High-Elevation Precipitation Patterns: Using Snow Measurements to Assess Daily Gridded Datasets across the Sierra Nevada, California*

JESSICA D. LUNDQUIST,[+] MIMI HUGHES,[#,@] BRIAN HENN,[+] ETHAN D. GUTMANN,[&] BEN LIVNEH,[#,@] JEFF DOZIER,[**] AND PAUL NEIMAN[@]

[+] *Civil and Environmental Engineering, University of Washington, Seattle, Washington*
[#] *Cooperative Institute for Research in Environmental Science, University of Colorado Boulder, Boulder, Colorado*
[@] *NOAA/Earth System Research Laboratory/Physical Sciences Division, Boulder, Colorado*
[&] *National Center for Atmospheric Research, Boulder, Colorado*
[**] *Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, California*

(Manuscript received 23 January 2015, in final form 17 April 2015)

ABSTRACT

Gridded spatiotemporal maps of precipitation are essential for hydrometeorological and ecological analyses. In the United States, most of these datasets are developed using the Cooperative Observer (COOP) network of ground-based precipitation measurements, interpolation, and the Parameter–Elevation Regressions on Independent Slopes Model (PRISM) to map these measurements to places where data are not available. Here, we evaluate two daily datasets gridded at $^1/_{16}°$ resolution against independent daily observations from over 100 snow pillows in California's Sierra Nevada from 1990 to 2010. Over the entire period, the gridded datasets performed reasonably well, with median total water-year errors generally falling within ±10%. However, errors in individual storm events sometimes exceeded 50% for the median difference across all stations, and in many cases, the same underpredicted storms appear in both datasets. Synoptic analysis reveals that these underpredicted storms coincide with 700-hPa winds from the west or northwest, which are associated with post-cold-frontal flow and disproportionately small precipitation rates in low-elevation valley locations, where the COOP stations are primarily located. This atmospheric circulation leads to a stronger than normal valley-to-mountain precipitation gradient and underestimation of actual mountain precipitation. Because of the small average number of storms (<10) reaching California each year, these individual storm misses can lead to large biases (~20%) in total water-year precipitation and thereby significantly affect estimates of statewide water resources.

1. Introduction

In the western United States, mountain snowpacks supply water for multiple purposes. Through forecasting and reservoir operations, the Sierra Nevada snowpack provides more than half the annual water supply and about 15% of the electrical power supply (Rheinheimer et al. 2012) for California's population of over 38 million (U.S. Census Bureau 2014). Estimates and predictions of this water supply on daily (flood and hydropower forecasting), seasonal (supply forecasting), and decadal (long-range planning) time scales typically depend on an ability to map point measurements of precipitation to entire watersheds. Spatial grids of precipitation also function as benchmarks to evaluate and downscale atmospheric model performance as well as drivers for hydrologic models, which are then calibrated and used to forecast streamflow.

Many gridded precipitation products have been developed, but their methodologies, at least as applied within the continental U.S. (CONUS), are remarkably similar (Table 1). Each starts with daily gauge observations as reported by the National Weather Service (NWS) and National Climatic Data Center (NCDC) Cooperative Observer (COOP) network, where volunteers report 24-h accumulated precipitation (Daly et al.

* Joint Institute for the Study of the Atmosphere and Ocean Contribution Number 2405.

Corresponding author address: Jessica D. Lundquist, Civil and Environmental Engineering, University of Washington, Box 352700, Seattle, WA 98195.
E-mail: jdlund@u.washington.edu

DOI: 10.1175/JHM-D-15-0019.1

TABLE 1. Gridded precipitation products in the CONUS.

Product (citation)	Spatial and temporal resolution (time period covered)	Source of daily precipitation data	Topographic gridding baseline for precipitation	Temperature gridding notes	Reported error statistics	Times cited*
PRISM (Daly et al. 1994, 2008)	4-km and 800-m data available daily (from Jan 1981 to present), monthly (from Jan 1895 to present), and as 30-yr climate normals (1961–90, 1971–2000, and 1981–2010)	Multiple data networks (see Daly et al. 2008; Daly 2013)	Empirical relationships to account for the influence of elevation, rain shadows, and coastal proximity	Empirical relationships to account for the influence of elevation, cold air pools (topographic ridge vs depression), and coastal proximity	Average cross-validation biases of <1 mm and mean annual error of 11% (for precipitation, using comparison against deliberately withheld stations), but mapped values for high-altitude terrain in the western United States often have >30% absolute error in annual precipitation.	GS: 2005; WS: 1177 (40% ecol., 34% hydr., and 34% atm.)
Daymet (Thornton et al. 1997)	1-km data available daily (1980–2013; with continuing updates each calendar year)	COOP stations and NRCS SNOTEL observations	Linear regression based on elevation; patterns are assumed constant in time	Linear lapse based on regression between station elevations and daily temperatures	Mean annual errors of 20%, with mean positive biases of 3%–7%; ±50% difference compared to PRISM in western United States (see Daly et al. 2008).	GS: 824; WS: 498 (62% ecol., 10% hydr., and 20% atm.)
M02 (Maurer et al. 2002)	1/8° (~12 km) data available daily (1949–2000; now available through 2010)	COOP stations with >20 years of data	PRISM 1961–90 monthly climatology	$-6.5°C\,km^{-1}$ constant lapse rate	For VIC simulations of California streamflow, RMSE = 68% and 46% and bias = 30% and −0.4% for monthly flow in the Tuolumne and Sacramento Rivers, respectively.	GS: 694; WS: 439 (23% ecol., 39% hydr., and 54% atm.)
HL05 (Hamlet and Lettenmaier 2005)	1/8° (~12 km) data available daily (1915–2003)	COOP stations with >5 years of data	PRISM 1961–90 monthly climatology; running mean is adjusted to match USHCN and HCCD** data	$-6.1°C\,km^{-1}$ constant lapse rate	For VIC streamflow simulations of the Sacramento and San Joaquin basins, there is "quite good for the most part" agreement and stationary error characteristics throughout the time series.	GS: 174; WS: 111 (24% ecol., 32% hydr., and 56% atm.)
NLDAS-2 (Cosgrove et al. 2003; Xia et al. 2012a,b)	1/8° (~12 km) data available hourly (from 1979 to present)	CPC daily gauge analysis (Higgins et al. 2000) disaggregated to hourly with NCEP stage II radar data	PRISM 1961–90 monthly climatology	$-6.5°C\,km^{-1}$ constant lapse rate; threshold air temperature of 0°C to separate rainfall from snowfall	Annual runoff biases for basins in the western United States using four different streamflow models are generally within ±75%.	GS: 1760 for search term "NLDAS" and 250 for Cosgrove et al. (2003); WS: 180 (16% ecol., 28% hydr., and 57% atm.)
H10 (Hamlet et al. 2010; Salathé et al. 2014)	1/16° (~6 km) data available daily (1916–2010; these dates are for California but the exact time period varies by region)	COOP stations with >5 years of data	PRISM 1971–2000 monthly climatology; running mean is adjusted to match USHCN and HCCD data	PRISM 1971–2000 monthly temperature climatology	—	GS: 23 for H10 and 0 for Salathé et al. (2014); WS: 0

TABLE 1. (*Continued*)

Product (citation)	Spatial and temporal resolution (time period covered)	Source of daily precipitation data	Topographic gridding baseline for precipitation	Temperature gridding notes	Reported error statistics	Times cited*
L13 (Livneh et al. 2013)	1/16° (~6 km) data available daily (1915–2011)	COOP stations with >20 years of data	PRISM 1961–90 monthly climatology	−6.5°C km^{-1} constant lapse rate	Compares well with M02, with some overprediction of modeled San Joaquin streamflow (numbers not reported)	GS: 14; WS: 4
NCEP stage IV (Lin and Mitchell 2005)	4-km data available	Gauge data modified by radar observations and with quality control by River Forecast Centers	PRISM 1961–90 monthly climatology, as employed in MM (Schaake et al. 2004)	—	—	GS: 17 000 for NCEP stage IV precipitation and 202 for Lin and Mitchell (2005); WS: Lin and Mitchell (2005) is not in Web of Science

* GS = Google Scholar search, accessed 29 Oct 2014. WS = Web of Science, accessed 27 Aug 2014. Citations from Web of Science are further described as percentage by discipline, where "ecol." represents the categories of environmental science or ecology and forestry; "hydr." represents the categories of water resources and engineering; "atm." represents atmospheric science; and other, less common fields are not reported.

** HCCD = Historical Canadian Climate Database.

2007). Some augment these observations with those from other networks, such as the National Resources Conservation Service (NRCS) SNOTEL observations (used in Daymet; Thornton et al. 1997) or radar-based observations and river forecaster modifications (NCEP stage IV analysis; Baldwin and Mitchell 1996; Lin and Mitchell 2005), where river forecasters in the western United States use the Mountain Mapper (MM; Schaake et al. 2004) to extend gauge observations across complex terrain (see www.cnrfc.noaa.gov/products/rfcprismuse.pdf). Some gridded products (Hamlet and Lettenmaier 2005, hereafter HL05; Hamlet et al. 2010, hereafter H10), designed for use in climate simulations, adjust the data for temporal continuity, forcing gridded 3-month running means calculated with all gauges to match running means from the longer record but sparser U.S. Historical Climatology Network (USHCN; Menne et al. 2009; Karl et al. 1990). Other products (Maurer et al. 2002, hereafter M02; Livneh et al. 2013, hereafter L13) attempt to ensure temporal stability by excluding stations with fewer than 20 years of total record. All the products use interpolation methods to map the station data to a grid (see Table 1), and all, except Daymet, use the Parameter–Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994, 2008) 30-yr monthly climate normals to adjust the spatial grid for effects of elevation and topographic orientation.

Despite their widespread use, few evaluations of the performance of gridded precipitation products exist. Generally, all available measures of precipitation are included in the product, leaving no stations for evaluation. A comparison of PRISM data to a weather model and to observations at a recently installed SNOTEL station, which was not included in PRISM, showed that gridded products could be off by a factor of 2 in one Colorado mountain range (Gutmann et al. 2012). Other evaluations have pointed out issues with precipitation undercatch, which is particularly problematic with snow (Goodison et al. 1998; M02; Yang et al. 2005; Rasmussen et al. 2012), wet-day statistics (Gutmann et al. 2014), and extreme event representations (Gervais et al. 2014) in gridded products. Alternatively, comparing streamflow simulations with observations may demonstrate that a precipitation product provides plausible information, but many model parameters can be tuned to mask precipitation bias (Kirchner 2006), thus resulting in a cascade of incorrect and compensatory values generating a simulated hydrograph that matches observed streamflow.

Here, we evaluate the H10 and L13 datasets over a 20-yr period in the state of California. We chose these two datasets because they are readily available at a high resolution (1/16°) over long time periods (1916–2010 and

1915–2011, respectively), which makes them easier to compare to point measurements, and they represent two gridding methodologies (H10 follows HL05 and L13 follows M02; see Table 1) that have been widely used and cited for hydroclimatic studies (Table 1). Because most gridded datasets utilize similar gauges and techniques, we presume that the results of studying these two broadly represent the performance of gridded precipitation products.

We propose daily snow measurements from snow pillows as an independent and underutilized resource for precipitation evaluation. In particular, we focus on a network of over 100 snow pillow stations in the Sierra Nevada, California, which are maintained by the California Department of Water Resources (CADWR). The SNOTEL network covers the Sierra Nevada sparsely compared to the CADWR network, and the SNOTEL data are included in the CADWR database. The SNOTEL network was used to validate the North American Land Data Assimilation System (NLDAS; Pan et al. 2003), which led to substantial improvements in precipitation gridding in the subsequent NLDAS product [NLDAS, version 2; see discussion in Cosgrove et al. (2003)]. Snow water equivalent (SWE) on 1 April at 427 California snow pillows and courses was used to manually adjust the PRISM monthly climatologies in the Sierra Nevada (C. Daly 2015, personal communication). However, to our knowledge, the daily accumulations at CADWR snow pillows have not been directly compared with any precipitation estimates.

While the present paper focuses on gridded precipitation, we use snow observations for independent evaluation. In any discussion of snow, a grid cell's temperature (or wet-bulb temperature; Marks et al. 2013) becomes critically important for distinguishing rain from snow and for determining rates of snowmelt. While it is beyond the scope of this paper to analyze gridded temperatures, we note that the datasets have different methodologies for adjusting temperature with elevation (see Table 1), with many using a constant lapse rate for the entire United States, which may be an erroneous assumption in some regions (e.g., Minder et al. 2010). We return to this issue in the discussion.

In this paper we 1) evaluate the performance of two fine-resolution gridded precipitation products at high elevations in the Sierra Nevada using increases in SWE as a surrogate for precipitation; 2) determine when, where, and why these gridded products fail to estimate high-elevation precipitation; and 3) discuss a practical path forward. Section 2 introduces the data used in this analysis. Section 3 describes the methodology, and section 4 presents results with regard to gridded data performance, underpredicted events, related synoptic weather patterns, and implications for annual water supply forecasting. Section 5 includes a discussion, and section 6 offers a summary and conclusions.

2. Data

a. *Hamlet (H10) and Livneh (L13) 1/16° gridded datasets*

H10 and Salathé et al. (2014) describe the 1/16° dataset developed for climate studies following the HL05 methodology. Precipitation values greater than 350 mm day^{-1} were removed, and stations were required to have 5 years of total data and at least 365 continuous days of data. PRISM normals were used from the 1971–2000 climatology. Rather than a constant lapse rate everywhere (as applied in many gridded datasets), they used PRISM maps to rescale maximum temperature T_{max} and minimum temperature T_{min}, taking care to preserve the observed diurnal temperature range on each day. For California, these data were produced for 1916–2010 and are available from the Climate Impacts Group (2014). As in HL05, USHCN, version 2, stations were used to correct temporal biases and shifts in the COOP-based gridded dataset.

L13 updated the M02 dataset to 1/16° resolution and extended it to encompass the 1915–2011 period. They followed the M02 methodology closely, using only stations with at least 20 years of valid data, and they used the 1961–90 PRISM climatology maps for monthly precipitation rescaling and a global −6.5°C km^{-1} lapse rate for temperature adjustment. Because of the longer time period than in the original M02 dataset, they were able to include more stations at some times. They also conducted some additional quality control of the input precipitation data, but none of the stations screened out were located in California (L13, their supplemental material).

b. *NCEP–NCAR reanalyses*

Gridded 700-hPa parameters for daily temperature and zonal and meridional winds from the NCEP–NCAR 40-Year Reanalysis Project (Kalnay et al. 1996) were obtained for the 2.5° grid cell centered at 37.5°N and 120°W, near the center of the Sierra Nevada. The 700-hPa level is at about 3000 m in elevation, which corresponds well with the elevation of snow observations in the Sierra Nevada. We used these data as an indicator of synoptic conditions and associated winds in the area, preferring them over finer-resolution reanalysis products as a broad first check because they are available over a much longer time period.

c. *Snow pillow measurements*

The CADWR manages a network of 125 snow pillows, 103 in the Sierra Nevada [Fig. 1; data available

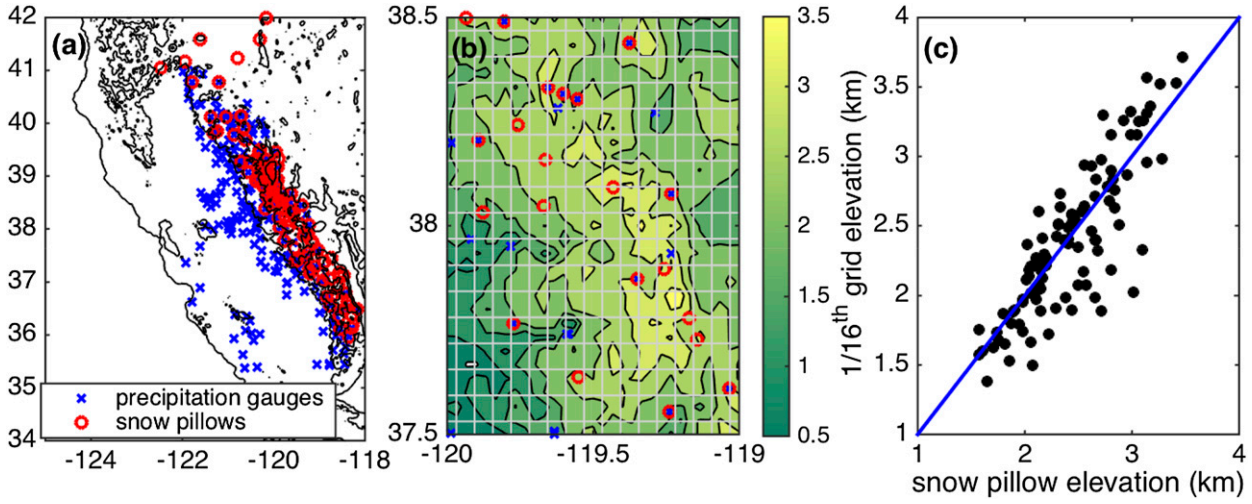

FIG. 1. (a) Map of precipitation gauge and snow pillow locations examined here, with California coastlines and 1-km elevation contours. (b) Enlarged 1° × 1° region from (a) in the center of the Sierra Nevada to illustrate station density, with 0.5-km elevation contours (black) and the 1/16° gridbox boundaries (light gray). (c) Scatterplot of snow pillow elevations and corresponding elevations from the 1/16° grid of H10 and L13.

from California Data Exchange Center (2014b)]. These are generally located in flat clearings and measure the weight of snow accumulating over an area of about 7 m^2 to determine SWE. Because pillows can experience several hours' delay in responding to changes in SWE (Beaumont 1965; Johnson and Marks 2004), data were analyzed at daily increments. All positive daily changes in measured snow water equivalent (+ΔSWE) were taken to be a measure of daily snowfall. An increase in SWE was attributed to snow falling on the pillow or to liquid water falling on snow already on the pillow and freezing into the snowpack, thereby increasing its density. In freezing locations where a snow pillow was collocated with a precipitation gauge, the timing and amount of +ΔSWE closely tracked the total accumulated precipitation. Exceptions occurred where the precipitation gauge suffered severe undercatch (in those cases +ΔSWE exceeds measured precipitation) or during warm rain events (when rainwater passes through the snowpack and drains away from the pillow and measured precipitation exceeds +ΔSWE). Snowmelt and/or sublimation can also decrease SWE. Wind redistribution of snow can either augment or decrease SWE, but this effect is slight because most California snow pillows are in sheltered locations (Farnes 1967). In summary, snow pillows are a reliable measure of high-elevation snowfall, and they do not suffer from the undercatch that standard precipitation gauges suffer in such environments (Yang et al. 2005). However, because the Sierra Nevada snowpacks are typically warm and isothermal, most rain falling on a snow pillow is not retained and therefore not measured (Lundquist et al. 2008).

d. Precipitation gauge data

Daily precipitation data (Fig. 1) were obtained from the CADWR and cooperating agencies, who manage a network of precipitation gauges throughout the state [data from California Data Exchange Center (2014a)]. Low-elevation sites consist of both tipping-bucket and accumulation reservoir gauges, while most higher-elevation sites use precipitation reservoir gauges with antifreeze.

e. Grid elevation comparison

For comparison, we selected the 1/16° grid cell containing each snow pillow. There was no consistent bias between the elevation of the grid cells (H10 and L13 use the same grid) and the snow pillows: the mean difference put the grid elevations 30 m lower than the pillows, and the median difference put the grid 17 m higher. There was a slight tendency for the pillows to be higher than the gridcell elevation at the lowest elevations and lower than the gridcell elevation at the highest elevations (Fig. 1c). This owes to the logistics of snow pillow siting: at lower elevations, water managers strive to find a site that measures more snow than rain (locally higher), while at higher elevations, the highest terrain is too steep, exposed, and/or inaccessible to place a snow pillow, which leads to locally lower site locations and an undersampling of the highest elevations (Rittger 2012). We plotted long-term mean gridded precipitation versus measured +ΔSWE as a function of local elevation difference and did not find any relation, so we do not expect elevation differences to influence our results.

3. Methodology

a. Snow pillow quality control

The snow pillow data were used to evaluate snowfall totals for both individual days and for water-year totals. Cases where +ΔSWE was unrealistic (>140 mm) or where the same exact +ΔSWE repeated for multiple days were removed from the dataset and labeled as missing. Because some stations exhibited spurious noise in the summer snow-free season, if the sum of +ΔSWE from 15 June to 15 September exceeded 200 mm in any given year, all data falling within that period were set to missing. While these dates were excluded completely from individual day comparisons, they were treated as zero values in annual totals for that station. Therefore, additional tests were conducted to determine whether a station's water-year total should be excluded for a given year. Stations with zero total snowfall over a year were excluded from that water year's analysis (assuming a broken sensor was unable to record snowfall). Stations with more than 30% of values for a year labeled as missing were excluded. Finally, to screen stations that broke (or were repaired and reinstated) midway through the year, we identified stations with a cumulative snowfall pattern that differed drastically from the rest of the stations in the Sierra Nevada. Specifically, we identified the date on which the median accumulation was 40% of the year's total and the date on which the median accumulation was 60% of the year's total. Any station that accumulated 100% of its annual total before the median reached 40% was presumed to have broken midseason and stopped recording. Likewise, any station that still had zero of its annual total on the date the median reached 60% was presumed to be broken during the first part of the water year and then repaired only in time to measure the last few snowfalls. In these cases, individual days were still included in the analysis, but that water year's total for those stations was set to missing.

b. Determination of differences between gridded P and +ΔSWE

Differences between the gridded datasets and observed snowfall were calculated in three ways: daily differences, water-year total differences, and event differences. Daily differences were calculated on all days with Sierra Nevada–wide median +ΔSWE > 0, for all stations where gridded minimum daily temperature reached or fell below 0°C. This threshold will tend to bias the comparison toward having more precipitation in the gridded products relative to +ΔSWE (see the discussion) and thus is conservative when evaluating underestimates of gridded precipitation. Statistics were calculated both for the median difference across all stations on a given day and for differences at all stations independently.

Water-year +ΔSWE totals were summed at all stations that passed the quality-control metrics discussed above. For the gridded datasets, water-year totals were calculated both for all precipitation regardless of temperature and for only precipitation falling on days with $T_{min} \leq 0$°C (assumed snowfall days). For both sets of temperature criteria, statistics were calculated both for the median difference across all stations in a given water year and for differences at all stations independently.

We selected time periods to examine for synoptic weather patterns and storm dynamics based on snowfall event characteristics. To do this, we defined a snowfall event as a 5-day consecutive sum. This is slightly longer than the 3-day snowfall event used by Serreze et al. (2001) but is designed to minimize issues of misregistration (discrepancies in the exact time period a snow pillow reports the snow gain) and to better capture the largest events, which tend to last longer (O'Hara et al. 2009). To minimize issues of temperature and rain versus snow in the event statistics, we only considered those stations at elevations above 2500 m (42 stations total) in selecting our top events (all stations were used in all other comparisons.) For 1) total snowfall and 2) gridded precipitation minus observed snowfall, and for both datasets, we first calculated the running 5-day sum at each individual station (to minimize local time registration issues) and then took the median across all stations above 2500 m. All differences were in terms of absolute numbers and not percentages in order to prioritize events that would contribute more toward annual totals.

Once this collection of events was created, we selected event types to maximize two criteria: 1) most snowfall and 2) most underpredicted snowfall by each gridded dataset, both ranked by the stationwide median 5-day sums. We looked at the 40 top 5-day totals for each metric and grouped consecutive or overlapping periods into one event. We then calculated the total value across the grouped period and ranked the grouped events into the top 15 for each category. In addition, any event with an atmospheric river (AR) identified as making landfall in California on at least one of the event days was considered an AR event, based on satellite observations of long, narrow plumes of enhanced integrated water vapor [after 1998, Neiman et al. (2008)] or vertically integrated water vapor transport (IVT) greater than 250 $kg\,m^{-1}\,s^{-1}$ in the North American Regional Reanalysis [NARR; Mesinger et al. 2006; before 1998, Rutz et al. (2014)].

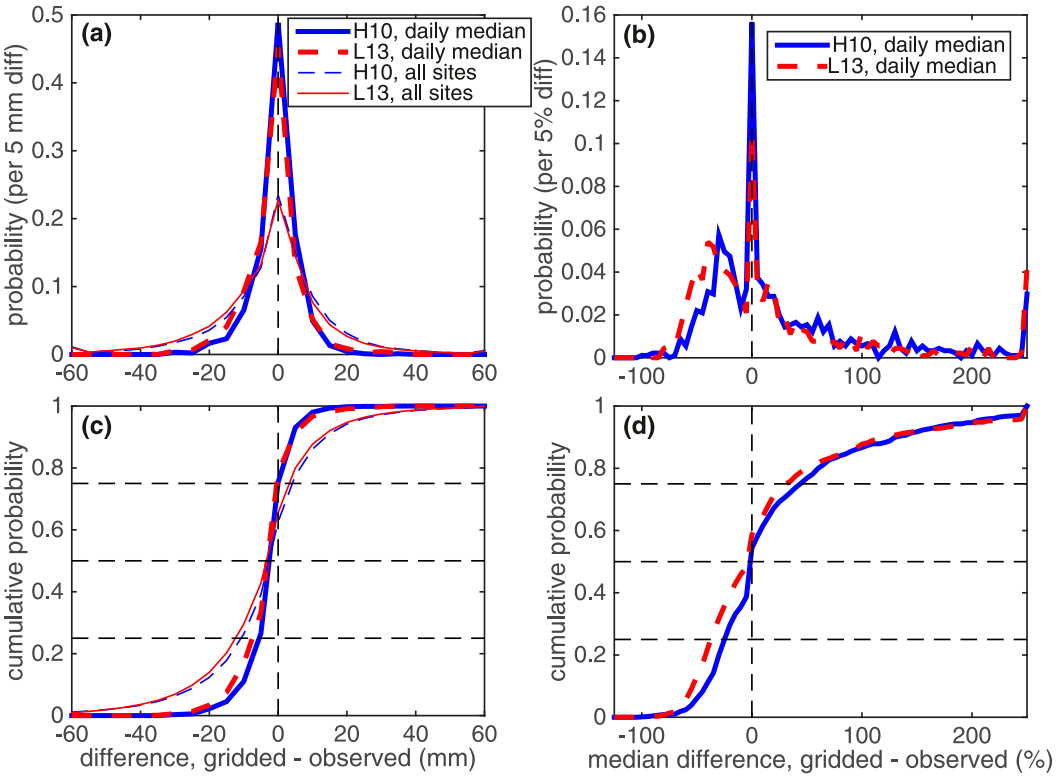

FIG. 2. (a),(b) Probability density functions and (c),(d) cumulative density functions of the difference between gridded precipitation and observed snowfall at California snow pillow locations for all days when the median state observed snowfall was >0 mm and for all locations where the gridded minimum temperature on that day was ≥0°C. Thick lines illustrate the distribution of the daily median difference across all locations, and the thin lines show the distribution of daily differences at all locations independently. Differences are expressed in both (left) absolute terms (mm) and (right) percentage of observed (%). Table 2 states the median values and interquartile ranges for each distribution.

c. *Wind and synoptic analysis*

Daily zonal and meridional 700-hPa winds were transformed into vectors and binned by originating direction (at 15° intervals). These bins were then weighted by the amount of precipitation (measured by gauges, measured by snowfall, or underpredicted by gridded datasets) to create wind-rose histograms of the predominant wind direction(s) contributing to total precipitation, total snowfall, or total gridded data underestimation.

Synoptic analysis was conducted by examining individual and composite maps of geopotential heights, winds, and IVT in the NARR (using plotting tools at www.esrl.noaa.gov/psd/cgi-bin/data/narr/plotday.pl) and by looking at NCEP Weather Prediction Center's archived daily weather maps (available at www.wpc.ncep.noaa.gov/dailywxmap/) for each day of the top six largest snowfall events and of the six underpredicted events that appeared in the top 15 of both gridded datasets. The latter include the surface analysis of fronts, along with station data (including prior 24-h accumulated precipitation in inches), for 0700 EST, which were visually inspected and assessed qualitatively.

4. Results

a. *Overall climatology and error statistics: Daily and water-year totals*

On a daily basis over the 20-yr period, the gridded datasets perform reasonably well, with a median underprediction of 3 mm, and over half the snowfall values differing from observed by less than 8 mm for the daily median and by less than 12 mm for individual stations (Fig. 2, Table 2). (For reference, the resolution of both a precipitation gauge and a snow pillow is 1/100 in., or 0.25 mm.) The median observed snow accumulation on any day with snowfall is 9 mm, and so the percent differences relative to observed were also calculated each day for Sierra Nevada–wide median differences (Figs. 2b,d; Table 2). While the mode of the difference was zero for both datasets, the medians indicated slight underprediction (−1% for H10 and −4% for L13). Interquartile ranges were from −25% to +44% for H10 and from −37% to +31% for L13. Both distributions had long tails of percent overprediction, likely representing days when very little snowfall was observed but

TABLE 2. Summary error statistics for difference of gridded minus observed.

Time	T	Sites	Dataset	Median (mm)	Median (% of observed)	Interquartile range (mm)	Interquartile range (% of observed)
Daily	All days with $T_{min} \leq 0°C$	Median across sites	H10	−3	−1	From −6 to 0	From −25 to 44
			L13	−3	−4	From −8 to 0	From −37 to 31
Daily	All days with $T_{min} \leq 0°C$	All sites	H10	−3	—	From −11 to 4	—
			L13	−3	—	From −12 to 3	—
Water-year totals	All days with $T_{min} \leq 0°C$	Median across sites	H10	0	−3	From −65 to 40	From −9 to 3
			L13	20	0	From −25 to 110	From −5 to 10
Water-year totals	All days with $T_{min} \leq 0°C$	All sites	H10	−15	—	From −200 to 146	—
			L13	20	—	From −168 to 214	—
Water-year totals	All T	Median across sites	H10	181	20	From 95 to 340	From 10 to 33
			L13	131	17	From 90 to 270	From 10 to 25
Water-year totals	All T	All sites	H10	181	—	From −12 to 416	—
			L13	131	—	From −78 to 374	—

when gridded precipitation (likely falling as rain rather than snow, despite $T_{min} \leq 0°C$) was mapped across the state. Issues of temperature and rain versus snow are included in the discussion.

The time series of median differences (only considering days with recorded snowfall) were slightly yet significantly antiautocorrelated (correlation coefficient of −0.21 for H10 and −0.14 for L13, both with p values <0.001) at a lag of 1 day. We interpret this antiautocorrelation to mean that there may have been some issues with the exact day when snowfall was recorded at the pillow versus in the gridded dataset, and therefore, we also explore annual and event statistics to eliminate any errors caused by temporal misregistration. Lag correlations of daily errors at lags beyond 1 day appeared to be random and generally were not statistically significant.

Over half of median water-year totals fell within ±10% of observed for both the H10 and L13 datasets (Fig. 3, Table 2), and about 90% of years fell within ±20% of observed. The spread was over twice as large when all stations were considered individually, and the distribution shifted over 100 mm or about 20% toward a positive bias when precipitation falling at all temperatures was summed (Fig. 3, Table 2). Water-year precipitation totals were generally greater for the H10 dataset than the L13 dataset when all temperatures were considered, but totals were greater for the L13 dataset than the H10 dataset when only summing precipitation falling on days with $T_{min} \leq 0°C$ (likely due to the steeper lapse rate and thus colder temperatures used in the L13 gridding).

Overestimation of grid precipitation versus observed snowfall may be because of issues with the grid's temperature and/or with the methodology used to distinguish rain from snow, rather than with actual problems with gridded total precipitation. However, underestimation of grid precipitation versus observed snowfall can only occur when grid precipitation is underestimated. Therefore, we focus the remainder of the study on cases of underestimation. In the distribution of water-year differences, the most severe underestimation occurred in all datasets in water-year 2008, with a low bias of −10% in both H10 and L13 when precipitation was summed across all temperatures and a low bias of −21% (H10) and −12% (L13) when only precipitation on days with $T_{min} \leq 0°C$ were summed. In section 4e, we examine the most prominent underpredicted storm events and their associated synoptics to better understand the gridded dataset bias in 2008.

b. *Most significant underpredicted snowfall events*

Of the top 15 most underpredicted events, six appeared in both the H10 and L13 datasets (Table 3). For these events, we considered the event total precipitation and +ΔSWE totals for the period of overlap and calculated the percent of underrepresentation (Fig. 4a). For comparison, we calculated the same for the six largest events at the same high-elevation stations (Fig. 4b). Only one event (January 2005) appears in both the top six largest snowfall events and the six underpredicted events. In general, the gridded datasets underestimated the median snowfall by 25%–50% (20–80 mm) in each of the underpredicted events, but tended to overestimate median snowfall by 10%–50% (15–90 mm) in the largest snowfall events. Overestimation may be due to an error in the gridded precipitation amount or to liquid precipitation not being measured by the snow pillows.

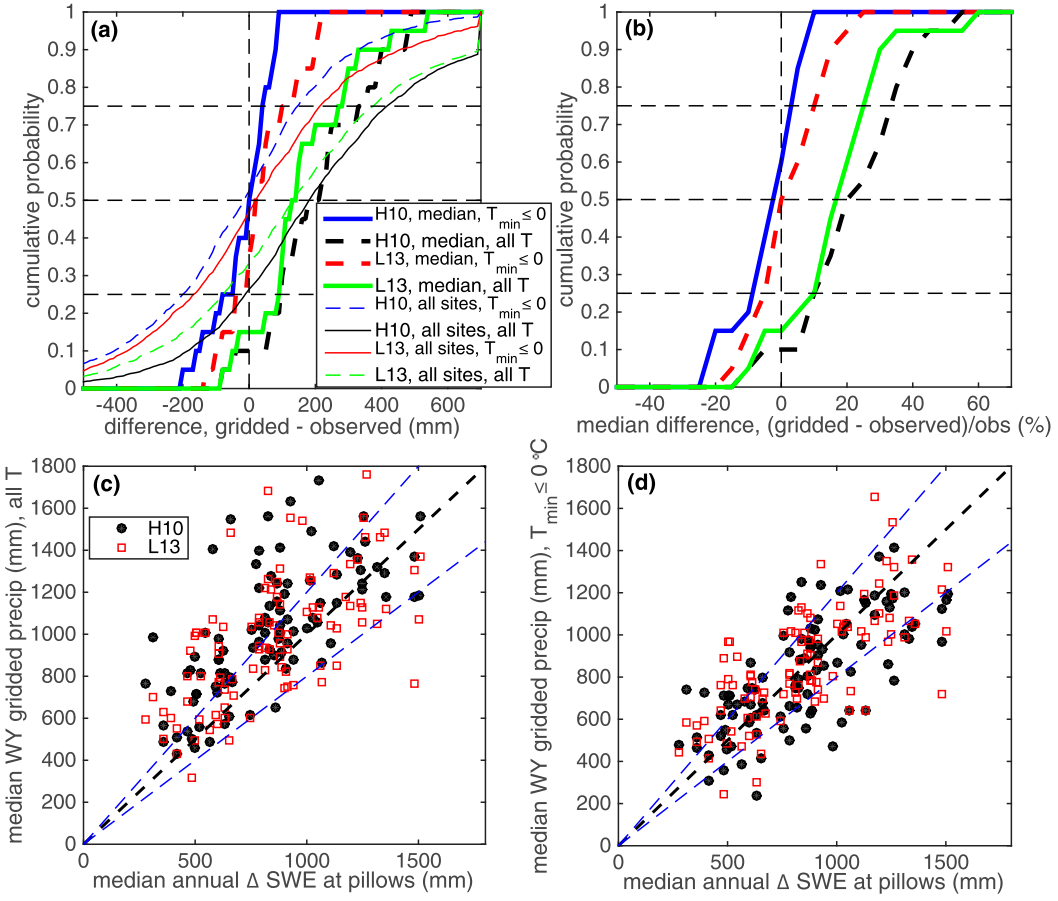

FIG. 3. (a) Cumulative distribution function of gridded minus observed water-year totals during 1991–2010 for median values across all sites (thick lines) and for all sites (thin lines), for all temperatures, and for only days when minimum temperature was ≤0°C for both H10 and L13 datasets. (b) As in (a), but for percent differences for median values across all sites. (c) Scatterplot of median values across all years for individual sites for water-year totals across all temperatures for H10 and L13 values vs observed. (d) As in (c), but for the sum of precipitation that fell on days when minimum temperature was ≤0°C. Black dashed line is 1:1 line, and blue dashed lines represent differences of ±20% from the 1:1 line.

Gridded temperatures were on average 2°C cooler in the six underpredicted events (Fig. 4a) than in the six biggest events (Fig. 4b). Although mean temperatures were below freezing in all of the events, the maximum temperature for each event was typically above freezing. Therefore, mixed-phase precipitation cannot be discounted. While the gridded datasets differed from each other, they were closer to each other than to the snow observations in almost all cases (Fig. 4).

c. *Wind patterns*

To understand the dynamics associated with these underpredicted events, we look at precipitation gauge-, snowfall-, and underprediction-weighted wind distributions (Fig. 5), using the median value across all sites for each day, and only considering days when the median measured snowfall was positive. Most total precipitation and snowfall occur when winds come from the southwest (as has been well documented in the literature; e.g., Pandey et al. 1999). However, the solid precipitation measured by the snow pillows has a distinct shift in the distribution compared to precipitation measured by traditional rain gauges on the same set of days. In particular, precipitation gauges measure 79% of total precipitation when winds come from the southwest (180°–270°), compared to the snow pillows only recording 65% of total snowfall. Snow pillows more frequently record gains in SWE during northwest winds (270°–360°) than precipitation gauges do: 32% of the total snow accumulated on pillows during northwest winds, compared to 20% of total precipitation for gauges. While this effect may appear slight in the direct accumulation data, it becomes much more pronounced in cases of underprediction at snow pillow locations (Fig. 5): 46% (H10) and 49%

TABLE 3. Top 15 largest observed snowfall events and top 15 underpredicted events for each dataset, with coincident underpredicted events (also shown in Fig. 4) marked in boldface. Superscript AR indicates that a landfalling atmospheric river was identified on at least one of the event days, as determined by the methodology of Neiman et al. (2008) for years after 1998. For years before 1998, superscript RBAR indicates that a landfalling atmospheric river was identified by visual inspection on at least one of the event days in NARR IVT using the threshold of Rutz et al. (2014).

Dates underpredicted by H10	Median total period underprediction (observed − H10 grid) (mm)	Dates underpredicted by L13	Median total period underprediction (observed − L13 grid) (mm)	Largest snowfall dates	Median total gain at sites above 2500-m elevation (mm)
From 23 Jan to 2 Feb 2008^{AR}	**66**	**From 21 Dec 2003 to 5 Jan 2004^{AR}**	**84**	From 28 Dec 2005 to 4 Jan 2006^{AR}	232
19–26 Feb 2008^{AR}	50	**From 27 Dec 2004 to 12 Jan 2005^{AR}**	**59**	5–12 Jan 2005^{AR}	217
From 21 Feb to 1 Mar 2007*	49	4–8 Jan 2008^{AR}	42	16–23 Feb 1996^{RBAR}	194
From 28 Dec 2003 to 5 Jan 2004^{AR}	**44**	18–24 Jan 2010^{AR}	40	From 27 Feb to 5 Mar 1991^{RBAR}	184
From 27 Feb to 6 Mar 2009^{AR}	**37**	From 28 Nov to 6 Dec 2001^{AR}	39	21–26 Jan 1997^{RBAR}	170
From 27 Nov to 5 Dec 2005 ^{AR}	**35**	**27–31 Jan 2008^{AR}**	**39**	6–12 Jan 1995^{RBAR}	166
21–28 Jan 1997^{RBAR}	33	13–17 Feb 2000^{AR}	33	1–8 Jan 2008^{AR}	166
2–6 Mar 2006^{AR}	**32**	**1–5 Dec 2005^{AR}**	**33**	9–14 Feb 2000^{AR}	162
19–23 Feb 1996^{RBAR}	30	16–20 Dec 2002^{AR}	33	5–11 Nov 2002^{AR}	158
22–26 Feb 2010^{AR}	29	4–8 Mar 1996^{RBAR}	33	7–13 Mar 1995^{RBAR}	157
18–23 Mar 2005^{AR}	28	10–14 Jan 1995^{RBAR}	33	From 27 Dec 2004 to 1 Jan 2005^{AR}	156
21–25 Jan 2009^{AR}	26	6–10 Mar 2002^{AR}	33	13–18 Dec 2002^{AR}	155
5–12 Jan 2005^{AR}	**26**	From 24 Feb to 1 Mar 2004^{AR}	29	8–12 Dec 1996^{RBAR}	142
2–6 Apr 2010^{AR}	25	**1–6 Mar 2009^{AR}**	**29**	2–6 Feb 1998^{AR}	137
11–15 Jan 2010^{AR}	12	**3–7 Mar 2006^{AR}**	**26**	18–22 Feb 1993**	134

* The 2007 event underpredicted by H10 appears associated with in situ sensor failure, as zero gridded precipitation occurred for a large Sierra Nevada region for the event.

** Non-plume-like blob of moisture approaching from the south.

(L13) of total underprediction occurs during winds from the northwest, compared to 48% (H10) and 45% (L13) occurring when winds come from the southwest (Fig. 5). In contrast, the wind distribution of cases with overprediction (not shown) looks identical to the precipitation gauge–weighted distribution.

These wind patterns illustrate that snow accumulation sometimes occurs when synoptic winds come from the northwest, but that relatively little gauge-based precipitation is measured during this time. This pattern leads to significant underprediction of precipitation falling at snow pillow locations by the gridded datasets (which are only based on precipitation gauge data) during these conditions. This problem can be better understood by looking at spatial patterns of precipitation gauge and snowfall measurements together, both in terms of absolute and relative values (Fig. 6). Much more total precipitation falls during southwest winds than northwest winds (Figs. 6a,b), and as such, these events dominate long-term spatial patterns, which climatologies such as PRISM are trained to match. While northwest-wind precipitation events produce less of the total precipitation, this effect is more pronounced at lower elevations than upper elevations. In other words, at many times, precipitation only occurs at higher elevations (with none recorded at lower elevations). Even when the median snowfall was greater than 0 mm, the median measured precipitation across 42 stations at elevations less than 200 m was 0 mm on 51% of days with southwest winds and 88% of days with northwest winds. These low-elevation stations are treated as a low-elevation precipitation index and used as the denominator in all plotted ratios in Figs. 6c and 6d. Only considering days with both measureable median snowfall and low-elevation precipitation, much more dramatic relative ratios occur during northwest winds (up to six times more mountain than valley precipitation; Fig. 6c) than southwest winds (approximately 2–3 times more mountain than valley precipitation; Fig. 6d). These ratios are likely due as much to the denominator of the ratio being small (very little valley precipitation) as to greater orographic enhancement processes (see the discussion). The basic

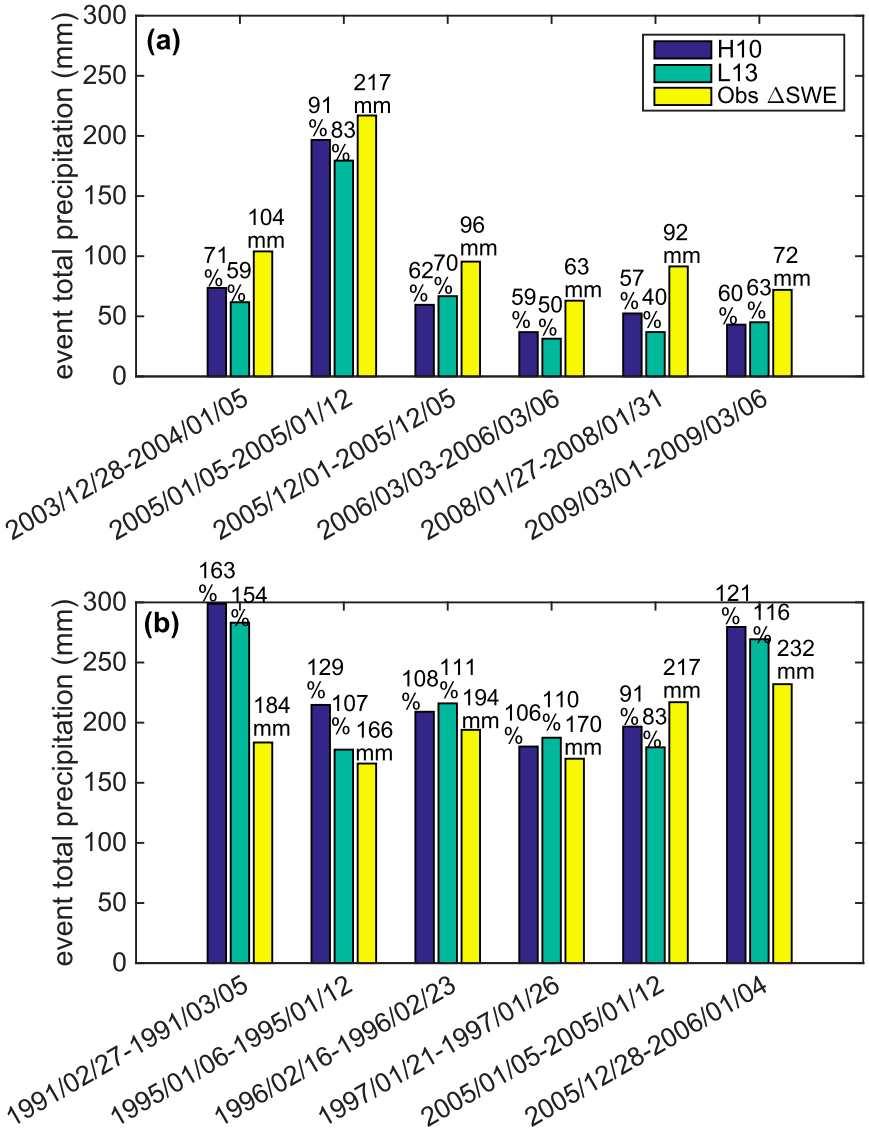

FIG. 4. Median (across all stations above 2500 m) storm total +ΔSWE (observed) and precipitation (from the gridded datasets; H10; L13) for (a) events ranking in the top 15 of underprediction in both datasets, considering only six events with overlapping dates; and (b) the six largest observed snowfall events, both presented in chronological order. Total observed snowfall is written, as well as the percentage of this total recorded in each of the gridded datasets.

methodology employed by both the H10 and L13 datasets uses the typical PRISM ratios (close to those shown in Fig. 6d) to predict precipitation at high elevation from low-elevation gauges, resulting in greater errors when conditions match those associated with northwest winds (Fig. 6c, also evident in Fig. 5).

d. Synoptics of underpredicted events

To better place these wind-related patterns into a synoptic context, we examined daily synoptic weather maps (see section 3c) for the top six underpredicted events that appeared in both datasets (Fig. 4) and for the top six largest snowfall events (Table 3), paying particular attention to large events that were not subject to underprediction. Distinct patterns emerged (Figs. 7, 8). Both the largest snowfall events and the most underestimated snowfall events were associated with storms with clear frontal signatures and with landfalling atmospheric rivers located in the warm sector of the storm (Figs. 7, 8; Table 3). Aside from stronger IVT in the largest snowfall events (as would be expected given their greater total snowfall amounts), the biggest differences

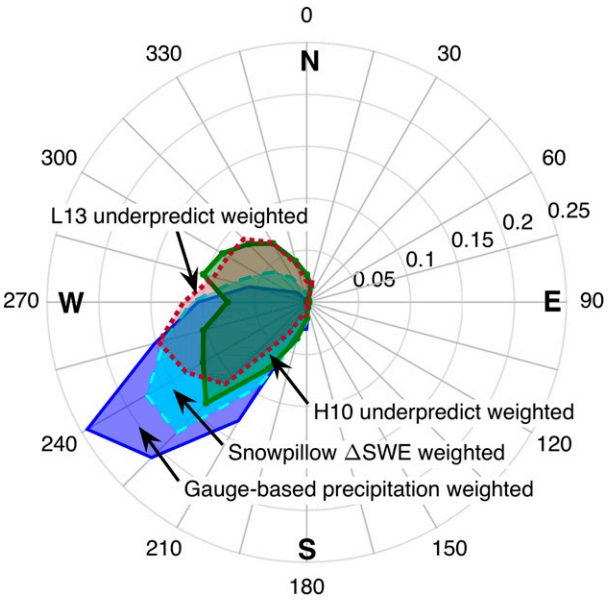

FIG. 5. Fraction (radial coordinates) of total gauge-based precipitation (dark blue solid line), snow pillow–measured snowfall (cyan dashed line), underprediction of snowfall by the H10 dataset (green line), and underprediction of snowfall by the L13 dataset (red dotted line), for each 15° wind direction window (axial gray lines that indicate the direction wind is coming from), as measured by daily 700-hPa reanalysis winds. All fractions are calculated using the median value across all sites for each day, and only days when median measured snowfall is >0 are considered.

were the storm movement and progression, which influenced the relative amount of time the California Sierra Nevada spent under the influence of the warm sector of the storm (moist conditions and winds from the southwest) as opposed to the cold sector of the storm (less moist conditions and winds from the northwest). Kingsmill et al. (2006) discuss in more detail the characteristics of these different storm sectors in California.

In a typical underpredicted storm event, the AR makes landfall in association with a 500-hPa low-pressure axis close to the coast (Fig. 7a), which taps into cool, moist air from the northwest in addition to warm, moist air from the southwest (Fig. 7e). The entire system tracks inland (Fig. 8a), generally moving across Nevada and/or Utah. The Sierra Nevada spends limited time in the warm sector but significantly more time in the cold sector (post–cold front), and northwest flow is evident both aloft (Fig. 7b) and in the integrated water vapor flux (Fig. 7f) on the last day of the event.

For comparison, in a storm with heavy snowfall that was better predicted by the gridded datasets, the low-pressure trough on the day of AR landfall is much broader, with an axis farther west of the coast (Fig. 7c), and moisture transport is uniformly from the southwest (Fig. 7g) with no northwest contribution. The surface low tracks north of California (Fig. 8b). In this situation, the Sierra Nevada is subjected to pre-warm-frontal warm sector, and pre-cold-frontal precipitation but spends limited time in the cold sector because of the tracking of the overall system (as the dashed red arrow in Fig. 8b shows). On the last day of the storm, the upper-level flow is westerly (Fig. 7d), and water vapor flux is still from the southwest (Fig. 7h). These results, both the synoptic situation and enhanced orographic ratios observed during northwest winds, have similarities to previously documented changes in Sierra Nevada orographic enhancement (Dettinger et al. 2004), and our explanation as well as theirs for why this occurs is included in the discussion.

e. *Contribution of underpredicted events to water-year totals*

Unlike in heavy rainfall (which can lead to flooding), one underpredicted snow storm may not impact state water resource management so long as subsequent storms and compensatory errors "erase" the error over the course of an entire water year, leading to accurate seasonal runoff prediction. While median statistics for all days and locations with $T_{min} \leq 0°C$ (Table 2; Figs. 2, 3) suggest that the gridded datasets are unbiased estimators, in some years, about 20% bias remains (Fig. 3). One of the top six underpredicted events in both datasets occurred in water-year 2008 (Fig. 4), and other 2008 storms appeared independently in the top six of each of the H10 and L13 datasets (Table 3). Over the course of the entire water year, 2008 had more total snowfall occurring during northwest winds than was typical for the entire 20-yr period, 44% (Fig. 9a) compared to 32% (Fig. 5). Water-year 2008 received a median observed snowfall of 782 mm. In the median, the gridded datasets both differed from this by −76 mm (−10%) when all precipitation was considered (Fig. 9b) and by −161 mm (−21%) and −96 mm (−12%) for H10 and L13, respectively, when summing only gridded precipitation on days with $T_{min} \leq 0°C$ (Fig. 9c). When considering individual stations and the $T_{min} \leq 0°C$ criteria, 44 (46%) and 36 (38%) of 95 working stations had worse than −20% annual underprediction for H10 and L13, respectively (Fig. 9). While there were site-specific variations, the overall pattern was regionally coherent, indicating Sierra Nevada–wide annual underprediction (Figs. 9d–f), with multiple sites exhibiting 20%–60% underprediction. As shown in Fig. 3, less than 10% of all 20 examined water years had median water-year total precipitation underprediction of this magnitude.

5. Discussion

Using CADWR daily snow pillow–measured SWE data as an evaluation tool for gridded precipitation products, we

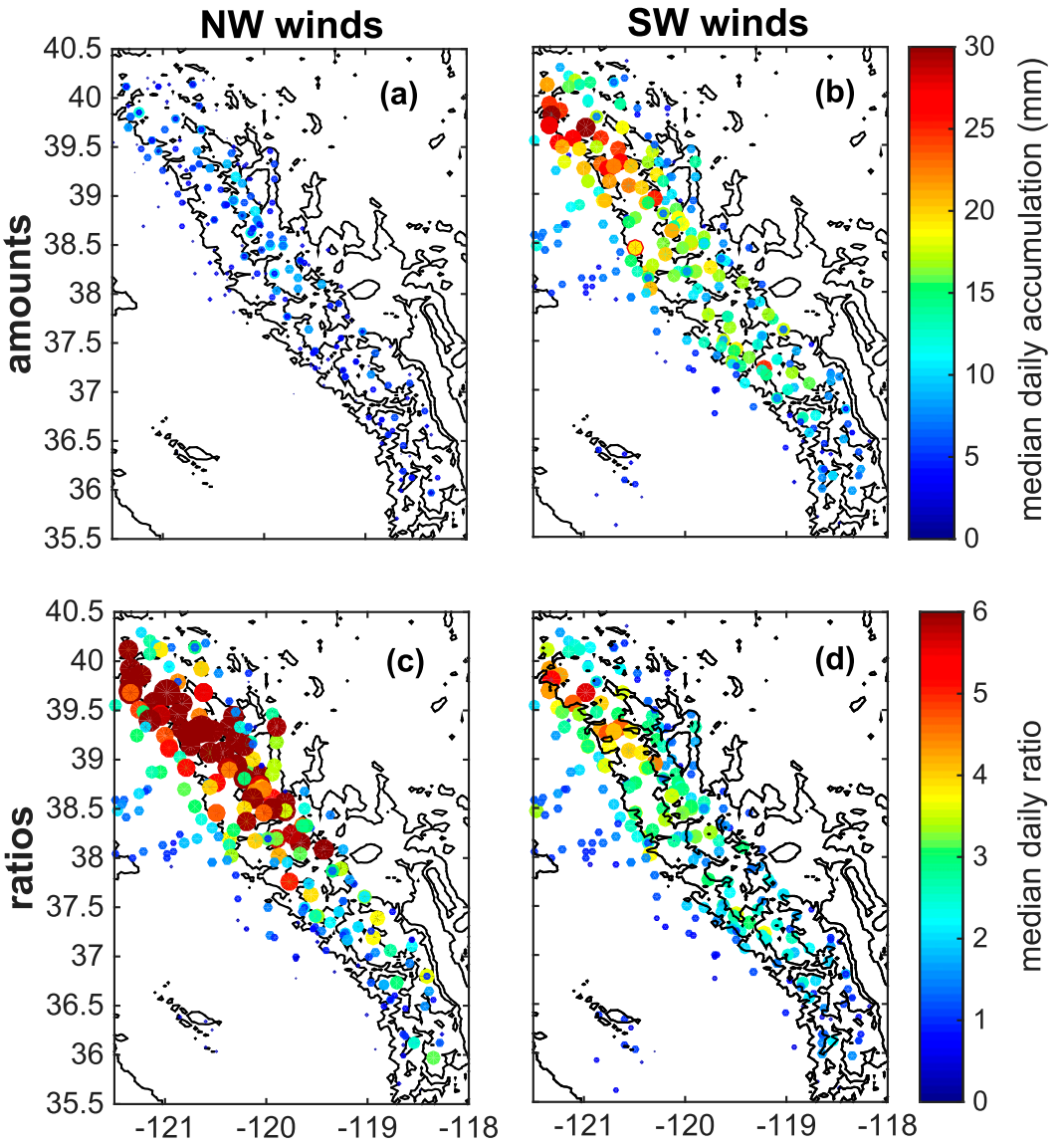

FIG. 6. (a),(b) Median observed daily Central Valley and Sierra Nevada precipitation and snowfall accumulation amounts and (c),(d) ratios relative to the median of stations at low elevations (<200 m, 42 stations total) for all snowfall events occurring under winds from the (a),(c) northwest quadrant (270°–360°; a total of 569 days) and (b),(d) southwest quadrant (180°–270°; a total of 721 days). Black lines show elevation contours at 1000 m increments. Only days with median $+\Delta$SWE > 0 were considered. Ratios (individual station daily total divided by median low-elevation station daily total) were calculated on each day with measureable low-elevation precipitation (defined as a median value >0). This only occurred on 22% of all days in the northwest quadrant and on 49% of all days in the southwest quadrant. Median ratios across these days are plotted.

find that, although the majority of errors are small, specific synoptic events can lead to statewide underprediction of high-elevation precipitation, and this can lead to −20% statewide, water-year total biases in some years. Here we expand on the ideas presented in the results by 1) addressing issues of undercatch; 2) discussing why northwest winds and more time spent in the cold sector of low-pressure systems lead to underprediction; 3) explaining why these events lead to water-year total biases in California despite their relatively low contribution to total precipitation; and 4) highlighting the pros and cons of using snow pillow measurements as a daily precipitation evaluation tool, with particular attention to issues of temperature and rain versus snow.

a. Undercatch issues

Because the information to accurately model undercatch is complex and not available for most gauge

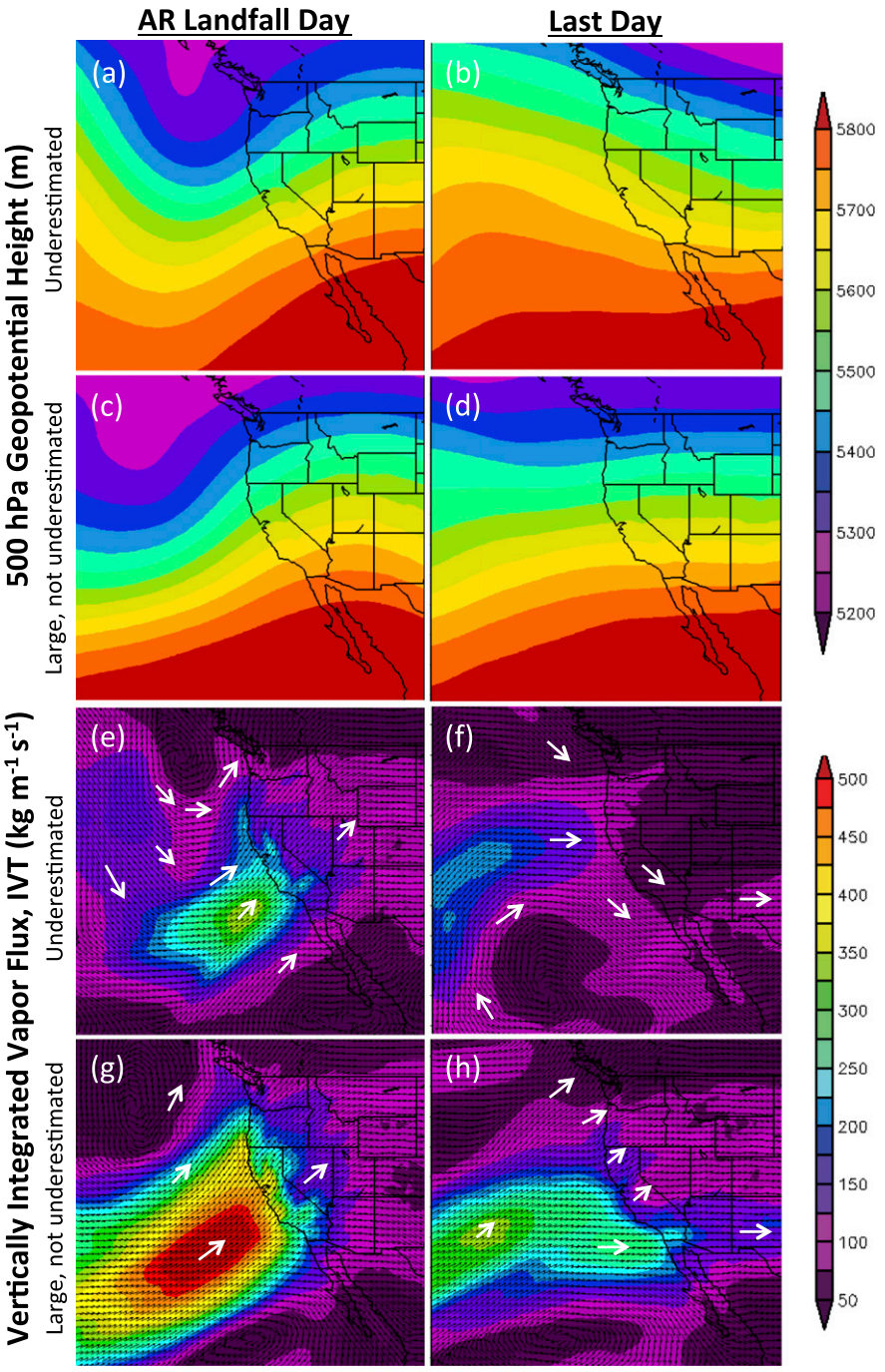

FIG. 7. Composite maps of (a)–(d) 500-hPa geopotential height and (e)–(h) IVT from NARR for a composite of daily average values from (a),(e) dates of AR landfall and (b),(f) the last date listed in the six most underestimated events (shown in Fig. 4a); and (c),(g) dates of AR landfall and (d),(h) the last date listed in five of the largest snowfall events (shown in Fig. 4b), excluding the 5–12 Jan 2005 event, which was underestimated. Dates for (a) and (e) are 2 Jan 2004, 8 Jan 2005, 2 Dec 2005, 5 Mar 2006, 26 Jan 2008, and 2 Mar 2009. (Note that the dates of overlap for the 26 Jan 2008 event in Fig. 4a start on 27 Jan 2008, but the AR made landfall the day before.) Dates for (b) and (f) are 5 Jan 2004, 12 Jan 2005, 5 Dec 2005, 6 Mar 2006, 31 Jan 2008, and 6 Mar 2009. Dates for (c) and (g) are 3 Mar 1991, 9 Jan 1995, 19 Feb 1996, 25 Jan 1997, and 2 Jan 2006. In cases where events had multiple AR landfalls, the one with the greatest water vapor transport was selected. Dates for (d) and (h) are 5 Mar 1991, 12 Jan 1995, 23 Feb 1996, 26 Jan 1997, and 4 Jan 2006. Imagery provided with assistance from the NOAA/ESRL/Physical Science Division, Boulder, Colorado (www.esrl.noaa.gov/psd/).

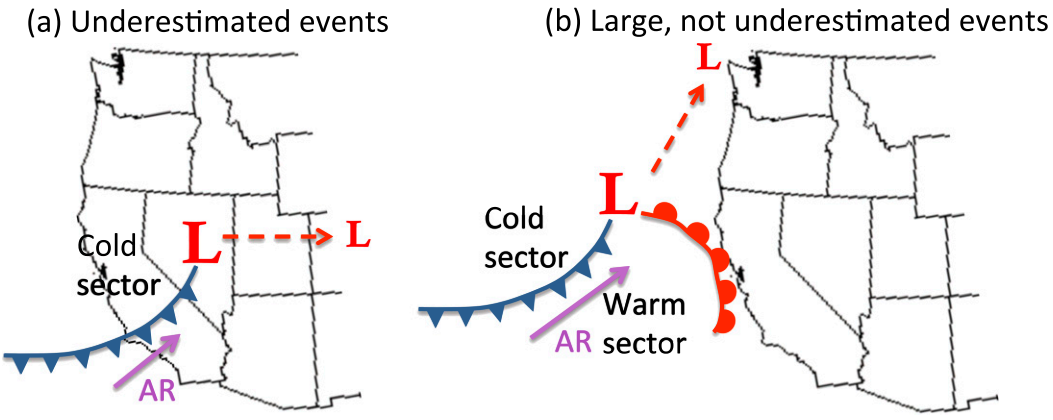

FIG. 8. Schematic of typical synoptic situations for (a) an event that is subject to severe underestimation of snowfall by gridded precipitation datasets and (b) a large snowfall event that is not subject to underestimation by gridded precipitation datasets.

locations, PRISM and the gridded datasets derived from it deliberately do not include an adjustment for precipitation gauge undercatch (M02). Undercatch can be significant (20%–50%) for snowfall, with errors generally increasing with wind speed (Goodison et al. 1998; Yang et al. 2005; Rasmussen et al. 2012). While this would lead us to expect precipitation underestimates in snow-dominated areas, there was no evidence that the most underpredicted events were associated with higher wind speeds or greater-than-usual undercatch issues. In fact, the upper quartiles of days underpredicted by the H10 and L13 datasets had smaller 700-hPa wind speeds than the upper quartiles of precipitation and snowfall events in general. One possible explanation is that Sierra Nevada snowfall tends to be warm and wet, which leads to higher catch efficiencies than typically observed in regions of colder and drier snow (Thériault et al. 2012). Another explanation is that the incorporation of snow pillow and course records in the PRISM climatology (Daly 2013; C. Daly 2015, personal communication) removed any consistent undercatch bias that would be detectable by snow pillows.

b. *Explanations for synoptic patterns related to snowfall underprediction*

Dettinger et al. (2004) looked at pairs of precipitation gauges in the central Sierra Nevada (near Lake Tahoe and Yosemite) and concluded that greater orographic enhancement (as defined by ratios of high- to low-elevation station precipitation) occurred on days with less southerly and more westerly winds, particularly in association with post-cold-frontal precipitation (their Fig. 9). They explained this as caused by the more westerly (cold sector) storms being more precisely perpendicular to the central Sierra Nevada topography, thereby causing maximum uplift. However, our results show that the pattern holds true across the entire Sierra Nevada (Fig. 6), which in general is more perpendicular to southwesterly flow and should therefore experience maximum uplift during the warm sector of the storm. For this reason, we prefer an explanation related to the different storm components [Houze (2012), his Fig. 26]. In the Houze (2012) literature review, multiple papers highlight drier, yet more turbulent, air and convective clouds associated with post-cold-frontal precipitation. We hypothesize that these convective clouds continue to produce snowfall in association with orographic uplift at higher elevations but do not generally produce precipitation in the Central Valley of California. This idea is supported by median low-elevation precipitation measurements of 0 mm during 88% of northwest-wind events with median snowfall measurements greater than zero. This could be due to the drier postfrontal air leading to a higher cloud base, such that precipitation may be evaporating or sublimating before it reaches the ground at lower elevations. Alternatively, or acting at the same time, the Sierra Barrier Jet (Neiman et al. 2010) often breaks down after the cold front's passage. Because the Sierra Barrier Jet generally acts to enhance uplift and precipitation over the valley upstream of the mountains, its disappearance likely decreases low-elevation precipitation relative to what is happening at higher elevations. In these situations, the low-elevation sites do not represent the high-elevation precipitation, and the statistical gridding methodology breaks down. Because the cold sector precipitation is generally a small fraction of storm totals, these errors can be neglected in many cases, but not all, as evidenced by the top six underpredicted events and by water-year 2008.

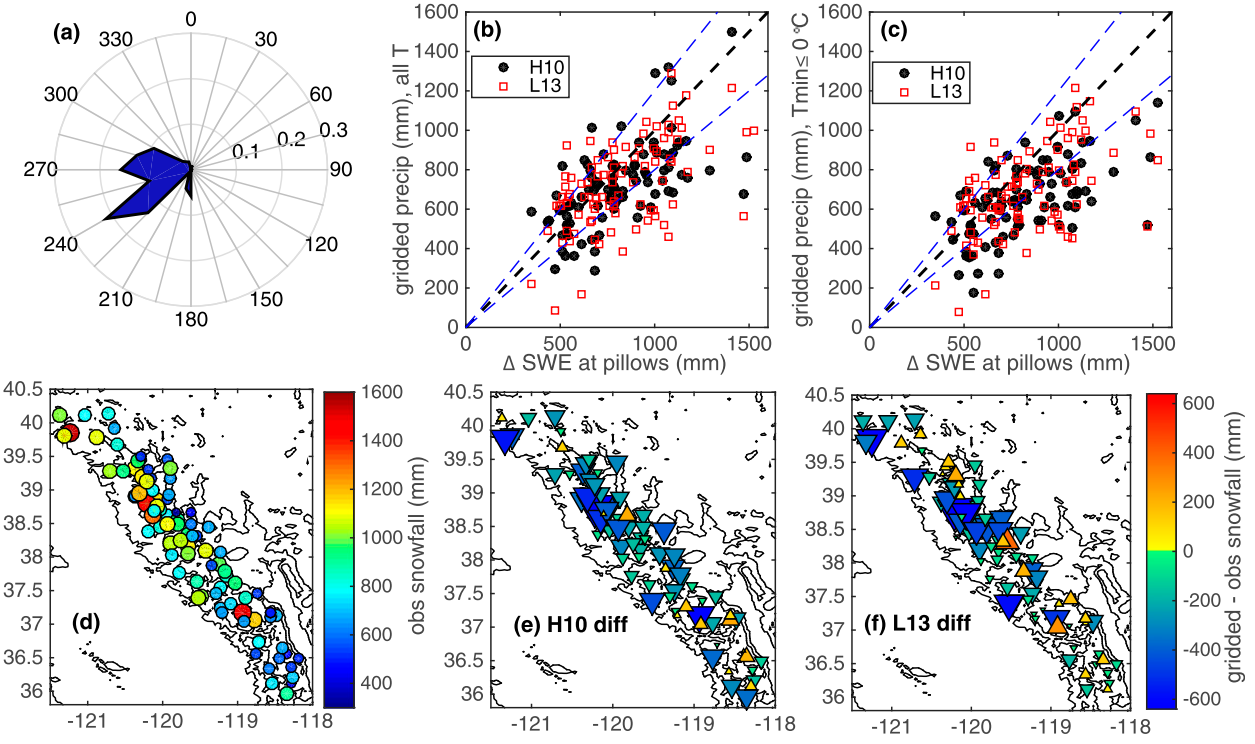

FIG. 9. Example of bias accumulated over water-year 2008. (a) Fraction of total water-year 2008 snowfall observed (radial coordinates) for each 15° wind direction window. Observed water-year snowfall vs gridded water-year precipitation for (b) precipitation at all temperatures and (c) only precipitation on days with $T_{min} \leq 0$°C, for the H10 and L13 datasets. Black dashed line is the 1:1 line, and blue dashed lines indicate ±20% deviations from that line. Spatial patterns of (d) total water-year snowfall observed and difference of annual gridded precipitation from that observed snowfall for the (e) H10 and (f) L13 gridded datasets.

c. *Why California is particularly susceptible to water-year total precipitation errors*

In California's Mediterranean climate, most precipitation falls between October and May. The Sierra Nevada averages ~10 snowstorms annually, but often one exceptionally large snowstorm makes up much of the annual total (O'Hara et al. 2009). Winter storms arrive from the Pacific Ocean, with the heaviest storms associated with narrow southwesterly streams of moisture, termed atmospheric rivers (Neiman et al. 2008; Ralph and Dettinger 2011). These events often contribute over 30% of the annual SWE (Guan et al. 2010, 2013). Based on SNOTEL data across the western United States, Serreze et al. (2001) found that the mean 3-day (72 h) largest event typically accounts for 10%–23% of the water content in total annual snowfall. Because the CADWR stations are in a separate network, Serreze et al. (2001) did not include them in their analysis, but the impact of the largest event, or just a few events, on California total snowfall is among the largest observed across the western United States. In the median across the 20 years we examined, over 50% of annual snowfall occurred in only 6–14 days (station median 10 days), the largest 3-day event provided 9%–25% (median 15%) of annual snowfall, and the largest 5-day event provided 12%–29% (median 19%) of annual snowfall (Fig. 10). Stations at lower elevations and toward the southern end of the Sierra Nevada had fewer days make up larger fractions of their annual totals (Fig. 10). Thus, because of the climatology of California, annual snowfall generally has a very small sample size of storms, and underestimating a single storm (or multiple storms, as occurred in 2008) can lead to significant total water-year biases.

d. *Snow pillow measurements as evaluation tools and complications due to temperature and rain versus snow*

Measurements from networks of snow pillows in the western United States comprise a widely available and practical source of information for evaluation of gridded precipitation datasets. Because pillows measure snow accumulation more accurately than precipitation gauges, they complement analyses based on gauges alone. For example, Pan et al. (2003) found that, on average, NLDAS precipitation was less than half of precipitation

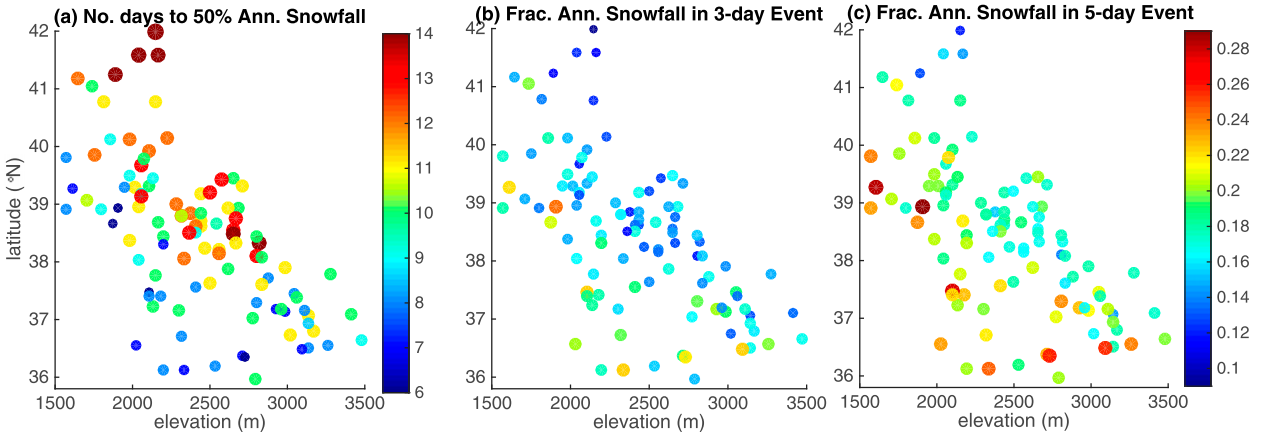

FIG. 10. (a) Median (1990–2010) number of days to reach 50% of total annual snowfall, and fraction of annual snowfall in largest consecutive (b) 3- and (c) 5-day snowfall event. All snow pillow data are plotted with their corresponding latitude and elevation to distinguish between stations.

observed at SNOTEL sites and that a hydrologic model forced with locally observed precipitation simulated snow accumulation and melt well, whereas one forced with the original gridded precipitation was biased low by over 50%, suggesting that most of the error in snow modeling arose from precipitation biases. This study led to key improvements in NLDAS, version 2, and to the dynamic NLDAS product (Cosgrove et al. 2003), which increased modeled precipitation at high elevations. However, to our knowledge, the new product has not been evaluated with data from snow pillows.

Our study focuses on gridded precipitation and particularly addresses events when gridded precipitation underpredicts observations at high elevations. However, the use of snow pillow measurements for evaluation forces us to consider uncertainty in both gridded fields of air temperature and in the methodology for distinguishing snowfall from rainfall. In cases where and when gridded precipitation is less than measured increases in SWE, we can identify an error in gridded total precipitation. However, any case when and where gridded total precipitation exceeds the measured snow increase may be due to a precipitation overestimate or to rainfall that was not measured by the snow pillow. When we try to isolate gridded snowfall from rainfall, results are subject to errors in the gridded temperature field or to errors in the temperature-based techniques used to distinguish rainfall from snowfall. In particular, both rainfall and snowfall often occur within the same day in the Sierra Nevada [see Lundquist et al. (2008) for a discussion of melting level], and multiple methodologies exist to estimate hourly temperatures from T_{max} and T_{min} (e.g., Waichler and Wigmosta 2003) and to distinguish fractions of rain and snow as a function of air temperature (e.g., U.S. Army Corps of Engineers 1956; Lundquist et al. 2008) or wet-bulb temperature (e.g., Marks et al. 2013). The simple method used here (to only consider precipitation on days when $T_{min} \leq 0°C$) is likely to overestimate the fraction of total precipitation falling as snow, because T_{max} may be well above the melting temperature. Thus, the good (unbiased) match between gridded precipitation when $T_{min} \leq 0°C$ and measured snow accumulation may, in reality, result from multiple errors making a right: for example, total precipitation may be underpredicted, as we would expect because of gauge undercatch, but the fraction of the total occurring as snow may be overpredicted, resulting in no net bias. This balance may also be influenced by overall storm temperatures, where warmer events are more likely to make snow pillow measurements underestimate total precipitation (due to rainfall at higher elevations) and colder events are more likely to make COOP gauge measurements underestimate total precipitation (due to undercatch of snowfall at lower elevations).

Differences between the H10 and L13 datasets when using local gridded $T_{min} \leq 0°C$ as a snowfall cutoff also arise because of how the datasets grid temperature. L13 uses a steeper lapse rate ($-6.5° km^{-1}$) than H10 (PRISM based) in most areas, which results in more locations and days meeting the $T_{min} \leq 0°C$ criteria for L13. This explains why, despite the overall precipitation (water-year totals for all temperatures) being larger for H10 than L13, this pattern reversed when only summing days and locations with $T_{min} \leq 0°C$ (Table 2). While a full analysis of combined temperature and precipitation errors is beyond the scope of this study, these issues highlight the challenges associated with relating errors in snow accumulation modeled using these datasets directly to errors in the datasets themselves. Here, we are only certain of errors related to underestimation of gridcell snowfall.

6. Conclusions

Gridded precipitation datasets incorporate nearly all available measurements of precipitation, and so their fidelity is hard to quantify even though it is often suspect in regions of complex terrain, where relatively few measurements are available. Here we evaluated two high-resolution (1/16° and daily) long-term datasets available over the continental United States (H10; L13) against daily snow pillow measurements of snow accumulation (+ΔSWE) at over 100 snow pillows across the Sierra Nevada, California, for the period 1990–2010. In general, over the entire period, the gridded datasets performed reasonably well, with over 50% of median errors on individual days falling between −37% and 44% and water-year total errors within ±10% (Table 2). However, errors in individual storm events sometimes exceeded 50% for the median difference across all stations, and in some years, these underpredicted storms led to 20% error in water-year total median statewide snowfall (e.g., water-year 2008, Fig. 9). Underprediction by the gridded datasets was associated with large-scale 700-hPa winds from the northwest and precipitation occurring during the cold sector of a frontal system. In these events, precipitation tends to be convective and less spatially organized than in the warm sector, such that much more precipitation occurs at higher elevations than in the low-elevation valleys, where most precipitation gauges are located. Because these events, in general, produce much less total precipitation than is produced during southwesterly flow through the warm sector of a storm, they are not well represented in long-term spatial climatology, as employed by PRISM gridding techniques. However, using the information presented in this study, they can be identified and flagged as periods likely to result in larger-than-usual errors, suggesting that snow pillow measurements could be used to supplement rain-gauge-based precipitation observations during such storms.

While the results presented here are specific to California, the basic principles could be applied to any mountain region. In general, precipitation accumulating at gauge-sparse regions (e.g., higher elevations) under conditions different from the climatologically predominant storm configuration is likely to not be well represented in gridded datasets. Precipitation accumulation during climatologically unusual conditions could be better represented by using a set of spatial pattern maps, each trained to a specific synoptic situation, rather than one PRISM climatology. In addition, high-resolution numerical weather models could be used to evaluate the spatial distribution of precipitation to further inform interpolation procedures. In principle, these practices could be incorporated into short-term and seasonal forecasting, as well as improve the representativeness of gridded datasets for understanding longer-term trends.

Acknowledgments. This work was supported by the National Science Foundation (NSF) Grant EAR-1344595 and by NOAA through their Hydrometeorology Testbed and through the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA10OAR4320148. The National Center for Atmospheric Research is sponsored by NSF (AGS-0753581). The opinions expressed herein are those of the authors and do not necessarily reflect those of the granting agencies. We thank Chris Daly and two anonymous reviewers who helped improve the manuscript. All data used in this study are publicly available. The H10 data are housed by the University of Washington Climate Impacts Group (http://cses.washington.edu/cig/data/wus.shtml). The L13 data are available from an ftp site (www.hydro.washington.edu/Lettenmaier/Data/livneh/livneh.et.al.2013.page.html). The California Department of Water Resources snow pillow data and precipitation gauge data are available from the California Data Exchange Center (CDEC; http://cdec.water.ca.gov). The NCEP–NCAR reanalyses data are hosted by NOAA/Earth System Research Laboratory (ESRL; www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html). The North American Regional Reanalysis (NARR) data were provided by the NOAA/OAR/ESRL/PSD, Boulder, Colorado, from their website (www.esrl.noaa.gov/psd/).

REFERENCES

Baldwin, M. E., and K. E. Mitchell, 1996: The NCEP hourly multisensor U.S. precipitation analysis. Preprints, *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., J95–J96.

Beaumont, R. T., 1965: Mt. Hood pressure pillow snow gage. *J. Appl. Meteor.,* **4,** 626–631, doi:10.1175/1520-0450(1965)004<0626:MHPPSG>2.0.CO;2.

California Data Exchange Center, 2014a: California Data Exchange Center—Precipitation. Accessed 10 May 2014. [Available online at http://cdec.water.ca.gov/snow_rain.html.]

——, 2014b: California Data Exchange Center—Snow. Accessed 10 May 2014. [Available online at http://cdec.water.ca.gov/snow/current/snow/index.html.]

Climate Impacts Group, 2014: Western US Hydroclimate Scenarios Project. Accessed 22 May 2014. [Available online at http://cses.washington.edu/cig/data/wus.shtml.]

Cosgrove, B. A., and Coauthors, 2003: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.,* **108,** 8842, doi:10.1029/2002JD003118.

Daly, C., 2013: Descriptions of PRISM spatial climate datasets for the conterminous United States. PRISM Doc., 14 pp. [Available online at http://www.prism.oregonstate.edu/documents/PRISM_datasets_aug2013.pdf.]

——, 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, doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

——, W. P. Gibson, G. H. Taylor, M. K. Doggett, and J. I. Smith, 2007: Observer bias in daily precipitation measurements at United States cooperative network stations. *Bull. Amer. Meteor. Soc.,* **88,** 899–912, doi:10.1175/BAMS-88-6-899.

——, and Coauthors, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.,* **28,** 2031–2064, doi:10.1002/joc.1688.

Dettinger, M., K. Redmond, and D. Cayan, 2004: Winter orographic precipitation ratios in the Sierra Nevada—Large-scale atmospheric circulations and hydrologic consequences. *J. Hydrometeor.,* **5,** 1102–1116, doi:10.1175/JHM-390.1.

Farnes, P. E., 1967: Criteria for determining mountain snow pillow sites. *Proc. 35th Annual Western Snow Conf.,* Boise, Idaho, Western Snow Conference, 59–62. [Available online at http://www.westernsnowconference.org/node/1280.]

Gervais, M., L. B. Tremblay, J. R. Gyakum, and E. Atallah, 2014: Representing extremes in a daily gridded precipitation analysis over the United States: Impacts of station density, resolution, and gridding methods. *J. Climate,* **27,** 5201–5218, doi:10.1175/JCLI-D-13-00319.1.

Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement intercomparison, final report. WMO/TD-872, 212 pp.

Guan, B., N. P. Molotch, D. E. Waliser, E. J. Fetzer, and P. J. Neiman, 2010: Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophys. Res. Lett.,* **37,** L20401, doi:10.1029/2010GL044696.

——, ——, ——, ——, and ——, 2013: The 2010/2011 snow season in California's Sierra Nevada: Role of atmospheric rivers and modes of large-scale variability. *Water Resour. Res.,* **49,** 6731–6743, doi:10.1002/wrcr.20537.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, M. P. 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, doi:10.1175/2011JCLI4109.1.

——, T. Pruitt, M. P. Clark, L. Brekke, J. R. Arnold, D. A. Raff, and R. M. Rasmussen, 2014: An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resour. Res.,* **50,** 7167–7186, doi:10.1002/2014WR015559.

Hamlet, A. F., and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature fields for the continental United States. *J. Hydrometeor.,* **6,** 330–336, doi:10.1175/JHM420.1.

——, and Coauthors, 2010: Final report for the Columbia Basin Climate Change Scenarios Project. Pacific Northwest (PNW) Hydroclimate Scenarios Project 2860, Climate Impacts Group, University of Washington. [Available online at http://warm.atmos.washington.edu/2860/.]

Higgins, R. W., W. Shi, E. Yarosh, and R. Joyce, 2000: *Improved United States Precipitation Quality Control System and Analysis.* NCEP/Climate Prediction Center Atlas 7, 40 pp.

Houze, R. A., 2012: Orographic effects on precipitating clouds. *Rev. Geophys.,* **50,** RG1001, doi:10.1029/2011RG000365.

Johnson, J. B., and D. Marks, 2004: The detection and correction of snow water equivalent pressure sensor errors. *Hydrol. Processes,* **18,** 3513–3525, doi:10.1002/hyp.5795.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Karl, T. R., C. N. Williams Jr., F. T. Quinlan, and T. A. Boden, 1990: United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data. Publ. 3404, Environmental Science Division, Oak Ridge National Laboratory, 389 pp.

Kingsmill, D. E., P. J. Neiman, F. M. Ralph, and A. B. White, 2006: Synoptic and topographic variability of Northern California precipitation characteristics in landfalling winter storms observed during CALJET. *Mon. Wea. Rev.,* **134,** 2072–2094, doi:10.1175/MWR3166.1.

Kirchner, J. W., 2006: Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.,* **42,** W03S04, doi:10.1029/2005WR004362.

Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology,* San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/refs/stage2-4.19hydro.pdf.]

Livneh, B., and Coauthors, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate,* **26,** 9384–9392, doi:10.1175/JCLI-D-12-00508.1.

Lundquist, J. D., P. J. Neiman, B. Martner, A. B. White, D. J. Gottas, and F. M. Ralph, 2008: Rain versus snow in the Sierra Nevada, California: Comparing Doppler profiling radar and surface observations of melting level. *J. Hydrometeor.,* **9,** 194–211, doi:10.1175/2007JHM853.1.

Marks, D., A. Winstral, M. Reba, J. Pomeroy, and M. Kumar, 2013: An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin. *Adv. Water Resour.,* **55,** 98–110, doi:10.1016/j.advwatres.2012.11.012.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States. *J. Climate,* **15,** 3237–3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

Menne, M. J., C. N. Williams, and R. S. Vose, 2009: The U.S. Historical Climatology Network monthly temperature data, version 2. *Bull. Amer. Meteor. Soc.,* **90,** 993–1007, doi:10.1175/2008BAMS2613.1.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.,* **87,** 343–360, doi:10.1175/BAMS-87-3-343.

Minder, J. R., P. W. Mote, and J. D. Lundquist, 2010: Surface temperature lapse rates over complex terrain: lessons from the Cascade Mountains. *J. Geophys. Res.,* **115,** D14122, doi:10.1029/2009JD013493.

Neiman, P. J., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger, 2008: Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the West Coast of North America based on eight years of SSM/I satellite observations. *J. Hydrometeor.,* **9,** 22–47, doi:10.1175/2007JHM855.1.

——, E. M. Sukovich, F. M. Ralph, and M. Hughes, 2010: A seven-year wind profiler-based climatology of the windward barrier jet along California's northern Sierra Nevada. *Mon. Wea. Rev.,* **138,** 1206–1233, doi:10.1175/2009MWR3170.1.

O'Hara, B. F., M. L. Kaplan, and S. J. Underwood, 2009: Synoptic climatological analyses of extreme snowfalls in the Sierra Nevada. *Wea. Forecasting,* **24,** 1610–1624, doi:10.1175/2009WAF2222249.1.

Pan, M., and Coauthors, 2003: Snow process modeling in the North American Land Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water equivalent. *J. Geophys. Res.,* **108,** 8850, doi:10.1029/2003JD003994.

Pandey, G. R., D. R. Cayan, and K. P. Georgakakos, 1999: Precipitation structure in the Sierra Nevada of California during winter. *J. Geophys. Res.,* **104,** 12 019–12 030, doi:10.1029/1999JD900103.

Ralph, F. M., and M. D. Dettinger, 2011: Storms, floods, and the science of atmospheric rivers. *Eos, Trans. Amer. Geophys. Union,* **92,** 265–266, doi:10.1029/2011EO320001.

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, doi:10.1175/BAMS-D-11-00052.1.

Rheinheimer, D. R., S. T. Ligare, and J. H. Viers, 2012: Water-energy sector vulnerability to climate warming in the Sierra Nevada: Simulating the regulated rivers of California's West Slope Sierra Nevada. California Energy Commission Publ. CEC-500-2012-016, 71 pp.

Rittger, K., 2012: Spatial estimates of snow water equivalent in the Sierra Nevada. Ph.D. thesis, Environmental Science and Management, University of California, Santa Barbara, 225 pp.

Rutz, J. J., W. J. Steenburgh, and F. M. Ralph, 2014: Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Mon. Wea. Rev.,* **142,** 905–921, doi:10.1175/MWR-D-13-00168.1.

Salathé, E. P., A. F. Hamlet, C. F. Mass, S.-Y. Lee, M. Stumbaugh, and R. Steed, 2014: Estimates of twenty-first-century flood risk in the Pacific Northwest based on regional climate model simulations. *J. Hydrometeor.,* **15,** 1881–1899, doi:10.1175/JHM-D-13-0137.1.

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. [Available online at https://ams.confex.com/ams/84Annual/techprogram/paper_72159.htm.]

Serreze, M. C., M. P. Clark, and A. Frei, 2001: Characteristics of large snowfall events in the montane western United States as examined using snowpack telemetry (SNOTEL) data. *Water Resour. Res.,* **37,** 675–688, doi:10.1029/2000WR900307.

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

Thornton, P. E., S. W. Running, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.,* **190,** 214–251, doi:10.1016/S0022-1694(96)03128-9.

U.S. Army Corps of Engineers, 1956: *Snow Hydrology: Summary Report of the Snow Investigations.* North Pacific Division, Corps of Engineers, 462 pp.

U.S. Census Bureau, 2014: State and county quick facts. Accessed 12 August 2014. [Available online at http://quickfacts.census.gov/qfd/states/06000.html.]

Waichler, S. R., and M. S. Wigmosta, 2003: Development of hourly meteorological values from daily data and significance to hydrological modeling at H. J. Andrews Experimental Forest. *J. Hydrometeor.,* **4,** 251–263, doi:10.1175/1525-7541(2003)4<251:DOHMVF>2.0.CO;2.

Xia, Y., and Coauthors, 2012a: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.,* **117,** D03109, doi:10.1029/2011JD016048.

——, and Coauthors, 2012b: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.,* **117,** D03110, doi:10.1029/2011JD016051.

Yang, D., D. Kane, Z. Zhang, D. Legates, and B. Goodison, 2005: Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophys. Res. Lett.,* **32,** L19501, doi:10.1029/2005GL024057.