Methodological Intercomparisons of Station-Based Gridded Meteorological Products: Utility, Limitations, and Paths Forward

ANDREW J. NEWMAN AND MARTYN P. CLARK

National Center for Atmospheric Research, Boulder, Colorado

RYAN J. LONGMAN AND THOMAS W. GIAMBELLUCA

Department of Geography and Environment, University of Hawai'i at Mānoa, Honolulu, Hawaii

(Manuscript received 31 May 2018, in final form 13 November 2018)

ABSTRACT

This study presents a gridded meteorology intercomparison using the State of Hawaii as a testbed. This is motivated by the goal to provide the broad user community with knowledge of interproduct differences and the reasons differences exist. More generally, the challenge of generating station-based gridded meteorological surfaces and the difficulties in attributing interproduct differences to specific methodological decisions are demonstrated. Hawaii is a useful testbed because it is traditionally underserved, yet meteorologically interesting and complex. In addition, several climatological and daily gridded meteorology datasets are now available, which are used extensively by the applications modeling community, thus an intercomparison enhances Hawaiian specific capabilities. We compare PRISM climatology and three daily datasets: new datasets from the University of Hawai'i and the National Center for Atmospheric Research, and Daymet version 3 for precipitation and temperature variables only. General conclusions that have emerged are 1) differences in input station data significantly influence the product differences, 2) explicit prediction of precipitation occurrence is crucial across multiple metrics, and 3) attribution of differences to specific methodological choices is difficult and limits the usefulness of intercomparisons. Because generating gridded meteorological fields is an elaborate process with many methodological choices interacting in complex ways, future work should 1) develop modular frameworks that allows users to easily examine the breadth of methodological choices, 2) collate available nontraditional high-quality observational datasets for true out-of-sample validation and make them publicly available, and 3) define benchmarks of acceptable performance for methodological components and products.

1. Introduction

Gridded representations of surface or near-surface meteorological observations (e.g., precipitation or temperature) are both widely used and highly uncertain. Spatiotemporally continuous gridded meteorological datasets are used extensively by many communities spanning numerical weather prediction (NWP) and remotely sensed product validation, land surface and hydrologic modeling, and ecology, as well as the vast climate impacts modeling across those fields (e.g., Day 1985; Franklin 1995; USBR 2012; Pierce et al. 2014; Liu et al. 2017; Derin et al. 2016). The broad user community of these products should be aware of 1) the differences across products, 2) product strengths and weaknesses,

and 3) the reasons for differences and deficiencies. More generally, the broad user community of these products should be aware that the gridded representations of surface or near-surface meteorological observations are uncertain.

Intercomparisons of observation-based products can be a potentially useful tool to understand differences across datasets and appreciate their uncertainties (e.g., Prein and Gobiet 2017; Henn et al. 2018a; Walton and Hall 2018). It is well known that in situ observations are uncertain, and gridded representations of point observations are contaminated by both measurement errors as well as errors associated with the spatial representativeness of the station network (Klein Tank et al. 2002; Clark and Slater 2006; Hofstra et al. 2009; Isotta et al. 2014; Schneider et al. 2014; Newman et al. 2015). However, it is only recently that this uncertainty has

Corresponding author: Andrew Newman, anewman@ucar.edu

DOI: 10.1175/JHM-D-18-0114.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

been acknowledged and evaluated in the aforementioned user communities (Liu et al. 2017; Behnke et al. 2016; Prein and Gobiet 2017; Beck et al. 2017, 2018; Raimonet et al. 2017; Gampe and Ludwig 2017; Laiti et al. 2018; Walton and Hall 2018). Appreciating these uncertainties allows for more robust evaluation of numerical weather prediction models (e.g., Liu et al. 2017; Prein and Gobiet 2017) and more robust calibration and application of impact models; see, for example, Bayesian methods in hydrology (Vrugt et al. 2008; Kuczera et al. 2010; Renard et al. 2010; Henn et al. 2018b).

This work is motivated by the general concept of comparison studies of complex modeling systems (Henderson-Sellers et al. 1993, 1995; Schlosser et al. 2000; Duan et al. 2006), their utility, and their limitations (Koster and Milly 1997; Nijssen et al. 2003; Clark et al. 2011, 2015). Gridded product generation is now a very complex process with multiple decisions made by developers at many points in the process. These decisions include which observations to use, observational density, the application of gauge undercatch, treatment of missing data, data homogenization, the choice of spatial interpolation algorithm, determining the orographic lapse rate, the prediction of precipitation occurrence, and how to identify and incorporate estimates of uncertainty. Interactions between the various methodological decisions, processing chain, etc., can constrain the strength of the conclusions (e.g., Raimonet et al. 2017; Henn et al. 2018a)—this means that we need to move beyond broad product comparisons, and instead use common frameworks and data to identify robust methodologies to improve our gridded representation of observed meteorology and their associated uncertainties.

The paper is organized as follows: we introduce the study area in section 2, discuss the available datasets and key methodological choices in section 3, and then perform a comparative analysis with a focus on what methodological choices may lead to agreement or lack thereof in section 4. To conclude, we discuss general implications for Hawaii, key methodological points elucidated here, and a path forward for more rigorous examination and selection of methodological choices for creating gridded datasets in section 5.

2. Study area

Hawaii is characterized by extreme spatial gradients in precipitation and temperature due to its complex topography and northeast trade wind environment (Giambelluca et al. 1986, 2011; Chu and Chen 2005; Frazier et al. 2016). Figure 1 provides a map of the Hawaiian island chain with the various island labeled for

FIG. 1. Hawaiian Islands with elevation shaded. From N19.

reference. Some locations on the windward side of the island chain receive up to 10 000 mm of precipitation annually, with measurable precipitation on nearly 9 in 10 days (Longman et al. 2019, hereafter L19; Newman et al. 2019, hereafter N19). On the lee sides of the islands, semiarid conditions prevail, with less than 250 mm precipitation annually and measurable precipitation occurring on 2 in 10 days in some regions. The vertical lapse rate of precipitation is also complex as it changes sign (from increasing with elevation to decreasing) around the trade wind inversion (TWI). The TWI, with a mean base height of 2150 m, is a persistent (~90% of the year) atmospheric feature that inhibits the vertical development of clouds to high elevation and results in arid conditions on the highest mountains of Hawaii (Longman et al. 2015). Temperature gradients are strongly controlled by elevation, with secondary influences from precipitation (reduced shortwave radiation) and the TWI. Finally, the diurnal temperature range has a complex lapse rate similar to precipitation with increases with increasing elevation below the TWI and decreases above. These intricacies make Hawaii a compelling scientific testbed to compare gridded precipitation and temperature products.

3. Datasets

Several daily precipitation and temperature datasets have recently been produced for the state of Hawaii. In this work, four gridded products available for Hawaii are examined. Three daily datasets of precipitation and temperature include the University of Hawai'i (UH) 250-m daily fields for 1990–2014 (Longman et al. 2018b; L19) an ensemble of 1-km daily grids for 1990–2014 (Newman et al. 2018; N19), and daily 1-km Daymet data from 1990 to 2014 (Daymet version 3; Thornton et al. 2017a). The Parameter-Elevation Regressions on Independent Slopes Model (PRISM) 450-m climate normal

values for 1971–2000 (Daly et al. 2006) are also included here for climatological comparisons. Summaries of key methodological choices for Hawaii across all four datasets are presented in Table 1. We review the basic methodological choices of the four datasets here to orient the readers, with more in-depth discussion throughout the intercomparison.

a. University of Hawai'i

The UH precipitation climatology is determined in a multistep procedure. First input data from the National Centers for Environmental Information [NCEI; e.g., Global Historical Climate Network Daily data (GHCN-D)] data feeds, many historical and current local networks not available on NCEI, radar precipitation estimates, atmospheric model output, and precipitation estimates from known plant physiological limits were compiled for the 30-yr period 1978–2007 (Giambelluca et al. 2011, 2013). Ordinary kriging was used on the in situ observations and virtual plant-based observations to produce one estimate of precipitation. Then, PRISM, the ordinary kriging estimate, radar, and atmospheric model estimates were weighted using their respective estimated uncertainties to develop the final precipitation distribution [see Giambelluca et al.'s (2011) methods section for complete details]. Therefore, the final precipitation climatology in the UH product is a complex blend of actual observations, synthetic observations, precipitation radar, atmospheric model output, and the PRISM estimates.

Temperature observations are more limited in Hawaii, and the UH development team chose to develop a piecewise linear regression framework that incorporates elevation and precipitation as predictors for the entire island chain. They develop one equation below the mean TWI height of 2150 m and one above (L19). Consequently, temperature is only a function of elevation and precipitation, with no x – y spatial information included outside of the indirect information from precipitation patterns. This means two locations on different islands will have the same temperature climatology if they have the same elevation and precipitation. Temperature data from 1990 to 2014 are used to develop the climatological fields.

Daily values in the UH product rely on the monthly climatologies to estimate daily anomalies (ratio anomalies for precipitation), which are then interpolated using inverse distance weighting (IDW) using the nearest five stations (L19). The monthly climatologies inform the orographic lapse rates and general spatial distribution of a variable, which is known as climatologically aided interpolation (CAI; e.g., Dawdy and Langbein 1960; Willmott and Robeson 1995). The input station network for the daily data are only stations with daily

data available within the 1990–2014 period, and it differs substantially from the climatological network in some regions for precipitation (L19; N19). The daily data include all stations available through NCEI as well as many local networks not available through NCEI. The UH product also includes partial station filling to reduce the number of missing data records. They follow Eischeid et al. (2000) to determine the filling method for a given station and do not make all stations serially complete (Longman et al. 2018a). This means a portion of the data are estimated and a given station may be available one day and not the next.

b. NCAR Ensemble

The NCAR Ensemble product (hereafter the Ensemble; Clark and Slater 2006; Newman et al. 2015; N19) uses only the in situ observations assembled in the UH product (Longman et al. 2018a) to develop its precipitation climatology for the same 1978–2008 period. The temperature climatology is developed by aggregating the daily temperature observations assembled by the UH effort, which spans the 1990–2014 period. Estimates of climatological monthly precipitation and temperature are developed using locally weighted (by distance) multiple linear regression on the station data with 25 stations considered for every grid point. Predictors for Hawaii include elevation, distance to coast, latitude, and longitude. The cross-validation errors of the regression are used as uncertainty estimates, which are then combined with spatially correlated random fields to generate 100 realizations of the monthly climatologies.

CAI is also used in the Ensemble to step from monthly climatologies to daily values using daily anomalies, again ratio anomalies for precipitation. The locally weighted regression uses only latitude and longitude as predictors for the anomaly interpolation as the orographic information is contained in the climatological fields. The probability of occurrence of precipitation is also predicted at each grid point using logistic regression with the same locally weighted predictor set. The daily input data for the Ensemble consist of the same raw station data as the UH daily product, NCEI data with substantial augmentation from local networks. Station filling is also performed here, using percentile matching and linear interpolation for 1- or 2-day temperature gaps to make stations serially complete. Only stations with serially complete data are used in the final product (Newman et al. 2015).

c. Daymet

Daymet (Thornton et al. 1997) operates on the daily time step only. Here, the climatologies are developed by averaging the 1990–2014 period in the daily product.

TABLE 1. Summary description of datasets including basic dataset characteristics, relevant data sources (specifically for Hawaii), and methodological choices.

Characteristic	Dataset			
	UH	NCAR Ensemble	Daymet	PRISM
Resolution	250 m	1 km	1 km	450 m
Time step	Daily, climatologically aided 1978–2007	Daily, climatologically aided 1978–2007	Daily	Monthly climatology 1971–2000
Precipitation climatology time period	NCEI monthly, local networks, synthetic point observations	GHCN-D, special local networks	—	GHCN-D, local networks, synthetic point observations
Climatological precipitation station data	Ordinary kriging of point obs; then combined with PRISM, model, and radar estimates	Multiple linear regression using location, elevation, distance to coast	—	PRISM facet method with manual adjustments
Climatological precipitation interpolation	GHCN-D, special local networks	GHCN-D, special local networks	—	GHCN-D, special local networks
Climatological temperature station data	1990–2014	1990–2014	—	1971–2000
Temperature climatology time period	Piecewise linear regression using elevation and precipitation	Locally weighted multiple linear regression using location, elevation, distance to coast	—	PRISM facet method with manual adjustments
Climatological temperature interpolation	Fill data where possible using regression relationships, not serially complete	Create static, serially complete station dataset	None, uses only available observations	—
Daily missing data filling	IDW interpolation of daily anomalies	Locally weighted multiple linear regression interpolation of daily anomalies	Truncated Gaussian filter on daily values	—
Daily precipitation/temperature interpolation method	No	Yes; explicit at each grid cell	Yes; threshold using weighted sum of occurrence vector	—

Daymet daily fields are developed using a Gaussian filter to establish locally varying station weights, which are then applied to a summation of weighted linear regression equations developed for all combinations of station pairs from all stations considered for a grid point (nearest 30 stations for temperature, 20 for precipitation; Thornton et al. 1997). The orographic lapse rate is defined as the weighted mean of the lapse rate for all station pairs for that grid point. Thus, the daily orographic lapse rates and precipitation/temperature distributions are determined independently at each grid point using only data for that day. Prediction of precipitation occurrence is performed at each grid point using a simple threshold (0.52) on the summation of the weighted occurrence vector (1 for nonzero precipitation, 0 for no precipitation). Daymet version 3 in Hawaii uses only data from the NCEI archives and performs no station filling, only using available observations for a given day.

d. PRISM

PRISM, which has been developed over the course of more than 20 years, distributes point measurements of monthly, seasonal, and annual precipitation to a gridded surface (Daly et al. 1994, 2002, 2003, 2008, 2017). PRISM has been characterized as a “knowledge based” interpolation algorithm because knowledge of atmospheric physics is used to inform a complex regression-based framework (Daly et al. 2002). For example, knowledge of processes such as varying amounts of orographic precipitation on certain aspects of mountains (windward vs leeward), or the potential presence of a temperature inversion are used to create weights for stations across topographic facets or aspects. This information is used to form a linear regression equation that determines the appropriate lapse rate from the relevant observations and the corresponding grid cell value. These lapse rates and amounts are determined for a specified climatological 30-yr normal period (Daly et al. 1994, 2002, 2008). Specifically for Hawaii, PRISM uses available data from the NCEI, local station networks (not available in standard NCEI data archives), and estimated station data using nearest neighbor or other approaches (Daly et al. 2006) for the 1971–2000 climate normal period.

4. Comparative analysis

The various gridded products were compared across metrics that are relevant to many applications including climatological distributions of precipitation, long-term daily precipitation occurrence, daily variability of precipitation, and extreme precipitation events, as well as daily mean temperature and diurnal temperature range.

We make no direct quantitative comparisons to observations as we cannot perform a true cross validation of the various products against a common set of independent observations. Instead, we present the most comprehensive observations available with the gridded products and compare and contrast their spatial similarity to the observations and within the products themselves. We quantitatively compare the product metrics through the use of aggregated gridpoint interproduct comparisons. Mean absolute difference (MAD) and spatial correlation between the various product combinations are computed to highlight systematic differences (MAD) and spatial similarity for each metric. These values are displayed for applicable products in each subsection, with MAD in the upper-right half of the table and spatial correlation in the lower-left half of the table. As an example, the Ensemble–UH spatial correlation is in the table cell given by the product (row, column) pair (Ensemble, UH), while the MAD is given in the (UH, Ensemble) pair. For these metrics the various products were all regridded using simple linear interpolation of nearby points to the PRISM grid resulting in 84 004 valid grid points. Additionally, references will be made to the individual islands and differences across them. Here we define the islands as Hawaii Island, Maui (for Maui only), Maui Nui (for Maui, Molokai, Lanai, and Kahoolawe), Oahu, and Kauai.

a. Mean precipitation distribution

Climatologies of the four gridded products and station observations are shown in Fig. 2. In all of the products, windward sides of each island have precipitation maxima, while the lee sides are much drier. On Hawaii Island and the southeast portion of Maui, precipitation increases to a point around 1500 m, then decreases above. This is caused by the presence of the TWI, which inhibits the vertical development of clouds and diverts atmospheric flow around the mountain rather than over (e.g., Leopold 1949; Sanderson 1993; Garza et al. 2012). On Kauai, Oahu, Molokai, and the northwest portion of Maui, the topography is low enough such that flow moves over the peaks, with precipitation maxima on those islands located near or over the peaks. Across the datasets, PRISM and the UH products are the most similar, followed by the Ensemble mean and finally the Daymet product. Interproduct spatial correlations and MAD confirm this; Daymet is the most dissimilar product (Table 2) of the four.

Several methodological choices underlie the differences seen in Fig. 2 and Table 2. First, PRISM uses a different climatological period than UH and the Ensemble. This is one possible explanation for the magnitude differences between PRISM and the UH and Ensemble

FIG. 2. Long-term daily mean precipitation (mm) for daily reporting observational sites and the four climatologies available for Hawaii. The gray contour is 2000 m. Note that the islands of Oahu and Kauai are not drawn to the same scales as the rest of Hawaii.

products. However, the spatial patterns are still very similar, which highlights the repeatability of the orographically forced precipitation patterns across Hawaii. Livneh et al. (2015) find similar results across the contiguous United States (CONUS), where changing climatological periods can have a large impact on amounts and possibly less on spatial patterns. The UH and PRISM are very similar despite the different climatological periods, likely due to the fact that the UH product incorporates the PRISM estimates into their climatological grids for the CAI.

A second explanation for differences in the map are the different choice of input stations used in the

methodologies. The Ensemble, PRISM, and UH products use similar sets of stations, but they are not exactly the same. For example, the Ensemble product does not use any synthetic observations like PRISM and UH (in their climatology). Daymet only uses the roughly 150 unique observations available in the GHCN-D (Daymet version 3 documentation; Thornton et al. 2017b), which is only 15% of the number of stations than available to the Ensemble or UH climatologies, which use up to 1000 unique real observations, and about 30% of the station network used in PRISM, which uses 422 real and synthetic observations (Daly et al. 2006). It is clear that Daymet is consistently drier than observations and the other products, except for Hawaii Island, even though the Daymet algorithm incorporates topographic effects. The input station network into Daymet is unable to resolve the observed precipitation maxima on nearly every island.

The intrinsic grid spacing will influence product performance as well. In our case, the products are all of sufficient resolution to resolve the spatial features in Hawaii, and thus we do not directly comment on the impacts of changes in this methodological choice. In general, the optimal resolution is a complex blend of the

TABLE 2. Spatial correlations of long-term mean daily precipitation between the various products on the lower-left half and MAD (mm) on the upper-right half of the matrix across the 84 004 valid grid points.

Product	PRISM	UH	Ensemble	Daymet
PRISM	—	0.8 mm	1.2 mm	1.6 mm
UH	0.98	—	0.7 mm	1.6 mm
Ensemble	0.92	0.94	—	1.7 mm
Daymet	0.77	0.76	0.75	—

FIG. 3. Long-term daily mean PoP for observational sites and the three daily time-step products available for Hawaii. Note that the islands of Oahu and Kauai are not drawn to the same scales as the rest of Hawaii.

physical processes occurring and subsequent spatial patterns, available observations, interpolation method, and finally user needs. Use of the variogram or spatial frequency analysis could inform an optimal resolution choice in conjunction with consideration of the other factors.

Finally, Daymet and the Ensemble show the strongest relationship with elevation as evinced by the valley ridge patterns on Kauai and Oahu in those products. This likely relates to the fact that the Ensemble and Daymet are not using smoothed topography in Hawaii, while PRISM does smooth the topography to create their facets. The UH climatology underlying this product uses ordinary kriging and combines other estimates such as PRISM, resulting in a smoother final climatology. Although Daymet has a strong precipitation–topographic relationship, this should not be conflated with high effective resolution, which may be coarser than perceived due to the sparse station network.

b. Precipitation occurrence

The daily probability of precipitation (PoP) is also important for impact modeling and is a higher-order statistic that can be difficult to capture when interpolating point measurements to a grid (e.g., Newman et al. 2015).

Across Hawaii, observations indicate very frequent precipitation ($>50\%$ of days) at most windward locations, with maximum precipitation areas receiving precipitation up to nearly 90% of days. Conversely, on lee sides and above the TWI, precipitation is very infrequent, occurring less than 20% of the time at many leeside stations (Fig. 3). Because PRISM has only a climatological amount product, no comparisons of PoP are possible. All three daily products have the highest PoP on the windward sides and near precipitation maxima, with the UH and Daymet products having the highest and lowest overall PoP, respectively, while the Ensemble lies between these end points. The spatial patterns of the products are generally similar, with spatial correlations ranging from 0.77 to 0.87. However, MAD values highlight that the UH product has a substantial offset from the Ensemble and Daymet (and observations qualitatively), while the Ensemble and Daymet have much smaller PoP differences of 0.16 on average (Table 3).

PoP is explicitly predicted at each grid point in Daymet and the Ensemble, while it is not in the UH product. The UH product uses simple inverse distance weighting to interpolation precipitation, which has the effect of increasing the occurrence of precipitation

TABLE 3. Spatial correlations of long-term daily mean PoP between the various products on the lower-left half and MAD (fractional) on the upper-right half of the matrix across the 84 004 valid grid points.

Product	UH	Ensemble	Daymet
UH	—	0.29	0.44
Ensemble	0.82	—	0.16
Daymet	0.77	0.87	—

across nearly all of the domain, particularly above the TWI and on lee slopes. The Daymet algorithm uses a set threshold value applied to the estimated gridcell PoP (the summation of the distance weighted occurrence vector). For Hawaii, this value may cause underestimation of occurrence in wet areas, but performs reasonably well on lee areas when qualitatively compared to the observations. The Ensemble uses locally weighted logistic regression at each grid point for each day to predict occurrence, which allows for more spatial flexibility.

c. Precipitation variability

Another higher-order statistic is the daily variability (standard deviation) of precipitation. Variability is particularly important for hydrologic modeling where precipitation intensity affects runoff generation. Observations show that absolute variability (mm) is larger where the most precipitation falls with the largest variability on Maui and Kauai. However, here we examine the coefficient of variation (CV), which is the standard deviation divided by the mean and thus normalizes the standard deviation to highlight regions with large variability relative to the mean value. Observed CV patterns highlight CV values generally less than two in rainy areas, with drier areas having CVs larger than four or five in some cases. All three gridded products are able to capture this general pattern and are similar with the Ensemble and UH products being most similar with a spatial correlation of 0.85 (Table 4). However, it appears the UH product generally underestimates the CV everywhere; Daymet may overestimate the CV in many places, particularly along the lee slopes, and the Ensemble may also over estimate CV in most lee regions (Fig. 4). Across individual islands the products are most dissimilar on Hawaii Island and Maui Nui, which also have the largest spatial variations in CV.

The spatial patterns of gridded CV highlight the methodological differences in interpolation methodology and input networks between the datasets. The UH and Ensemble products use the exact same input network, yet the UH product has a vastly different CV. IDW interpolation is known to smooth amounts (e.g., Fig. 2) and increase occurrence (e.g., Fig. 3) between observation points, which would serve to reduce CV in

TABLE 4. Spatial correlations of long-term mean daily coefficient of variation of precipitation between the various products on the lower-left half and MAD on the upper-right half of the matrix across the 84 004 valid grid points.

Product	UH	Ensemble	Daymet
UH	—	1.48	1.79
Ensemble	0.85	—	1.03
Daymet	0.77	0.76	—

those regions. It is noteworthy that the observation sites are visible in the UH product when they are sparse, as seen on Hawaii Island with high CV bullseyes. While Daymet predicts occurrence at each grid point and thus more realistically represents PoP, the limited station network, particularly on Hawaii Island, results in an unrealistic CV distribution. Finally, the Ensemble predicts PoP and permits amounts at any given grid point to exceed the observed values due to the inclusion of uncertainty. This appears to result in a more realistic CV distribution.

d. Extreme precipitation

Last, extreme precipitation¹ events and their spatial distribution are particularly important for many aspects of impact modeling (e.g., Gutmann et al. 2014) and may be difficult to capture in gridded products (Gervais et al. 2014). Here we use two metrics to examine extremes across Hawaii. The 99.9th percentile (1 day in 3 years) of all days, and the fraction of precipitation coming from heavy precipitation days, here set as days greater than the 95th percentile of rainy days, denoted as R95pTot (Karl et al. 1999; Gampe and Ludwig 2017; Walton and Grieco 2017). Rainy days are defined as days with ≥ 1 mm of accumulation.

1) 99.9TH PERCENTILE

For the 99.9th percentile shown in Fig. 5, observations highlight that the largest values occur in the rainiest areas, with maximum observed values on Maui and Kauai over 300 mm. Again, the gridded products are able to capture the general distribution of the 99.9th percentile, with the UH and Ensemble products showing the most similarity (Table 5) to each other and the available observations (Fig. 5). Daymet is the most dissimilar of the three products, with underestimation of this extreme on Kauai and Maui and overestimation on the highest terrain of Maui and Hawaii Island. It is also clear that the Ensemble overestimates the 99.9th percentile in heavy precipitation regions. The ensemble and

¹ Note that we use the descriptor extreme for uncommon (e.g., 99.9th percentiles) but not exceedingly rare events.

FIG. 4. Long-term mean daily coefficient of variation for precipitation for observational sites and the three daily time-step products available for Hawaii. Note that the islands of Oahu and Kauai are not drawn to the same scales as the rest of Hawaii.

UH products have good agreement across all individual islands, while Daymet disagrees in its spatial pattern much more across Maui Nui and Kauai.

Again, the interpolation methods and input stations are the key methodological differences. As discussed throughout, Daymet has the sparsest input network, and that clearly influences the ability of that interpolation algorithm to reproduce precipitation maxima (Fig. 2) and any higher-order statistics related to amount. The standard observations available through GHCN-D are particularly lacking on Maui Nui and Kauai. Conversely, the UH product is able to reproduce the extreme values using highly localized IDW, the dense observation network with additional daily observations designed to capture extreme rainfall gradients (e.g., Longman et al. 2018a), and CAI. Although Daymet effectively uses a distance-weighted interpolation scheme as well, its weights are less localized than the UH product, which likely reinforces the underestimation seen in Daymet. Last, even though the Ensemble and UH products use the exact same input stations, slight differences in the station filling commingled with large algorithmic differences result in substantially different estimates of

the 99.9th percentile precipitation both in magnitude and spatial distribution.

2) R95PTOT

Finally, we examine the fraction of precipitation coming from precipitation days greater than the 95th percentile of rainy days, to the total precipitation from all rainy days in Fig. 6 and Table 6. A rainy day is defined as a day with ≥ 1 mm accumulation (CLIMDEX project, index 27; CLIMDEX 2018; Karl et al. 1999; Gampe and Ludwig 2017; Walton and Grieco 2017). This percentile-based metric offers insight into integrated behavior of a product for heavy precipitation days, or the tail of the precipitation distribution, rather than a specific percentile metric such as the 99.9th percentile. It is defined as

$$R95pTot = \frac{\sum_{d=1}^n R_d > P95}{\sum_{d=1}^n R_d}, \quad (1)$$

where n is the number of rainy days, R_d is the accumulation on the d th day, and P95 is the value of

FIG. 5. Long-term daily 99.9th percentile precipitation (mm) for observational sites and the three daily time-step products available for Hawaii. Note that the islands of Oahu and Kauai are not drawn to the same scales as the rest of Hawaii.

precipitation at the 95th percentile when considering only rainy days.

Heavy precipitation contributes less to total precipitation in the areas of least and largest accumulation, with a maximum in fractional contribution in areas with moderate average precipitation rates ($1\text{--}8\text{ mm day}^{-1}$, maximum around 2.5 mm day^{-1}) in the observations. All three products generally reproduce this feature. Qualitative comparisons to observations highlight that the Ensemble generally overestimates, Daymet often underestimates, and the UH product most closely recreates the observed R95pTot distribution. This is borne out in the differences across the products, where the

Daymet and Ensemble products are the most different in terms of spatial correlation and MAD. Additionally, all three products generally disagree in the spatial representation of R95pTot with a maximum spatial correlation of 0.64 between the Ensemble and the UH product. Across the individual islands, positive correlation for Hawaii Island for Daymet–UH and Ensemble–Daymet (0.64, 0.35) and fairly large anticorrelation (-0.54 , -0.56) for Maui Nui, with smaller positive and negative correlations for Oahu and Kauai, are seen. This results in the near-zero aggregate spatial correlation for Daymet–Ensemble and Daymet–UH.

This highlights how methodological choices modify the integrated contribution of the tail of the precipitation distribution. The addition of noise and the transformation steps in the Ensemble system increases the contribution of extreme events to total precipitation (Figs. 5, 6), while the Daymet system reduces the contribution of extreme events (Figs. 5, 6) as compared to the observations. The UH product may best match observations (Figs. 5, 6), but the values between observations in sparsely observed areas may still be in question because of the known limitations of IDW interpolation.

TABLE 5. Spatial correlations of long-term daily 99.9th percentile precipitation between the various products on the lower-left half and MAD (mm) on the upper-right half of the matrix across the 84 004 valid grid points.

Product	UH	Ensemble	Daymet
UH	—	62.2 mm	39.4 mm
Ensemble	0.83	—	67.1 mm
Daymet	0.55	0.58	—

FIG. 6. Long-term daily fractional contribution of precipitation > 95 th percentile for rainy days (R95pTot) for observational sites and the three daily time-step products available for Hawaii. Note that the islands of Oahu and Kauai are not drawn to the same scales as the rest of Hawaii.

For a more in-depth analysis of how this product performs spatially, see [L19](#).

e. Mean temperature distribution

Mean daily air temperature is primarily a function of elevation and precipitation ([Giambelluca et al. 2014](#)) in Hawaii. Daily air temperature is highest at the low elevations on the lee sides of the islands and lowest at the highest elevations, as seen in [Fig. 7](#). As compared to precipitation, many fewer temperature observations are available across the state. Here, we visually compare and contrast to the long-term average of daily data for

117 stations for the 1990–2014 period. Although only roughly 100 long-term temperature stations are available in Hawaii, all products are very similar in their spatial distribution patterns with spatial correlations all at or above 0.98 ([Table 7](#)). This similarity is most likely explained by the strong control elevation exerts on temperature. Two groupings of similar MAD values are seen across the products: all are less than 1.1 K, but PRISM-based comparisons have roughly twice the MAD than similar comparisons between the other datasets (e.g., PRISM vs UH MAD of 0.7 K and Daymet vs UH MAD of 0.4 K).

Methodological differences across the datasets for climatological mean temperature lie primarily in the interpolation techniques. PRISM uses weighted linear regression to determine lapse rate equations with the weights formed from complex relationships across many variables; Daymet also determines a weighted lapse rate equation, but their weighting is determined by distance; the UH product uses a constant piecewise regression equation set across the entire state to determine the climatological pattern; and the Ensemble uses locally weighted multiple linear regression with elevation as

TABLE 6. Spatial correlations of long-term daily fractional contribution of precipitation > 95 th percentile for rainy days (R95pTot) between the various products on the lower-left half and MAD on the upper-right half of the matrix across the 84 004 valid grid points.

Product	UH	Ensemble	Daymet
UH	—	0.05	0.05
Ensemble	0.64	—	0.09
Daymet	0.44	0.11	—

FIG. 7. Long-term daily mean air temperature ($^{\circ}\text{C}$) for observational sites and the four climatological products available for Hawaii. Note that the islands of Oahu and Kauai are not drawn to the same scales as the rest of Hawaii.

one of the predictors for its climatology. Although these approaches are quite different, because they all incorporate elevation as a predictor, they all arrive at nearly the same solution. The other primary difference between the products is the climatological period used. PRISM uses the 1971–2000 period, while the other products use the 1990–2014 period in their CAI (UH and Ensemble), and they are all aggregated up to long-term mean values over the 1990–2014 period. Because the MAD values for PRISM are roughly twice as large as the MAD values for the other products, it could be inferred that the climatological period differences account for

roughly half the differences, while interpolation and station network differences may account for the other half.

f. Mean diurnal temperature range distribution

Mean daily diurnal temperature range (DTR) also varies by elevation across the island chain but is more tightly linked to precipitation and the presence of the TWI than mean temperature. Figure 8 highlights the available observations as well as the four representations of long-term DTR. From the observations, it is clear that windward sides and stations near the coast have lower DTR, which increases toward the TWI, with a decrease in DTR above the TWI. This is because areas above the TWI behave more similar to the free atmosphere, with a less developed boundary layer and more constant near-surface air temperature. The maximum DTR occurs just below the TWI because these areas are further from the coastline and typically less cloudy and moist (Giambelluca et al. 2014; N19).

It is immediately clear the products differ more substantially for DTR than for mean daily air temperature, with spatial correlations ranging from 0.58 to 0.75 and

TABLE 7. Spatial correlations of long-term mean daily 2-m air temperature between the various products on the lower-left half and MAD (K) on the upper-right half of the matrix across the 84 004 valid grid points.

Product	PRISM	UH	Ensemble	Daymet
PRISM	—	0.7 K	1.1 K	0.7 K
UH	0.99	—	0.6 K	0.4 K
Ensemble	0.98	0.99	—	0.6 K
Daymet	0.99	0.99	0.99	—

FIG. 8. Long-term daily mean diurnal temperature range (K) for observational sites and the four climatological products available for Hawaii. Note that the islands of Oahu and Kauai are not drawn to the same scales as the rest of Hawaii.

MAD ranging from 0.8 to 1.4 K (Table 8). Additionally, there is large variation in the individual island spatial correlations, with deviations between the products most evident on Kauai and Maui Nui. The UH product produces a very uniform spatial variation in DTR across the islands with a maximum right around their mean TWI level in the piecewise regression framework. Daymet produces its maximum DTR at the highest elevations. PRISM has a more uniform DTR across all islands, but does highlight higher DTR just below the TWI on Hawaii Island. The Ensemble DTR varies widely across each island; specifically, on Hawaii Island,

the Ensemble places the maximum DTR near the highest elevations, but primarily on the saddle region between and just below the highest peaks.

As discussed for mean air temperature, the regression framework for each product and the input networks are different. This likely contributes a large portion of the spatial pattern differences between the products, especially the UH product. The inability of Daymet to identify the decrease in DTR above the TWI most likely stems from a lack of available observations above the TWI, and thus the weighted lapse rate determination is dominated by station pairs below the TWI. The Ensemble may have a similar issue: it considers 25 stations per grid point and with few high-elevation observations, and stations below the TWI may unduly influence the regression.

The other major methodological difference is that the Ensemble directly estimates DTR, while the other products estimate maximum and minimum temperatures. As discussed in Newman et al. (2015), discrepancies in DTR could arise when independently estimating maximum and minimum temperatures. Here the direct estimation of DTR clearly gives a different picture than estimating

TABLE 8. Spatial correlations of long-term mean daily 2-m diurnal temperature range between the various products on the lower-left half and MAD (K) on the upper-right half of the matrix across the 84 004 valid grid points.

Product	PRISM	UH	Ensemble	Daymet
PRISM	—	1.2 K	1.4 K	0.8 K
UH	0.74	—	1.3 K	1.0 K
Ensemble	0.58	0.63	—	1.2 K
Daymet	0.68	0.75	0.69	—

maximum and minimum temperatures, then computing DTR as the difference. This methodological choice should be examined in more detail moving forward, as DTR is included in many empirical estimates of short-wave radiation (e.g., Thornton and Running 1999), which is relevant for any impact modeling computing the surface energy balance.

5. Summary discussion

This study seeks to provide the broad user community of gridded near-surface observation-based meteorological fields with knowledge of differences across products and understanding of why interproduct differences exist. Hawaii provides a useful testbed to highlight the value and limitations of intercomparison studies. Hawaii now has multiple climatological and daily products available, but limited understanding of their differences, strengths, and weaknesses. This research is important because of the tremendous ecological diversity across the islands and potential climate change impacts (Myers et al. 2000; Sakai et al. 2002). Here we compared four available products (PRISM, UH, Daymet, and Ensemble) across various climate indices that are relevant to many impact modelers.

Key generalizable conclusions emerging from this intercomparison are that 1) the selection of observations, especially as it affects the overall quality and spatial density, significantly influences the results and should be more carefully considered by both end users and developers; 2) basic statistics for both precipitation and temperature (e.g., mean value) have more interproduct and observation–product agreement than higher order statistics (frequency of occurrence or extreme events) in agreement with past studies (Gutmann et al. 2014; Gervais et al. 2014); and 3) attribution of differences to specific methodological choices is difficult.

Our final conclusion emphasizes the primary research challenge: many studies have examined multiple products with metrics similar to those used here (e.g., Gampe and Ludwig 2017; Henn et al. 2018a; Walton and Hall 2018; Beck et al. 2018) and others have developed novel methods of product evaluation for specific applications such as hydrologic modeling (Quintero et al. 2016; Beck et al. 2017; Laiti et al. 2018), yet all come to the same conclusion that it is difficult to attribute differences in products to specific methodological choices. Product X has certain characteristics while product Y has different characteristics, which may be because of reason A or B. While this may identify the best product for the particular user, the general inability to identify specific reasons for interproduct differences limits the usefulness of these types of broad intercomparisons.

Because gridded meteorological product generation has become a complex process with many methodological choices that influence how other choices behave, disentangling and subsequently improving products will only become more difficult without more controlled experiments and comparisons.

Future work is needed to systematically examine individual products or product components. Initial and subsequent product development initiatives generally (but not always) include some type of cross-validation information at the input stations (e.g., Daly et al. 1994; Clark and Slater 2006; Newman et al. 2015; Thornton et al. 2017b; L19). This provides summary statistics of product performance for the method of validation selected by the product development team. Many other studies have examined individual development decisions systematically, particularly the impact of observational density on precipitation sampling (e.g., Rudolf et al. 1994; Hofstra et al. 2010; Schneider et al. 2014; Beguería et al. 2016), and the impact of the spatial interpolation algorithm (e.g., Ly et al. 2011; Wagner et al. 2012; Contractor et al. 2015). Overall product cross validation neglects uncertainty contributions associated with representativeness issues (e.g., sparse input network) and individual methodological choice contributions to the total uncertainty, while evaluating the fidelity of individual components neglects the interplay between methodological development decisions.

To move forward, we need to 1) develop a modular framework that allows users to easily examine the breadth of methodological choices (e.g., Chamberlin 1965; Clark et al. 2011, 2015), 2) collate available high-quality observational datasets for true out-of-sample validation (e.g., Daly 2006; Daly et al. 2017) and make them publicly available (e.g., Longman et al. 2018a), and 3) define application-specific benchmarks of acceptable performance for methodological components and products. This could also include expert review of methodological choices and product outputs (e.g. Daly 2006). We use this study as a call to motivate the community to make progress on these three fronts—the work proposed here will help us improve our understanding of the myriad choices made in product development and help identify the most appropriate methods for specific applications.

Acknowledgments. The U.S. Army Corps of Engineers (USACE) Climate Preparedness and Resilience program funded this work. Color maps used here are provided by Colorbrewer (<http://colorbrewer2.org/>; precipitation difference), the Generic Mapping Tools (GMT; <http://gmt.soest.hawaii.edu/>; temperature difference and precipitation), ESRI (PoP), and the GRID-Arendal project (<http://www.grida.no/>, temperature) via the cpt-city colormap archive (<http://soliton.vm.bytemark.co.uk/pub/cpt-city/>)

and the NCAR Command Language (NCL; https://www.ncl.ucar.edu/Document/Graphics/color_table_gallery.shtml). We would like to acknowledge high-performance computing support from Cheyenne (<https://doi.org/10.5065/D6RX99HX>) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Finally, we thank two anonymous reviewers for their constructive comments that helped us improve this manuscript.

REFERENCES

- Beck, H. E., and Coauthors, 2017: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201–6217, <https://doi.org/10.5194/hess-21-6201-2017>.
- , and Coauthors, 2018: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2018-481>.
- Beguería, S., S. M. Vicente-Serrano, M. Tomais-Burguera, and M. Maneta, 2016: Bias in the variance of gridded data sets leads to misleading conclusions about changes in climate variability. *Int. J. Climatol.*, **36**, 3413–3422, <https://doi.org/10.1002/joc.4561>.
- Behnke, R., S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, 2016: Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecol. Appl.*, **26**, 1338–1351, <https://doi.org/10.1002/15-1061>.
- Chamberlin, T. C., 1965: The method of multiple working hypotheses. *Science*, **148**, 754–759, <https://doi.org/10.1126/science.148.3671.754>.
- Chu, P.-S., and H. Chen, 2005: Interannual and interdecadal precipitation variations in the Hawaiian Islands. *J. Climate*, **18**, 4796–4813, <https://doi.org/10.1175/JCLI3578.1>.
- Clark, M. P., and A. G. Slater, 2006: Probabilistic quantitative precipitation estimation in complex terrain. *J. Hydrometeorol.*, **7**, 3–22, <https://doi.org/10.1175/JHM474.1>.
- , D. Kavetski, and F. Fenicia, 2011: Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resour. Res.*, **47**, W09301, <https://doi.org/10.1029/2010WR009827>.
- , and Coauthors, 2015: A unified approach to hydrologic modeling: Part 1. Modeling concept. *Water Resour. Res.*, **51**, 2498–2514, <https://doi.org/10.1002/2015WR017198>.
- CLIMDEX, 2018: Climate extremes indices. Accessed 2 April 2018, <https://www.climdex.org/indices.html>.
- Contractor, S., L. V. Alexander, M. G. Donat, and N. Herold, 2015: How well do gridded datasets of observed daily precipitation compare over Australia? *Adv. Meteorol.*, **2015**, 325718, <http://dx.doi.org/10.1155/2015/325718>.
- Daly, C., 2006: Guidelines for assessing the suitability of spatial climate data sets. *Int. J. Climatol.*, **26**, 707–721, <https://doi.org/10.1002/joc.1322>.
- , R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteorol.*, **33**, 140–158, [https://doi.org/10.1175/1520-0450\(1994\)033<0140:ASTMFM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2).
- , W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Res.*, **22**, 99–113, <https://doi.org/10.3354/cr022099>.
- , E. H. Helmer, and M. Quinones, 2003: Mapping the climate of Puerto Rico, Vieques, and Culebra. *Int. J. Climatol.*, **23**, 1359–1381, <https://doi.org/10.1002/joc.937>.
- , J. Smith, M. Doggett, M. Halbleib, and W. Gibson, 2006: High-resolution climate maps for the Pacific Basin Islands, 1971–2000. National Park Service Pacific West Regional Office Rep., 59 pp.
- , M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. A. Pasteris, 2008: Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, <https://doi.org/10.1002/joc.1688>.
- , M. E. Slater, J. A. Roberti, S. H. Laseter, and L. W. Swift Jr., 2017: High-resolution precipitation mapping in a mountainous watershed: ground truth for evaluating uncertainty in a national precipitation dataset. *Int. J. Climatol.*, **37**, 124–137, <https://doi.org/10.1002/joc.4986>.
- Dawdy, D. R., and W. B. Langbein, 1960: Mapping mean areal precipitation. *Hydrol. Sci. J.*, **5**, 16–23, <https://doi.org/10.1080/02626666009493176>.
- Day, G. N., 1985: Extended streamflow forecasting using NWSRFS. *J. Water Resour. Plann. Manage.*, **111**, 157–170, [https://doi.org/10.1061/\(ASCE\)0733-9496\(1985\)111:2\(157\)](https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)).
- Derin, Y., and Coauthors, 2016: Multiregional satellite precipitation products evaluation over complex terrain. *J. Hydrometeorol.*, **17**, 1817–1836, <https://doi.org/10.1175/JHM-D-15-0197.1>.
- Duan, Q., and Coauthors, 2006: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.*, **320**, 3–17, <https://doi.org/10.1016/j.jhydrol.2005.07.031>.
- 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. Meteorol.*, **39**, 1580–1591, [https://doi.org/10.1175/1520-0450\(2000\)039<1580:CASCND>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<1580:CASCND>2.0.CO;2).
- Franklin, J., 1995: Predictive vegetation mapping: geographical modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.*, **19**, 474–499, <https://doi.org/10.1177/030913339501900403>.
- Frazier, A. G., T. W. Giambelluca, H. F. Diaz, and H. L. Needham, 2016: Comparison of geostatistical approaches to spatially interpolate month-year rainfall for the Hawaiian Islands. *Int. J. Climatol.*, **36**, 1459–1470, <https://doi.org/10.1002/joc.4437>.
- Gampe, D., and R. Ludwig, 2017: Evaluation of gridded precipitation data products for hydrological applications in complex topography. *Hydrology*, **4**, 53, <https://doi.org/10.3390/hydrology4040053>.
- Garza, J. A., P. S. Chu, C. W. Norton, and T. A. Schroeder, 2012: Changes of the prevailing trade winds over the islands of Hawaii and the North Pacific. *J. Geophys. Res.*, **117**, D11109, <https://doi.org/10.1029/2011JD016888>.
- 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, <https://doi.org/10.1175/JCLI-D-13-00319.1>.
- Giambelluca, T. W., M. A. Nullet, and T. A. Schroeder, 1986: Rainfall atlas of Hawai'i. Rep. R76, Hawaii Division of Water and Land Development, Department of Land and Natural Resources, 267 pp.
- , Q. Chen, A. G. Frazier, J. P. Price, Y.-L. Chen, P.-S. Chu, and J. K. Eischeid, 2011: The rainfall atlas of Hawai'i. Final Rep., Department of Geography, University of Hawai'i at Manoa, 72 pp., <http://rainfall.geography.hawaii.edu/assets/files/PDF/RF%20Atlas%20Final%20Report.pdf>.
- , —, —, —, —, J. K. Eischeid, and D. M. Delparte, 2013: Online rainfall atlas of Hawai'i. *Bull. Amer. Meteor. Soc.*, **94**, 313–316, <https://doi.org/10.1175/BAMS-D-11-00228.1>.

- , and Coauthors, 2014: Evapotranspiration of Hawai'i. Final Rep. submitted to the U.S. Army Corps of Engineers–Honolulu District and the Commission on Water Resource Management, 178 pp., <http://evapotranspiration.geography.hawaii.edu/assets/files/PDF/ET%20Project%20Final%20Report.pdf>.
- 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, <https://doi.org/10.1002/2014WR015559>.
- Henderson-Sellers, A., Z.-L. Yang, and R. E. Dickinson, 1993: The Project for Intercomparison of Land-Surface Schemes. *Bull. Amer. Meteor. Soc.*, **74**, 1335–1349, [https://doi.org/10.1175/1520-0477\(1993\)074<1335:TPFIOL>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<1335:TPFIOL>2.0.CO;2).
- , A. J. Pitman, P. K. Love, P. Irannejad, and T. H. Chen, 1995: The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.*, **76**, 489–503, [https://doi.org/10.1175/1520-0477\(1995\)076<0489:TPFIOL>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<0489:TPFIOL>2.0.CO;2).
- Henn, B., A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, 2018a: An assessment of differences in gridded precipitation datasets in complex terrain. *J. Hydrol.*, **556**, 1205–1219, <https://doi.org/10.1016/j.jhydrol.2017.03.008>.
- , M. P. Clark, D. Kavetski, A. J. Newman, M. Hughes, B. McGurk, and J. D. Lundquist, 2018b: Spatiotemporal patterns of precipitation inferred from streamflow observations across the Sierra Nevada mountain range. *J. Hydrol.*, **556**, 993–1012, <https://doi.org/10.1016/j.jhydrol.2016.08.009>.
- Hofstra, N., M. Haylock, M. New, and P. D. Jones, 2009: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *J. Geophys. Res.*, **114**, D21101, <https://doi.org/10.1029/2009JD011799>.
- , M. New, and C. McSweeney, 2010: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dyn.*, **35**, 841–858, <https://doi.org/10.1007/s00382-009-0698-1>.
- Isotta, F. A., and Coauthors, 2014: The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *Int. J. Climatol.*, **34**, 1657–1675, <https://doi.org/10.1002/joc.3794>.
- Karl, T. R., N. Nicholls, and A. Ghazi, 1999: CLIVAR/GCOS/ WMO workshop on indices and indicators for climate extremes: Workshop summary. *Climatic Change*, **42**, 3–7, <https://doi.org/10.1023/A:1005491526870>.
- Klein Tank, A., and Coauthors, 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol.*, **22**, 1441–1453, <https://doi.org/10.1002/joc.773>.
- Koster, R. D., and P. C. D. Milly, 1997: The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *J. Climate*, **10**, 1578–1591, [https://doi.org/10.1175/1520-0442\(1997\)010<1578:TIBTAR>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1578:TIBTAR>2.0.CO;2).
- Kuczera, G., B. Renard, M. Thyer, and D. Kavetski, 2010: There are no hydrological monsters, just models and observations with large uncertainties! *Hydrol. Sci. J.*, **55**, 980–991, <https://doi.org/10.1080/02626667.2010.504677>.
- Laiti, L., S. Mallucci, S. Piccolroaz, A. Bellin, D. Zardi, A. Fiori, G. Nikulin, and B. Majone, 2018: Testing the hydrological coherence of high-resolution gridded precipitation and temperature data sets. *Water Resour. Res.*, **54**, 1999–2016, <https://doi.org/10.1002/2017WR021633>.
- Leopold, L. B., 1949: The interactions of trade wind and sea breeze in Hawaii. *J. Meteor.*, **6**, 312–320, [https://doi.org/10.1175/1520-0469\(1949\)006<0312:TIOTWA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1949)006<0312:TIOTWA>2.0.CO;2).
- Liu, C., and Coauthors, 2017: Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dyn.*, **49**, 71–95, <https://doi.org/10.1007/s00382-016-3327-9>.
- Livneh, B., T. J. Bohn, D. W. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, D. R. Cayan, and L. Brekke, 2015: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Sci. Data*, **2**, 150042, <https://doi.org/10.1038/sdata.2015.42>.
- Longman, R. J., H. F. Diaz, and T. W. Giambelluca, 2015: Sustained increase in lower tropospheric subsidence over the central tropical North Pacific. *J. Climate*, **28**, 8743–8759, <https://doi.org/10.1175/JCLI-D-15-0006.1>.
- , and Coauthors, 2018a: Compilation of climate data from heterogeneous networks across the Hawaiian Islands. *Sci. Data*, **5**, 180012, <https://doi.org/10.1038/sdata.2018.12>.
- , A. G. Frazier, A. J. Newman, T. W. Giambelluca, A. Kagawa-Viviani, and M. P. Clark, 2018b: Deterministic gridded (250 m) daily precipitation and temperature for the Hawaiian Islands (1990–2014). University of Hawai'i Dept. of Geography and Environment and UCAR/NCAR-CISL-RDA, <https://doi.org/10.5065/D6X065VV>.
- , —, —, —, D. Schanzebach, A. Kagawa-Viviani, H. Needham, J. R. Arnold, and M. P. Clark, 2019: High-resolution gridded daily rainfall and temperature for the Hawaiian Islands (1990–2014). *J. Hydrometeorol.*, **20**, 489–508, <https://doi.org/10.1175/JHM-D-18-0112.1>.
- Ly, S., C. Charles, and A. Degre, 2011: Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrol. Earth Syst. Sci.*, **15**, 2259–2274, <https://doi.org/10.5194/hess-15-2259-2011>.
- Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. Da Fonseca, and J. Kent, 2000: Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858, <https://doi.org/10.1038/35002501>.
- Newman, A. J., and Coauthors, 2015: Gridded ensemble precipitation and temperature estimates for the contiguous United States. *J. Hydrometeorol.*, **16**, 2481–2500, <https://doi.org/10.1175/JHM-D-15-0026.1>.
- , M. P. Clark, R. J. Longman, and T. W. Giambelluca, 2018: Ensemble gridded (1 km) daily precipitation and temperature for the Hawaiian Islands (1990–2014). UCAR/NCAR-CISL-RDA, <https://doi.org/10.5065/D6SB44JV>.
- , —, —, E. Gilleland, T. W. Giambelluca, and J. R. Arnold, 2019: Use of daily station observations to produce high-resolution gridded probabilistic precipitation and temperature time series for the Hawaiian Islands. *J. Hydrometeorol.*, **20**, 509–529, <https://doi.org/10.1175/JHM-D-18-0113.1>.
- Nijssen, B., and Coauthors, 2003: Simulation of high latitude hydrological processes in the Torne–Kalix basin: PILPS Phase 2(e): 2. Comparison of model results with observations. *Global Planet. Change*, **38**, 31–53, [https://doi.org/10.1016/S0921-8181\(03\)00004-3](https://doi.org/10.1016/S0921-8181(03)00004-3).
- Pierce, D. W., D. R. Cayan, and B. L. Thrasher, 2014: Statistical downscaling using localized constructed analogs (LOCA). *J. Hydrometeorol.*, **15**, 2558–2585, <https://doi.org/10.1175/JHM-D-14-0082.1>.
- Prein, A., and A. Gobiet, 2017: Impacts of uncertainties in European gridded precipitation observations on regional

- climate analysis. *Int. J. Climatol.*, **37**, 305–327, <https://doi.org/10.1002/joc.4706>.
- Quintero, F., W. F. Krajewski, R. Mantilla, S. Small, and B. Seo, 2016: A spatial–dynamical framework for evaluation of satellite rainfall products for flood prediction. *J. Hydrometeorol.*, **17**, 2137–2154, <https://doi.org/10.1175/JHM-D-15-0195.1>.
- Raimonet, M., L. Oudin, V. Thieu, M. Silvestre, R. Vautard, C. Rabouille, and P. Le Moigne, 2017: Evaluation of gridded meteorological datasets for hydrological modeling. *J. Hydrometeorol.*, **18**, 3027–3041, <https://doi.org/10.1175/JHM-D-17-0018.1>.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks, 2010: Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. *Water Resour. Res.*, **46**, W05521, <https://doi.org/10.1029/2009WR008328>.
- Rudolf, B., H. Hauschild, W. Rueth, and U. Schneider, 1994: Terrestrial precipitation analysis: Operational method and required density of point measurements. *Global Precipitations and Climate Change*, M. Desbois and F. Deisalmund, Eds., Springer, 173–186, https://doi.org/10.1007/978-3-642-79268-7_10.
- Sakai, A. K., W. L. Wagner, and L. A. Mehrhoff, 2002: Patterns of endangerment in the Hawaiian flora. *Syst. Biol.*, **51**, 276–302, <https://doi.org/10.1080/10635150252899770>.
- Sanderson, M., 1993: *Prevailing Trade Winds: Climate and Weather in Hawaii*. University of Hawaii Press, 127 pp.
- Schlosser, C. A., and Coauthors, 2000: Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS phase 2(d). *Mon. Wea. Rev.*, **128**, 301–321, [https://doi.org/10.1175/1520-0493\(2000\)128<0301:SOABGH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0301:SOABGH>2.0.CO;2).
- Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.*, **115**, 15–40, <https://doi.org/10.1007/s00704-013-0860-x>.
- Thornton, P. E., and S. W. Running, 1999: An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agric. For. Meteorol.*, **93**, 211–228, [https://doi.org/10.1016/S0168-1923\(98\)00128-9](https://doi.org/10.1016/S0168-1923(98)00128-9).
- , —, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.*, **190**, 214–251, [https://doi.org/10.1016/S0022-1694\(96\)03128-9](https://doi.org/10.1016/S0022-1694(96)03128-9).
- , M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, and R. B. Cook, 2017a: Daymet: Daily Surface Weather Data on a 1-km Grid for North America, version 3. ORNL DAAC, accessed 7 February 2019, <https://doi.org/10.3334/ORNLDAAAC/1328>.
- , —, and R. S. Vose, 2017b: Daymet: Annual tile summary cross-validation statistics for North America, version 3. ORNL DAAC, accessed 11 May 2017, <https://doi.org/10.3334/ORNLDAAAC/1348>.
- USBR, 2012: Colorado River basin water supply and demand study. U.S. Bureau of Reclamation, 1697 pp., <https://www.usbr.gov/lc/region/programs/crbstudy.html>.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson, 2008: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resour. Res.*, **44**, W00B09, <https://doi.org/10.1029/2007WR006720>.
- Wagner, P. D., P. Fiener, F. Wilken, S. Kumar, and K. Schneider, 2012: Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *J. Hydrol.*, **464–465**, 388–400, <https://doi.org/10.1016/j.jhydrol.2012.07.026>.
- Walton, D., and M. Grieco, 2017: Evaluating Extreme precipitation in gridded datasets: A case study over California. *2017 Fall Meeting*, New Orleans, LA, Amer. Geophys. Union, Abstract A41C-2295.
- , and A. Hall, 2018: An assessment of high-resolution gridded temperature datasets over California. *J. Climate*, **31**, 3789–3810, <https://doi.org/10.1175/JCLI-D-17-0410.1>.
- Willmott, C. J., and S. M. Robeson, 1995: Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.*, **15**, 221–229, <https://doi.org/10.1002/joc.3370150207>.