

Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States

ANDREW J. NEWMAN, MARTYN P. CLARK, AND JASON CRAIG

National Center for Atmospheric Research, Boulder, Colorado*

BART NIJSSEN

Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington

ANDREW WOOD, ETHAN GUTMANN, AND NAOKI MIZUKAMI

National Center for Atmospheric Research, Boulder, Colorado*

LEVI BREKKE

Bureau of Reclamation, U.S. Department of Interior, Denver, Colorado

JEFF R. ARNOLD

Institute for Water Resources, U.S. Army Corps of Engineers, Seattle, Washington

(Manuscript received 4 February 2015, in final form 1 July 2015)

ABSTRACT

Gridded precipitation and temperature products are inherently uncertain because of myriad factors, including interpolation from a sparse observation network, measurement representativeness, and measurement errors. Generally uncertainty is not explicitly accounted for in gridded products of precipitation or temperature; if it is represented, it is often included in an ad hoc manner. A lack of quantitative uncertainty estimates for hydrometeorological forcing fields limits the application of advanced data assimilation systems and other tools in land surface and hydrologic modeling. This study develops a gridded, observation-based ensemble of precipitation and temperature at a daily increment for the period 1980–2012 for the conterminous United States, northern Mexico, and southern Canada. This allows for the estimation of precipitation and temperature uncertainty in hydrologic modeling and data assimilation through the use of the ensemble variance. Statistical verification of the ensemble indicates that it has generally good reliability and discrimination of events of various magnitudes but has a slight wet bias for high threshold events (>50 mm). The ensemble mean is similar to other widely used hydrometeorological datasets but with some important differences. The ensemble product produces a more realistic occurrence of precipitation statistics (wet day fraction), which impacts the empirical derivation of other fields used in land surface and hydrologic modeling. In terms of applications, skill in simulations of streamflow in 671 headwater basins is similar to other coarse-resolution datasets. This is the first version, and future work will address temporal correlation of precipitation anomalies, inclusion of other data streams, and examination of topographic lapse rate choices.

1. Introduction

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Andrew J. Newman, NCAR, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: anewman@ucar.edu

Historical quantitative meteorological datasets for use in land surface and hydrologic modeling are available from a multitude of sources for various temporal and spatial extents and spatial resolutions (Huffman et al. 2001, 2009; Maurer et al. 2002; Lin and Mitchell

2005; Hamlet and Lettenmaier 2005; Mesinger et al. 2006; Daly et al. 2008; Haylock et al. 2008; Saha et al. 2010; Abatzoglou 2013; Xia et al. 2012; Livneh et al. 2013; Thornton et al. 2014). These products provide estimates of many different fields ranging from precipitation only (e.g., Lin and Mitchell 2005) to full meteorological fields and surface energy terms (e.g., Xia et al. 2012).

These quantitative precipitation and temperature estimates are inherently uncertain. Gauge-based estimates lack the ability to capture the true spatial variability of precipitation due to low station density, particularly in complex terrain, and often suffer from gauge undercatch (Yang et al. 1998) or other sampling errors (Clark and Slater 2006). Ground-based precipitation radar estimates suffer from sampling uncertainty as a result of terrain blocking, curvature of the earth, beam spreading, and uncertainties in the rain rate–reflectivity relationship (Austin 1987; Westrick et al. 1999; Young et al. 1999; Dinku et al. 2002; Schiemann et al. 2011). Even improved estimates from polarimetric radars are subject to sometimes large uncertainty (Ryzhkov et al. 2003; Cifelli et al. 2011). Estimates of air temperature from station networks have similar causes of uncertainty as gauge-based precipitation estimates; they lack sufficient station density to resolve the spatial variability of temperature, particularly in complex terrain (e.g., Stahl et al. 2006).

In general, the aforementioned datasets are deterministic and do not provide uncertainty estimates or ensemble output for any of the fields estimated. There are some notable exceptions, especially in the remote sensing precipitation estimation community where retrievals are inherently highly uncertain. KIRSTETTER ET AL. (2015) has developed a new probabilistic precipitation methodology for ground-based radar, and many studies have modeled or evaluated errors in satellite rainfall estimates (Morrissey and Green 1998; Krajewski et al. 2000; Huffman et al. 2001, 2009; Bauer et al. 2002; Hossain and Anagnostou 2006; Bellerby 2007; Tian and Peters-Lidard 2007; AghaKouchak et al. 2009; Habib et al. 2009; Tian et al. 2009, 2013; Sapiano and Arkin 2009; Shen et al. 2010; KIRSTETTER ET AL. 2013a,b; Maggioni et al. 2014; Gebregiorgis and Hossain 2015). In one specific example, Krajewski et al. (2000) discuss the uncertainty estimates subsequently used by Huffman et al. (2001, 2009) and note that they are developed in a somewhat ad hoc manner. Additionally, ensemble rainfall products have been developed over the last decade for satellite and ground scanning radar precipitation estimates (Bellerby and Sun 2005; Teo and Grimes 2007; AghaKouchak et al. 2010; Bellerby 2013; Nogueira and Barros 2013; Alemohammad 2014; Greatrex et al. 2014; Skinner et al. 2015). Nogueira and

Barros (2013) use stage IV data to generate 1-km downscaled precipitation ensembles through modified fractal Brownian surfaces (e.g., Bindlish and Barros 1996) in an effort to provide high-resolution information in an ensemble product. However, this product is constrained to the stage IV gridbox mean and does not truly provide an uncertainty estimate. Also, Alemohammad (2014) develops a Bayesian ensemble approach for a single satellite sensor with realistic uncertainty estimates.

Uncertainties in hydrologic model forcing are a major source of hydrologic simulation error (Kavetski et al. 2003; Nijssen and Lettenmaier 2004; Renard et al. 2010; Skinner et al. 2015), and proper uncertainty specification is necessary for robust ensemble prediction. Additionally, the lack of robust uncertainty estimates (if any at all) for observed fields limits the successful application of data assimilation in hydrology (Moradkhani et al. 2005; Clark et al. 2006). Ideally all quantitative meteorological datasets, not just highly uncertain remotely sensed precipitation retrievals, should incorporate uncertainty estimates as an integrated and formalized component of dataset generation that will begin to improve our understanding and characterization of forcing uncertainties.

Here we extend the work of Clark and Slater (2006) to develop a daily, station-based, ensemble dataset of precipitation and temperature at $1/8^{\circ}$ resolution for the contiguous United States (CONUS), northern Mexico, and southern Canada for the period 1980–2012. The dataset is initially intended for use in forcing land surface and hydrologic models and hydrologic data assimilation studies but could also be used for atmospheric model validation studies. The paper is organized as follows. Section 2 summarizes the probabilistic interpolation system, which uses the methods of Clark and Slater (2006) with minor modifications. Next, general dataset performance, comparisons to several widely used deterministic datasets, and dataset availability are described in sections 3–6. Finally, section 7 provides concluding remarks and avenues for future improvements.

2. Probabilistic interpolation

Station data from the Global Historical Climatology Network-Daily dataset (GHCN-Daily; Menne et al. 2012a,b) and the U.S. Natural Resources Conservation Service (NRCS) Snowpack Telemetry (SNOWTEL) observation network were compiled for this analysis. The GHCN-Daily data were supplied by the U.S. National Climatic Data Center (NCDC) and originate from a variety of networks within the United States including the Cooperative Observer (COOP) network; the Community Collaborative Rain, Hail, and Snow (CoCoRaHS) network; and the various automated airport weather

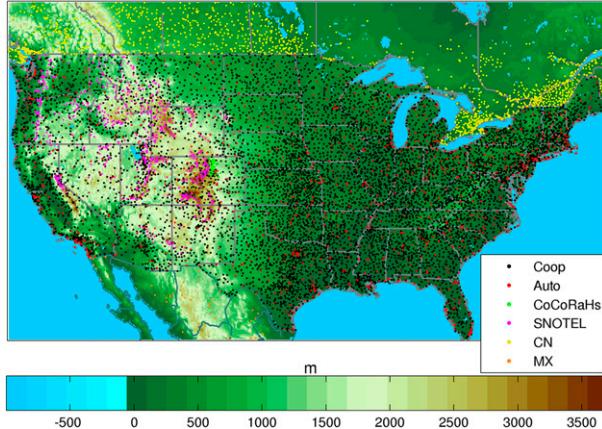


FIG. 1. Domain extent, elevation, and all stations with either precipitation, temperature, or both measurements used in the ensemble generation. Elevation data come from the NLDAS-2, which is derived from GTOPO30 (<https://ita.cr.usgs.gov/GTOPO30>). Auto denotes automatic observation stations (e.g., Automatic Surface Observing System), CN denotes stations in Canada, and MX denotes stations in Mexico.

stations [e.g., Automated Surface Observing System (ASOS)]. Observations from other countries are also in the GHCN-Daily, with observations from Canada and Mexico included in this study (Fig. 1).

a. Spatial interpolation

A serially complete station dataset was used to generate a $1/8^{\circ}$ (~ 12 km) resolution, gridded daily ensemble of precipitation (Prcp_{ens}), daily mean temperature T_{mean} , and diurnal temperature range T_{range} dataset on all nonwater pixels in Fig. 1, which is phase 2 of the North American Land Data Assimilation System (NLDAS-2; Xia et al. 2009, 2012) domain and the grid used here. It covers all river basins in the CONUS (including contributing areas in Canada and Mexico). The T_{range} is defined as the difference between the daily maximum temperature T_{max} and the daily minimum temperature T_{min} . Details of the generation of the serially complete station dataset can be found in the appendix. After all processing, over 12 000 unique stations with precipitation, temperature, or both types of observations were included in the analysis. The 12 153 stations provided precipitation observations and 8953 stations provided temperature observations. The T_{range} was used in this work under the assumption that T_{mean} and the $T_{\text{max}} - T_{\text{min}}$ difference would generate more reliable station-derived lapse rates, and to avoid cases in which $T_{\text{max}} < T_{\text{min}}$. It is straightforward to calculate T_{max} and T_{min} from T_{mean} and T_{range} . The general methodology of Clark and Slater (2006) was followed with the following steps.

- 1) Distance-dependent weights were computed at each grid cell for all stations within a given radius from Clark and Slater (2006):

$$w_{i_{\text{sta}}, i_{\text{sta}}} = \left[1 - \left(\frac{d_{i_{\text{sta}}}}{\text{MAXD}} \right)^3 \right]^3. \quad (1)$$

The weight $w_{i_{\text{sta}}, i_{\text{sta}}}$ is the weight for the current station used to populate diagonal elements in the diagonal weight matrix \mathbf{W} as used in the local regression equations defined in Eqs. (3) and (6); $d_{i_{\text{sta}}}$ is the distance of the current station to the grid point being considered; and MAXD is a specified maximum distance. Stations with distances larger than MAXD received zero weight. In this work, we limited the number of stations considered at each grid point n_{sta} to the 30 closest stations and computed the weights with MAXD set to 100 km if all 30 stations were within 100 km. If there were fewer than 30 stations within 100 km, then MAXD was reset to 1 km larger than the thirtieth station distance and the weights were computed using the new MAXD for that specific grid point. Roughly 90% of the grid points in the CONUS have 30 stations within a 100-km radius. In this work, these differences result in the weight matrix always being set to $n_{\text{sta}} \times n_{\text{sta}}$ (30×30) with no zero-weighted stations.

- 2) Multivariate, locally weighted logistic regression was used to estimate the probability of precipitation (PoP) at each grid point using the vector of precipitation yes/no occurrence from the n_{sta} closest stations, the locally varying weights \mathbf{W} , and spatial attributes (initially latitude, longitude, and elevation, but any number of attributes can be incorporated):

$$\text{PoP}_{i_{\text{grd}}} = \frac{1}{1 + \exp(-\mathbf{Z}_{i_{\text{grd}}} \boldsymbol{\beta})}, \quad (2)$$

where $\mathbf{Z}_{i_{\text{grd}}}$ is the row vector of spatial attributes for the target grid cell and $\boldsymbol{\beta}$ is the column vector of regression coefficients determined iteratively following Loader (1999):

$$\boldsymbol{\beta}_{\text{new}} = \boldsymbol{\beta}_{\text{old}} + (\mathbf{X}^T \mathbf{W} \mathbf{V} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} (\mathbf{P}_{\text{occur}} - \boldsymbol{\pi}), \quad (3)$$

where \mathbf{X} is the design matrix of spatial attributes from the n_{sta} nearest stations, $\mathbf{P}_{\text{occur}}$ is the vector of precipitation occurrence (zero or one), \mathbf{V} is the $n_{\text{sta}} \times n_{\text{sta}}$ variance matrix, and $\boldsymbol{\pi}$ is the n_{sta} vector of estimated PoP at each station. The variable $\boldsymbol{\beta}$ is initialized as a column vector of ones.

- 3) Multivariate, locally weighted linear regression was used to determine the estimate of transformed precipitation at each grid point (PCP) using transformed

nonzero station precipitation \mathbf{P}'_{sta} , the locally varying weights, and spatial attributes. In this paper, the transformation of nonzero precipitation for each time step at each grid point was performed using a simple power-law transformation following

$$\mathbf{P}'_{\text{sta}_{\text{grid}}} = (\mathbf{P}_{\text{sta}_{\text{grid}}})^{1/\text{Tran}}, \quad (4)$$

where \mathbf{P}_{sta} is the vector original quantity (nonzero precipitation), \mathbf{P}'_{sta} is the transformed vector quantity, and Tran is the power-law transformation coefficient and was set to 4.0 after sensitivity testing. Then the PCP at a given grid cell is given as

$$\text{PCP}_{i_{\text{grid}}} = \mathbf{Z}_{i_{\text{grid}}} \boldsymbol{\beta}^a, \quad (5)$$

where $\boldsymbol{\beta}^a$ is the vector of regression coefficients estimated by

$$\boldsymbol{\beta}^a = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{P}'_{\text{sta}}, \quad (6)$$

where again \mathbf{X} is the design matrix of station attributes, \mathbf{W} is the weight matrix, and \mathbf{P}'_{sta} is the transformed precipitation amounts.

As noted above, we have limited consideration to the 30 nearest stations. Therefore, all vectors and matrices such as \mathbf{W} , \mathbf{X} , \mathbf{P}'_{sta} , $\mathbf{P}_{\text{occur}}$, etc., are vectors or matrices with a maximum dimension of 30. At any given grid cell the locally weighted logistic and linear regression are applied to that cell using the unique vectors and matrices for the 30 closest stations to that grid cell.

- 4) Uncertainty estimates E for the precipitation amounts were determined at each grid point by summing the regression residuals. This effectively combines spatial representativeness and individual station errors. From [Clark and Slater \(2006\)](#), E at a grid point $E_{i_{\text{grid}}}$ is computed as

$$E_{i_{\text{grid}}} = \left[\frac{\sum_{i_{\text{sta}}=1}^{n_{\text{sta}}} w_{i_{\text{sta}}, i_{\text{sta}}} (\text{PCP}_{i_{\text{sta}}} - \mathbf{P}'_{\text{sta}_{i_{\text{sta}}}})}{\sum_{i_{\text{sta}}=1}^{n_{\text{sta}}} w_{i_{\text{sta}}, i_{\text{sta}}}} \right]^{1/2}. \quad (7)$$

Steps 1, 3, and 4 were then applied to T_{mean} and T_{range} to generate gridpoint estimates of temperature including uncertainty.

Two other modifications to the original [Clark and Slater \(2006\)](#) methodology were made for this paper. [Clark and Slater \(2006\)](#) computed climatological cumulative density functions (CDFs) of monthly precipitation for each station and then transformed the data

into normal space before the regression steps using those CDFs. In this work, this step was dropped and only the power-law transformation was used.

Additionally, this work added topographic slope in the west–east and south–north directions to the regression as a first step toward introducing an automated method to account for windward and leeward slope precipitation differences. Smoothed topography using a 2D Gaussian low-pass filter with a size of 180 km and a standard deviation of 24 km was used to compute the two slope components. This was done to remove the high-frequency noise associated with the full-resolution topography and provide a closer match to precipitation enhancement along mountain ranges, with broad enhanced and decreased precipitation regions on the windward and leeward sides. No consideration for slope directions outside of the cardinal directions was given. For example, mountains with slopes of southwest–northeast direction will have positive slopes in west–east and south–north, and the regression can weight each component independently. Finally, for areas with little topographic slope, the regression equations used only latitude, longitude, and elevation as predictors. The change in sign of slope across mountain ranges allowed the locally weighted regression to account for relationships along slopes and to weight stations with dissimilar slopes properly, similar to the Parameter-Elevation Regressions on Independent Slopes (PRISM) method ([Daly et al. 2008](#)).

The differences in ensemble performance for precipitation between this version and the original stem from those changes in the methodology. Moving away from a climatologically based inverse-quantile transformation removed the assumption of a stationary climate when transforming observed precipitation and slightly improved computational efficiency. Additionally, limiting the regression to consider the 30 closest stations and recomputing weights where necessary slightly changed the character of the locally weighted regression. The move to a time-step-based, power-law transformation prior to the regression steps rather than building climatologically based CDFs and performing an inverse-quantile transform means that the transformed precipitation distributions (i.e., \mathbf{P}'_{sta}) are less Gaussian in some cases.

b. Ensemble generation

Each ensemble member was generated by using spatially correlated random fields (SCRFs) sampled from the standard normal distribution ([Clark and Slater 2006](#)). The SCRFs are generated using a conditional simulation approach ([Johnson 1987](#)). This method generates random numbers progressively for points on the grid, conditioning the new random numbers on the

previously generated values. It allows for spatial correlation structure using a spatial correlation length:

$$c_{ij} = c_o \exp\left(-\frac{d_{ij}}{C_{\text{len}}}\right), \quad (8)$$

where c_{ij} is the correlation of the current grid point, the initial correlation value c_o and spatial correlation length C_{len} are estimated from the station data themselves, and d_{ij} is the distance (km) between the i th and j th grid points. A nested grid approach following [Fang and Tacher \(2003\)](#) was implemented by [Clark and Slater \(2006\)](#) and also used here to increase the computational efficiency of the [Johnson \(1987\)](#) method.

An SCRF was generated independently for each day for each of the estimated variables: precipitation, mean temperature, and diurnal range. Precipitation occurrence at each grid point was determined using the regression PoP and the cumulative probability (CP) of the precipitation SCRF. Precipitation occurred at a grid point if CP was greater than PoP (step 2). If precipitation was deemed to occur, the precipitation amount was determined by first rescaling CP by

$$\text{CS}_{i_{\text{grd}}} = \frac{\text{CP}_{i_{\text{grd}}} - (1 - \text{PoP}_{i_{\text{grd}}})}{\text{PoP}_{i_{\text{grd}}}}, \quad (9)$$

where CS is the rescaled CP of the SCRF. Then, the corresponding standard normal deviate of CS was found (RN) and the transformed precipitation amount was determined as

$$\text{Prcp}'_{\text{ens},i_{\text{grd}}} = \text{PCP}_{i_{\text{grd}}} + \text{RN}_{P,i_{\text{grd}}} (E_{P,i_{\text{grd}}}), \quad (10)$$

where $\text{Prcp}'_{\text{ens}}$ is the power-law transformed precipitation amount for the current ensemble member at the current grid point, $\text{RN}_{P,i_{\text{grd}}}$ is the gridpoint normal deviate for the precipitation SCRF, and again PCP is the regression-estimated transformed precipitation gridpoint amount [step 3] and E is the gridpoint uncertainty estimate [step 4, Eq. (7)]. Finally, the actual precipitation amount (Prcp_{ens}) for the current ensemble member and grid point is

$$\text{Prcp}_{\text{ens},i_{\text{grd}}} = (\text{Prcp}'_{\text{ens},i_{\text{grd}}})^{\text{Tran}}. \quad (11)$$

Similarly, the mean temperatures and diurnal temperature ranges were modified by their respective temperature error terms and SCRFs (directly without the power-law transformation or rescaling of the SCRF):

$$T_{\text{ens}} = T + \text{RN}_{T,i_{\text{grd}}} (E_{T,i_{\text{grd}}}), \quad (12)$$

where T_{ens} is either T_{mean} or T_{range} for the current ensemble member at the current grid point, T is the

regression gridpoint temperature, $\text{RN}_{T,i_{\text{grd}}}$ is the gridpoint normal deviate for the SCRF of T_{mean} or T_{range} , and $E_{T,i_{\text{grd}}}$ is the gridpoint uncertainty estimate for either temperature variable.

Spatial correlation lengths (i.e., C_{len}) used in the generation of the SCRFs are determined from the station data for each meteorological season separately [December–February (DJF; winter), March–May (spring), June–August (JJA; summer), and September–November (fall)]. Temporal autocorrelation, determined from the station data, is included in the SCRFs for temperature using the temperature serial autocorrelation with a 1-day lag via

$$\text{SCRF}_t = \text{SCRF}_{t-1}(\rho_{-1}) + \sqrt{1 - \rho_{-1}^2}(\text{SCRF}), \quad (13)$$

where t is the current time step and ρ_{-1} is the lag-1 temperature autocorrelation. The precipitation SCRF is conditioned using Eq. (13) by setting ρ to the mean observed cross correlation between precipitation and the diurnal temperature range as well using the current T_{range} SCRF in place of SCRF_{t-1} . As a result, days with large diurnal temperature ranges are less likely to have precipitation, which is observed in the data since in general nonprecipitation days tend to be days with less cloud cover. This does not necessarily hold in all seasons or all precipitation events and is an area that should be considered for future research. [Figure 2](#) shows example SCRFs for T_{range} for two time steps and precipitation for one. Since temperature has a high autocorrelation, the current and previous T_{range} SCRFs have high spatial similarity. Precipitation and T_{range} are negatively correlated, but only weakly. Thus, in general where the T_{range} SCRF is positive, the precipitation SCRF is more negative, but with less spatial similarity.

An important result of using SCRFs to generate gridpoint values is that this allows for gridpoint events at locations without observations that are more extreme (e.g., more precipitation) than any of the observations (plus the topographic correction). This is an important distinction from any interpolation-based schemes, in which gridpoint events are limited to the maximum observed value (plus topographic correction). This should improve the representation of extreme events in the ensemble ([Gervais et al. 2014](#); [Gutmann et al. 2014](#)).

3. Ensemble characteristics

Precipitation statistics using leave-one-out cross validation were computed over 1980–2012 for roughly half (about 5500) of the 12 000+ stations to assess the behavior and characteristics of the ensemble. The results are summarized using reliability and discrimination diagrams. These diagrams are efficient summaries of

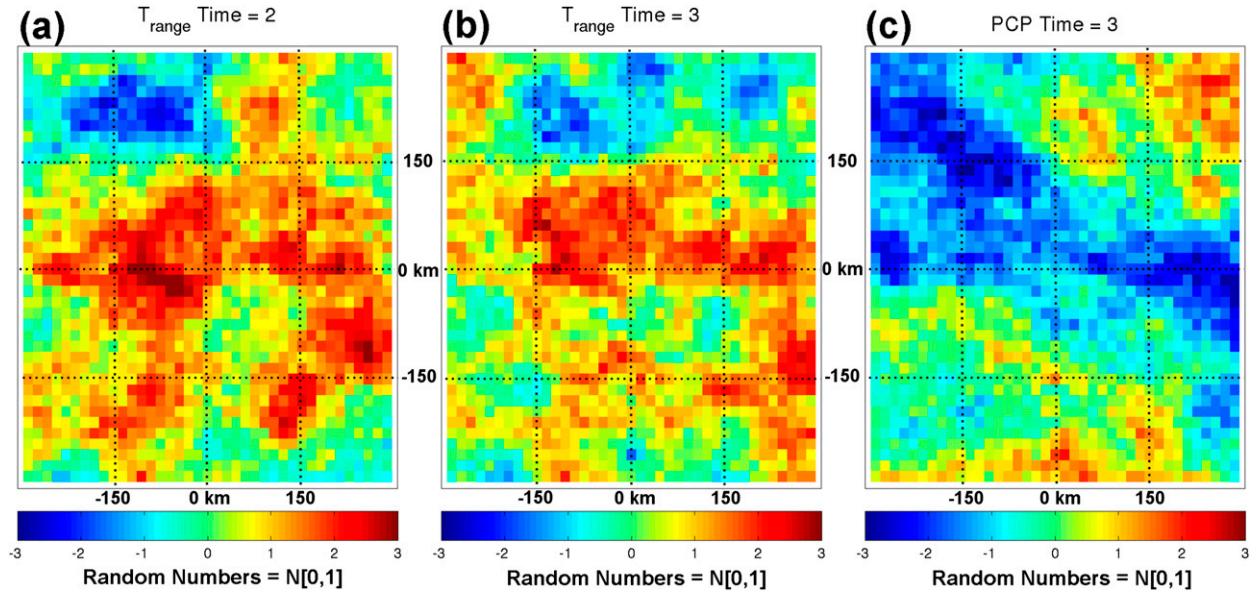


FIG. 2. (a),(b) Example temperature spatially correlated random fields for two time steps and (c) the corresponding precipitation spatially correlated random field for the same time step as (b). Note that spatial correlation lengths in this example do not match the actual ensemble product.

probabilistic (ensemble) skill (Wilks 2006). Reliability diagrams display a calibration function of the ensemble, which is a line plot of the conditional probability of an event (precipitation above a threshold in this case) as a function of the estimated probability (Wilks 2006), the ensemble probability in this case. A perfectly reliable ensemble would have the same estimated probability for every observed probability for every event threshold (the calibration function will follow the 1–1 line). A dry bias (underestimation of events) will find the calibration function above the 1–1 line while a wet bias (overestimation of events) will find the calibration below the 1–1 line.

Discrimination diagrams display the distribution of the observed conditional likelihood for events and nonevents versus the ensemble estimated probability. Good discrimination between events and nonevents is seen through maximization in separation of the two likelihood distributions. In the limit, perfect nonevent distribution likelihoods should be maximized at zero probability and event distribution likelihoods should be maximized at a probability of 1 (Wilks 2006).

a. Precipitation statistics

Discrimination diagrams (Wilks 2006; Figs. 3a–d) indicate that the ensemble estimates a high probability of event occurrence when there is an observed event greater than 0 mm, and has good discrimination between events and nonevents, seen through the large separation of the event and nonevent probability distributions. For larger

precipitation event thresholds, the ensemble is able to assign high probabilities of occurrence, except for the highest probabilities of occurrence for events >50 mm (Fig. 3d), and displays slightly less overlap in the two distributions for the highest threshold compared to Clark and Slater (2006).

Reliability diagrams (Wilks 2006; Figs. 3e–h) show that the ensemble has good reliability for all precipitation events (threshold >0 mm, Fig. 3e). As the precipitation event threshold is increased (Figs. 3f–h), a very slight wet bias at low probabilities of occurrence develops along with a slight dry bias for higher probabilities (>0.5), resulting in a slightly underconfident ensemble for probabilities between ~0.2 and 0.8. For the larger events, a slight wet bias is again present with a now-reduced dry bias for the higher probabilities of occurrence (Fig. 3g), with a slight wet bias at the highest probability. At the 50-mm threshold a persistent slight wet bias is present along with increased sampling uncertainty (Fig. 3h). This performance is quite similar to Clark and Slater (2006) for the lowest threshold and in general comparable at the higher event thresholds, except the consistent small dry bias for larger thresholds and higher probabilities of occurrence. Clark and Slater (2006) note good calibration for all precipitation events with some deviations (both wet and dry biases) as the precipitation threshold is increased.

Figure 4 gives the difference between the ensemble and observations for a random subset of stations across CONUS from the leave-one-out cross-validation station

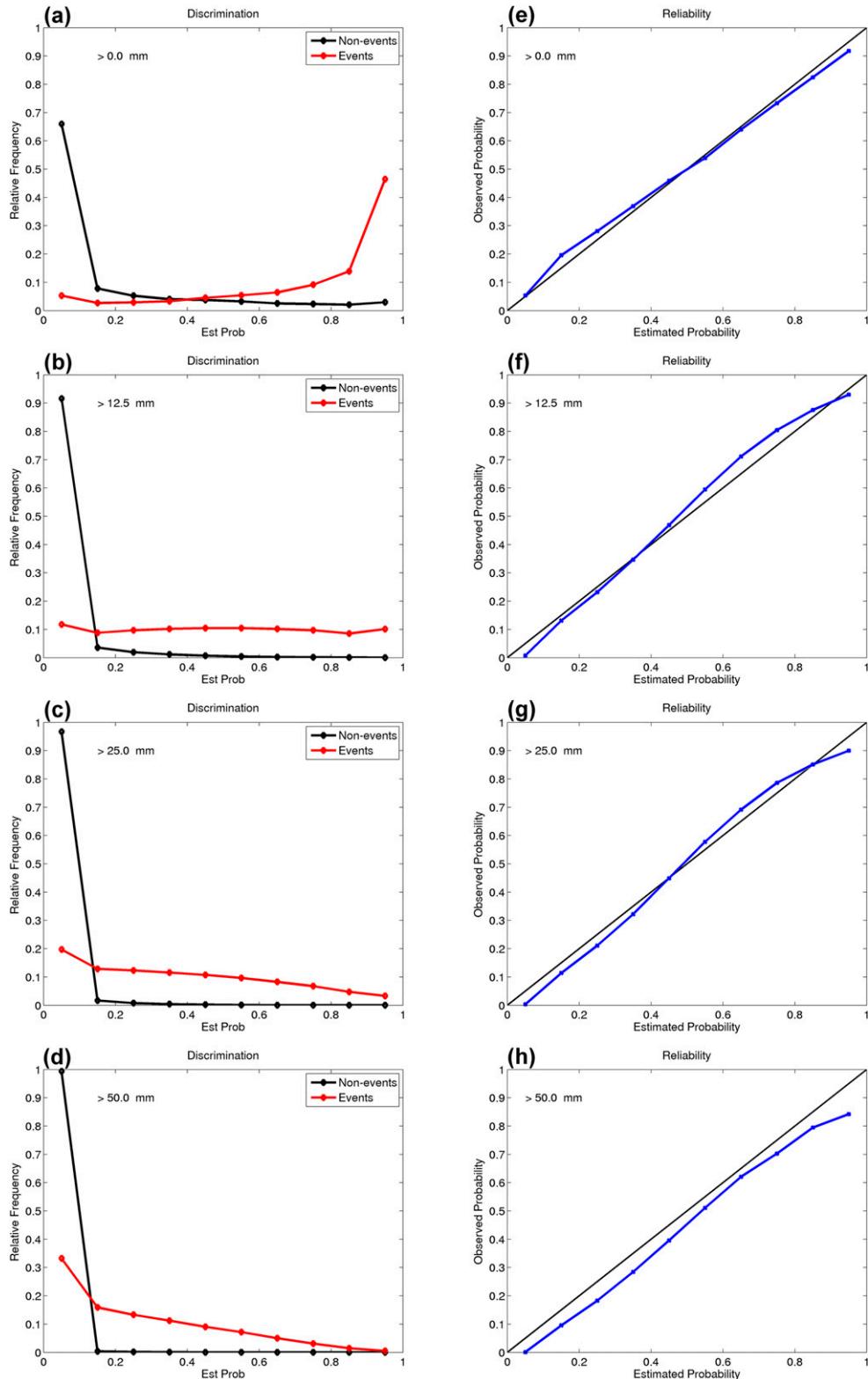


FIG. 3. (a)–(d) Discrimination and (e)–(h) reliability diagrams for precipitation event thresholds of >0.0 , >12.7 , >25.4 , and >50.0 mm.

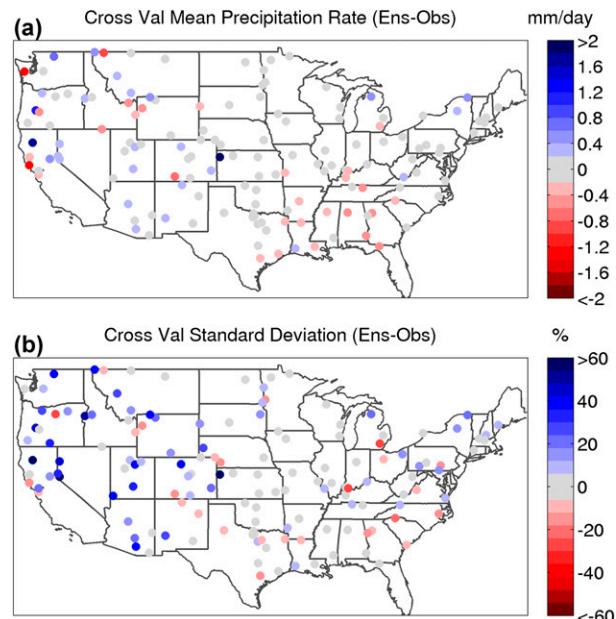


FIG. 4. Randomly selected subset of stations across CONUS from the leave-one-out cross validation in Fig. 3. (a) Mean precipitation difference (mm day^{-1}) and (b) time series std dev difference (%).

set. The ensemble tends to slightly overestimate precipitation in the western United States and slightly underestimate it in the southeastern United States (Fig. 4a). The ensemble also slightly overestimates the standard deviation of the station time series across the western United States with less of a tendency to underestimate station standard deviation across the eastern and southeastern United States (Fig. 4b).

b. Spatial example

Monthly precipitation for June 1993 for two ensemble members is shown in Figs. 5a and 5b. Overall, the spatial patterns are very consistent between members, with little to no precipitation in the southwestern United States and two areas of higher precipitation totals along the U.S. Gulf of Mexico coastline and the central United States. A more detailed examination of the two members in the high accumulation zones shows that ensemble member 75 (Fig. 5b) has more precipitation in parts of the central United States, while ensemble member 11 (Fig. 5a) has generally higher accumulations along the coastline. Comparisons to the ensemble mean field (Fig. 5c) highlight the more discrete nature of the individual members with areas of deviations from the ensemble mean. Finally, the ensemble standard deviation highlights areas of higher or lower uncertainty (Fig. 5d). The two high-precipitation areas generally have similar monthly accumulations (~ 200 – 300 mm), but the Gulf Coast region has standard deviations as large as 70–80 mm ($\sim 30\%$), while the central United States generally has ensemble standard deviations of 25–50 mm ($\sim 10\%$ – 20%). This uncertainty information available through the ensemble will enable land data assimilation schemes to more properly account for the forcing data uncertainty in time and space.

4. Comparisons to deterministic forcing datasets

Ensemble mean spatial output is compared to three widely used deterministic forcing datasets: NLDAS-2

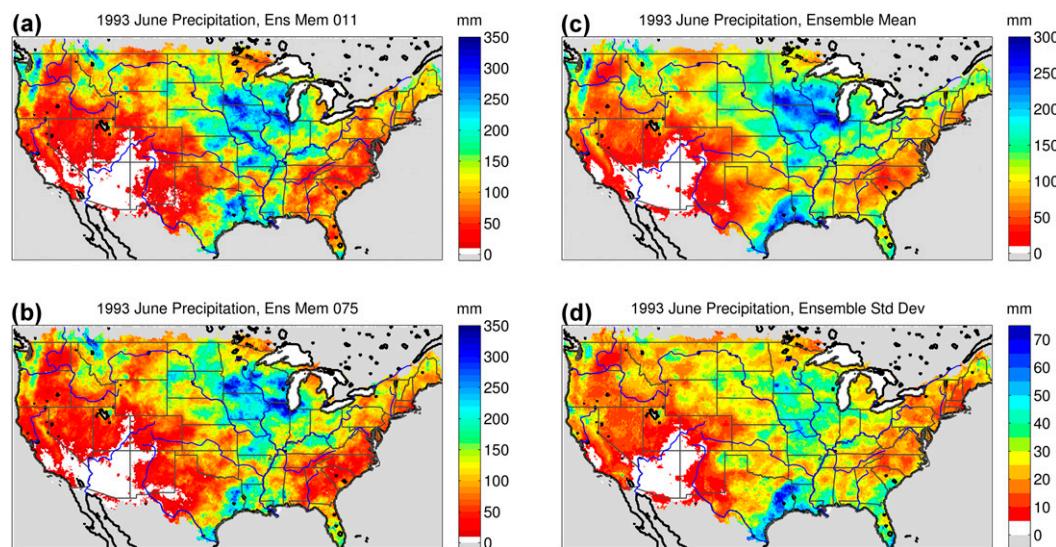


FIG. 5. (a),(b) Example monthly precipitation totals for June 1993 from two ensemble members, along with the monthly ensemble (c) mean and (d) std dev.

TABLE 1. Summary of basic methodological choices related to the ensemble dataset and the three deterministic datasets used for comparison.

Characteristic	Dataset			
	Ensemble	Maurer	NLDAS	Daymet
Native resolution	1/8°	1/8°	1/8°	1 km
Time step	Daily	Daily	Hourly	Daily
Precipitation station data	GHCN-Daily, SNOTEL	NCDC summary of the day	CPC daily analysis (GHCN-Daily)	GHCN-Daily, SNOTEL
Missing data filling (precipitation)	Create static, serially complete station dataset	None, uses available observations	None, uses available observations	None, uses available observations
Precipitation occurrence prediction	Yes	No	No	Yes
Precipitation interpolation method	Probabilistic interpolation	Inverse distance with station distribution correction	Optimal interpolation	Truncated Gaussian filter
Precipitation lapse rate	Derived from stations	PRISM	PRISM	Derived from stations
Temperature input data	GHCN-Daily, SNOTEL	GHCN-Daily	NARR 32 km	GHCN-Daily, SNOTEL
Temperature lapse rate	Derived from stations	Constant -6.5 K km^{-1}	Constant -6.5 K km^{-1} (using delta of NARR topography to 12-km topography)	Derived from stations

(Xia et al. 2012), Daymet (Thornton et al. 2014), and Maurer et al. (2002, 2013; hereafter termed “Maurer”). These deterministic datasets all make different methodological choices in their dataset development. A few of the important choices are summarized here and in Table 1. The NLDAS precipitation is generated by disaggregating the Climate Prediction Center (CPC) daily 1/8° precipitation analysis using CPC hourly gauge data, U.S. radar data, satellite precipitation products, and North American Regional Reanalysis (NARR) precipitation data. Note that the CPC 1/8° daily precipitation analysis uses PRISM (Daly et al. 1994, 2008) to account for topographic effects. Daymet is produced at daily time steps using observational data and only uses the observations themselves to estimate the precipitation and temperature topographic corrections (Thornton et al. 1997, 2014) and includes SNOTEL data. Finally, the Maurer dataset uses bilinear interpolation between observations, gridcell precipitation is scaled by the PRISM 1961–90 climate mean, and temperature is adjusted for elevation using a constant temperature lapse rate of -6.5 K km^{-1} . All three datasets use all available observations for a given day; they do not create a static, serially complete station dataset before any spatial interpolation is performed.

As noted above, the ensemble uses GHCN-Daily and SNOTEL observations, determines the precipitation and temperature lapse rates each day from the station data, and uses a static serially complete station network to create the gridded fields (Table 1). The various methodological choices each dataset

makes will introduce differences in the respective final products. We stress the term differences here, as the “truth” is unknown, especially in the sparsely observed complex terrain of the western United States. However, all three datasets have been widely used to force land surface or hydrologic models and have been shown to allow for realistic hydrological and land surface simulations.

a. Precipitation across datasets

The mean daily precipitation fields across the four datasets are very similar, with all products capturing the increased precipitation in the southeastern United States, arid regions farther west, and areas of enhanced precipitation corresponding to mountainous regions in the western United States and the Pacific Northwest (Fig. 6). Differences between the four products are generally small with a few noteworthy regions of consistently positive or negative differences, almost exclusively confined to regions with complex terrain. In general, the ensemble mean has more precipitation over most of the Intermountain West than Maurer (Fig. 7a), except in select locations along the U.S. West Coast. The ensemble mean is also slightly less than Maurer throughout most of the southeastern United States. NLDAS has similar or less precipitation than Maurer over most of the United States, with only a few regions having positive differences (Fig. 7b), while Daymet generally has more precipitation than Maurer (Fig. 7c) or than the ensemble mean (Fig. 7e). However, the ensemble product may still have too much lee-side precipitation as compared to the other

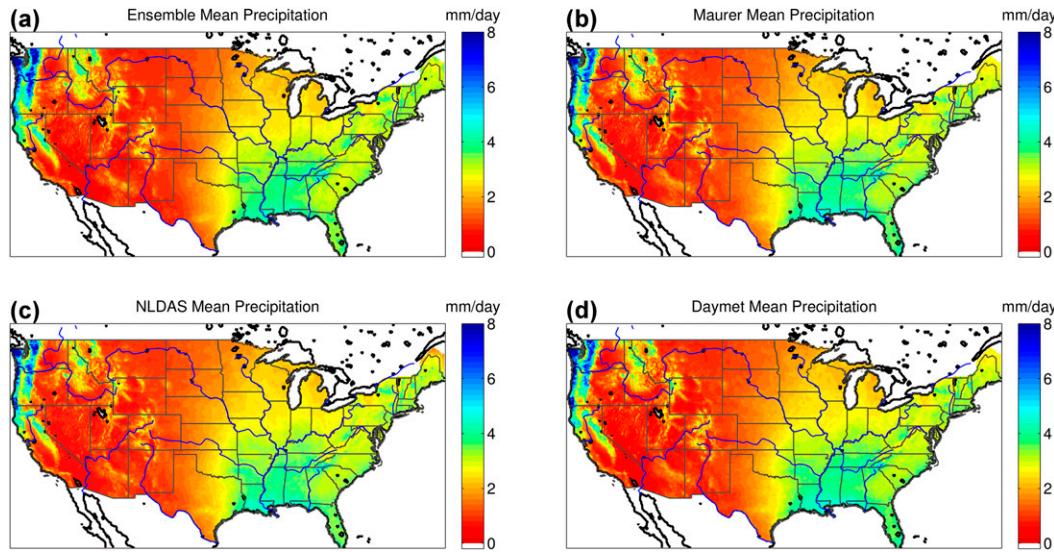


FIG. 6. Mean daily precipitation rate from the (a) ensemble, (b) Maurer, (c) NLDAS, and (d) Daymet datasets.

products, even with the inclusion of the topographic slope in the regression.

Figure 8a shows the frequency of precipitation differences between Maurer and the other three datasets. Daymet is wettest compared to Maurer, while NLDAS and Maurer are most similar because of their similar

datasets and PRISM corrections. The ensemble minus Maurer differences have the most spread, which could be expected considering the ensemble is the most distinctly different dataset (Table 1). Overall, the ensemble differences mostly lie within the other interdataset differences (Fig. 8b), with the largest differences occurring

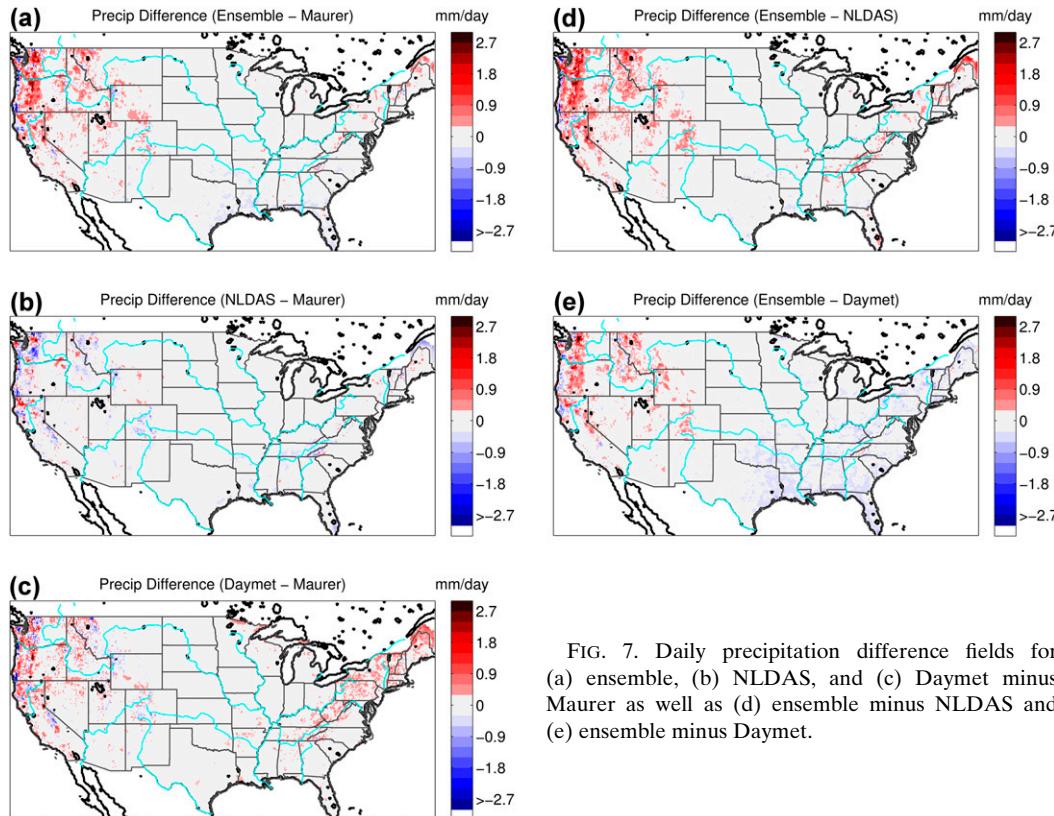


FIG. 7. Daily precipitation difference fields for (a) ensemble, (b) NLDAS, and (c) Daymet minus Maurer as well as (d) ensemble minus NLDAS and (e) ensemble minus Daymet.

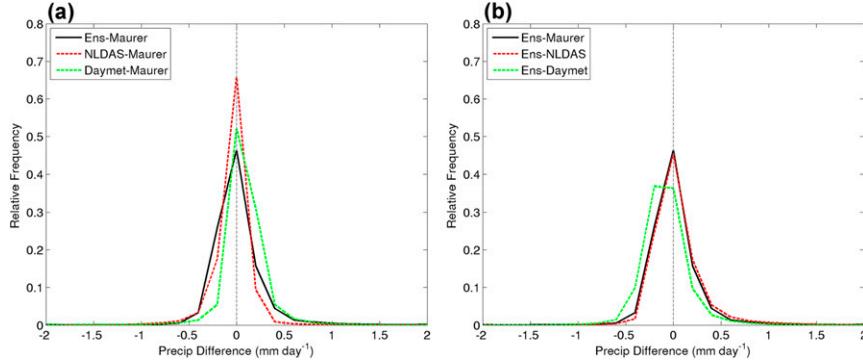


FIG. 8. PDFs of the long-term mean precipitation differences from the (a) ensemble, NLDAS, and Daymet minus Maurer and (b) ensemble minus Maurer, NLDAS, and Daymet.

in complex terrain where the largest uncertainty resides (Fig. 7).

The differences across the datasets arise from the inclusion of different stations, topographic correction differences, and rescaling to climate normals (Table 1). Maurer and Daymet do not include station filling as a preprocessing step as done in the ensemble, choosing to use only available station data for each day (with Maurer never including SNOTEL), while NLDAS uses the daily $1/8^{\circ}$ gridded CPC analysis, again from available data only. Maurer and NLDAS use PRISM topographic precipitation corrections while Daymet and the ensemble develop their own topographic corrections from the data. Additionally, the Maurer dataset is scaled to match the PRISM (Daly et al. 1994) climate mean state for the period 1961–90 while the other datasets are not (Table 1).

b. Focused comparisons

For the rest of the paper we focus only on the Maurer dataset since it is very widely used and is developed at the same grid resolution as the ensemble product from similar input datasets.

1) SEASONAL PRECIPITATION

Segregating precipitation differences by season (Fig. 9) highlights larger differences across seasons than in the annual mean (Fig. 7a vs Fig. 9). In the Intermountain West, the absolute differences are largest during the high-precipitation winter season and smallest during the dry summer season. The ensemble is generally slightly drier than Maurer during the summer season across the eastern two-thirds of the United States, while it is similar or slightly wetter during the winter season. Figure 9 highlights that the ensemble product gives similar estimates to Maurer during large-scale winter precipitation over the eastern United States, but slightly lower mean precipitation rates than Maurer during the summer when precipitation is more intermittent and small scale. A general slight

underestimate of precipitation was indicated in Fig. 4, and Fig. 9 highlights that a majority of the underestimation in the ensemble likely occurs during the summer. This likely stems from the use of a regression algorithm, while Maurer effectively uses bilinear interpolation [synergraphic mapping system (SYMAP) algorithm (Shepard 1968, 1984)]. Highly localized events may be underpredicted to a greater extent when using regression versus interpolation. Future improvements of the ensemble product will work to address this potential issue.

2) WET DAY FRACTION

Wet day fraction can be a substantial source of uncertainty in hydrologic model forcing data when the other

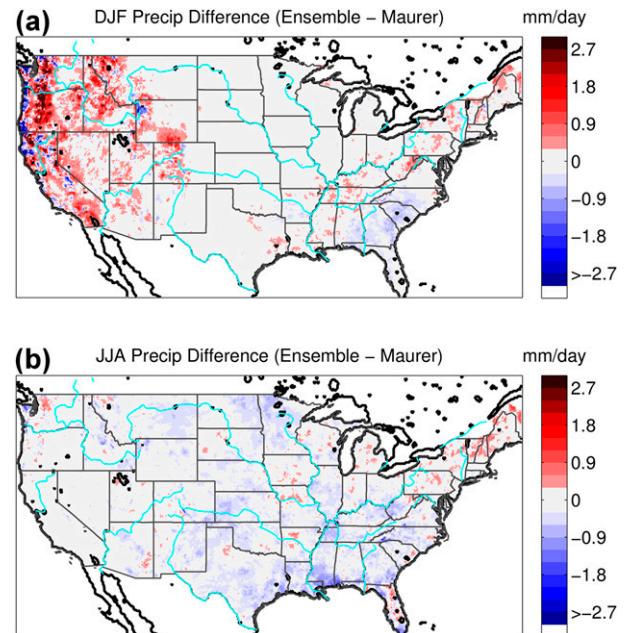


FIG. 9. Ensemble mean minus Maurer daily precipitation difference for only days during (a) DJF and (b) JJA.

meteorological variables (i.e., humidity and radiative fluxes) are derived from precipitation and temperature data with an empirical algorithm such as the mountain microclimate simulation model (MTCLIM; [Hungerford et al. 1989](#); [Bohn et al. 2013](#)). In MTCLIM, shortwave radiation estimation can be affected greatly by wet day fraction because shortwave radiation is reduced by 25% on a rain day as compared to a no-rain day.

The long-term wet day fraction (WF) for the Maurer data and each ensemble member is determined by summing all days with nonzero precipitation and dividing by the total number of days. For the ensemble, the mean of the 100 ensemble members is taken to produce an ensemble mean WF, which is compared to the WF from the Maurer dataset. Both datasets show higher WF in the northeastern United States as well as the high terrain of the western United States and in the Pacific Northwest ([Figs. 10a,b](#)). The ensemble system generally has a lower WF ([Fig. 10c](#)) than Maurer. This is specifically for WF, where linear interpolation increases the number of days with precipitation for grid points between stations when either end point has nonzero precipitation. However, locally the ensemble system has a higher WF in the mountainous regions (e.g., Colorado Rocky Mountains) because of the inclusion of SNOTEL station data and the use of SCRFs in the spatial interpolation, neither of which were used in construction of the Maurer dataset ([Fig. 10c](#)). The overall lower WF in the ensemble system matches the findings in [Gutmann et al. \(2014\)](#), who argued that gridded precipitation products such as Maurer often have WF values that are too high on grid points between stations.

Seasonal WF differences show that the ensemble generally has less summer WF than Maurer, while winter WF differences are generally smaller, except in the Intermountain West ([Figs. 10d,e](#)). The ensemble consistently has more WF than Maurer in the northeastern United States ([Figs. 10c–e](#)). This is likely attributable to the differences in the input station data. Also, the Maurer dataset has a noticeably noisier WF “speckle pattern” than the ensemble product ([Figs. 10a,b](#)), with the effect of the individual stations and their radius of influence more readily apparent because of the use of the synergistic mapping system algorithm ([Shepard 1968, 1984](#)) in the Maurer dataset. The ensemble mean WF is smoother because of the averaging of the 100 realizations.

3) TEMPERATURE

Annual mean temperature differences between the two datasets are generally minimal across the CONUS, except in regions of complex terrain, particularly the high terrain of the western United States. In these regions, the ensemble dataset has higher mean temperatures at high elevations and slightly lower mean temperatures at

lower elevations than the Maurer dataset ([Figs. 10f–h](#)). This dichotomy is even more prevalent in DJF ([Fig. 10i](#)) than in JJA ([Fig. 10j](#)) or in the annual mean, with some mountain ranges having a DJF mean temperature up to 5 K higher than in the Maurer dataset. In most of the Intermountain West, the ensemble mean annual temperature is generally 1–3 K higher than the Maurer mean annual temperature.

This agrees with [Mizukami et al. \(2014\)](#), who found that the Maurer dataset is consistently colder than NLDAS (which does use a constant temperature lapse rate, but incorporates temperature data from the NARR, which is a moderately high-resolution reanalysis product, [Table 1](#)), especially at higher elevations. The differences in temperatures in mountainous regions are due to the different treatments of the temperature lapse rate. Again, Maurer assumes a constant temperature lapse rate, whereas in the ensemble the lapse rate is determined from the input data and allowed to vary by day and grid cell ([Table 1](#)). This indicates that the -6.5 K Km^{-1} lapse rate is generally too large in the winter but more realistic in the summer (e.g., [Stahl et al. 2006](#)). [Oyler et al. \(2015\)](#) have shown there are temperature biases in SNOTEL data that can be traced to sensor changes in the network, especially in DJF, even after basic quality control, which may be enhancing the difference between the two methods in the Intermountain West. Since we have not homogenized the input datasets, this spurious trend (and others) will be present in the ensemble output; thus, it is not recommended to use this dataset for climate trend analysis.

5. Example applications

a. Performance in a hydrologic modeling system

To provide an additional comparison to deterministic forcing datasets, the ensemble mean forcing was run through a hydrologic modeling system consisting of the SNOW-17 temperature index snow model ([Anderson 1973](#)) and the Sacramento soil moisture accounting (SAC-SMA) conceptual hydrologic model ([Burnash et al. 1973](#)). The hydrologic modeling system was optimized using the shuffled complex evolution (SCE) global optimization routine ([Duan et al. 1992](#)) for 671 basins ([Newman et al. 2014, 2015a](#)). [Figure 11](#) gives the results of the calibration and validation periods for the Daymet, NLDAS, Maurer, and ensemble mean forcing in terms of the Nash–Sutcliffe efficiency (NSE; [Nash and Sutcliffe 1970](#)). It can be seen that the CDFs for the 1-km dataset (Daymet) have consistently higher NSE values, while the various 12-km resolution datasets (NLDAS, Maurer, and ensemble) have essentially indistinguishable calibration and validation period performance over

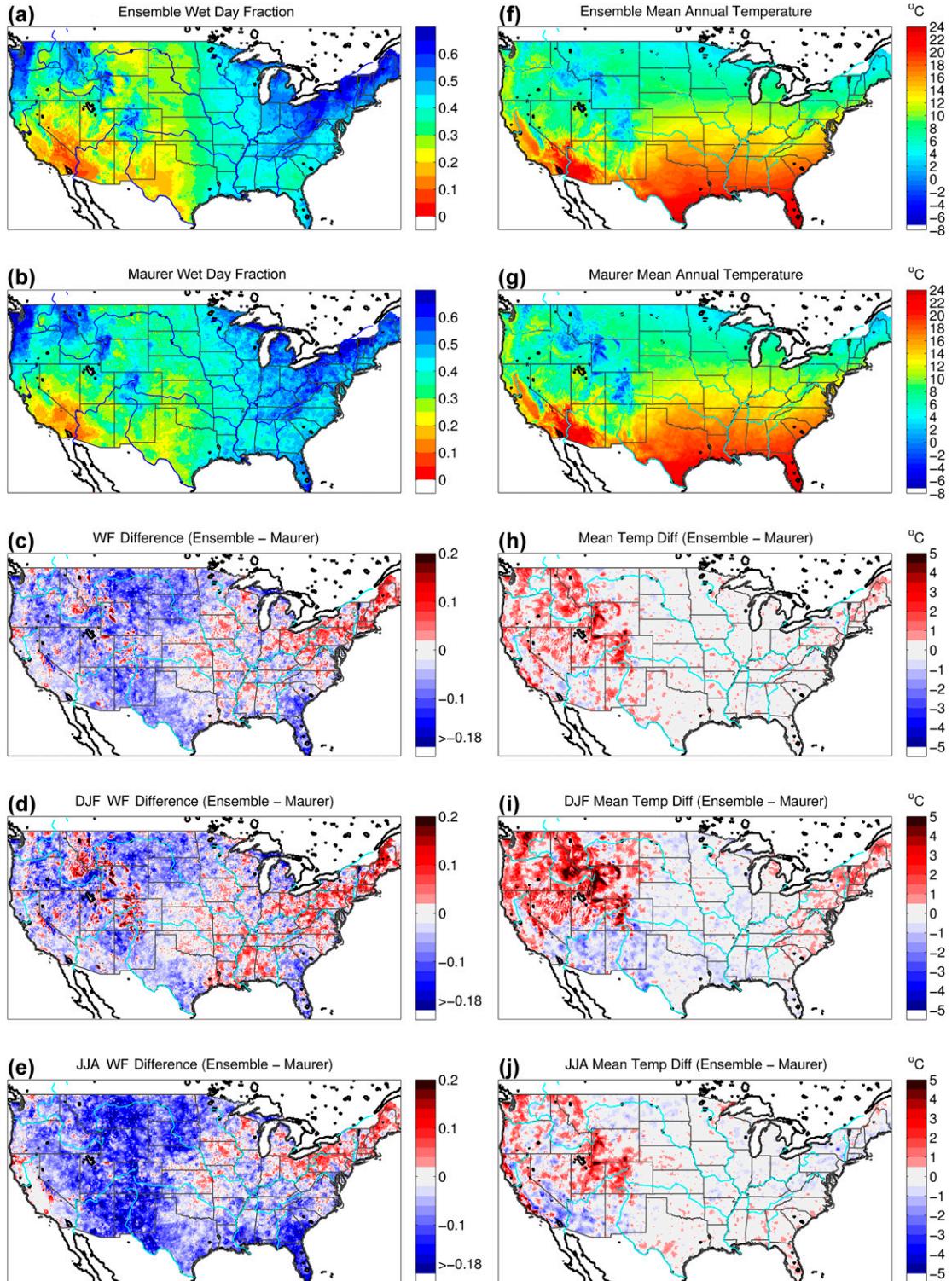


FIG. 10. (a)–(e) Ensemble and Maurer long-term WF and long-term, DJF, and JJA dates only difference fields. (f)–(j) Ensemble and Maurer long-term T_{mean} and long-term, DJF, and JJA dates only difference fields.

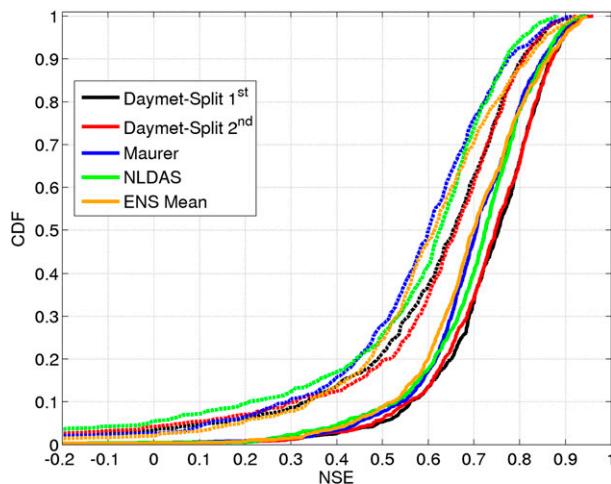


FIG. 11. Calibration (solid) and validation (dashed) NSE CDFs for 671 basins described in Newman et al. (2014, 2015a).

this large basin set. A note of caution is that the ensemble mean dataset has a very high wet day fraction. This has little consequence in the conceptual model used here, but it does impact energy balance models such as VIC (see Mizukami et al. 2015). Calibration using ensembles such as this can build on methods described by Skinner et al. (2015).

b. Example basin simulation

An ensemble forcing dataset is particularly useful for estimating forcing uncertainties as discussed previously, but it is also useful for estimating model state uncertainties by providing many equally plausible spinup periods for a land surface or hydrologic model. Here we provide a very brief example application (Fig. 12) of the ensemble dataset to the Crystal River in central Colorado using a very simple temperature index snow model (SNOW-17) and conceptual hydrologic model (SAC-SMA). Basin mean forcing data generation methodology and model parameters were taken from Newman et al. (2014, 2015a). The ensemble spread nearly always envelopes the three deterministic datasets for daily mean temperature (Fig. 12a); however, the ensemble mean climatology is slightly warmer than all three deterministic datasets. Occasionally, the Maurer dataset lies below the coldest ensemble member, particularly in winter, because of the use of the constant temperature lapse rate. While the ensemble characterizes uncertainty given its methodological choices, a different set of methodological choices (e.g., constant temperature lapse rate as in Maurer) can create estimates that lie outside the range of the ensemble. Further examination of uncertainty in the context of methodological choices is outside the scope of this initial ensemble dataset.

For modeled snow water equivalent (SWE), the deterministic datasets are again nearly always enveloped within the ensemble spread. In this case, the comparison is less straightforward because the temperature index model parameters were optimized using daily streamflow calculated with the Daymet forcing dataset (Fig. 12b). The model includes a frozen precipitation gauge correction factor (~ 2 in this case) that amplifies precipitation differences. For reference, the ensemble 1 April SWE variability is roughly 10% of the ensemble 1 April climatology in this example.

6. Dataset availability

The final dataset is freely available online (<http://dx.doi.org/10.5065/D6TH8JR2>; see <http://ral.ucar.edu/projects/hap/flowpredict/subpages/pqpe.php> for more information; Newman et al. 2015b). The 100-member ensemble for the $1/8^{\circ}$ grid (Fig. 1) includes daily precipitation, T_{mean} , and T_{range} and is provided in climate and forecast (CF) compliant netCDF format. It spans the entire NLDAS-2 grid and covers the time period from 1 January 1980 through 31 December 2012.

7. Summary

This paper addresses the need for uncertainty estimates in observationally derived gridded precipitation and temperature datasets by generating a station-based, ensemble precipitation and temperature dataset on a $1/8^{\circ}$ grid over the CONUS, with coverage extended into northern Mexico and southern Canada to include all watersheds draining into the CONUS. The ensemble has good reliability for low-probability thresholds but is slightly underconfident and has a small dry bias for high probabilities for nearly all event thresholds (Fig. 3). Discrimination of events and nonevent likelihood distributions is generally very good, and the ensemble is able to assign nonzero probabilities to large precipitation events (Fig. 3). Figure 4 indicates that future improvements to the topographic corrections in the western United States may be needed as well as further investigation into the impact of the number of stations included in the locally weighted regression or noise estimates across the eastern United States where a majority of the precipitation is localized and intense. Comparisons to three widely used deterministic products Maurer et al. (2002), NLDAS-2 (Xia et al. 2012), and Daymet (Thornton et al. 2014) find that the general spatial patterns and magnitudes of precipitation are similar across all the datasets and interdataset differences are comparable (Figs. 6–8).

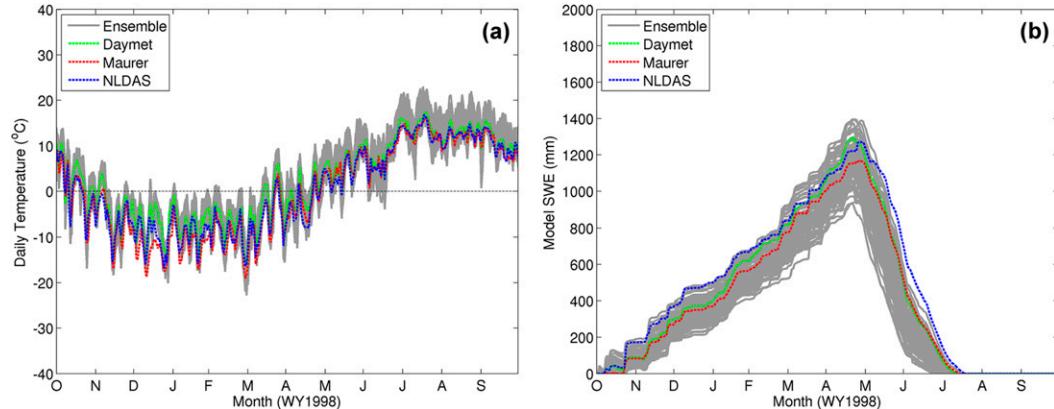


FIG. 12. (a) Crystal River basin mean daily air temperature from all ensemble members, Daymet, Maurer, and NLDAS for water year 1998, and (b) example modeled SWE for the same period.

Focused comparisons of seasonal precipitation (Fig. 9), WF, and temperature with the Maurer dataset show some key differences. Since the ensemble temperature lapse rates are derived from the station data, this dataset has higher T_{mean} values than Maurer et al. (2002), especially in high elevations during winter (DJF; Fig. 10). The higher temperatures may also be due to the inclusion of SNOTEL observations, which may have increased temperatures over the last 20 years (Oyler et al. 2015). Additionally, precipitation is not scaled by the PRISM climatology, and when combined with SNOTEL data, differences exist in the higher terrain of the western United States. Improved WF fields are produced in the ensemble framework, which has consequences in the empirical derivation of other fields used in land surface and hydrologic modeling like net shortwave down radiation (Hungerford et al. 1989; Bohn et al. 2013; Mizukami et al. 2014). The WF field is smoother and generally has fewer days of precipitation throughout the domain, except in a few locations in the eastern United States and Intermountain West (Fig. 10).

We suggest two main avenues of research for these types of products. First, there is substantial scope for additional methodological improvements, such as examination of the underestimation of summer precipitation in the southeastern United States, improvement in topographic corrections, improvement of the representation of the temporal autocorrelation for precipitation, and improvement of the cross correlation between temperature and precipitation. Specifically, precipitation persistence is applied to the anomaly field (precipitation amount), while a more suitable scheme may be to use the temporal persistence in precipitation occurrence. This could be modeled using a Markov chain (Gabriel and Neumann 1962).

The second avenue consists of leveraging additional data sources in a data fusion framework to create

more complete and robust ensemble datasets. With proper processing, radar, satellite, and even atmospheric model precipitation products can be incorporated in this type of ensemble system. Recent radar rainfall work has shown vast improvements and the ability to better quantify uncertainty at each step and produce merged radar–rain gauge analyses on smaller scales (e.g., Schiemann et al. 2011). Finally, more sparse observations of wind and radiation could be combined with short-term deterministic or ensemble model wind and radiation data to produce ensembles containing all necessary fields for energy-balance land surface modeling in a more dynamically consistent manner than currently possible.

Acknowledgments. This work is funded by the Bureau of Reclamation, U.S. Department of the Interior; the U.S. Army Corps of Engineers Climate Preparedness and Resilience Programs; and the National Science Foundation. The authors thank Kyoko Ikeda for her help with various GHCN-Daily processing steps. The NLDAS-2 data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

APPENDIX

Station Data Processing

a. Quality control

Extensive quality-control procedures are performed on the GHCN-Daily data (Durre et al. 2008, 2010). These procedures include spatial consistency checks, multiday temperature streak checks, climatological outlier checks, and others described in detail in

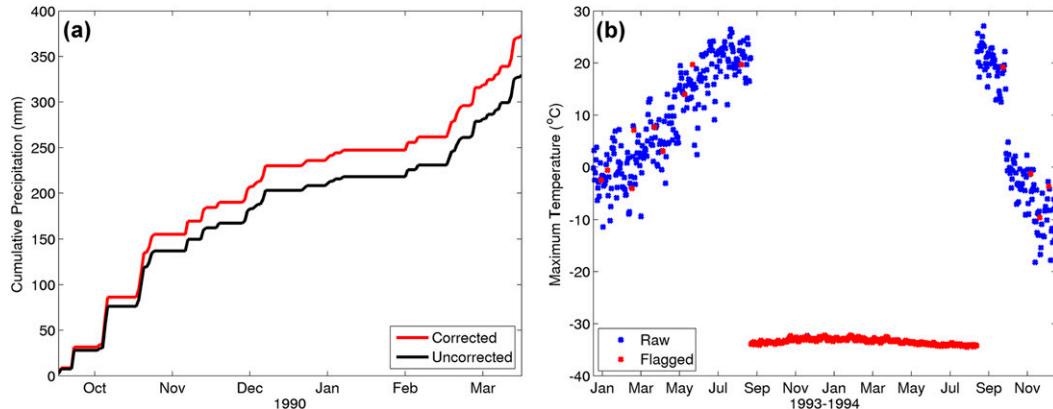


FIG. A1. Example SNOTEL (a) winter precipitation undercatch correction and (b) temperature quality control.

Durre et al. (2008). However, potential errors with the data still exist, including systematic precipitation gauge undercatch during snowfall (Goodison et al. 1998; Yang et al. 1998) or changes in station equipment impacting temperature readings (Menne et al. 2009, 2010). Additionally, unlike the U.S. Historical Climate Network, GHCN-Daily data are not homogenized to account for station movement or urbanization issues (e.g., Menne and Williams 2005). Homogenization was not performed in this work as this dataset is not intended for climate trend analysis.

NRCS SNOTEL data used in this dataset did not undergo any prior quality control, and thus the general methods of Serreze et al. (1999) were applied. They are the following for temperature: 1) a gross error check where $T_{\max} > 45^{\circ}\text{C}$ and $T_{\min} < -45^{\circ}\text{C}$ are flagged as missing, 2) temperature runs of two or more days with the exact same T_{\max} or T_{\min} are flagged as missing, and 3) a standard deviation check where T_{\max} or T_{\min} values outside three standard deviations of the station mean climatology for that date are flagged as missing.

For precipitation 1) daily values >250 mm are flagged as missing and 2) winter season (October–March) precipitation are corrected using a regional bulk undercatch ratio [see the appendix of Serreze et al. (1999) for more details]. Figure A1 presents an example SNOTEL site with the gauge undercatch correction applied (Fig. A1a) and another with various temperature issues (Fig. A1b). Again, no homogenization is performed on the SNOTEL data as with the GHCN-Daily data. Recent work (Oyler et al. 2015) highlights the dangers of using nonhomogenized time series for climate trend analysis.

b. Serially complete time series

As was discussed in section 4, elevation lapse rates for precipitation and temperature are determined directly from the available station data for each time step. This requires the station data to be consistent in time and be serially complete to avoid large changes in derived lapse rates as stations are added or removed or report missing data for a given event. To meet this

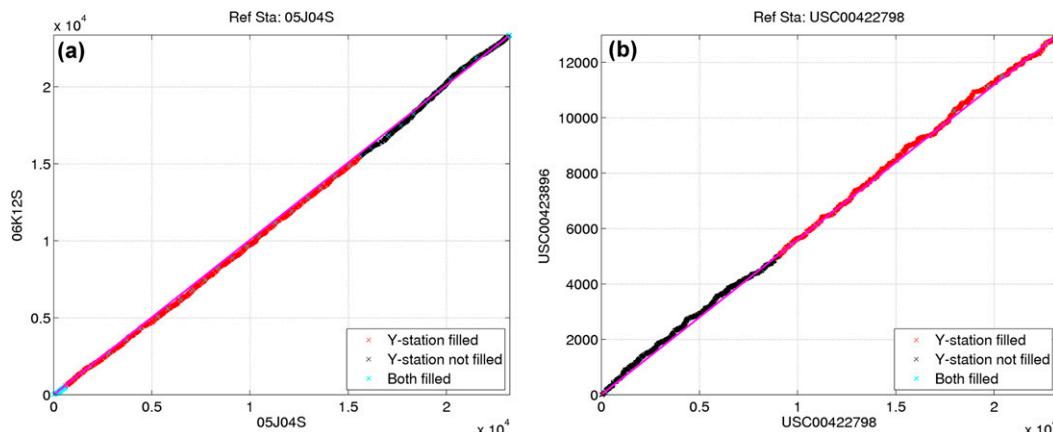


FIG. A2. Example double mass curves for two stations with filled and nonfilled portions noted.

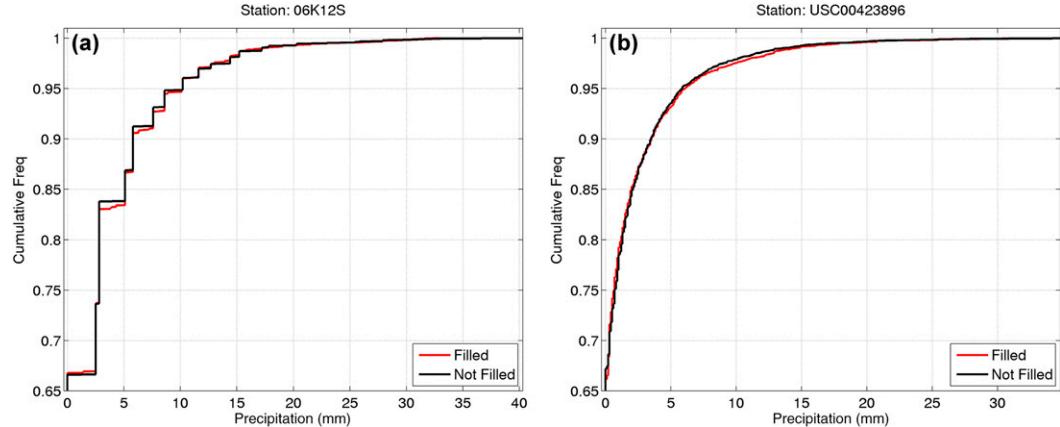


FIG. A3. Example CDFs for two different stations for the filled and nonfilled periods.

requirement, a serially complete station dataset was developed from the quality-controlled station data described above.

Initial station screening consisted of rejecting stations with less than 10 years of valid data during 1980–2012, with the remaining stations selected as target stations to be made serially complete (Eischeid et al. 2000). For a given target station, two data-filling methods are used. First, for missing data gaps of one or two consecutive days in T_{\max} or T_{\min} , simple linear interpolation is used to fill missing values. Second, for data gaps larger than 2 days in temperature and all missing values in precipitation a quantile matching method is used (Panofsky and Brier 1968). This process consists of the following steps for temperature and precipitation.

- 1) The closest 30 stations with at least 10 years of overlap with each target station are found.
- 2) For each day of the year (DOY), Spearman rank correlations (Lehmann and D'Abreu 1998) of the overlap stations with the target station are computed using a 31-day window for the valid overlap data.
- 3) The overlap stations are ranked from highest to lowest rank correlation for the current DOY.
- 4) Missing values at the target station for the current DOY are filled from the highest ranking overlap station with corresponding valid data based on quantile matching:
 - (i) CDFs for the target and overlap station are computed for the 31-day window around the current DOY.
 - (ii) The cumulative probability of the overlap station value is computed from the overlap station CDF.
 - (iii) That cumulative probability is used in the target station CDF to determine the fill value for the target station.

If there are no overlap stations with valid data for a given target station missing value, then the target station is rejected. Using rank correlations and CDF matching removes the step of segregating stations by their reporting time used in Eischeid et al. (2000) because it accounts for the differences in the respective station distributions (Panofsky and Brier 1968; Wood et al. 2002). The methodology employed here produces consistent time series during the filled and nonfilled periods (Fig. A2). This is also a limitation of the method, in that direct CDF matching assumes the CDFs are stationary from the overlap to missing record periods (Teegavarapu 2014; Fig. A3). Overall, 12 124 stations for precipitation and 9374 stations for temperature are included in the serially complete dataset with 24.4% (25.6%) of the total precipitation (temperature) data filled and 59% (55%) of the precipitation (temperature) stations containing less than 25% filled data (Fig. A4).

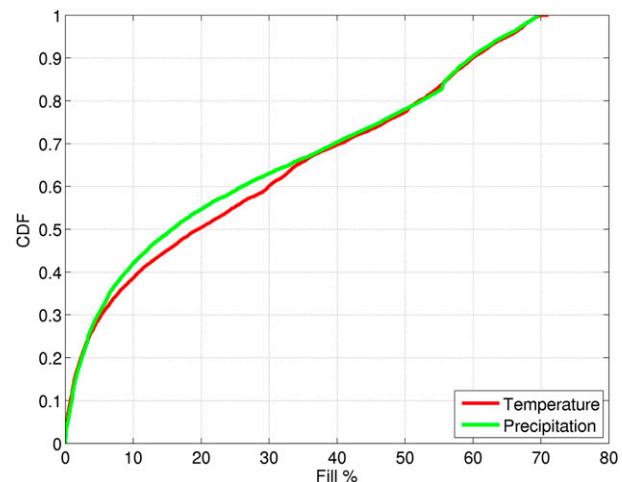


FIG. A4. CDFs of station fill percentage for temperature and precipitation.

REFERENCES

- Abatzoglou, J. T., 2013: Development of gridded surface meteorological data for ecological applications and modeling. *Int. J. Climatol.*, **33**, 121–131, doi:10.1002/joc.3413.
- AghaKouchak, A., N. Nasrollahi, and E. Habib, 2009: Accounting for uncertainties of the TRMM satellite estimates. *Remote Sens.*, **1**, 606–619, doi:10.3390/rs1030606.
- , E. Habib, and A. Bárdossy, 2010: A comparison of three remotely sensed rainfall ensemble generators. *Atmos. Res.*, **98**, 387–399, doi:10.1016/j.atmosres.2010.07.016.
- Alemohammad, S. H., 2014: Characterization of uncertainty in remotely-sensed precipitation estimates. Ph.D. thesis, Massachusetts Institute of Technology, 156 pp. [Available online at http://hamed.mit.edu/sites/default/files/Alemohammad_PhD_Thesis.pdf.]
- Anderson, E. A., 1973: National Weather Service River Forecast System—Snow accumulation and ablation model. NOAA Tech. Memo. NWS HYDRO-17, 87 pp. [Available online at <ftp://ftp.wcc.nrcs.usda.gov/wntsc/H&H/snow/AndersonHYDRO17.pdf>.]
- Austin, P. M., 1987: Relation between measured radar reflectivity and surface rainfall. *Mon. Wea. Rev.*, **115**, 1053–1070, doi:10.1175/1520-0493(1987)115<1053:RBMRAA>2.0.CO;2.
- Bauer, P., J.-F. Mahfouf, W. S. Olson, F. S. Marzano, S. D. Michele, A. Tassa, and A. Mugnai, 2002: Error analysis of TMI rainfall estimates over ocean for variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2129–2214, doi:10.1256/003590002320603575.
- Bellerby, T., 2007: Satellite rainfall satellite rainfall uncertainty estimation using an artificial neural network. *J. Hydrometeor.*, **8**, 1397–1412, doi:10.1175/2007JHM846.1.
- , 2013: Ensemble representation of uncertainty in Lagrangian satellite rainfall estimates. *J. Hydrometeor.*, **14**, 1483–1499, doi:10.1175/JHM-D-12-0121.1.
- , and J. Sun, 2005: Probabilistic and ensemble representations of the uncertainty in an IR/microwave satellite precipitation product. *J. Hydrometeor.*, **6**, 1032–1044, doi:10.1175/JHM454.1.
- Bindlish, R., and A. P. Barros, 1996: Aggregation of digital terrain data using a modified fractal interpolation scheme. *Comput. Geosci.*, **22**, 907–917, doi:10.1016/S0098-3004(96)00049-0.
- Bohn, T. J., B. Livneh, J. W. Oyler, S. W. Running, B. Nijssen, and D. P. Lettenmaier, 2013: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. *Agric. For. Meteor.*, **176**, 38–49, doi:10.1016/j.agrformet.2013.03.003.
- Burnash, R. J. C., R. L. Ferral, and R. A. McGuire, 1973: A generalized streamflow simulation system: Conceptual models for digital computers. Joint Federal and State River Forecast Center, U.S. National Weather Service, and California Department of Water Resources Tech. Rep., 204 pp.
- Cifelli, R., V. Chandrasekar, S. Lim, P. C. Kennedy, Y. Wang, and S. A. Rutledge, 2011: A new dual-polarization radar rainfall algorithm: Application in Colorado precipitation events. *J. Atmos. Oceanic Technol.*, **28**, 352–364, doi:10.1175/2010JTECHA1488.1.
- Clark, M. P., and A. G. Slater, 2006: Probabilistic quantitative precipitation estimation in complex terrain. *J. Hydrometeor.*, **7**, 3–22, doi:10.1175/JHM474.1.
- , —, A. P. Barrett, L. E. Hay, G. J. McCabe, B. Rajagopalan, and G. H. Leavesley, 2006: Assimilation of snow-covered area information into hydrologic and land-surface models. *Adv. Water Resour.*, **29**, 1209–1221, doi:10.1016/j.advwatres.2005.10.001.
- 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, doi: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, doi:10.1002/joc.1688.
- Dinku, T., M. Anagnostou, and M. Borgia, 2002: Improving radar-based estimation of rainfall over complex terrain. *J. Appl. Meteor.*, **41**, 1163–1178, doi:10.1175/1520-0450(2002)041<1163:IRBEOR>2.0.CO;2.
- Duan, Q., S. Sorooshian, and V. K. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, **28**, 1015–1031, doi:10.1029/91WR02985.
- Durre, I., M. J. Menne, and R. S. Vose, 2008: Strategies for evaluating quality assurance procedures. *J. Appl. Meteor. Climatol.*, **47**, 1785–1791, doi:10.1175/2007JAMC1706.1.
- , —, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, doi:10.1175/2010JAMC2375.1.
- Eischeid, J. K., P. A. Pasteris, H. F. Dias, M. S. Plantico, and N. J. Lott, 2000: Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteor.*, **39**, 1580–1591, doi:10.1175/1520-0450(2000)039<1580:CASND>2.0.CO;2.
- Fang, J., and L. Tacher, 2003: An efficient and accurate algorithm for generating spatially-correlated random fields. *Commun. Numer. Methods Eng.*, **19**, 801–808, doi:10.1002/cnm.621.
- Gabriel, K. R., and J. Neumann, 1962: A Markov chain model for daily rainfall occurrences at Tel Aviv. *Quart. J. Roy. Meteor. Soc.*, **88**, 90–95, doi:10.1002/qj.49708837511.
- Gebregiorgis, A. S., and F. Hossain, 2015: How well can we estimate error variance of satellite precipitation data around the world? *Atmos. Res.*, **154**, 39–59, doi:10.1016/j.atmosres.2014.11.005.
- 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 intercomparison. Instruments and Observing Methods Rep. 67, 212 pp. [Available online at <https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-67-solid-precip/WMOTd872.pdf>.]
- Greatrex, H., D. Grimes, and T. Wheeler, 2014: Advances in the stochastic modeling of satellite-derived rainfall estimates using a sparse calibration dataset. *J. Hydrometeor.*, **15**, 1810–1831, doi:10.1175/JHM-D-13-0145.1.
- Gutmann, E., 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.
- Habib, E., A. Henschke, and R. F. Adler, 2009: Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA. *Atmos. Res.*, **94**, 373–388, doi:10.1016/j.atmosres.2009.06.015.
- Hamlet, A. F., and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature

- fields for the continental U.S. *J. Hydrometeor.*, **6**, 330–336, doi:10.1175/JHM420.1.
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, doi:10.1029/2008JD010201.
- Hossain, F., and E. N. Anagnostou, 2006: A two-dimensional satellite rainfall error model. *IEEE Trans. Geosci. Remote Sens.*, **44**, 1511–1521, doi:10.1109/TGRS.2005.863866.
- Huffman, G. J., R. F. Adler, M. Morrissey, D. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multi-satellite observations. *J. Hydrometeor.*, **2**, 36–50, doi:10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2.
- , —, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:10.1029/2009GL040000.
- Hungerford, R. D., R. Nemani, S. W. Running, and J. C. Coughlan, 1989: MT-CLIM: A mountain microclimate simulation model. U.S. Forest Service Research Paper INT-414, 52 pp. [Available online at http://www.fs.fed.us/rm/pubs_int/int_rp414.pdf.]
- Johnson, M. E., 1987: *Multivariate Statistical Simulation*. Wiley, 230 pp.
- Kavetski, D. N., S. W. Franks, and G. Kuczera, 2003: Confronting input uncertainty in environmental modelling. *Calibration of Watershed Models*, Q. Duan et al., Eds, Water Science and Application Series, Vol. 6, Amer. Geophys. Union, 49–68.
- Kirstetter, P.-E., Y. Hong, J. J. Gourley, M. Schwaller, W. Petersen, and J. Zhang, 2013a: Comparison of TRMM 2a25 products, version 6 and version 7, with NOAA/NSSL ground radar-based national mosaic QPE. *J. Hydrometeor.*, **14**, 661–669, doi:10.1175/JHM-D-12-030.1.
- , N. Viltard, and M. Gosset, 2013b: An error model for instantaneous satellite rainfall estimates: Evaluation of BRAINTMI over West Africa. *Quart. J. Roy. Meteor. Soc.*, **139**, 894–911, doi:10.1002/qj.1964.
- , J. J. Gourley, Y. Hong, J. Zhang, S. Moazamigoodarzi, C. Langston, and A. Arthur, 2015: Probabilistic precipitation rate estimates with ground-based radar networks. *Water Resour. Res.*, **51**, 1422–1442, doi:10.1002/2014WR015672.
- Krajewski, W. F., G. J. Ciach, J. R. McCollum, and C. Bacotiu, 2000: Initial validation of the global precipitation climatology project monthly rainfall over the United States. *J. Appl. Meteor.*, **39**, 1071–1086, doi:10.1175/1520-0450(2000)039<1071:IVOTGP>2.0.CO;2.
- Lehmann, E. L., and H. J. M. D'Abra, 1998: *Nonparametrics: Statistical Methods Based on Ranks*. Rev. Ed. Prentice-Hall, 463 pp.
- 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 <https://ams.confex.com/ams/Annual2005/webprogram/Paper83847.html>.]
- Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 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.
- Loader, C., 1999: *Local Regression and Likelihood*. Springer, 308 pp.
- Maggioni, V., M. R. P. Sapiano, R. F. Adler, Y. Tian, and G. J. Huffman, 2014: An error model for uncertainty quantification in high-time-resolution precipitation products. *J. Hydrometeor.*, **15**, 1274–1292, doi:10.1175/JHM-D-13-0112.1.
- Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset 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.
- , —, —, —, and —, 2013: Gridded meteorological data: 1949–2010. Accessed 12 September 2013. [Available online at http://www.engr.scu.edu/~emaurer/gridded_obs/index_gridded_obs.html.]
- Menne, M. J., and C. N. Williams Jr., 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Climate*, **18**, 4271–4286, doi:10.1175/JCLI3524.1.
- , —, 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.
- , —, and M. A. Palecki, 2010: On the reliability of the U.S. surface temperature record. *J. Geophys. Res.*, **115**, D11108, doi:10.1029/2009JD013094.
- , I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012a: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, doi:10.1175/JTECH-D-11-00103.1.
- , and Coauthors, 2012b: Global Historical Climatology Network - Daily (GHCN-Daily), version 2.92. NOAA National Climatic Data Center, accessed 17 September 2013, doi:10.7289/V5D21VHZ.
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, doi:10.1175/BAMS-87-3-343.
- Mizukami, N., M. P. Clark, A. G. Slater, L. D. Brekke, M. M. Elsner, J. R. Arnold, and S. Gangopadhyay, 2014: Hydrological implications of difference large-scale meteorological model forcing datasets in mountainous regions. *J. Hydrometeor.*, **15**, 474–488, doi:10.1175/JHM-D-13-036.1.
- , and Coauthors, 2015: Implications of the methodological choices for hydrologic portrayals of climate change over the contiguous United States: Statistically downscaled forcing data and hydrologic models. *J. Hydrometeor.*, doi:10.1175/JHM-D-14-0187.1, in press.
- Moradkhani, H., K. Hsu, H. Gupta, and S. Sorooshian, 2005: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resour. Res.*, **41**, W05012, doi:10.1029/2004WR003604.
- Morrissey, M. L., and J. S. Green, 1998: Uncertainty analysis of satellite rainfall algorithms over the tropical Pacific. *J. Geol. Res.*, **103**, 19 569–19 576, doi:10.1029/98JD00309.
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models. Part I: A discussion of principles. *J. Hydrol.*, **10**, 282–290, doi:10.1016/0022-1694(70)90255-6.
- Newman, A. J., K. Sampson, M. P. Clark, A. Bock, R. J. Viger, and D. Blodgett, 2014: A large-sample watershed-scale hydroeteorological dataset for the contiguous USA. UCAR/NCAR, doi:10.5065/D6MW2F4D.
- , and Coauthors, 2015a: Development of a large-sample watershed-scale hydroeteorological dataset for the contiguous USA: Dataset characteristics and assessment of regional variability in hydrologic model performance. *Hydroclim. Earth Syst. Sci.*, **19**, 209–223, doi:10.5194/hess-19-209-2015.
- , M. P. Clark, J. Craig, B. Nijssen, A. Wood, and E. Gutmann, 2015b: Gridded ensemble precipitation and temperature estimates over the contiguous United States. UCAR/NCAR, doi:10.5065/D6TH8JR2.

- Nijssen, B., and D. P. Lettenmaier, 2004: Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *J. Geophys. Res.*, **109**, D02103, doi:10.1029/2003JD003497.
- Nogueira, M., and A. P. Barros, 2013: The Integrated Precipitation and Hydrology Experiment—Hydrologic Applications for the Southeast US (IPHEX-H4SE): Part III: High-resolution ensemble rainfall products. Rep. EPL-2013-IPHEX-H4SE-3, EPL/Duke University, 38 pp. [Available online at <http://hdl.handle.net/10161/8969>.]
- Oyler, J. W., S. Z. Dobrowski, A. P. Ballantyne, A. E. Klene, and S. W. Running, 2015: Artificial amplification of warming trends across the mountains of the western United States. *Geophys. Res. Lett.*, **42**, 153–161, doi:10.1002/2014GL062803.
- Panofsky, H. A., and G. W. Brier, 1968: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University, 224 pp.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyre, and S. W. Franks, 2010: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resour. Res.*, **46**, W05521, doi:10.1029/2009WR008328.
- Ryzhkov, A. V., S. Giangrande, and T. Schuur, 2003: Rainfall measurements with the polarimetric WSR-88D radar. National Severe Storms Laboratory, 98 pp. [Available online at http://docs.lib.noaa.gov/noaa_documents/OAR/National_Severe_Storms_Laboratory/Rainfall_measurements_Polarimetric_WSR88D_Radar.pdf.]
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.
- Sapiano, M. R. P., and P. A. Arkin, 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *J. Hydrometeor.*, **10**, 149–166, doi:10.1175/2008JHM1052.1.
- Schiemann, R., R. Erdin, M. Willi, C. Frei, M. Berenguer, and D. Sempere-Torres, 2011: Geostatistical radar–raingauge combination with nonparametric correlograms: Methodological considerations and application in Switzerland. *Hydrol. Earth Syst. Sci.*, **15**, 1515–1526, doi:10.5194/hess-15-1515-2011.
- Serreze, M. C., M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. L. Pulwarty, 1999: Characteristics of the western U.S. snowpack from snowpack telemetry (SNOWTELE) data. *Water Resour. Res.*, **35**, 2145–2160, doi:10.1029/1999WR900090.
- Shen, Y., A. Xiong, Y. Wang, and P. Xie, 2010: Performance of high-resolution satellite precipitation products over China. *J. Geophys. Res.*, **115**, D02114, doi:10.1029/2009JD012097.
- Shepard, D. S., 1968: A two-dimensional interpolation function for irregularly spaced data. *Proceedings of the 23rd Association for Computing Machinery Conference*, ACM, 517–524, doi:10.1145/800186.810616.
- , 1984: Computer mapping: The SYMAP interpolation algorithm. *Spatial Statistics and Models*, G. L. Gaile and C. J. Willmott, Eds., D. Reidel, 133–145.
- Skinner, C. J., T. J. Bellerby, H. Greatrex, and D. I. F. Grimes, 2015: Hydrological modeling using ensemble satellite rainfall estimates in a sparsely gauged river basin: The need for whole ensemble calibration. *J. Hydrol.*, **522**, 110–122, doi:10.1016/j.jhydrol.2014.12.052.
- Stahl, K., R. D. Moore, J. A. Floyer, M. G. Aspin, and I. G. McKendry, 2006: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agric. For. Meteor.*, **139**, 224–236, doi:10.1016/j.agrformet.2006.07.004.
- Teegavarapu, R. S. V., 2014: Statistical corrections of spatially interpolated missing precipitation data estimates. *Hydro. Processes*, **28**, 3789–3808, doi:10.1002/hyp.9906.
- Teo, C. K., and D. I. F. Grimes, 2007: Stochastic modelling of rainfall from satellite data. *J. Hydrol.*, **346**, 33–50, doi:10.1016/j.jhydrol.2007.08.014.
- 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.
- , M. M. Thornton, B. W. Mayer, N. Wilhelm, Y. Wei, R. Devarakonda, and R. B. Cook, 2014: Daymet: Daily surface weather data on a 1-km grid for North America, version 2. Oak Ridge National Laboratory Distributed Active Archive Center, accessed 15 November 2014, doi:10.3334/ORNLDaac/1219.
- Tian, Y., and C. D. Peters-Lidard, 2007: Systematic anomalies over inland water bodies in satellite-based precipitation estimates. *Geophys. Res. Lett.*, **34**, L14403, doi:10.1029/2007GL030787.
- , and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates. *J. Geophys. Res.*, **114**, D24101, doi:10.1029/2009JD011949.
- , G. J. Huffman, R. F. Adler, L. Tang, M. Sapiano, V. Maggioni, and H. Wu, 2013: Modeling errors in daily precipitation measurements: Additive or multiplicative? *Geophys. Res. Lett.*, **40**, 2060–2065, doi:10.1002/grl.50320.
- 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, doi:10.1175/1520-0477(1999)080<2289:TLOTWR>2.0.CO;2.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.
- Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long-range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.*, **107**, 4429, doi:10.1029/2001JD000659.
- Xia, Y., and Coauthors, 2009: NLDAS_FORA0125_H version 002: NLDAS primary forcing data L4 hourly 0.125 x 0.125 degree. Goddard Earth Sciences Data and Information Services Center, accessed 15 September 2013, doi:10.5067/6J5LHHOHZHN4.
- , and Coauthors, 2012: 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.
- Yang, D., B. E. Goodison, J. R. Metcalfe, V. S. Golubev, R. Bates, T. Pangburn, and C. L. Hanson, 1998: Accuracy of NWS 8" standard nonrecording precipitation gauge: Results and application of WMO intercomparison. *J. Atmos. Oceanic Technol.*, **15**, 54–68, doi:10.1175/1520-0426(1998)015<0054:AONSNP>2.0.CO;2.
- Young, C. B., B. R. Nelson, A. A. Bradley, J. A. Smith, C. D. Peters-Lidard, A. Kruger, and M. L. Baek, 1999: An evaluation of NEXRAD precipitation estimates in complex terrain. *J. Geophys. Res.*, **104**, 19 691–19 703, doi:10.1029/1999JD900123.