

Selecting the Optimal Fine-Scale Historical Climate Data for Assessing Current and Future Hydrological Conditions

M. A. STERN,^a L. E. FLINT,^b A. L. FLINT,^b R. M. BOYNTON,^c J. A. E. STEWART,^c J. W. WRIGHT,^d AND J. H. THORNE^c

^a *U.S. Geological Survey, Sacramento, California*

^b *Earth Knowledge, Inc., Davis, California*

^c *Department of Environmental Science and Policy, University of California, Davis, Davis, California*

^d *Pacific Southwest Research Station, USDA Forest Service, Davis, California*

(Manuscript received 4 March 2021, in final form 13 December 2021)

ABSTRACT: High-resolution historical climate grids are readily available and frequently used as inputs for a wide range of regional management and risk assessments, including water supply, ecological processes, and as baseline for climate change impact studies that compare them to future projected conditions. Because historical gridded climates are produced using various methods, their portrayal of landscape conditions differ, which becomes a source of uncertainty when they are applied to subsequent analyses. Here we tested the range of values from five gridded climate datasets. We compared their values to observations from 1231 weather stations, first using each dataset's native scale, and then after each was rescaled to 270-m resolution. We inputted the downscaled grids to a mechanistic hydrology model and assessed the spatial results of six hydrological variables across California, in 10 ecoregions and 11 large watersheds in the Sierra Nevada. PRISM was most accurate for precipitation, ClimateNA for maximum temperature, and TopoWx for minimum temperature. The single most accurate dataset overall was PRISM due to the best performance for precipitation and low air temperature errors. Hydrological differences ranged up to 70% of the average monthly streamflow with an average of 35% disagreement for all months derived from different historical climate maps. Large differences in minimum air temperature data produced differences in modeled actual evapotranspiration, snowpack, and streamflow. Areas with the highest variability in climate data, including the Sierra Nevada and Klamath Mountains ecoregions, also had the largest spread for snow water equivalent, recharge, and runoff.

SIGNIFICANCE STATEMENT: Gridded historical climate datasets are vital inputs for hydrological and other models used to quantify current water supply, drought risk, and other ecosystem processes. They are also compared to future climate projections to assess the future climate change risk. Numerous interpolated climate datasets are available with varying resolution and methods, yet all are based on climate station data and represent the same historical record at and in between each station. We found significant disagreement between historical gridded datasets, including the one used to bias correct the current climate change projections used by the state of California. Some datasets had large biases compared to station data, especially in snow-dominated regions, leading to large disagreements in modeled monthly streamflow.

KEYWORDS: North America; Hydrology; Hydrologic models; Model evaluation/performance; Uncertainty; Downscaling

1. Introduction

High-resolution historical gridded climate datasets are widely used to drive and calibrate hydrological and ecological models, which are commonly used for water supply forecasting, drought indices, and wildfire prediction models, and are important for other process-based correlation studies. Yet, historical gridded climate datasets may vary in their portrayal of landscape conditions, creating a need to determine their

accuracy and the differences among them. Previous studies have compared various climate products for different geographies in the United States and globally for hydrological or ecological models or strictly for methodological comparisons and have commonly found some agreement among datasets but also systematic biases and substantial amounts of uncertainty across different regions (Behnke et al. 2016; Walton and Hall 2018; Eum et al. 2014; Sun et al. 2014; Yin et al. 2015), resulting in differences in estimated crop yield (Parkes et al. 2019), ecological models (Behnke et al. 2016; Tang et al. 2018), drought metrics (Ahmadalipour and Moradkhani 2017), and hydrological models (Asong et al. 2020; Muche et al. 2020; Eum et al. 2014).

Many hydroclimatic comparisons only consider precipitation in their comparison (i.e., Henn et al. 2018; Hughes et al. 2020; Lundquist et al. 2015; Muche et al. 2020), and comparisons are rarely done at a fine scale (<1 km) or using down-scaled inputs. In addition, previous work in this area has

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-21-0045.s1>.

Corresponding author: Michelle Stern, mstern@usgs.gov

DOI: 10.1175/JHM-D-21-0045.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

mainly focused on lumped-parameter hydrologic models like the Variable Infiltration Capacity (VIC; Liang et al. 1994) or the Soil and Water Assessment Tool (SWAT; Arnold et al. 1998), which minimize the potential benefit of finer-scale climate data to the accuracy of snowpack and streamflow results. Dynamic models such as the Weather Research and Forecasting (WRF; Skamarock et al. 2008) are powerful due to their ability to couple two-way atmospheric and hydrological processes, yet can be limited in areas of complex topography and climate since they are generally not run at fine-scale resolutions (<1 km) for long periods or large areas due to high computational costs.

Newman et al. (2015) found that differences among gridded climate datasets could be attributed to specific methodological choices that can limit the usefulness of intercomparisons. But, many comparisons have been conducted because of the importance of hydrology and climate as the driving information for models that inform resource planning, ecosystem dynamics, public safety, and other objectives. This is especially relevant for studies involving future climate scenarios. For example, Tang et al. (2018) found that the baseline climate dataset can contribute 40% of the total variation among future projections in a species distribution modeling framework. They emphasized the importance of considering the sensitivity of future projections of species distributions to the choice of baseline climate information, especially in mountainous environments with complex climatic gradients.

There are two main methods used to create gridded climate surfaces 1) by using weather station data and interpolating with geostatistical techniques or 2) with reanalysis products that use atmospheric models that assimilate historical observations like remote sensing and station data. Typically, the reanalysis products are too spatially coarse to use in fine-scale applications, so they are downscaled using geostatistical techniques or by a regional climate model. Walton and Hall (2018) examined air temperature from station-based, reanalysis, and hybrid gridded datasets and found that reanalysis-based datasets have systematic biases relative to station data and were therefore less suitable for capturing climatologies than the station-based datasets that capture variable lapse rates. Therefore, this study focuses only on station-based datasets.

We were motivated to conduct this study to select the most accurate historical climate dataset for assessment of future climate change risks for water availability and forestry programs in California. We approached this evaluation by selecting five commonly used, fine-scale gridded historical climate datasets to assess their accuracy in portraying precipitation and minimum and maximum temperature recorded at weather stations, and to compare modeled differences among their outputs when each was run through the Basin Characterization Model (Flint et al. 2021a), a mechanistic landscape hydrology model. Our dataset selection criteria were based on a literature review of publicly available, station-based climate datasets commonly used in hydrologic modeling. We selected datasets that were available at a monthly time step from 1980 to 2013 at a relatively fine spatial resolution (<6 km) and that covered the conterminous United States at a minimum. We also downscaled the datasets to 270-m

resolution to quantify the impact of scale on errors and bias against station data arrayed across a complex hydroclimatic landscape. To examine the uncertainty introduced to landscape hydrology, we inputted the downscaled gridded datasets to a fine-scale distributed-parameter water balance model. To do this, we investigated various input climate data differences in long-term hydrological variables and monthly streamflow timing and volume.

2. Methods

We conducted five analyses. First, we tested the accuracy of maximum monthly air temperature, minimum monthly air temperature, and monthly precipitation from five historical-climate gridded datasets to the records from 1231 weather stations. Second, we downscaled four of the gridded datasets to 270 m and repeated the accuracy assessment to test for the effect of scale. Third, we ran the four downscaled datasets through a mechanistic hydrology model to portray differences in the effects of the climate data on landscape-scale hydrology. We compared the six hydrologic variables resulting from the input of the four climate datasets: recharge, runoff, soil storage, 1 April snowpack, actual evapotranspiration, and climatic water deficit. We assessed the differences in outputs at the statewide level for 10 ecoregions and also compared the six hydrologic variables and streamflow for 11 major watersheds defined by the California Department of Water Resources (CDWR 2007) as full natural flow (FNF) basins in California's Sierra Nevada, a vital source of water for the state.

a. Study area

Our study area consists of California and the basins that flow into it, a region with diverse elevation, geology, hydrology, and atmospheric processes. Precipitation in California varies more from year to year than anywhere else in the conterminous United States (Dettinger 2016), from 0 to 1625 mm yr^{-1} . A critical component of water supply in California is the snowpack. One of the greatest influences on streamflow sensitivity to climate change is the location of the rainfall/snow elevation line (Miller et al. 2003), which has a substantial effect on snowpack. Therefore, accurately representing precipitation and temperature across elevation gradients is essential for developing accurate hydrological models in California.

We divided our study area into 10 U.S. Environmental Protection Agency (EPA)-defined North America Level III Ecoregions (ecoregions, https://gaftp.epa.gov/EPADDataCommons/ORD/Ecoregions/cec_na/NA_LEVEL_III.pdf) (Fig. 1) that represent climate and ecological variation across the state. Ecoregions were used to summarize accuracy statistics for precipitation and minimum temperature against station data and to compare spatial differences in long-term average climate and hydrological variables across California. Due to a lack of station density, the Sonoran Desert, Mojave Basin and Range, and Central Basin and Range were combined and named Southern Basin and Range. The California Coastal Sage, Chaparral, and Oak Woodlands region was renamed Southern and Central California Plains and Hills. Additionally, 11

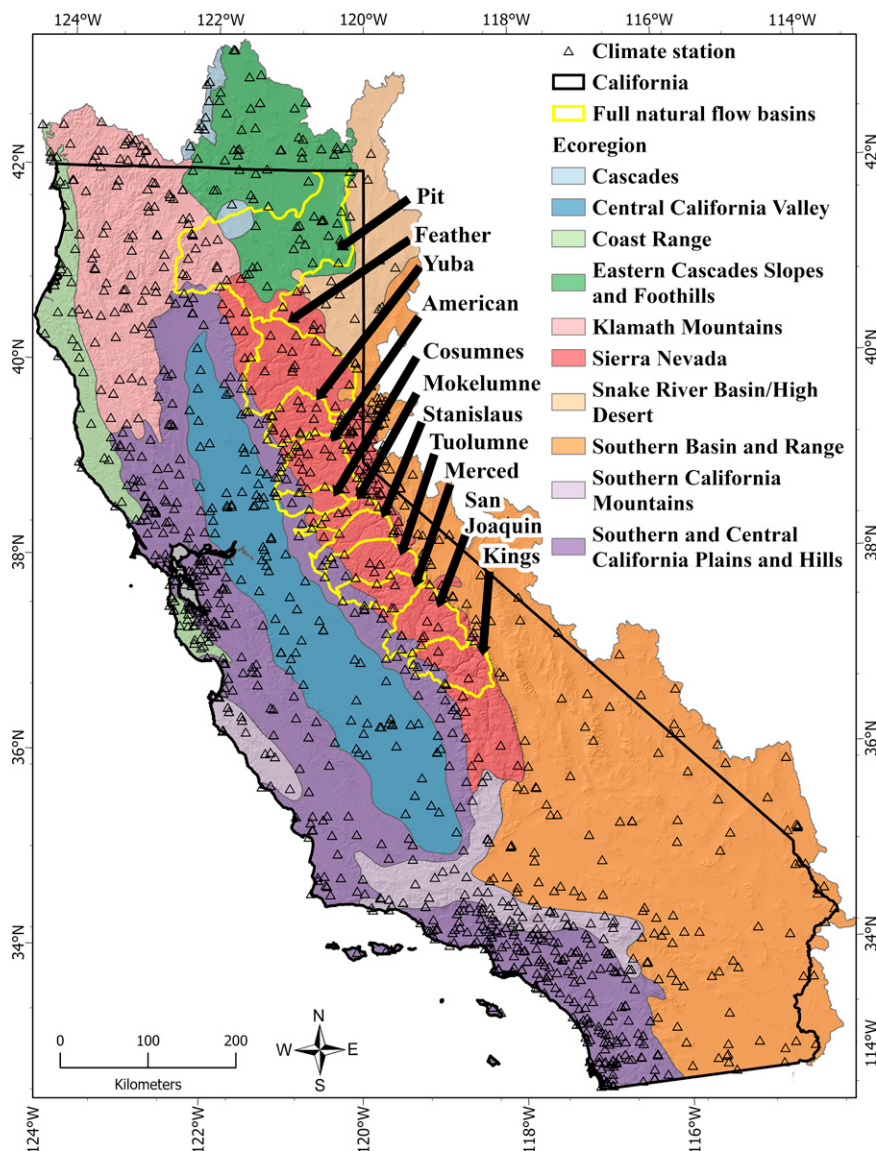


FIG. 1. Study area, climate station locations, U.S. Environmental Protection Agency (EPA)-defined North America Level III ecoregions (ecoregions), and California Department of Water Resources (CDWR) full natural flow (FNF) hydrological basins.

watersheds located in the Sierra Nevada, the state's critical water supply region, were delineated to assess hydrological uncertainty resulting from climate datasets. The 11 FNF basins (Fig. 1) were used to summarize disagreement between climate datasets and the amplification of uncertainty through hydrologic variables and assess variability in monthly streamflow.

b. Climate datasets

The five climate datasets compared in this study are widely used in hydrological, ecological, and many other applications in North America (Table 1). Each dataset fundamentally was created by interpolating station data and distributing values spatially

over the landscape. However, each dataset has a different spatial resolution, time step, variables, footprint, lapse rate (topographic induced variability), and interpolation method. Each dataset interpolates a different set of station data from various networks across North America and has performed its own quality controls and data processing including homogenization, the length of record threshold to include a weather station, and removal of outliers and erroneous data. Homogenization of station data can correct for changes in time-of-day observation practices, instrumentation calibrations, and location changes. Most of the datasets included as many stations as were available on any given day; and therefore, stations can come in and out of being included. Some interpolation methods are sensitive to this, and it

TABLE 1. Gridded station-based climate datasets used in this study, including descriptive information and citation. ppt = precipitation, tmn = minimum temperature, tmx = maximum temperature, m = meters, km = kilometers, CONUS = continental United States, LST = land surface temperature.

	PRISM (M3)	TopoWX	DayMet (Version 3)	Livneh	ClimateNA
Native resolution	4 km and 800 m	800 m	1 km	6 km	Scale-free/1 km
Available period	1895–2021	1948–2016	1980–2017	1915–2013	1901–2021
Time step	Daily, monthly	Daily, monthly	Daily, monthly	Daily, monthly	Monthly
Variables available	ppt, tmn, tmx	tmn, tmx	ppt, tmn, tmx	ppt, tmn, tmx	ppt, tmn, tmx
Lapse rate	Determined based on nearby stations	Determined based on nearby stations	Determined based on nearby stations	6.5°C km ⁻¹	Variable
Footprint	CONUS	North America	North America, Hawaii, Puerto Rico	CONUS, Canada south of 53°N, Mexico	North America
Downscaling/ interpolation method	Elevation regression model with stations weighted based on multiple physical factors	Regression kriging for climate normals with auxiliary predictors including lat, lon, elev, satellite LST; moving window geographically weighted regression	Truncated Gaussian filter combined with elevation regression	SYMAP algorithm: inverse distance weighting with directional adjustment	Bilinear interpolation and local elevation adjustments (PRISM or WorldClim for 30-yr means)
Data available	prism.oregonstate.edu; Flint et al. (2021b)	www.scribhub.org/resources/topowx/	www.daymet.ornl.gov	www.psl.noaa.gov/data/gridded/data.livneh.html	climatena.ca

can cause large biases or shifts in climatologies (Walton and Hall 2018). We briefly introduce each gridded dataset. For more detailed information on each dataset, the reader is directed to the provided citations.

1) PRISM

Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994, 2008) is a widely used gridded climate dataset that incorporates distance and clustering of stations, elevation, coastal proximity, topographic position, effective terrain height, topographic facet, and vertical layer in its interpolation algorithm. PRISM is available for the conterminous United States (CONUS) at a daily or monthly time step with an 800-m or 4-km resolution. PRISM data used in this study were the recently updated M3 monthly dataset (https://prism.oregonstate.edu/documents/PRISM_datasets.pdf), down-scaled from 4 km to 270 m. The (M3) release in October 2019 included NCAR–NCEP Reanalysis grids and employed a day-shifting algorithm, among other improvements, to reduce spatial anomalies and improve the accuracy of high-elevation air temperature and the ability to render unusual temperature inversions.

2) TOPOWX

TopoWX (Oyler et al. 2015) is a temperature-only dataset that incorporates daily station data, atmospheric reanalysis data, and MODIS land temperature data. TopoWX is available at an 800-

m resolution for North America for 1948–2016. TopoWX is the only dataset considered in this study that performed homogenization by filling missing values using neighboring observations and spatial regressions. Homogenization of station data can be problematic and over smooth trend fields like in the Hamlet dataset (Hamlet and Lettenmaier 2005), which is not considered in this study due to these known issues. However, oversmoothing has been determined to not be an issue with TopoWX (Walton and Hall 2018). Although TopoWX only provides minimum and maximum temperature, it was included in this study because it is not often used in climate comparison studies, provides a fine-scale temperature dataset from homogenized station data, and can be used to test with other precipitation data to drive a hydrologic model.

3) DAYMET

The DayMet (Thornton et al. 1997) dataset consists of a 1-km resolution daily or monthly climate product for North America for 1980–2017. DayMet employs a smooth curve to fit the Global Historical Climatology Network-Daily (GHCN-D; Menne et al. 2012) station data to a grid using the weighted average of nearby stations. The weights are determined by a truncated Gaussian filter with a radius that varies continuously throughout the domain to adjust for varying station density. Minimum and maximum temperature values are adjusted for elevation using a linear temperature–elevation relationship.

4) LIVNEH

The Livneh (Livneh et al. 2015) dataset is an updated and extended version of the Maurer dataset (Maurer et al. 2002), increasing the period from 1950–2000 to 1915–2013 and the spatial resolution from 12 to 6 km. This dataset is available at a daily or monthly time step and covers CONUS, Mexico, and Canada south of 53°N. The gridding methodology employs the synergistic mapping system (SYMAP; Shepard 1984), in which temperature is calculated as a weighted average using the four nearest stations. The weights are determined by inverse distance weighting and a directional adjustment. Livneh uses a fixed lapse rate of $6.5^{\circ}\text{C km}^{-1}$ and is the only dataset in this study that did not vary the lapse rate based on nearby stations.

5) CLIMATE_{NA}

Climate_{NA} (Wang et al. 2016) is a “scale-free” climate product that utilizes PRISM and WorldClim normals (1971–2000 averages), and station data from across North America, and is available from 1901 to 2021. Climate_{NA} uses a bilinear interpolation of station data and local elevation adjustments using the long-term averages to produce monthly data. Climate_{NA} is commonly used in ecological applications, and its online and stand-alone application makes point extraction user friendly. Point extractions of monthly time series at station locations for statistical analyses were done using the Climate_{NA} executable version 6.21. Extracting monthly gridded data from the application for the study area was attempted but was not possible at the scale and resolution of this study due to technical problems with the program. Due to this technical issue, downscaling monthly climate grids was not possible for this study. Long-term mean (1981–2010) gridded maps of temperature and precipitation were available online at 1-km resolution and were compared to other long-term average maps.

c. Station data

To quantitatively test each gridded dataset for accuracy, we used a novel, station-based and serially complete dataset of daily precipitation and maximum and minimum temperature for North America [Serially Complete Dataset North America (SCDNA); Tang et al. 2020] from 1979 to 2018. To reduce the inherent uncertainty in station data for validation, we used a published dataset that has employed extensive quality control and gap-filling techniques to develop a robust, temporally continuous time series for each station in North America. Raw gauge data in the SCDNA originated from the GHCN-D, the Global Surface Summary of the Day (GSOD), the Environment and Climate Change Canada (ECCC), and the Livneh et al. (2015) database compiled for stations in Mexico. Although these raw datasets are used in the other gridded datasets, the SCDNA is uniquely equipped to quantify the accuracy of the chosen gridded datasets since the gridded datasets do not perform the same quality assurance measures or employ gap-filling or homogenization for temporal consistency. We extracted data for 1231 stations from the SCDNA (Fig. 1), averaged the daily temperature data into monthly values and summed precipitation data monthly, to compare with

each monthly gridded climate product. Other qualitative and quantitative comparisons are considered in this study to better understand differences between each climate dataset and the uncertainty these differences can add to hydrological modeling.

d. Watershed model

The Basin Characterization Model (BCM; Flint et al. 2021a) is a distributed fine-scale (270 m) water balance model that requires temporally and spatially continuous climate grids to calculate the water balance for each grid cell on a monthly time step. The BCM calculates unimpaired or natural hydrologic variables from downscaled climate data and potential evapotranspiration (PET) calculated using a modified Priestley–Taylor equation (Flint and Childs 1991; Priestley and Taylor 1972) together with topographic shading and cloudiness (Bristow and Campbell 1984). The snow accumulation and melt calculated in the BCM is based on the SNOW-17 model (Anderson 2006) used operationally by the National Weather Service (NWS) for flood forecasting.

BCM hydrological variables have been examined and applied in numerous publications in the western United States (Curtis et al. 2021, Flint et al. 2013, Thorne et al. 2012, Flint et al. 2012, Torregrosa et al. 2020). The California-wide BCM (Flint et al. 2021b) has been calibrated/validated to solar radiation, potential evapotranspiration, actual evapotranspiration (Reitz et al. 2017), snow water equivalent (SWE) to 99 National Resources Conservation Service (NRCS) Snowpack Telemetry (SNOTEL) stations (wcc.nrcs.usda.gov/snow/), and over 100 streamflow gauges from the USGS National Water Information System (U.S. Geological Survey 2021) across California. Distributed SNOW-17 parameter maps were developed by spatially interpolating the calibrated points using deterministic variables to provide snow parameter sets for each grid cell. A full description of the BCM and calibration procedures can be found in the software manual (Flint et al. 2021a).

e. Analyses and model comparisons

In addition to assessing each of the gridded datasets at their native resolutions, spatial downscaling was applied to a 270-m resolution using procedures described in Flint and Flint (2012). Spatial downscaling produces fine-scale monthly climate grids using a modified gradient and inverse distance squared (GIDS) calculation (Nalder and Wein 1998) based on northing, easting, and elevation parameters. This method incorporates the topographic and elevational effects on the climate and has been shown to improve or maintain accuracy against station data (Flint and Flint 2012). The downscaled data were tested relative to the station data and were also used as the inputs for the BCM in order to compare the landscape hydrology consequences of the use of four of the gridded climate datasets.

Each climate dataset was spatially downscaled to 270-m resolution from their respective native resolutions to match the BCM grid resolution. The California-wide BCM model was run separately using each downscaled climate dataset as input for water

years 1981–2010. No other changes were made to the BCM input variables or parameters. Although the BCM snowpack and vegetation parameters were initially calibrated using PRISM data, the deterministic input variables including soil data, vegetation, and geology remained constant between runs. Whereas climate dataset comparisons to station data can assess accuracy, the hydrologic variable comparisons aim to show relative uncertainty between datasets and not accuracy against streamflow. California-wide, long-term comparisons between datasets of climate inputs and BCM hydrological variables used average annual grids from 1981 to 2010. We calculated the differences among the gridded datasets values for three climate variables (precipitation, minimum and maximum temperature) and six hydrological variables (actual evapotranspiration, 1 April snowpack, soil storage, recharge, runoff, and climatic water deficit), derived from the BCM. The interdataset range was calculated by finding the highest and lowest value of each gridded climate input and hydrological output variable for every grid cell in the study area. A grid cell value of zero indicated a perfect agreement between all of the gridded datasets.

The California-wide BCM for each dataset was extracted for the 11 FNF basins, and interdataset spread was summarized for climate inputs, recharge, runoff, climatic water deficit, and 1 April snowpack from water years 1981 to 2010. The BCM outputs were postprocessed in an Excel spreadsheet to produce unimpaired streamflow estimates from 2000 to 2010. This period was chosen because it had a range of dry and wet conditions across California.

f. Statistical metrics

Five statistical metrics were calculated for each native and downscaled pixel coincident with SCDNA observations, when available (1980–2013). Gridded dataset pixels were extracted using bilinear interpolation. Monthly mean absolute error (MAE), mean bias error (MBE), root mean squared error (RMSE), mean percent error (MPE), and the coefficient of determination (R^2) were calculated for each gridded dataset against the precipitation and temperature records from the 1231 SCDNA locations (Fig. 1). The best-fit value for MAE, MBE, RMSE, and MPE is 0, and the best fit for R^2 is 1. The MAE statistic shows the average absolute error of the gridded datasets compared with station data, considering systematic and random errors. This statistic is often used because the errors are not likely to be normally distributed. The MBE statistic quantifies the extent to which the gridded dataset overestimates or underestimates the expected precipitation or temperature value.

RMSE quantifies the standard deviation of the difference between the gridded dataset and the station data and describes how spread out the error residuals are around a best fit regression line. MPE is another measure of prediction accuracy and can show percent bias in gridded station datasets, however, due to the arbitrary zero point of air temperature, temperature values were converted from Celsius to kelvins for MPE calculations. The R^2 is the proportion of variance of station data that is explained by the gridded dataset and thus quantifies the best fit linear relationship between

each gridded dataset and station data. Statistics were calculated for each station and summarized over the study area and summarized by ecoregion to look at regional differences and bias. Cumulative long-term (1980–2013) precipitation data at each station were extracted from each gridded dataset to assess how bias accumulates over time from station data.

3. Results

a. Comparisons to station data

There is no single “best” dataset for all three variables in California. However, for precipitation and minimum temperature, some gridded datasets outperformed others (Fig. 2). The single dataset with the lowest amount of error for precipitation and air temperature combined was PRISM; although TopoWX, ClimateNA, and DayMet performed better for temperature, DayMet and ClimateNA were considerably less accurate for precipitation, and TopoWX does not provide precipitation. PRISM had the lowest error values across all metrics for precipitation, in some cases by a factor of 2. The MAE, MBE, RMSE, and MPE for PRISM precipitation were $2.4 \text{ mm month}^{-1}$, $2.0 \text{ mm month}^{-1}$, $4.4 \text{ mm month}^{-1}$, and 7.4%, respectively (Fig. 2). The next most accurate dataset for precipitation was Livneh, followed by DayMet and ClimateNA. The MPE of 6-km Livneh was reduced by a factor of 2 in the downscaled version but was still higher than the downscaled PRISM. Although each gridded dataset had a high R^2 value for all stations averaged across the study area, other statistical indices indicated greater differences between datasets. Downscaled versions of each gridded dataset showed improved or similar statistical performance than their respective native resolutions against the SCDNA dataset.

Maximum temperature results showed very little variation in accuracy across datasets, and overall, very low errors (Fig. 2). Compared to the other datasets, the 6-km Livneh had a slightly higher MAE, MBE, RMSE, and MPE with values of 0.5°C , -0.5°C , 0.6°C , and 0.2%, respectively. The other datasets ranged $0.1^\circ\text{--}0.2^\circ\text{C}$ lower in error and bias than Livneh. ClimateNA most closely aligned with the station data for maximum temperature, although only slightly. Minimum temperature showed larger differences between datasets than maximum temperature. Again, the 6-km Livneh fared the worst with a MAE of 1.3°C , MBE of -1.3°C , RMSE of 1.4°C , and MPE of 0.5%. TopoWX performed the best against station data for minimum temperature, with the lowest MAE, MBE, RMSE, and MPE of 0.1°C , 0.0°C , 0.2°C , and 2%, respectively.

Mean bias error for each native resolution dataset was summarized for precipitation and minimum temperature by ecoregion to assess topo-climatic differences between the datasets that can inform their applicability for local watershed or ecological modeling. Precipitation bias for each dataset shows that regional differences can be much larger than differences averaged over California (Fig. 3a). The largest negative (wetter) bias values were found in the Coast Range and Cascades, and the largest positive (drier) bias values were found in the

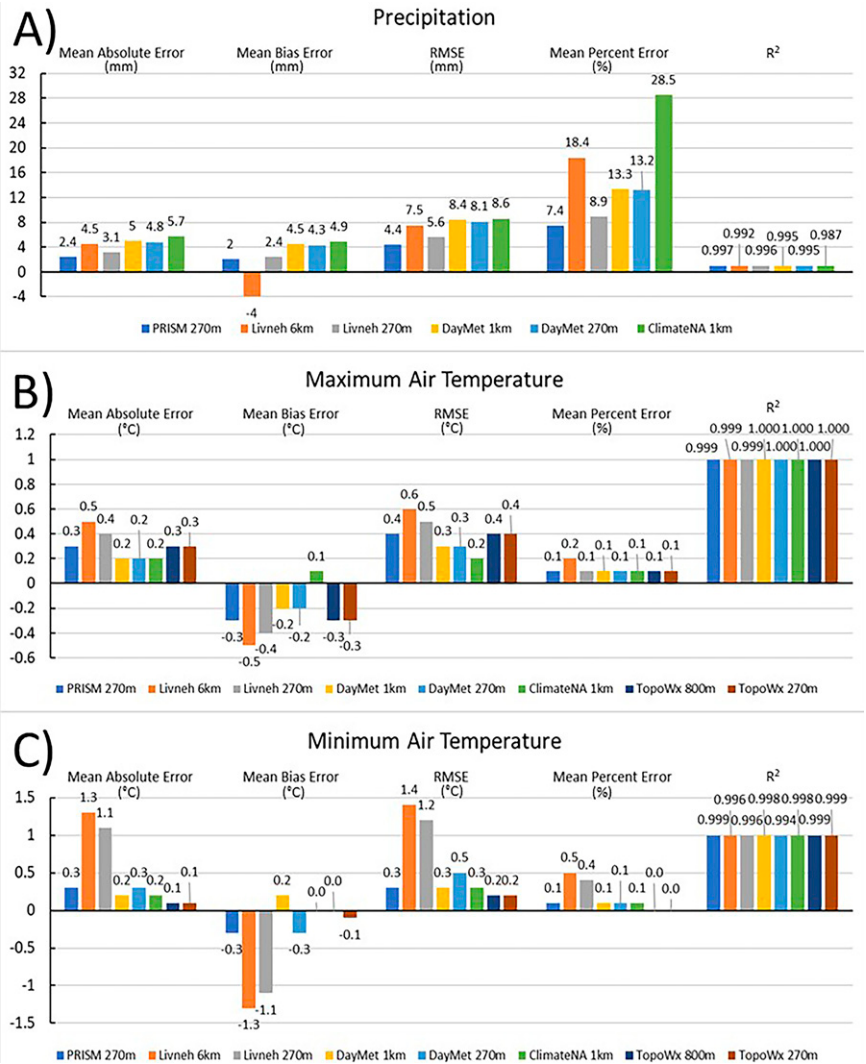


FIG. 2. Statistics for each native and downscaled gridded climate dataset against monthly station data for (a) precipitation, (b) maximum air temperature, and (c) minimum air temperature from 1980 to 2013. Note the different y-axis scale for each variable. RMSE = root mean squared error, R^2 = coefficient of determination, mm = millimeters, °C = degrees Celsius, m = meters, km = kilometers. Statistics are available in tabular format as Table S1 in the online supplemental material.

Klamath Mountains, Eastern Cascades Slopes and Foothills, and Snake River Basin/High Desert ecoregions (Fig. 3a). Overall, precipitation bias was closest to zero in the Southern Basin and Range and the Central California Valley.

Minimum temperature bias was similar for PRISM, DayMet, and ClimateNA, with a slightly cool bias in the Southern Basin and Range and less of a cool bias through the Central Valley and Sierra Nevada (Fig. 3b). The Coast Range and Klamath Mountains were warmer for PRISM and ClimateNA, and slightly cooler for DayMet. TopoWX showed a warm bias in the Sierra Nevada and most of

northern California with a slightly cool bias for the rest of the ecoregions. Livneh showed the highest values of bias with a cool bias of over 1 degree in every ecoregion except the Central California Valley and the Southern and Central California Plains and Hills.

The average amount of deviation from each climate station for precipitation in California for each gridded dataset from 1980 to 2013 shows that each dataset overestimates precipitation at a station on average by the end of the 33-yr period, from roughly 800 to 2000 mm (4.1%–10.3%) cumulatively, or 24–60 mm yr⁻¹ (Fig. 4). A deviation value of zero would

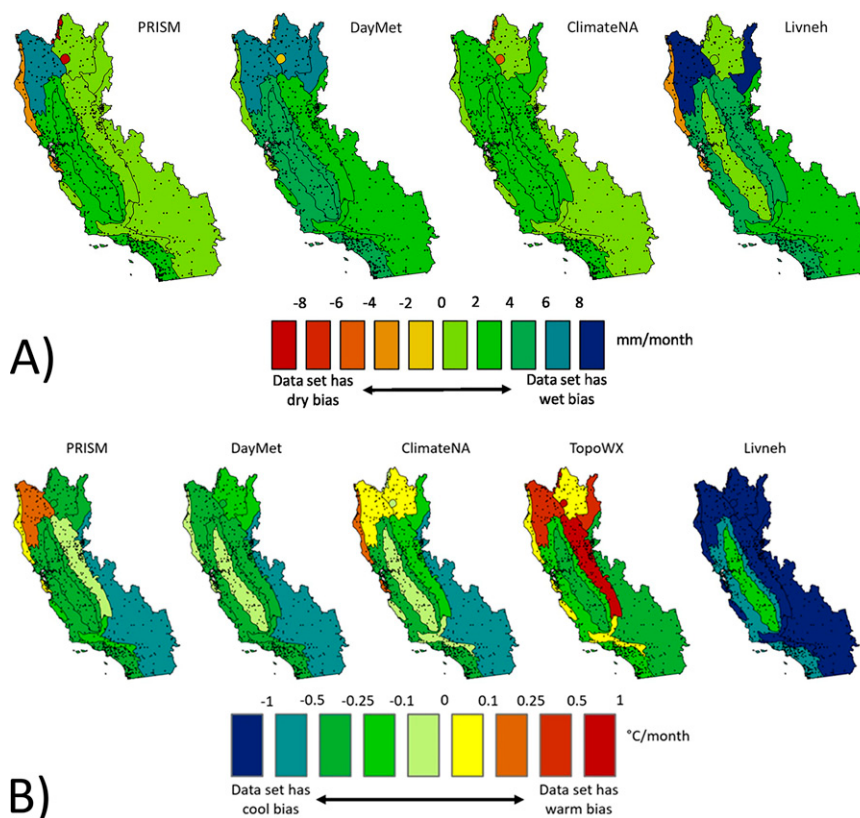


FIG. 3. Average monthly model divergence from station data (bias) summarized by ecoregion for (a) precipitation and (b) minimum temperature from 1980 to 2013. TopoWX does not provide precipitation data. Station locations are shown as black dots.

indicate a perfect match to the station cumulative values. The cumulative deviation shows ClimateNA with the highest amount of bias over time especially in the 2011–13 period, and PRISM with the lowest amount of bias (Fig. 4). PRISM bias

over this period was roughly 800 mm (4.1%), less than half the bias in DayMet, Livneh, or ClimateNA. DayMet and Livneh had about 1800 mm (9.5%) and 1700 mm (8.8%) of bias by the end of the period, respectively.

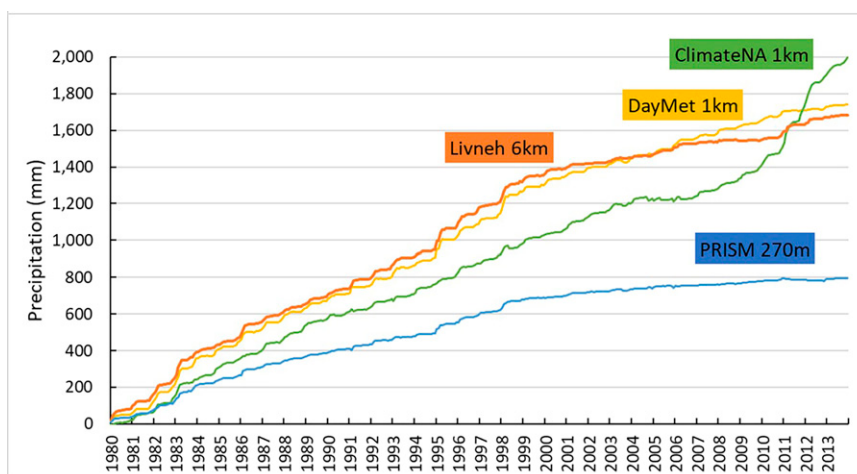


FIG. 4. Cumulative deviation of monthly precipitation for each gridded precipitation dataset extracted at each climate station from 1980 to 2013, presented as absolute values. A value of 0 indicates a perfect match to station data. The scale of each gridded dataset is indicated with km = kilometers, m = meters.

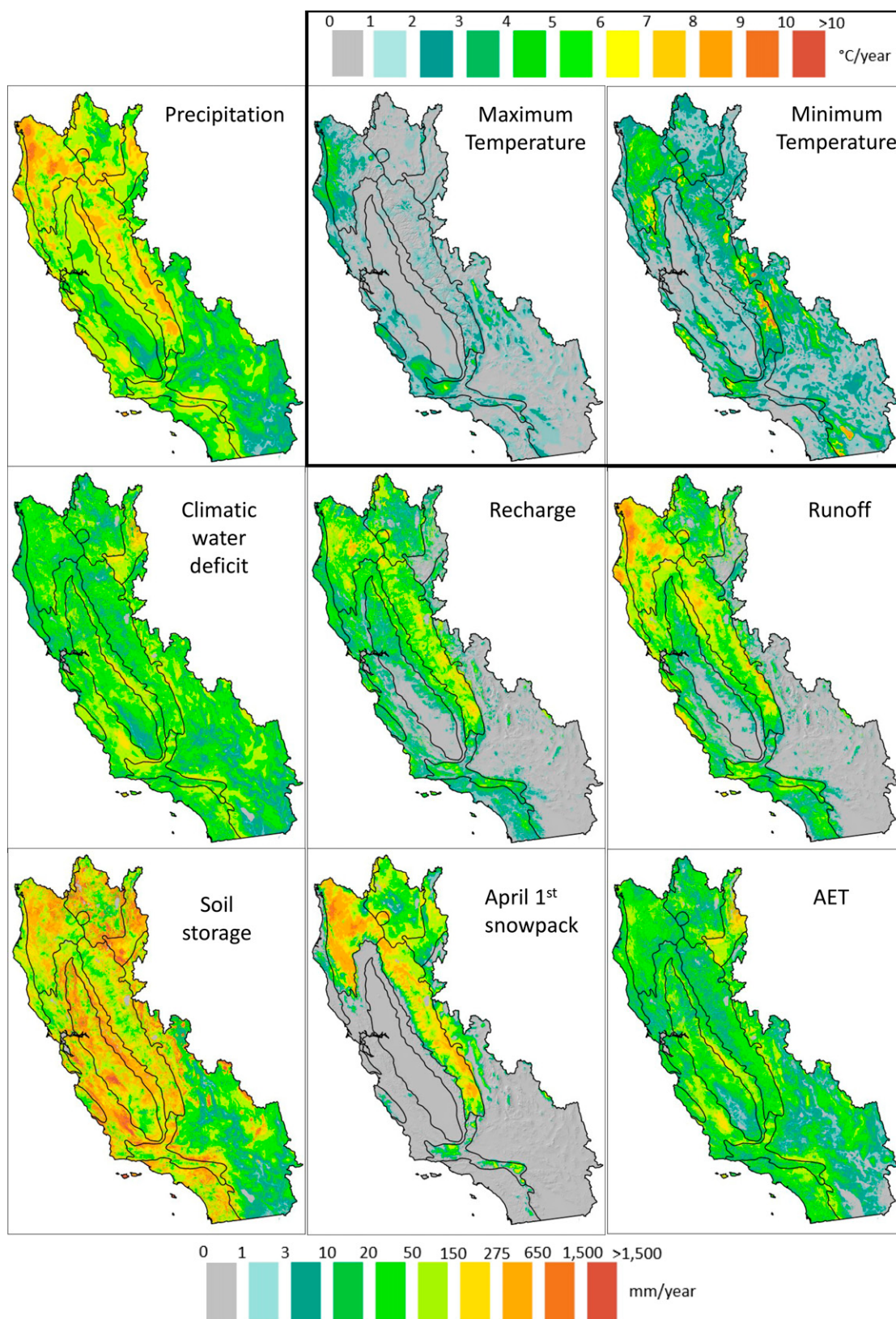


FIG. 5. Interdataset spread in (top) climate inputs (temperature and precipitation) and (middle),(bottom) Basin Characterization Model (BCM) hydrological model variables for water year 1981–2010 annual averages. Warm colors show the highest amount of interdataset spread (less agreement); cooler colors indicate higher agreement between datasets.

b. Comparisons of long-term gridded data

Four of the gridded climate datasets were extracted for the study area and downscaled to 270 m to provide climate inputs to the BCM. Climate NA was excluded because gridded climates were not available. The interdataset range of values was calculated on a per-pixel basis for precipitation, minimum and maximum temperature, recharge, runoff, climatic water deficit, soil storage, 1 April snowpack (SWE), and actual evapotranspiration for every 270-m grid cell for PRISM, Day-Met, Livneh, and TopoWX (hybrid air temperature with PRISM precipitation climate inputs) (Fig. 5). The interdataset range identifies where the datasets agree and disagree the most on the landscape. Precipitation had the largest interdataset range in the Klamath Mountains and Sierra Nevada ecoregions but also showed large differences in the Coast Range, Cascades, and Southern California Mountains. There was little-to-no interdataset spread in the Southern Basin and Range and the Central California Valley ecoregions. Maximum temperature showed the smallest level of discord for most grid cells, and disagreement was mostly in coastal mountain ecoregions and areas with high topographic complexity within the Sierra Nevada and Mojave Desert. Minimum temperature showed large interdataset ranges in the Sierra Nevada, Klamath Mountains, and some parts of the Southern Basin and Range ecoregion.

Ecoregions such as in the Sierra Nevada and Klamath Mountains had the highest disagreements among gridded climate datasets, which led to greater differences in hydrological variables (Fig. 5, rows 2 and 3). The hydrological variable with the highest amount of interdataset spread was the soil storage variable, with large differences in all regions. Runoff showed higher interdataset spread than recharge. Recharge and runoff showed similar spatial patterns of interdataset spread, mirroring the precipitation results, with highest disagreement in the Sierra Nevada and Klamath Mountains ecoregions. The differences in precipitation and minimum temperature led to large discrepancies in 1 April snowpack, a value generally used as an indicator of available snow water equivalent to carry the water supply through the dry summer months. Climatic water deficit and actual evapotranspiration showed interdataset spread across all regions, and spatial patterns indicate they show a more complex relationship to precipitation and air temperature differences.

c. Comparisons of hydrological model results for FNF basins

Results from the BCM for each climate dataset were summarized by FNF basin from 1981 to 2010, and the average range of potential outcomes for each variable were calculated for the 11 FNF basins. The 1 April snowpack had the largest range among datasets (Fig. 6), showing the effects of different precipitation and air temperature estimates in a snowpack-dominated hydrological regime. Runoff and recharge were the next two hydrological variables with the largest interdataset range for the FNF basins. Climatic water deficit and actual evapotranspiration differences ranged between 15 and 50 mm yr⁻¹, and potential evapotranspiration differences ranged from 14 to 31 mm yr⁻¹

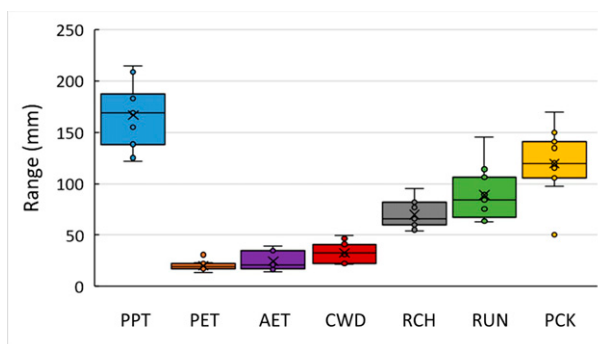


FIG. 6. Average interdataset spread across climate datasets for hydrological variables for 11 FNF basins from water years 1981–2010. The “X” indicates the mean of the data, the middle line is the median. The top of the box represents the third quartile, and the bottom of the box represents the first quartile. The maximum value is indicated by the highest horizontal bar, and the minimum value is indicated by the lowest horizontal bar. Other dots outside the whiskers are outliers. PPT = precipitation, PET = potential evapotranspiration, AET = actual evapotranspiration, CWD = climatic water deficit, RCH = recharge, RUN = runoff, PCK = 1 April snowpack.

depending on climate dataset. Potential evapotranspiration is based on maximum temperature, a climate variable that had less disagreement between climate datasets. In contrast, actual evapotranspiration is vegetation specific, and when run through the BCM incorporates precipitation, temperature, spatially explicit soil moisture holding capacity, and bedrock permeability, which means it can be limited by available water, resulting in a higher range. Disagreement between datasets of precipitation and air temperature lead to the highest uncertainty in model estimates of snowpack, recharge, and runoff, highlighting the importance of choosing the most accurate dataset due to the sensitivity of hydrological variables to climate inputs.

To assess watershed-level differences in hydrological outputs, 1981–2010 average interdataset spread for air temperature and precipitation (Fig. 7a), and four hydrological variables: climatic water deficit, recharge, runoff, and 1 April snowpack (Fig. 7b), are shown for the 11 FNF basins (Fig. 1). Figure 7a shows an interdataset spread of minimum temperature of 2°–4°C, depending on the basin. Maximum temperature varied much less, with 0.8°–1.4°C interdataset spread. The FNF basin with the smallest disagreement between datasets is the Cosumnes for minimum temperature, and Yuba for maximum temperature. Differences among datasets for precipitation ranged from 122 mm yr⁻¹ in the Mokelumne to 215 mm yr⁻¹ for the Yuba basin. Basins with the largest differences in precipitation corresponded to higher differences between recharge and runoff (Fig. 7b), and in the Feather, interdataset spread of snowpack was higher than precipitation due to the combined differences in precipitation and air temperature. Differences among datasets for climatic water deficit were small but were highest for the San Joaquin and Kings. Although interdataset spread in runoff was higher than recharge in the northern Sierra basins, most of the southern

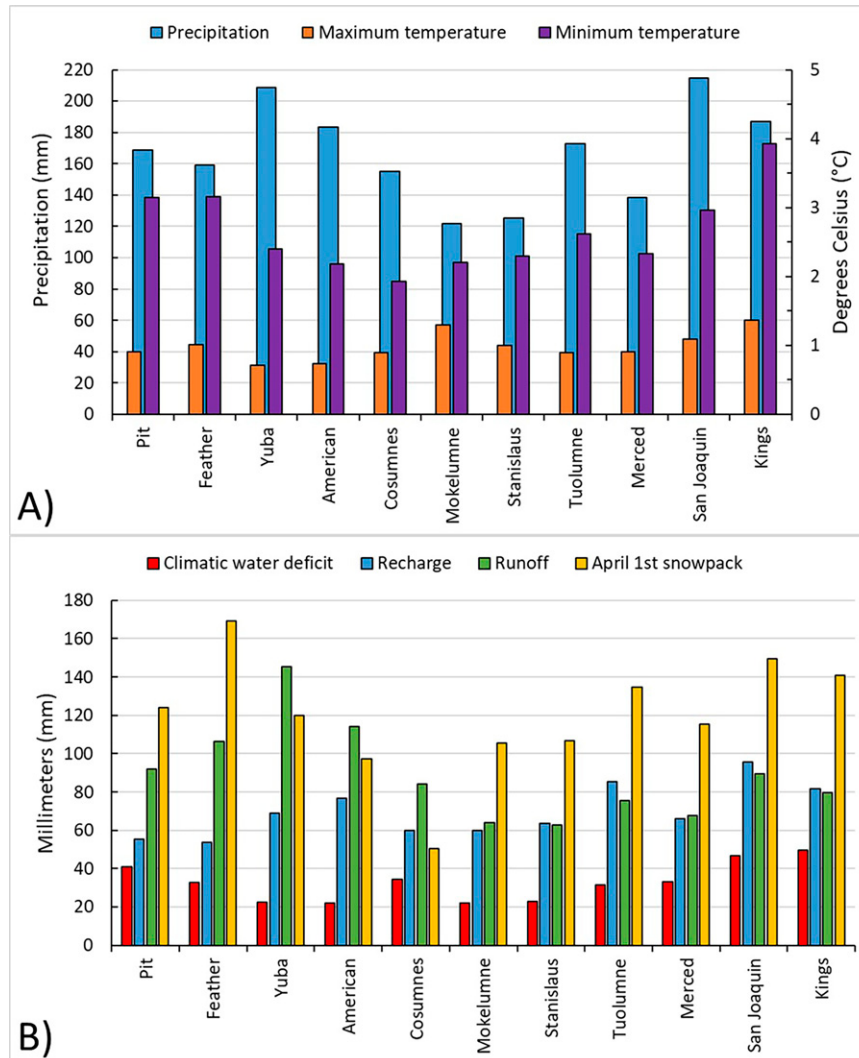


FIG. 7. Interdataset spread between PRISM, Livneh, TopoWX, and DayMet for (a) climate inputs (air temperature and precipitation) and (b) four hydrological variables (climatic water deficit, recharge, runoff, and 1 April snowpack) from the Basin Characterization Model for each full natural flow (FNF) basin from water years 1981 to 2010. Note: TopoWX includes PRISM precipitation and TopoWX air temperature.

Sierra basins showed higher variability in recharge compared to runoff.

Average monthly streamflow for the 11 FNF basins by each climate dataset from 2000 to 2010 is shown in Fig. 8. Results show the average hydrograph varies by a factor of 2 in the months of December, January, and February. The months with the highest agreement between climate datasets and the least sensitivity to differences in input data are August, September, and October. The TopoWX dataset ran through the BCM was a hybrid of TopoWX minimum and maximum temperature and PRISM precipitation. The comparison of two datasets with the same precipitation was done to highlight the importance of considering air temperature in addition to precipitation for hydrological modeling, especially in areas that

experience significant snowpack. Due only to changes in air temperature, the PRISM dataset resulted in higher streamflow than the hybrid TopoWX dataset for the months of March, April, and May, and was on average 4% higher for the FNF basins.

Depending on the dataset used to drive the BCM, annual streamflow volumes ranged $\pm 15\%$ due to changes in precipitation and snowpack timing. Figure 8b shows the interdataset spread, largely influenced by the differences in the Livneh climate dataset. The dataset with the highest amount of modeled mean streamflow from all basins in 2000–10 was DayMet, followed by PRISM, TopoWX hybrid, and then Livneh. Even though precipitation bias was higher than PRISM, Livneh produced less streamflow than the other datasets and

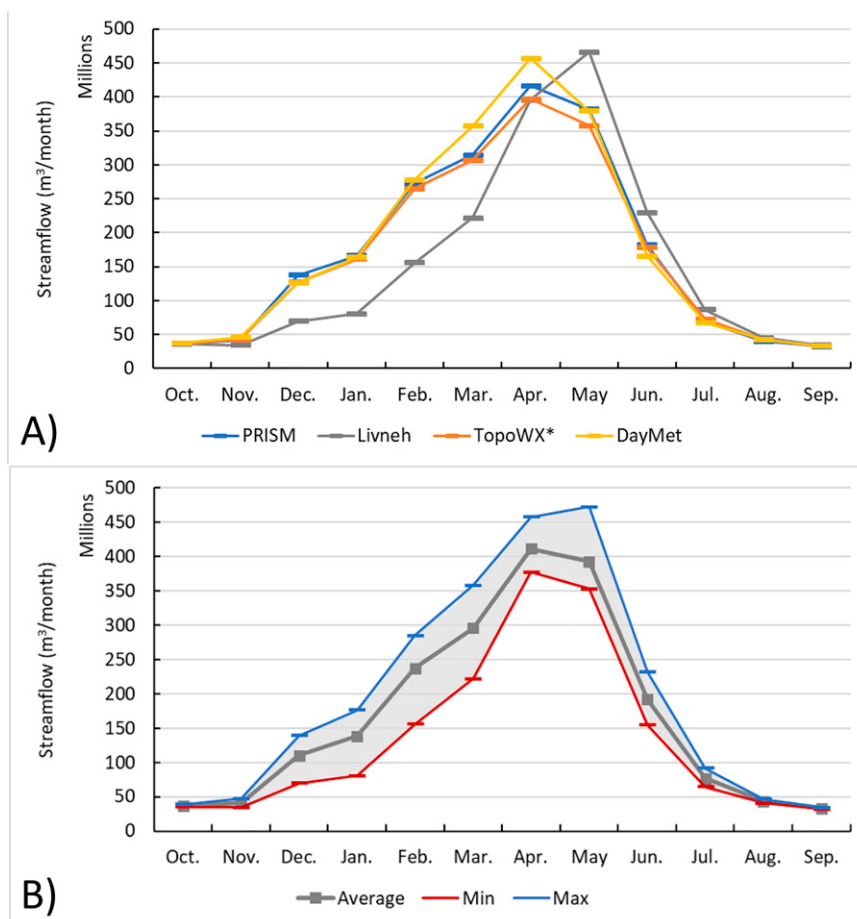


FIG. 8. Average monthly streamflow for all basins in million cubic meters per month for water years 2000–10. (a) shows streamflow for each climate dataset, (b) shows the range in monthly BCM streamflow from all climate datasets. Note: TopoWX includes PRISM precipitation and TopoWX air temperature.

underestimated the FNF modeled estimates by about 8%. The interdataset spread can range up to 70% of the mean monthly streamflow in December and January, with an average of 35% of the mean streamflow for all months.

4. Discussion

Accuracy statistics of PRISM, Livneh, DayMet, ClimateNA, and TopoWX indicated that there was not one single dataset that was the most accurate for precipitation and air temperature. However, PRISM had the overall lowest precipitation errors and did not have high errors in air temperature although it did not perform the best compared to other datasets. Although not a fully independent accuracy assessment due to an earlier, unprocessed version of the SCDNA station data being used as one component in the creation of the gridded climate products we tested, a comparison to station data such as presented here is a commonly used method to assess the ability of a gridded dataset to represent measured data or the suitability of a gridded climate dataset for a

specific research goal, application, or study area (Behnke et al. 2016; Eum et al. 2014; Asong et al. 2020; Muche et al. 2020). In this study, we used a climate station dataset, SCDNA, for validation that employs additional quality assurance measures and gap filling techniques to include a more appropriate validation dataset. Consistent with other gridded climate datasets, SCDNA has not been corrected for observations in biases like station undercatch due to wind and topographic effects. Underrepresentation of high elevation locations or areas with complex terrain from the GHCN-D stations may introduce some bias into the statistical results by not capturing enough extreme locations across California.

Downscaling the climate datasets to 270 m improved or resulted in similar statistics as the native resolution, depending on the dataset and climate variable. The Livneh dataset showed the largest improvement when downscaled over the native 6-km resolution, especially for precipitation, indicating that downscaling can improve the accuracy of coarse-resolution datasets and does not degrade the accuracy of higher-resolution datasets. The ClimateNA precipitation dataset at 1-km resolution

performed worse than Livneh but had lower errors for air temperature. There were not enough coarse-resolution datasets in this study to conclude that higher-resolution datasets are more accurate for precipitation or air temperature, but in a region of complex elevation and climatology, downscaling from a few kilometers' resolution to 270 m can improve the representation of elevational gradients and local-scale climate patterns.

Uncertainty in input climate and hydrology data can arise from multiple sources. McMillan et al. (2018) identified five uncertainty categories: measurement, derived data, interpolation, scaling, and data management uncertainty. Measurement uncertainty exists in precipitation and air temperature data due to many potential reasons like sensor error, calibration or installation errors, datalogger or technical issues, and lack of appropriate quality control and assurance procedures. Derived data uncertainty is when a hydrologic quantity is estimated using a proxy measurement, such as using river stage to calculate streamflow. Interpolation uncertainty exists when a measurement is interpolated in space or time due to underlying assumptions in the interpolation methods. Scaling uncertainty exists when a process measured at one scale is used to approximate the process at a different scale, such as if a local rain gauge does not represent the surrounding area due to hyperlocalized conditions. Data management uncertainty exists in recorded values or metadata due to many potential human or computing errors. It is important to consider all aspects of uncertainty in climate data and subsequent gridded data products and to minimize these uncertainties whenever possible.

There are numerous gridded climate products available at varying spatial resolutions, time steps, footprints, and interpolation schemes and assumptions. Although these datasets contain uncertainties (Newman et al. 2015), gridded climate datasets are often treated as ground truth without acknowledging problems with station data or assumptions made to interpolate a spatially and temporally complete climate product (Walton and Hall 2018). Still, gridded historical climate data products have been shown to produce better hydrological simulation than using station measurements alone (Ledesma and Futter 2017), particularly for low-density climate station arrays that are not able to accurately represent regional temperature and precipitation patterns required for accurate hydrological simulations (Ledesma and Futter 2017).

Quantitative uncertainty statistics at the California scale may not represent gridded dataset accuracy at a local scale or away from station locations. A regional or local comparison can help identify the most appropriate dataset for a study area. For example, a study located in the Central California Valley ecoregion may not require as careful a consideration for which climate dataset to use because the interdataset spread is very low for precipitation and temperature. However, a study in the Sierra Nevada or Klamath Mountains ecoregion requires thoughtful selection because the choice of gridded climate dataset can greatly impact the results of statistical analyses or models due to the large interdataset spread and higher precipitation and minimum temperature bias, depending on which dataset is chosen. Large differences in precipitation and air temperature led to the highest

uncertainty in snowpack, recharge, runoff, and less so, actual evapotranspiration and climatic water deficit.

Differences between datasets can be attributed to which climate stations are used and if station data are gap-filled or homogenized, as well as to methodological differences. Each gridded dataset we examined uses a different interpolation method and grid cell resolution, which can greatly impact climate dataset accuracy. DayMet had reasonable statistics across California but has been found to show large jumps on a daily time step due to missing station data (Walton and Hall 2018). As shown in the minimum temperature bias result, Livneh had a large cold bias compared to station data, likely attributed to the fixed lapse rate that is not appropriate for areas of complex terrain like the Sierra Nevada and Klamath Mountains ecoregions. For that reason, Livneh may not be suitable for studies in snow-dominated regions or most of California due to topographic differences. As an example of how errors might impact climate change projections, the Livneh dataset was used to bias correct the General Circulation Models selected to portray future climate conditions in California's 4th Climate Change Assessment (Pierce et al. 2015; <https://www.climateassessment.ca.gov/>). The Livneh minimum temperature bias of about 1°C cooler than weather stations could mean future warming is underestimated by that amount. This underestimate could cascade through hydrological processes, leading to modeled overestimates of when Sierra Nevada snowpack might vanish, when yearly melt-out might occur, and underestimates of how long future seasonal dry periods might be.

TopoWX was warmer in the Sierra Nevada and Klamath Mountains ecoregions, which led to higher streamflow in the winter and summer months compared to PRISM, when PRISM precipitation was used for both simulations. TopoWX was also the most accurate air temperature dataset against station data and can be used to combine with precipitation to improve streamflow estimates. However, there are currently (2021) no plans to update the TopoWX dataset beyond 2018 and is therefore limited in potential applications. ClimateNA was statistically accurate for temperature but had large errors in precipitation and extracting high-resolution, statewide monthly gridded datasets for input to the BCM was not possible, limiting the available comparisons to other climate datasets. ClimateNA and TopoWX are not commonly included in comparison studies, potentially due to some of these limitations.

Choosing a gridded climate dataset for climate change studies is an important consideration with implications for hydrological and ecological modeling and depends on the local availability of historical and future datasets, and the purpose and application of the study. Although additional uncertainty is introduced by using future climate scenarios, minimizing historical biases can limit the uncertainty propagated through hydrological or ecological models while projecting into the future from climate forcing. Downscaling technique, bias correction method, and resolution can introduce large uncertainties into climate change modeling studies. In addition, availability of all needed climate variables, and historical datasets that are kept up to date are important considerations for regional assessments. Results from this study can be used to guide the decision-making process to

determine the gridded climate dataset that is most suitable for a region and to quantify the implications for hydrological modeling in California and provide guidance for sensitivity analyses to select climate datasets in any location.

5. Conclusions

We assessed the accuracy of five commonly used historical-climate gridded datasets and the associated uncertainty in landscape hydrologic variables entrained by dataset choice. Each dataset was downscaled to a common 270-m resolution to determine if accuracy improved, and to match the BCM model grid for modeled hydrologic comparisons. Downscaling coarser-scale climate grids resulted in improved or similar statistical accuracy to station data, especially for precipitation. Livneh, DayMet, ClimateNA, and PRISM generally overestimated monthly precipitation compared to the SCDNA stations, a subsample of stations used in each of the datasets. The PRISM precipitation dataset was the most accurate for California and had low errors for minimum and maximum air temperature. TopoWX was the most accurate for minimum temperature and had very low errors for maximum temperature but does not provide precipitation data. The five gridded datasets all performed similarly for maximum temperature averaged across the state. Although Livneh was the second most accurate for precipitation, it was highly inaccurate for minimum temperature, and had the highest errors for maximum temperature, resulting in the most anomalous streamflow result. ClimateNA was the most accurate for maximum air temperature, but had high precipitation errors. For these reasons, the PRISM dataset was the overall most accurate and realistic dataset compared with station data for California, which we are now using as baseline in other studies.

Variability in precipitation and air temperature affected hydrologic variables differentially. Recharge and runoff differed in areas where precipitation occurs in higher volumes and where precipitation differed the most between datasets. Climatic water deficit and actual evapotranspiration showed higher agreement between datasets in areas where high interdataset spread occurred for precipitation, indicating these estimates may be more robust against differences in climate datasets. Coincident locations of high interdataset spread for precipitation and minimum air temperature led to even higher interdataset spread of 1 April snowpack values. Differences in climate inputs for watersheds in the Sierra Nevada resulted in large potential uncertainties in streamflow estimates, with an interdataset spread of up to 70% of the monthly streamflow, and on average 35%. Differences in streamflow due to air temperature only between the TopoWX hybrid and PRISM datasets averaged 4% month⁻¹, with higher dataset spread in December, May, and August of 7%–8%.

Acknowledgments. The authors declare no conflict of interest. This work was supported by Agreement 19-IA-11272138-010 with the USDA Forest Service, and in part by the USDA Forest Service through CALFIRE Agreement 8CA04059. Any use of trade, product or firm name is for descriptive purposes only and does not imply endorsement

by the U.S. Government. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA Forest Service determination or policy. This article has been peer reviewed and approved for publication consistent with USGS Fundamental Science Practices (<https://pubs.usgs.gov/circ/1367>).

Data availability statement. All datasets used in this study are available online from the citations listed in section 2 (methods) and corresponding references below.

REFERENCES

- Ahmadalipour, A., and H. Moradkhani, 2017: Analyzing the uncertainty of ensemble-based gridded observations in land surface simulations and drought assessment. *J. Hydrol.*, **555**, 557–568, <https://doi.org/10.1016/j.jhydrol.2017.10.059>.
- Anderson, E. A., 2006: Snow accumulation and ablation model—SNOW-17. User's manual, National Weather Service, 61 pp., http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/22snow17.pdf.
- Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams, 1998: Large area hydrologic modeling and assessment Part I: Model development 1. *J. Amer. Water Res. Assoc.*, **34**, 73–89, <https://doi.org/10.1111/j.1752-1688.1998.tb05961.x>.
- Asong, Z. E., M. E. Elshamy, D. Princez, H. S. Wheeler, J. W. Pomeroy, A. Pietroniro, and A. Cannon, 2020: High-resolution meteorological forcing data for hydrological modelling and climate change impact analysis in the Mackenzie River Basin. *Earth Syst. Sci. Data*, **12**, 629–645, <https://doi.org/10.5194/essd-12-629-2020>.
- 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>.
- Bristow, K. L., and G. S. Campbell, 1984: On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agric. For. Meteorol.*, **31**, 159–166, [https://doi.org/10.1016/0168-1923\(84\)90017-0](https://doi.org/10.1016/0168-1923(84)90017-0).
- CDWR, 2007: California Central Valley unimpaired flow data. 4th ed. California Dept. of Water Resources, 50 pp., https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/bay_delta_plan/water_quality_control_planning/docs/sjrf_sprinfo/dwr_2007a.pdf.
- Curtis, J. A., L. E. Flint, M. A. Stern, J. Lewis, and R. D. Klein, 2021: Amplified impact of climate change on fine-sediment delivery to a subsiding coast, Humboldt Bay, California. *Estuaries Coasts*, **44**, 2173–2193, <https://doi.org/10.1007/s12237-021-00938-x>.
- Daly, C., 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).
- , M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.*, **28**, 2031–2064, <https://doi.org/10.1002/joc.1688>.
- Dettinger, M., 2016: Historical and future relations between large storms and droughts in California. *San Francisco Estuary Watershed Sci.*, **14**, 21, <https://doi.org/10.15447/sfews.2016v14iss2art1>.

- Eum, H. I., Y. Dibikey, T. Prowse, and B. Bonsal, 2014: Inter-comparison of high-resolution gridded climate data sets and their implication on hydrological model simulation over the Athabasca Watershed, Canada. *Hydrol. Processes*, **28**, 4250–4271, <https://doi.org/10.1002/hyp.10236>.
- Flint, A. L., and S. W. Childs, 1991: Use of the Priestley-Taylor evaporation equation for soil water limited conditions in a small forest clearcut. *Agric. For. Meteorol.*, **56**, 247–260, [https://doi.org/10.1016/0168-1923\(91\)90094-7](https://doi.org/10.1016/0168-1923(91)90094-7).
- Flint, L. E., and A. L. Flint, 2012: Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis. *Ecol. Process.*, **1**, 2, <https://doi.org/10.1186/2192-1709-1-2>.
- , —, B. J. Stolp, and W. R. Danskin, 2012: A basin-scale approach for assessing water resources in a semiarid environment: San Diego region, California and Mexico. *Hydrol. Earth Syst. Sci.*, **16**, 3817–3833, <https://doi.org/10.5194/hess-16-3817-2012>.
- , —, J. H. Thorne, and R. Boynton, 2013: Fine-scale hydrologic modeling for regional landscape applications: The California Basin Characterization Model development and performance. *Ecol. Processes*, **2**, 25, <https://doi.org/10.1186/2192-1709-2-25>.
- , —, and M. A. Stern, 2021a: The Basin Characterization Model—A regional water balance software package. USGS Techniques and Methods 6-H1, 85 pp., <https://doi.org/10.3133/tm6H1>.
- , —, and M. A. Stern, 2021b: The Basin Characterization Model—A regional water balance software package (BCMv8) data release and model archive for hydrologic California, water years 1896–2020, U.S. Geological Survey, accessed 30 March 2021, <https://doi.org/10.5066/P9PT36UI>.
- Hamlet, A. F., and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature fields for the continental United States. *J. Hydrometeorol.*, **6**, 330–336, <https://doi.org/10.1175/JHM420.1>.
- Henn, B., A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, 2018: An assessment of differences in gridded precipitation datasets in complex terrain. *J. Hydrol.*, **556**, 1205–1219, <https://doi.org/10.1016/j.jhydrol.2017.03.008>.
- Hughes, M., J. D. Lundquist, and B. Henn, 2020: Dynamical downscaling improves upon gridded precipitation products in the Sierra Nevada, California. *Climate Dyn.*, **55**, 111–129, <https://doi.org/10.1007/s00382-017-3631-z>.
- Ledesma, J. L., and M. N. Futter, 2017: Gridded climate data products are an alternative to instrumental measurements as inputs to rainfall-runoff models. *Hydrol. Processes*, **31**, 3283–3293, <https://doi.org/10.1002/hyp.11269>.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res. Atmos.*, **99**, 14 415–14 428, <https://doi.org/10.1029/94JD00483>.
- 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>.
- Lundquist, J. D., M. Hughes, B. Henn, E. D. Gutmann, B. Livneh, J. Dozier, and P. Neiman, 2015: High-elevation precipitation patterns: Using snow measurements to assess daily gridded datasets across the Sierra Nevada, California. *J. Hydrometeorol.*, **16**, 1773–1792, <https://doi.org/10.1175/JHM-D-15-0019.1>.
- Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251, [https://doi.org/10.1175/1520-0442\(2002\)015<3237:ALTHBD>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2).
- McMillan, H. K., I. K. Westerberg, and T. Krueger, 2018: Hydrological data uncertainty and its implications. *Wiley Interdiscip. Rev.: Water*, **5**, e1319, <https://doi.org/10.1002/wat2.1319>.
- Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the global historical climatology network-daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, <https://doi.org/10.1175/JTECH-D-11-00103.1>.
- Miller, N. L., K. E. Bashford, and E. Strem, 2003: Potential impacts of climate change on California hydrology 1. *J. Amer. Water Res. Assoc.*, **39**, 771–784, <https://doi.org/10.1111/j.1752-1688.2003.tb04404.x>.
- Muche, M. E., and Coauthors, 2020: Comparison and evaluation of gridded precipitation datasets in a Kansas agricultural watershed using SWAT. *J. Amer. Water Res. Assoc.*, **56**, 486–506, <https://doi.org/10.1111/1752-1688.12819>.
- Nalder, I. A., and R. W. Wein, 1998: Spatial interpolation of climatic normals: Test of a new method in the Canadian boreal forest. *Agric. For. Meteorol.*, **92**, 211–225, [https://doi.org/10.1016/S0168-1923\(98\)00102-6](https://doi.org/10.1016/S0168-1923(98)00102-6).
- 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>.
- Oyler, J. W., A. Ballantyne, K. Jencso, M. Sweet, and S. W. Running, 2015: Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *Int. J. Climatol.*, **35**, 2258–2279, <https://doi.org/10.1002/joc.4127>.
- Parkes, B., T. P. Higginbottom, K. Hufkens, F. Ceballos, B. Kramer, and T. Foster, 2019: Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environ. Res. Lett.*, **14**, 11, <https://doi.org/10.1088/1748-9326/ab5eb6>.
- Pierce, D. W., D. R. Cayan, E. P. Mauer, J. T. Abatzoglou, and K. C. Hegewisch, 2015: Improved bias correction techniques for hydrological simulations of climate change. *J. Hydrometeorol.*, **16**, 2421–2442, <https://doi.org/10.1175/JHM-D-14-0236.1>.
- Priestley, C. H. B., and R. J. Taylor, 1972: On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Wea. Rev.*, **100**, 81–92, [https://doi.org/10.1175/1520-0493\(1972\)100<0081:OTAOSH>2.3.CO;2](https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2).
- Reitz, M., W. E. Sanford, G. B. Senay, and J. Cazenias, 2017: Annual estimates of recharge, quick-flow runoff, and evapotranspiration for the conterminous United States using empirical regression equations. *J. Amer. Wat. Res. Assoc.*, **53**, 961–983, <https://doi.org/10.1111/1752-1688.12546>.
- Shepard, D. S., 1984: Computer mapping: The SYMAP interpolation algorithm. *Spatial Statistics Models*, Springer, 133–143, https://doi.org/10.1007/978-94-017-3048-8_7.
- Skamarock, W., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp.
- Sun, Q., C. Miao, Q. Duan, D. Kong, A. Ye, Z. Di, and W. Gong, 2014: Would the ‘real’ observed dataset stand up? A critical examination of eight observed gridded climate datasets for China. *Environ. Res. Lett.*, **9**, 015001, <https://doi.org/10.1088/1748-9326/9/1/015001>.

- Tang, G., M. P. Clark, A. J. Newman, A. W. Wood, S. M. Papalexiou, V. Vionnet, and P. H. Whitfield, 2020: SCDNA: A serially complete precipitation and temperature dataset for North America from 1979 to 2018. *Earth Syst. Sci. Data*, **12**, 2381–2409, <https://doi.org/10.5194/essd-2020-92>.
- Tang, Y., and Coauthors, 2018: Uncertainty of future projections of species distributions in mountainous regions. *PLOS ONE*, **13**, e0189496, <https://doi.org/10.1371/journal.pone.0189496>.
- Thorne, J. H., R. Boynton, L. Flint, A. Flint, and T. N'goc Le, 2012: Development and application of downscaled hydroclimatic predictor variables for use in climate vulnerability and assessment studies. California Energy Commission Publ. CEC-500-2012-010, 86 pp., https://ca.water.usgs.gov/pubs/Flint_BMC_California.pdf.
- 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, [https://doi.org/10.1016/S0022-1694\(96\)03128-9](https://doi.org/10.1016/S0022-1694(96)03128-9).
- Torregrosa, A., L. E. Flint, and A. L. Flint, 2020: Hydrologic resilience from summertime fog and recharge: A case study for Coho Salmon recovery planning. *J. Amer. Water Res. Assoc.*, **56**, 134–160, <https://doi.org/10.1111/1752-1688.12811>.
- U.S. Geological Survey, 2021: National Water Information System: Web Interface. U.S. Geological Survey, <https://doi.org/10.5066/F7P55KJN>.
- Walton, D., 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>.
- Wang, T., A. Hamann, D. Spittlehouse, and C. Carroll, 2016: Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLOS ONE*, **11**, e0156720, <https://doi.org/10.1371/journal.pone.0156720>.
- Yin, H., M. G. Donat, L. V. Alexander, and Y. Sun, 2015: Multi-dataset comparison of gridded observed temperature and precipitation extremes over China. *Int. J. Climatol.*, **35**, 2809–2827, <https://doi.org/10.1002/joc.4174>.