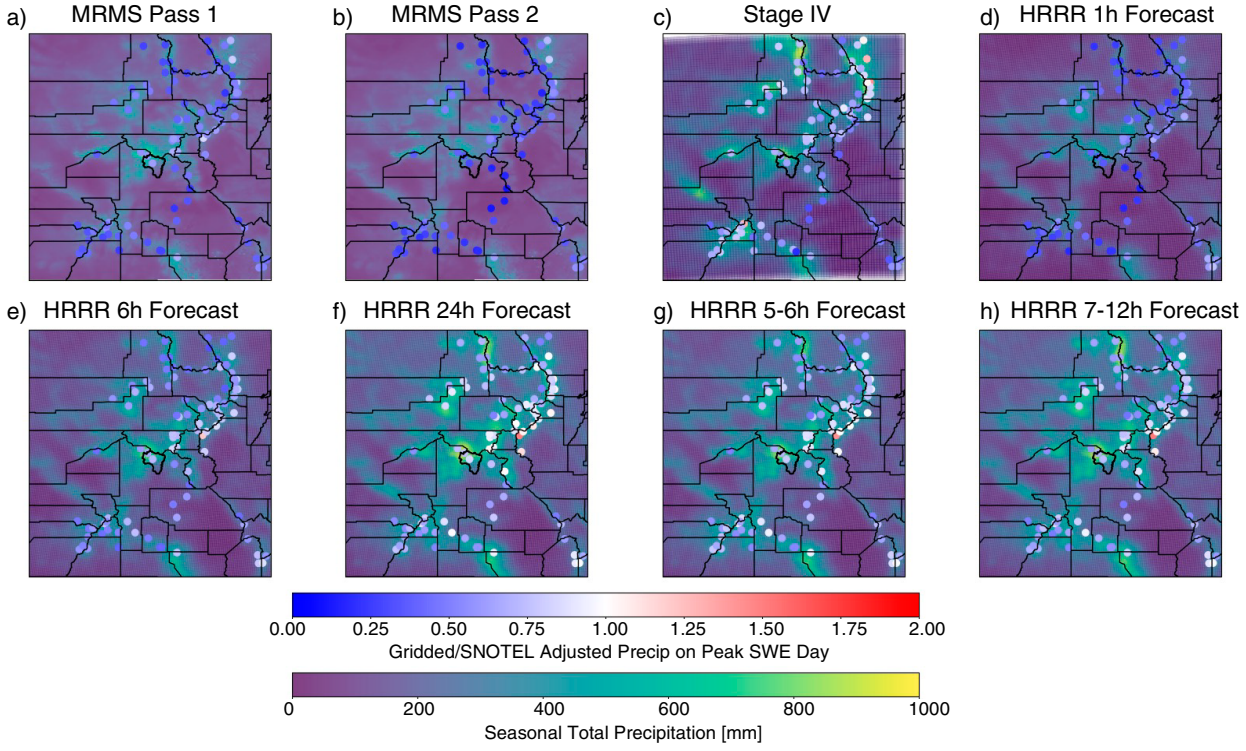

FIG. 3. The 1 Oct–30 Apr total precipitation from the gridded products: (a) MRMS pass 1, (b) MRMS pass 2, (c) Stage IV, (d) HRRR 1-h forecast, (e) HRRR 6-h forecast, (f) HRRR 24-h forecast, (g) HRRR forecast hours 5–6, and (h) HRRR forecast hours 6–12. Overlaid is ratio of seasonal total QPE from the gridded products to seasonal total QPE from the SNOTEL snow-adjusted gauge product.

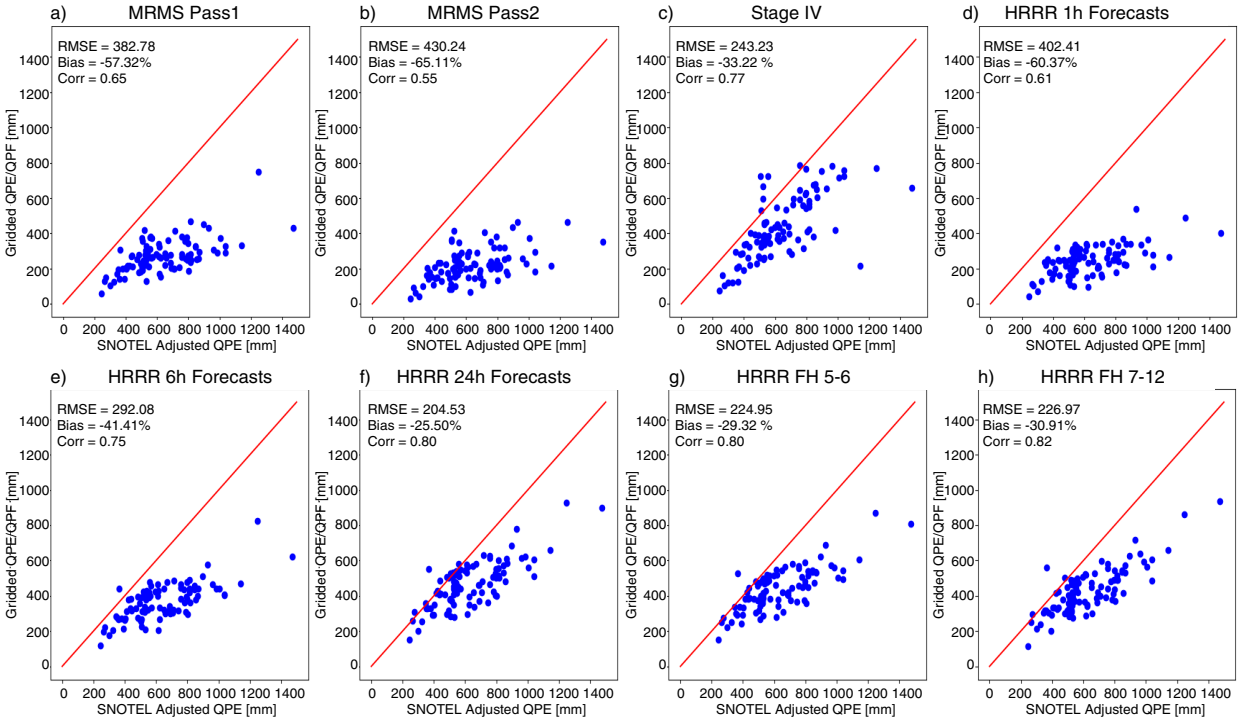

FIG. 4. Scatterplots of WY2022 seasonal total precipitation from the SNOTEL snow-adjusted gauge product in comparison with gridded QPE and QPF products.

the point measurement to the model grid being related to surrounding terrain (Meromy et al. 2013), it is reasonable to question whether the biases shown in Figs. 3 and 4 are related to differences in the mean gridbox elevation and elevation of the SNOTEL sites. Over the Sierra Nevada mountains in California, Lundquist et al. (2015) found that, on average, product grid boxes had higher elevations than coincident snow pillows operated by the California Department of Water Resources (CADWR), though there was a tendency for the snow pillow sites to be located at a higher elevation than the grid cell at lower elevations. Figure 5 shows scatterplots of the SNOTEL site elevation in comparison with the average elevation of the nearest HRRR 3-km grid box, with shading representing the bias of the HRRR relative to SNOTEL at each site. HRRR gridbox elevations are somewhat evenly distributed between being both higher and lower than the SNOTEL sites, with a slight tendency for the HRRR grid boxes to have higher average elevation. On average, HRRR grid boxes are 41 m higher than the SNOTEL sites, with a median difference of HRRR grid boxes being 23.5 m higher. As HRRR forecast length/lead time increases, there does appear to be some tendency for those grid boxes with higher average elevation than the coincident SNOTEL site to reduce the magnitude of the low bias relative to SNOTEL gauge-adjusted precipitation more than those grid boxes with lower average elevation, and all of the grid boxes with slight positive biases at 24-, 5–6-, and 6–12-h forecasts have higher average elevations than the SNOTEL sites with which they are compared.

When combined with the available SNOTEL and other operational gauge data in the ERW, the SPLASH and SAIL observations provide seven data points over the relatively small basin area of 300 km^2 and an opportunity to examine the spatial variability of gridded product performance in more detail. Figure 6 shows time series of cool-season precipitation from available gauges within the basin, while Fig. 7 shows the same for SNOTEL sites. At every site besides Gunnison (Fig. 6c) and Crested Butte (Fig. 7a), MRMS pass 2 and HRRR FH1 are nearly identical, while at Gunnison and Crested Butte, the differences between the two datasets are small. This is likely due to the lack of radar data in the basin (Fig. 1c) and the use of HRRR FH1 in place of observation in MRMS pass 2. Pass 1, with lower latency is unable to rely on the most recent HRRR model run, using instead later forecast hours from previous model runs. These two products (pass 2 and FH1) also largely underestimate seasonal precipitation at every site except at Gunnison, which is in the valley and receives less precipitation overall. The mean fractional bias is 0.38 for pass 2 and 0.42 for FH1 at all gauge and SNOTEL sites but GUN01. HRRR forecasts at longer lead times or forecast lengths consistently result in more seasonal precipitation than those at shorter lead times/forecast lengths. Stage IV also has an average low bias of 0.69 relative to the gauges (excluding GUN01, where it has 2.2 times as much seasonal precipitation as the gauge). This mismatch is somewhat surprising at the CBUC2 and GUN01 sites, which are part of gauge networks that are likely to have been included in the Stage IV estimates.

The ASO completed aerial measurements over the ERW in both 2022 and 2023. ASO flights are generally scheduled near the expected date SWE will reach its peak for a given basin (median date peak SWE in the ERW occurs 9 April at

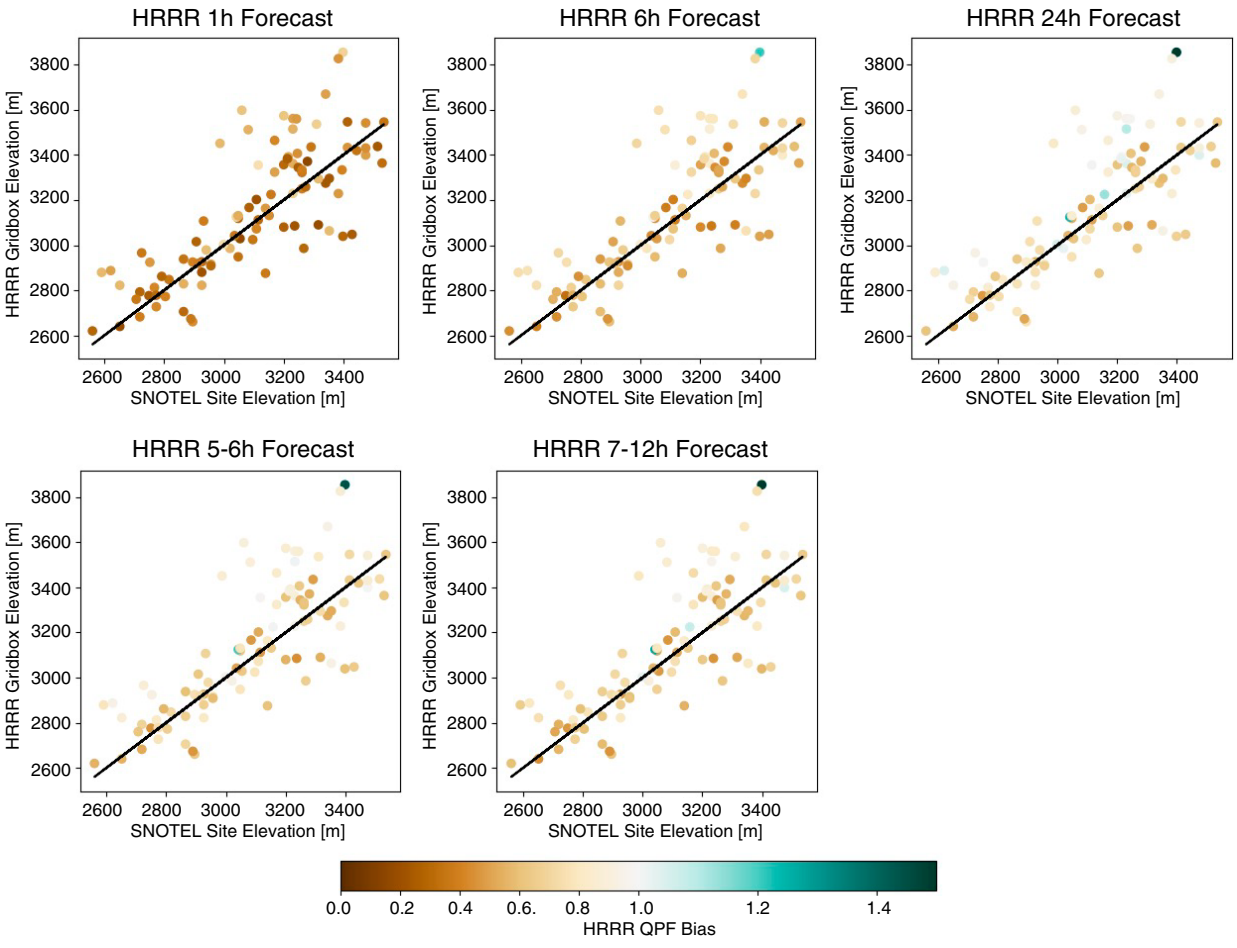

FIG. 5. HRRR 3-km gridbox elevation in comparison with elevation of SNOTEL sites. Color shading indicates HRRR seasonal precipitation bias relative to SNOTEL.

Crested Butte and 19 April at Schofield Pass). In 2022, the ASO flight over the ERW took place on 21 April, whereas in 2023 the flight occurred on 1 April. These data provide a high-resolution, instantaneous view of seasonal SWE accumulated to the date of the flight, but it is important to note that this instantaneous snapshot of SWE on the ground accounts for any melting and losses due to sublimation that have taken place prior to the flight, whereas the accumulated precipitation from the gridded QPE/QPF datasets does not. Therefore, precipitation estimates and forecasts should be equal to or greater than observed SWE. Areas where snow has already melted in the lower elevations of the watershed may show positive biases in accumulated precipitation relative to the ASO data, and we cannot make any conclusions about the biases in the gridded products in these areas. As can be seen in Fig. 7, in 2022 a large melt event occurred beginning on 17 April. Because the ASO flights in 2022 include both accumulation and removal of SWE, we can only assess the level of underestimation from the gridded products.

The left two columns of Fig. 8 show that MRMS and HRRR FH1 mostly underestimate precipitation on the periphery of the basin (i.e., along the ridges defining the basin boundaries). At lower elevations, all precipitation estimates and forecasts were higher than the observed SWE, which does not necessarily indicate that the precipitation was overestimated here due to intermittent snowpacks in some areas, and more rapid melt rates at lower elevations. Due to the large melt event on 17 April, we cannot make any conclusions about the high-biased areas, but we can note the underestimation of MRMS pass 1 and 2 and HRRR FH1 in the high terrain of the basin. Stage IV, which is more gauge-based and incorporates human input has biases near the basin edges that are closer to 1 than the other gridded products that rely more heavily on radar data. Figures 8i–p show the gridded product biases versus the ASO flight from 1 April 2023. In this case, the SNOTEL data do not indicate the presence of melt prior to the flight (Fig. S4 in the online supplemental material), and all products exhibit low biases in most of the basin, with a few products showing some areas near the middle of the basin (i.e., lower elevations by the river) with unbiased or slightly high biased total precipitation. For MRMS pass 1 and pass 2 and HRRR 1-h forecasts, average biases relative to the ASO over the ERW in WY2023 are similar to those relative to the SNOTEL and gauges

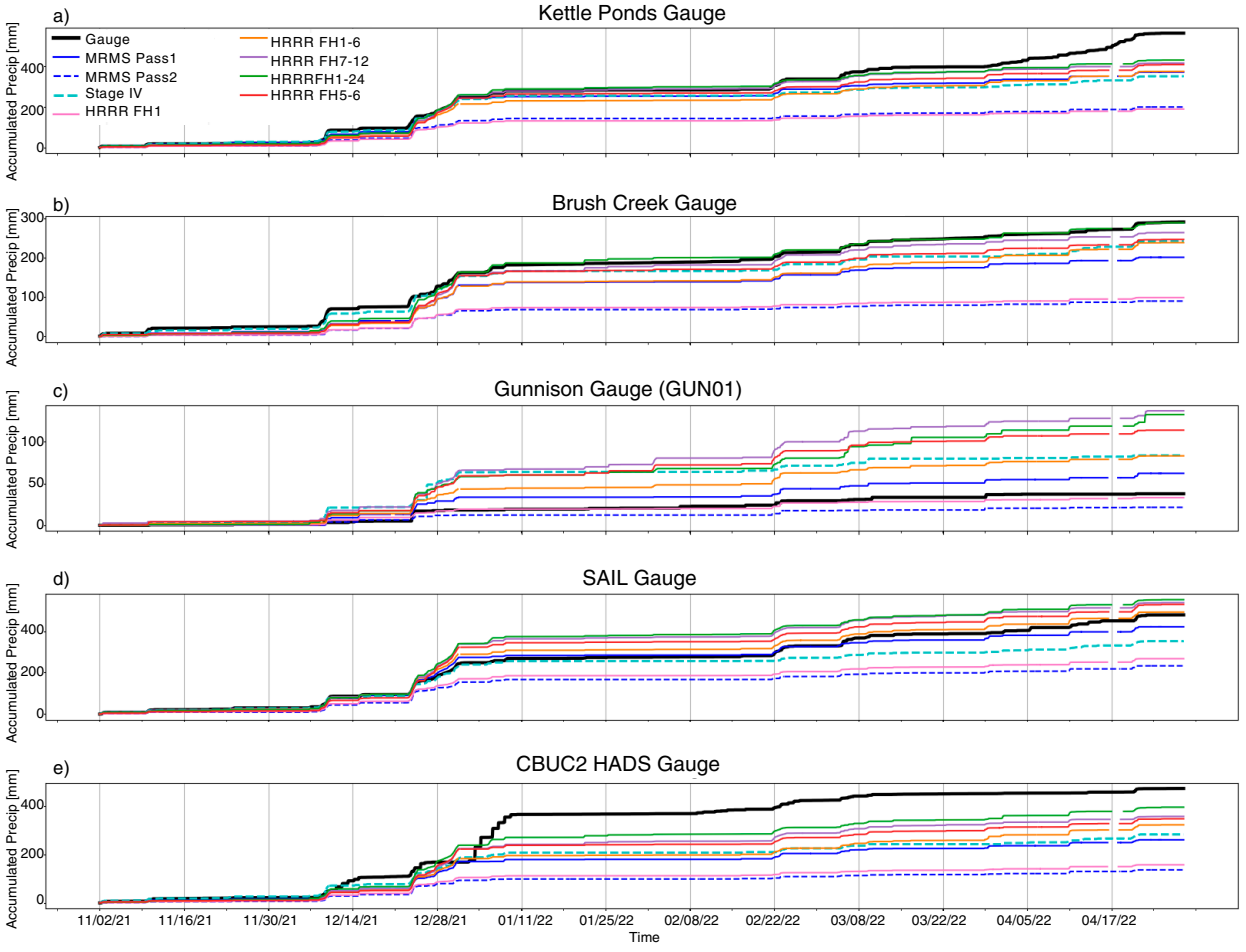

FIG. 6. Cold-season precipitation accumulation time series from (a),(b) SPLASH; (d) SAIL; and (c),(e) other gauge networks in the ERW in comparison with gridded QPE and QPF products.

(0.62, 0.32 and 0.36, respectively), whereas Stage IV and HRRR 24- and 6–12-h forecasts have biases closest to 1 (0.96 for Stage IV and 0.91 for HRRR).

For most of the domain, WY2023 featured more frequent and heavier snowfall events than WY2022 (Fig. S6 in the online supplemental material) leading to the difference in seasonal precipitation between the two years. Biases in gridded QPE and QPF products have been shown to be dependent on storm characteristics. Therefore, we compare both water years 2022 and 2023 for differences in how well the gridded products represent seasonal precipitation in the study domain. Although the two water years were marked by very different precipitation characteristics, the biases from the gridded precipitation products at SNOTEL sites are highly correlated (average $r^2 = 0.77$) from year to year, and, with few exceptions, are consistently less than 1.0 and centered around 0.5 for the MRMS products and early HRRR forecast hours (Fig. 9). This implies that the biases are not related to the characteristics of the precipitation, but more likely to the general strengths and weaknesses of the data that go into each product (i.e., radar quality, DA).

4. Discussion

As mentioned previously, because the SNOTEL SWE and ASO measurements are cumulative and may include periods of melt and sublimation, we can only draw conclusions about the low biases exhibited by the gridded products when compared with them. While there is still the possibility of undercatch error in the gauge measurements, this is ameliorated to some degree by using the snow-adjusted product, but overestimation of the gridded precipitation datasets relative to the SNOTEL gauge data may still be a result of underestimation by the in situ. However, the underestimation of the gridded precipitation versus the reference products can only occur when the precipitation is in fact underestimated by either the QPE or QPF gridded products (Lundquist et al. 2015). Overall, the gridded QPE/QPF products exhibited low biases relative to the in situ measurements, with few locations exhibiting no or high biases, mostly at longer forecast lead times. Both MRMS products and HRRR 1-h forecasts had bias values on the order of 0.4, whereas Stage IV and HRRR 24-, 5–6-, and 6–12-h forecasts all had bias values on the order of 0.7.

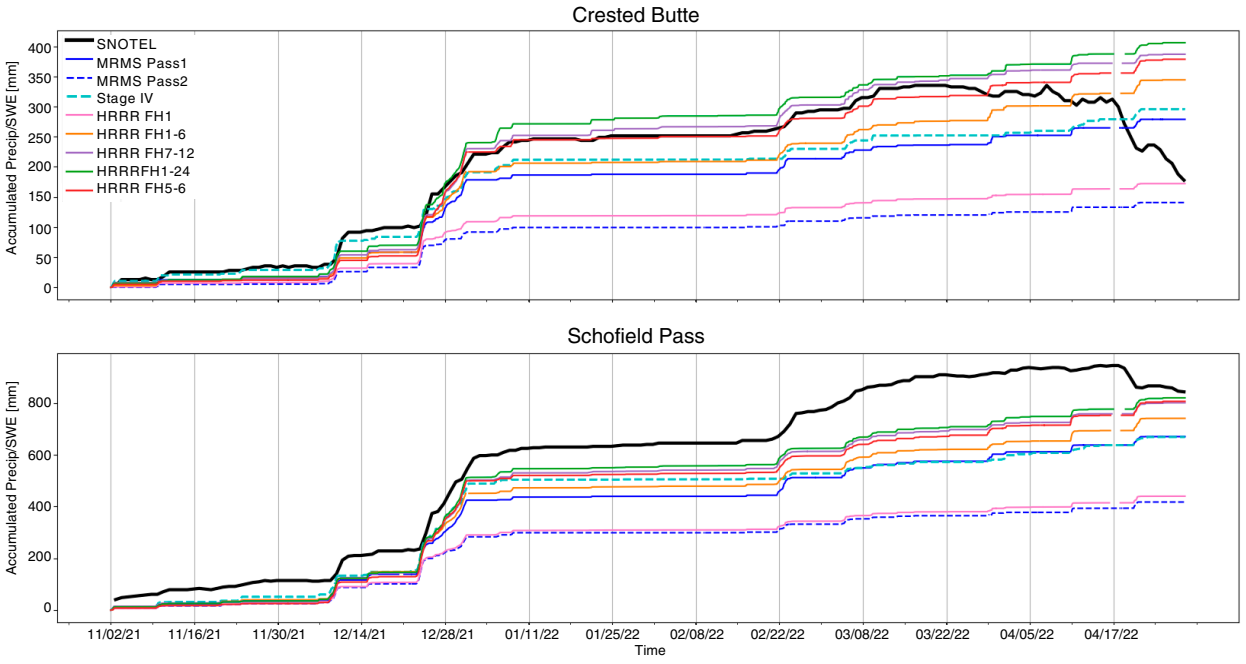

FIG. 7. As in Fig. 6, but in comparison with SNOTEL.

The strong relationship between the MRMS and HRRR products (i.e., that each product may incorporate elements of the other in producing its precipitation estimate/forecast) is evident in the results. Given the lack of reliable radar coverage over much of the study domain, and the tendency for light to moderate snow to exhibit fairly low radar reflectivity (<35 dBZ; Matrosov 1992; NWS 2023) neither the HRRR nor the MRMS is able to make much use of this information. In the absence of radar reflectivity information, HRRR is reliant on modeled temperature tendencies to produce precipitation in the early hours of the forecast period and appears to be producing either no or too little precipitation. MRMS, meanwhile, in the absence of radar observations and presence of frozen precipitation, is using the first hour of the HRRR forecast in its QPE product. This results in these products (i.e., pass-2 and HRRR 1-h forecasts) having very similar biases (~0.35) relative to SNOTEL, gauges, and the ASO. The lower latency of pass 1 means that the HRRR run for that corresponding hour is unavailable, resulting in different bias characteristics for pass 1 and pass 2. The reduction in bias in the later HRRR forecast hours and at longer lead times implies that as the influence of the model DA decreases and the influence of model physics increases, the model is able to more appropriately represent snowfall. The impacts of model DA versus physics on snowfall forecasts is a topic of continued research.

Underestimation of frozen precipitation in complex terrain by operational gridded products has many implications for the hydrologic community. For example, a hydrologic model forced over the cold season using any of the operational gridded products would be more likely to produce an underestimate of streamflow (Palladino et al. 2023). In a region where every drop of water is important, large biases in precipitation and streamflow complicate the job of the water managers.

5. Conclusions

In this study, the idea that operational NWP models may be better than observation-derived products when estimating cold-season precipitation was examined in the mountains of Colorado. While previous research has suggested that model simulations may more accurately represent rangewide seasonal to annual precipitation (and the ensuing modeled hydrologic response), these studies included only well-calibrated longer-term simulations. Our research sought to determine whether the same overall benefits are derived in shorter-term forecasts from the operational HRRR model.

The answer to our first research question, "Is the operational HRRR model as good if not better than observationally based gridded QPE at estimating wintertime precipitation in Colorado?" depends on the answer to our second question, "Are there ranges of forecast hours, lead times, or forecast lengths that provide precipitation forecasts that are better suited for forcing hydrologic models?". This analysis showed that the MRMS QPE and early forecast hours of the HRRR exhibited low biases when compared with SNOTEL, gauges, and airborne SWE measurements. Longer forecast periods of 6 and 24 h, as well as longer forecast lead times of 6 and 12 h produced seasonal precipitation totals with smaller biases. Forecasts with the least influence from the early forecast hours also have the lowest RMSE relative to the SNOTEL. These results imply that the lack of quality radar data has a large impact on high-resolution gridded precipitation information in the mountains of Colorado. They also imply that

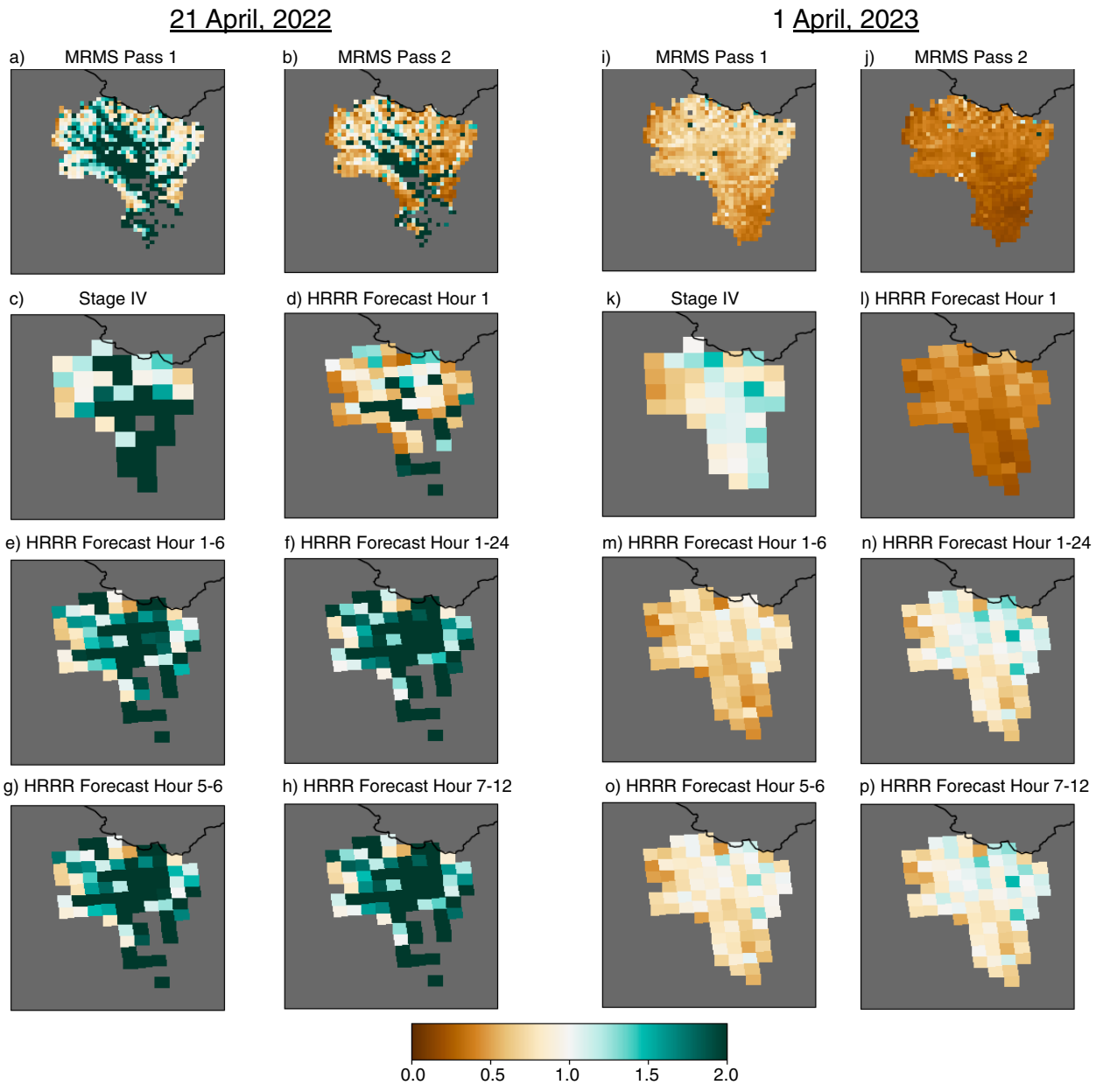

FIG. 8. Ratio of gridded QPE/QPF total precipitation to ASO observed SWE in the ERW on the date of the ASO flight for (a)–(h) 2022 and (i)–(p) 2023.

DA may influence the snowfall forecast bias at the shortest lead times of a given forecast cycle, however, model spinup may also be a factor in the evolution of the low bias with lead time. The impact of DA as compared with other model processes on precipitation forecasts at varying forecast lengths and lead times remains under investigation.

We also set out to investigate how storm characteristics impact the performance of the gridded QPE and QPF. WY2022 and WY2023 were very different years for cold-season precipitation in the upper Colorado River basin. However, the biases in seasonal total precipitation from the gridded QPE and HRRR QPF were consistently low at all SNOTEL sites for both years, indicating that storm characteristics are not the most important factor influencing the performance of these products. Rather, the included data and methods of blending various data types into a multisensor product are the more likely cause of the biases.

As western U.S. populations increase and water supplies remain stagnant or decrease, it is important to understand our abilities and limitations in measuring and predicting how much water is being deposited and stored in mountain snowfall. In addition, both HRRR and MRMS play an important

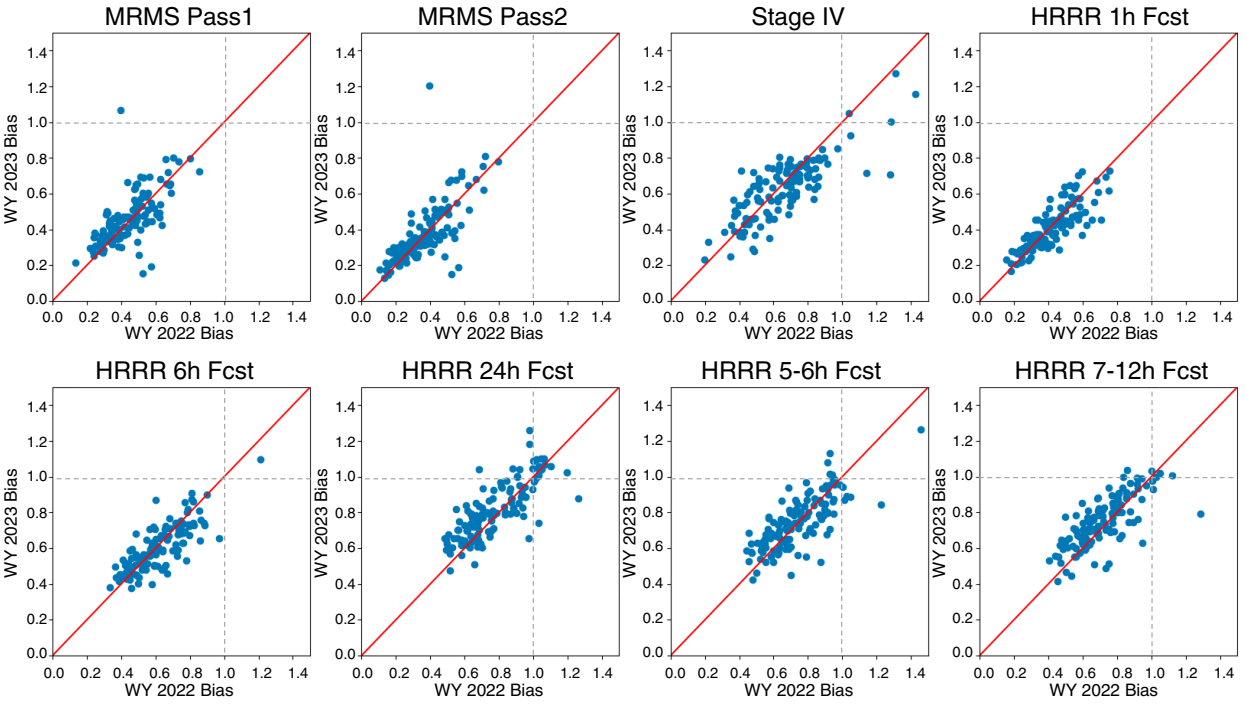

FIG. 9. Bias in 1 Oct–30 Apr total precipitation from the gridded products for WY2023 relative to WY2022. Gray dashed lines indicate neutral bias (1.0) in both years.

role in providing precipitation and atmospheric information to operational water models. Therefore, understanding their ability to represent cold-season precipitation will provide useful context to water managers regarding the potential uncertainties in hydrologic forecasts produced using these data.

Acknowledgments. This work was supported by the NOAA Physical Sciences Laboratory, which led the SPLASH field campaign. Many thanks are given to Tilden Meyers for providing KPS and BCK data; Rachel Palladino, NOAA Hollings Scholar, for foundational analysis on precipitation product and streamflow model performance; and Trevor Alcott for helpful comments on this paper. This work benefited from multiple discussions with the SPLASH research community.

Data availability statement. MRMS Multisensor and Radar Only QPE is available from the MRMS website (https://www.nssl.noaa.gov/projects/mrms/) or AWS (https://registry.opendata.aws/noaa-mrms-pds/). HRRR data can also be obtained from AWS (https://noaa-hrrr-bdp-pds.s3.amazonaws.com/index.html). SNOTEL data were acquired from the NRCS Report Generator V2 (https://wcc.sc.egov.usda.gov/reportGenerator/). SPLASH data are available to the public through Zenodo (https://zenodo.org/communities/splash), with additional data and visualization capabilities available on the SPLASH website (https://psl.noaa.gov/splash/). SAIL gauge data are available via the Atmospheric Radiation Measurement (ARM) program Data Discovery portal (https://adc.arm.gov/discovery/#/). CBUC2 data can be obtained from Mesowest (https://mesowest.utah.edu), and GUN01 can be obtained from CoAgMet (https://coagmet.colostate.edu/station/gun01_main.html). ASO data are available online (https://data.airbornesnowobservatories.com/).

REFERENCES

Anderson, M. L., Z.-Q. Chen, M. L. Kavvas, and A. Feldman, 2002: Coupling HEC-HMS with atmospheric models for prediction of watershed runoff. *J. Hydrol. Eng.*, **7**, 312–318, https://doi.org/10.1061/(ASCE)1084-0699(2002)7:4(312).

Bartsotas, N. S., E. N. Anagnostou, E. I. Nikolopoulos, and G. Kallos, 2018: Investigating satellite precipitation uncertainty over complex terrain. *J. Geophys. Res. Atmos.*, **123**, 5346–5359, https://doi.org/10.1029/2017JD027559.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Bytheway, J. L., M. R. Abel, R. Cifelli, K. Mahoney, and J. M. English, 2022: Demonstrating a probabilistic quantitative precipitation estimate for evaluating precipitation forecasts in complex terrain. *Wea. Forecasting*, **37**, 45–64, https://doi.org/10.1175/WAF-D-21-0074.1.

Cao, Q., T. H. Painter, W. R. Currier, J. D. Lundquist, and D. P. Lettenmaier, 2018: Estimation of precipitation of the OLYMPEX domain during winter 2015/16. *J. Hydrometeor.*, **19**, 143–160, https://doi.org/10.1175/JHM-D-17-0076.1.

Cao, Y., and A. P. Barros, 2023: Topographic controls on active microwave behavior of mountain snowpacks. *Remote Sens. Environ.*, **284**, 113373, https://doi.org/10.1016/j.rse.2022.113373.

Caron, M., and W. J. Steenburgh, 2020: Evaluation of recent NCEP operational model upgrades for cool-season precipitation forecasting over the western conterminous United States.

Wea. Forecasting, **35**, 857–877, https://doi.org/10.1175/WAF-D-19-0182.1.

Currier, W. R., T. Thorson, and J. D. Lundquist, 2017: Independent evaluation of frozen precipitation from WRF and PRISM in the Olympic Mountains. *J. Hydrometeor.*, **18**, 2681–2703, https://doi.org/10.1175/JHM-D-17-0026.1.

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

——, M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, https://doi.org/10.1002/joc.1688.

de Boer, G., and Coauthors, 2023: Supporting advancement in weather and water prediction in the upper Colorado River basin: The SPLASH campaign. *Bull. Amer. Meteor. Soc.*, **104**, E1853–E1874, https://doi.org/10.1175/BAMS-D-22-0147.1.

Derin, Y., and Coauthors, 2016: Multiregional satellite precipitation products evaluation over complex terrain. *J. Hydrometeor.*, **17**, 1817–1836, https://doi.org/10.1175/JHM-D-15-0197.1.

Dowell, D. C., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updated 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.

Du, J., 2011: NCEP/EMC 4KM gridded data (GRIB) stage IV data. Version 1.0. UCAR/NCAR–Earth Observing Laboratory, accessed 30 May 2023, https://doi.org/10.5065/D6PG1QDD.

Ebert, E. E., J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, **88**, 47–64, https://doi.org/10.1175/BAMS-88-1-47.

English, J. M., D. D. Turner, T. I. Alcott, W. R. Moninger, J. L. Bytheway, R. Cifelli, and M. Marquis, 2021: Evaluating operational and experimental HRRR model forecasts of atmospheric river events in California. *Wea. Forecasting*, **36**, 1925–1944, https://doi.org/10.1175/WAF-D-21-0081.1.

Feldman, D. R., and Coauthors, 2023: The Surface Atmosphere Integrated Field Laboratory (SAIL) campaign. *Bull. Amer. Meteor. Soc.*, **104**, E2192–E2222, https://doi.org/10.1175/BAMS-D-22-0049.1.

Fleming, S. W., L. Zukiewicz, M. L. Strobel, H. Hofman, and A. G. Goodbody, 2023: SNOTEL, the soil climate analysis network, and water supply forecasting at the natural resources conservation service: Past, present and future. *J. Amer. Water Resour. Assoc.*, **59**, 585–599, https://doi.org/10.1111/1752-1688.13104.

Förster, K., G. Meon, T. Marke, and U. Strasser, 2014: Effect of meteorological forcing and snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany). *Hydrol. Earth Sys. Sci.*, **18**, 4703–4720, https://doi.org/10.5194/hess-18-4703-2014.

——, J. Garvelmann, G. Meißl, and U. Strasser, 2018: Modeling forest snow processes with a new version of WaSiM. *Hydrol. Sci. J.*, **63**, 1540–1557, https://doi.org/10.1080/02626667.2018.1518626.

Ge, S., J. Silverstein, J. Eklund, P. Limerick, and D. Stewart, 2023: Fixing the flawed Colorado River compact. *Eos*, **104**, https://doi.org/10.1029/2023EO230232.

Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation intercomparison. Instruments and Observing Methods Rep. 67, WMO/TD-872, 212 pp., https://library.wmo.int/viewer/28336/download?file=wmo-td_872.pdf&type=pdf&navigator=1.

Gouttevin, I., V. Vionnet, Y. Seity, A. Boone, M. Lafaysse, Y. Deliot, and H. Merzisen, 2023: To the origin of a winter-time screen-level temperature bias at high altitude in a kilometric NWP model. *J. Hydrometeor.*, **24**, 53–71, https://doi.org/10.1175/JHM-D-21-0200.1.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, M. Pl Clark, J. Dudhia, and G. Thompson, 2012: A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *J. Climate*, **25**, 262–281, https://doi.org/10.1175/2011JCLI4109.1.

Henn, B., A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, 2018: An assessment of differences in gridded precipitation datasets in complex terrain. *J. Hydrol.*, **556**, 1205–1219, https://doi.org/10.1016/j.jhydrol.2017.03.008.

Ikeda, K., and Coauthors, 2010: Simulation of seasonal snowfall over Colorado. *Atmos. Res.*, **97**, 462–477, https://doi.org/10.1016/j.atmosres.2010.04.010.

James, E. P., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updated convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, https://doi.org/10.1175/WAF-D-21-0130.1.

Jennings, K. S., M. M. Arienzo, M. Collins, B. J. Hatchett, A. W. Nolin, and G. Aggett, 2023: Crowdsourced data highlight precipitation phase partitioning variability in rain-snow transition zone. *Earth Space Sci.*, **10**, e2022EA002714, https://doi.org/10.1029/2022EA002714.

Kim, D., B. Nelson, and D.-J. Seo, 2009: Characteristics of reprocessed Hydrometeorological Automated Data System (HADS) hourly precipitation data. *Wea. Forecasting*, **24**, 1287–1296, https://doi.org/10.1175/2009WAF2222227.1.

Kirk, J. P., and T. W. Schmidlin, 2018: Large precipitation event variability among headwater SNOTEL sites and impacts on streamflow in the upper Colorado River basin. *Phys. Geogr.*, **39**, 445–470, https://doi.org/10.1080/02723646.2017.1345579.

Kirstetter, P.-E., J. J. Gourley, Y. Hong, J. Zhang, S. Moasamigoodarzi, C. Langston, and A. Arthur, 2015: Probabilistic precipitation estimates with ground-based radar networks. *Water Resour. Res.*, **51**, 1422–1442, https://doi.org/10.1002/2014WR015672.

Kochendorfer, J., and Coauthors, 2022: How well are we measuring snow post-SPICE? *Bull. Amer. Meteor. Soc.*, **103**, E370–E388, https://doi.org/10.1175/BAMS-D-20-0228.1.

Kunstmann, H., and C. Stadler, 2005: High resolution distributed atmospheric-hydrological modeling for Alpine catchments. *J. Hydrol.*, **314**, 105–124, https://doi.org/10.1016/j.jhydrol.2005.03.033.

Li, D., M. L. Wrzesien, M. Durand, J. Adam, and D. P. Lettenmaier, 2017: How much runoff originates as snow in the western United States, and how will that change in the future? *Geophys. Res. Lett.*, **44**, 6163–6172, https://doi.org/10.1002/2017GL073551.

——, D. P. Lettenmaier, S. A. Margulis, and K. Andreadis, 2019: The value of accurate high-resolution and spatially continuous snow information to streamflow. *J. Hydrometeor.*, **20**, 731–749, https://doi.org/10.1175/JHM-D-18-0210.1.

Li, J., H. Chen, P. Li, and X. Jiang, 2023: Evaluating the precipitation biases over the western periphery of the Sichuan basin

by the ECMWF operational forecast model. *Wea. Forecasting*, **38**, 1481–1496, https://doi.org/10.1175/WAF-D-22-0218.1.

Lin, Y., and K. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analysis: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2, https://ams.confex.com/ams/pdfpapers/83847.pdf.

Lundquist, J., M. R. Abel, E. Gutmann, and S. Kapnick, 2019: Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. *Bull. Amer. Meteor. Soc.*, **100**, 2473–2490, https://doi.org/10.1175/BAMS-D-19-0001.1.

——, ——, B. Henn, E. D. Gutmann, B. Livneh, J. Dozier, and P. Neiman, 2015: High-elevation precipitation patterns: Using snow measurements to assess daily gridded datasets across the Sierra Nevada, California. *J. Hydrometeor.*, **16**, 1773–1792, https://doi.org/10.1175/JHM-D-15-0019.1.

Maddox, R. A., J. Zhang, J. J. Gourley, and K. W. Howard, 2002: Weather radar coverage over the contiguous United States. *Wea. Forecasting*, **17**, 927–934, https://doi.org/10.1175/1520-0434(2002)017<0927:WRCOTC>2.0.CO;2.

Marks, D., J. Domingo, D. Susong, T. Link, and D. Garen, 1999: A spatially distributed energy balance snowmelt model for application in mountain basins. *Hydrol. Processes*, **13**, 1935–1959, https://doi.org/10.1002/(SICI)1099-1085(199909)13:12/13<1935::AID-HYP868>3.0.CO;2-C.

Martinaitis, S. M., S. B. Cocks, Y. Qi, B. T. Kaney, J. Zhang, and K. Howard, 2015: Understanding winter precipitation impacts on automated gauge observations within a real-time system. *J. Hydrometeor.*, **16**, 2345–2363, https://doi.org/10.1175/JHM-D-15-0020.1.

——, and Coauthors, 2020: A physically based multisensor quantitative precipitation estimation approach for gap-filling radar coverage. *J. Hydrometeor.*, **21**, 1485–1511, https://doi.org/10.1175/JHM-D-19-0264.1.

——, S. B. Cocks, M. J. Simpson, A. P. Osborne, S. S. Harkema, H. M. Grams, J. Zhang, and K. W. Howard, 2021: Advancements and characteristics of gauge ingest and quality control within the Multi-Radar Multi-Sensor system. *J. Hydrometeor.*, **22**, 2455–2474, https://doi.org/10.1175/JHM-D-20-0234.1.

Matrosov, S. Y., 1992: Radar reflectivity in snowfall. *IEEE Trans. Geosci. Remote Sens.*, **30**, 454–461, https://doi.org/10.1109/36.142923.

Meromy, L., N. P. Molotch, T. E. Link, S. R. Fassnacht, and R. Rice, 2013: Subgrid variability of snow water equivalent at operational snow stations in the western USA. *Hydrol. Processes*, **27**, 2383–2400, https://doi.org/10.1002/hyp.9355.

Meyer, J. D. D., J. Jin, and S.-Y. Wang, 2012: Systematic patterns of the inconsistency between snow water equivalent and accumulated precipitation as reported by the snowpack telemetry network. *J. Hydrometeor.*, **13**, 1970–1976, https://doi.org/10.1175/JHM-D-12-066.1.

Molotch, N. P., and R. C. Bales, 2005: Scaling snow observations from the point to the grid element: Implications for observation network design. *Water Resour. Res.*, **41**, W11421, https://doi.org/10.1029/2005WR004229.

Nelson, B. R., O. P. Prat, D.-J. Seo, and E. Habib, 2016: Assessment and implications of NCEP stage IV quantitative precipitation estimates for product intercomparisons. *Wea. Forecasting*, **31**, 371–394, https://doi.org/10.1175/WAF-D-14-00112.1.

NRCS, 2023: SNOTEL sensors. Accessed 21 November 2023, https://www.nrcs.usda.gov/wps/portal/wcc/home/dataAccessHelp/faqs/snotelSensors.

NWS, 2001: II.1—HRAPGRID Hydrologic Rainfall Analysis Project (HRAP) grid system. NOAA, 2 pp., https://www.weather.gov/media/owp/oh/hrl/docs/21hrapgrid.pdf.

——, 2023: WSR-88D dual-polarization radar decision aid. NOAA, 10 pp., https://training.weather.gov/wdtd/courses/dualpol/Outreach/DualPol-Flipchart.pdf.

Painter, T. H., and Coauthors, 2016: The airborne snow observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sens. Environ.*, **184**, 139–152, https://doi.org/10.1016/j.rse.2016.06.018.

Palladino, R. K., W. R. Currier, and M. R. Hughes, 2023: Evaluation of the National Water Model in the East River Watershed. *37th Conf. on Hydrology*, Denver, CO, Amer. Meteor. Soc., J6B.3, https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/416348.

Raleigh, M. S., J. D. Lundquist, and M. P. Clark, 2015: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrol. Earth Syst. Sci.*, **19**, 3153–3179, https://doi.org/10.5194/hess-19-3153-2015.

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.

Schaake, J., A. Henkel, and S. Cong, 2004: Application of PRISM climatologies for hydrologic modeling and forecasting in the western U.S. *18th Conf. on Hydrology*, Seattle, WA, Amer. Meteor. Soc., 5.3, https://ams.confex.com/ams/84Annual/techprogram/paper_72159.htm.

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

Smiatek, G., H. Kunstmann, and J. Werhahn, 2012: Implementation and performance analysis of a high resolution coupled numerical weather and river runoff prediction model system for an Alpine catchment. *Environ. Modell. Software*, **38**, 231–243, https://doi.org/10.1016/j.envsoft.2012.06.001.

Sun, Q., C. Maio, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A review of global precipitation datasets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, https://doi.org/10.1002/2017RG000574.

Tang, G., and Coauthors, 2023: The impact of meteorological forcing uncertainty on hydrological modeling: A global analysis of cryosphere basins. *Water Resour. Res.*, **59**, e2022WR033767, https://doi.org/10.1029/2022WR033767.

Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, **37**, L24407, https://doi.org/10.1029/2010GL046008.

Villarini, G., and W. F. Krajewski, 2009: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surv. Geophys.*, **31**, 107–129, https://doi.org/10.1007/s10712-009-9079-x.

Wayand, N. E., A. F. Hamlet, M. R. Abel, S. I. Feld, and J. D. Lundquist, 2013: Intercomparison of meteorological forcing data from empirical and mesoscale model sources in the North Fork American River basin in northern Sierra Nevada, California. *J. Hydrometeor.*, **14**, 677–699, https://doi.org/10.1175/JHM-D-12-0102.1.

Wen, Y., P. Kirstetter, J. J. Gourley, Y. Hong, A. Behrangi, and Z. Flamig, 2017: Evaluation of MRMS snowfall products

over the western United States. *J. Hydrometeor.*, **18**, 1707–1713, https://doi.org/10.1175/JHM-D-16-0266.1.

Westrick, K. J., C. F. Mass, and B. A. Colle, 1999: The limitations of the WSR-88D radar network for quantitative precipitation measurement over the coastal western United States. *Bull. Amer. Meteor. Soc.*, **80**, 2289–2298, https://doi.org/10.1175/1520-0477(1999)080<2289:TLOTWR>2.0.CO;2.

——, P. Storck, and C. F. Mass, 2002: Description and evaluation of a hydrometeorological forecast system for mountainous watersheds. *Wea. Forecasting*, **17**, 250–262, https://doi.org/10.1175/1520-0434(2002)017<0250:DAEOAH>2.0.CO;2.

Weygandt, S. S., S. G. Benjamin, M. Hu, C. R. Alexander, T. G. Smirnova, and E. P. James, 2022: Radar reflectivity-based model initialization using specified latent heating (Radar-LHI) within a diabatic digital filter or pre-forecast integration. *Wea. Forecasting*, **37**, 1419–1434, https://doi.org/10.1175/WAF-D-21-0142.1.

Zhang, J., Y. Qi, K. Howard, C. Langston, and B. Kaney, 2012: Radar quality index (RQI)—A combined measure for beam blockage and VPR effects in a national network. *IAHS Publ.*, **351**, 388–393.

——, and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

——, L. Tang, S. Cocks, P. Zhang, A. Ryzhkov, K. Howard, C. Langston, and B. Kaney, 2020: A dual-polarization radar synthetic QPE for operations. *J. Hydrometeor.*, **21**, 2507–2521, https://doi.org/10.1175/JHM-D-19-0194.1.

——, A. P. Osborne, and S. M. Martinaitis, 2023: MRMS QPE analysis of record. *37th Conf. on Hydrology*, Denver, Colorado, Amer. Meteor. Soc., 13A.2, https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/412812.