

## RESEARCH ARTICLE

# The Office of Water Prediction's Analysis of Record for Calibration, version 1.1: Dataset description and precipitation evaluation

Greg Fall<sup>1</sup>  | David Kitzmiller<sup>2</sup> | Sandra Pavlovic<sup>2</sup> | Ziya Zhang<sup>3</sup> | Nathan Patrick<sup>4</sup> | Michael St. Laurent<sup>3</sup> | Carl Trypaluk<sup>3</sup> | Wanru Wu<sup>2</sup> | Dennis Miller<sup>2</sup>

<sup>1</sup>Office of Water Prediction, NOAA/NWS, Chanhassen, Minnesota, USA

<sup>2</sup>Office of Water Prediction, NOAA/NWS, Silver Spring, Maryland, USA

<sup>3</sup>Center for Water Resources, RTI International, Fort Collins, Colorado, USA

<sup>4</sup>California-Nevada River Forecast Center, NOAA/NWS, Sacramento, California, USA

**Correspondence**

Greg Fall, Office of Water Prediction, NOAA/NWS, Chanhassen, MN, USA.  
Email: [gregory.fall@noaa.gov](mailto:gregory.fall@noaa.gov)

**Abstract**

Hydrologic models operated by the National Weather Service call for an accurate, consistent, high-resolution, multi-decade, continental-scale record of hydrometeorological fields to serve as forcing data for model calibration. To serve this purpose, the Analysis of Record for Calibration was developed, and version 1.1 of the dataset is described in this study. Geospatial and scientific requirements, methods used in dataset generation, and input data sources are described. Given the prominent role of precipitation in model calibration, accurate and consistent precipitation is a particularly high priority for the analysis. To evaluate the analysis from this perspective, its daily precipitation is compared with surface observing stations over 43 years. The analysis exhibits low bias compared with other similar products. It also displays non-stationary bias behavior after 2015 due to the lack of a climatological constraint, as well as frequent occurrences of heavy-to-extreme precipitation that are often difficult to verify. These findings should be taken into account when the product is used for model calibration.

**KEYWORDS**

hydrology, meteorology, modeling, calibration, model forcing, precipitation, observations, geospatial analysis, geostatistics, time-series analysis, quality assurance/quality control (QAQC)

## 1 | INTRODUCTION

The United States (U.S.) National Weather Service (NWS) relies on several operational land surface and hydrologic modeling capabilities to fulfill its water prediction mission. Implementing any such model requires, in advance of operations, a calibration activity (Anderson, 2002) in which retrospective simulations are performed to adjust model parameters for soil, runoff, groundwater, vegetation, snow, reservoir, and channel properties (Lahmers et al., 2021; Wang et al., 2011; Whittaker et al., 2010)—or, in the case of conceptual models, perhaps less explicitly

Paper No. JAWR-22-0123-P of the *Journal of the American Water Resources Association* (JAWR).

Discussions are open until six months from publication.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 Research Triangle Institute. Journal of the American Water Resources Association published by Wiley Periodicals LLC on behalf of American Water Resources Association. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

## Research Impact Statement

To improve water prediction in the United States, the Analysis of Record for Calibration offers a multi-decade driving dataset for hydrologic and land surface model calibration, validation, and reanalysis.

physical parameters (Khakbaz et al., 2012; Smith et al., 2003)—followed by validation to confirm that those parameters are effective in minimizing error measures for retrospective streamflow predictions (Gupta et al., 2005), ordinarily covering time periods distinct from those used for calibration (Refsgaard, 1997; Shen et al., 2022).

The execution of these models, both in the retrospective calibration/validation phase and in real-time operations, relies upon hydrometeorological input data (e.g., Newman et al., 2015). Traditional operational hydrology requires basin-averaged temperature and precipitation time series (Khakbaz et al., 2012), whereas continental-scale hydrology often demands a more complete set of spatially and temporally discretized hydrometeorological forcing elements to drive land surface models (Xia et al., 2012).

To meet the calibration requirements of regional operational hydrology as practiced at NWS River Forecast Centers (RFCs) and of continental scale hydrology as practiced at the NWS Office of Water Prediction's (OWP's) National Water Center (NWC), a high-resolution, multi-decade, CONUS-scale record of hydrometeorological elements is needed. The dataset developed for this purpose at OWP is known as the *Analysis of Record for Calibration* (AORC), often specified in this study as AORC-CONUS to distinguish it from counterparts that cover other regions. Beyond the CONUS domain, the AORC additionally consists of forcing data to support the calibration of model configurations for Hawaii, Puerto Rico and the U.S. Virgin Islands, and Alaska. However, this study concerns only the AORC-CONUS domain, which includes the contiguous U.S., southern Canada, and northern Mexico, on a geographic grid at 30-arc-second (approximately 800 m) resolution. AORC temporal coverage is hourly, from February 1979 to the present (more than 44 years at the time of this writing).

The remainder of this introduction discusses the history behind the creation of the AORC-CONUS dataset and summarizes its requirements. The study that follows provides additional details on the dataset, including its spatial domain, its input data sources, and the methods used in its generation, in particular how those methods evolve in time alongside changes in the availability and quality of input data sources. Finally, the study provides an evaluation of AORC daily precipitation from 1979 to 2021, illustrating both the overall accuracy of AORC precipitation and the ways in which it responds to the evolution of its sources and methods.

## 1.1 | Background

### 1.1.1 | Planning and preparation (2010–2015)

In the early 2010s, the National Weather Service identified the need for a continental-scale, climatology-constrained, multi-decade, near-surface weather record, encompassing at minimum the contiguous United States (CONUS) and contributing watersheds and consistent in its quality, that is, in its bias and error structure. This *analysis of record* would drive distributed land surface and hydrologic models for the purpose of calibrating such models as operated at NWS RFCs. Around the same time, the NWS began planning for a CONUS-scale hydrologic modeling capability to be operated at a national center. In the context of that planned capability, which became the National Water Model (NWM), the analysis of record was imagined not just as a specifically historical product, but also as the final step in—and logical conclusion of—a robust operational process to generate spatially discretized (i.e., gridded) driving data for operational water forecasts. Consequently, the analysis of record would consist of both a historical gridded dataset assembled from the best available data from recent decades, and a near-real-time component updated at short lags (from days to weeks) relative to real time. The transition from historical to near-real-time datasets was expected to coincide with the initiation of operations for the first NWM release calibrated using the analysis of record.

### 1.1.2 | AORC version 1.0 (2015–2020)

With the establishment of the Office of Water Prediction in 2015 and the development and implementation of the NWM in 2016, the assembly of the analysis of record continued, and it came to be known as the AORC. AORC-CONUS version 1.0 was completed in late 2019, and the first calibration of the NWM using AORC-CONUS 1.0 was performed in preparation for the release of NWM 2.1, which entered into operations on April 20, 2021.

**TABLE 1** Improvements made for the CONUS Analysis of Record for Calibration (AORC-CONUS), version 1.1.

Element or topic	Period	Implemented improvement
Precipitation	1979–2010	Fully replaced AORC with NLDAS-2 for any hour in which values >280 mm/h or otherwise invalid (e.g., negative) grid values were detected in AORC-CONUS 1.0
Precipitation	2012–2021	Replaced NLDAS-2 daily precipitation over Ontario and the Great Lakes with CaPA for May 2012 through December 2015; use of CaPA continues thereafter but mainly replaces CPC daily precipitation
Specific humidity	1979–2021	Replaced zero values in AORC-CONUS 1.0 with a minimum non-zero threshold
Domain	1979–2021	Extended the domain to match that of NLDAS-2 for all elements
Land-water mask	1979–2021	Filled in missing values over water bodies in AORC-CONUS 1.0 with NLDAS-2 for all elements
Missing files	1979–2021	Used NLDAS-2 where any hourly files were missing in AORC-CONUS 1.0

Note: “NLDAS-2” refers to forcing data from the North American Land Data Assimilation System project, phase 2 (Xia et al., 2009). “CaPA” refers to the Government of Canada’s 10-km Regional Deterministic Precipitation Analysis (RDPA) based on the Canadian Precipitation Analysis (CaPA) system (Fortin et al., 2018). “CPC” refers to the NOAA Climate Prediction Center’s Global Unified Gauge-Based Analysis of Daily Precipitation (Xie et al., 2007).

### 1.1.3 | AORC version 1.1 (2020–present)

After receiving feedback from several NWS RFCs and from the NWM 2.1 calibration/validation team, the Office of Water Prediction performed numerous improvements to produce AORC-CONUS version 1.1; these are listed in Table 1. As those changes were implemented, the AORC-CONUS 1.1 team also established a near-real-time updating system, which produces hourly AORC grids at a 9-day lag relative to real time. NWM version 3.0, scheduled for release in 2023, is calibrated with AORC-CONUS version 1.1.

## 1.2 | Requirements

### 1.2.1 | Geophysical requirements

The geophysical requirements of the AORC are summarized in Table 2. Fulfillment of these requirements provides a dataset that can be used to drive most land surface and hydrologic models relevant to CONUS watersheds at a regional or continental scale, with sufficient spatial detail to produce re-analyses of land surface conditions (including snowpack, soil moisture, and runoff) and streamflow simulations such as those used in operational hydrologic forecasting. A 44-year (and expanding) time period is sufficient to satisfy the calibration requirements of most hydrologic modeling systems currently in use (Boulmaiz et al., 2020), including the NWM (Lahmers et al., 2021).

Existing datasets such as the North American Land Data Assimilation System project, phase 2 (NLDAS-2; Xia et al., 2009) and the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset (Daly et al., 2008) partially fulfill the requirements of the AORC, but no single existing product sufficiently fulfills them all. For example, the grid resolution of NLDAS-2 (0.125°, about 12 km) is too coarse; PRISM provides only a subset of the needed hydrometeorological elements, and its primary coverage is restricted to the CONUS. To fulfill all the geophysical requirements of the AORC, multiple datasets are systematically assembled using an approach designed to take best advantage of the favorable qualities of each constituent.

### 1.2.2 | Scientific requirements

Operational hydrologic forecasts, including those within different cycles of the NWM, draw their forcing data from a variety of numerical weather prediction (NWP) systems. Given such diversity, it must be assumed that these forcing data, and their counterparts used for model calibration, are unbiased, have covariance structures similar to that of observed quantities, and possess favorable error statistics. Depending on hydrologic model adaptability and calibration strategy (Oudin et al., 2006; Wang et al., 2023), the extent to which this is or is not the case for the AORC dataset determines its viability for calibration, as bias and error in calibration forcings may lead to uncertain or poorly specified model parameters, directly impacting model reliability and performance (Renard et al., 2010). While accuracy is the key scientific requirement of a forcing dataset intended for hydrologic model calibration, temporal consistency of bias and error behavior across the periods for which hydrologic model calibration and validation are performed is of comparable importance (Mizukami & Smith, 2012). As this study shows,

**TABLE 2** Summary of geophysical requirements for the AORC over the CONUS domain.

Category		Requirement	
Spatial coverage		CONUS RFC region boundaries: –125° E to –67° E longitude, 24° N to 53° N latitude	
Spatial resolution		Grid/raster: $\leq 1 \text{ km}^2$	
Temporal coverage		1979–present	
Temporal resolution		1 h	
	Name	Abbreviation	Vertical level
Hydrometeorological elements	Total precipitation	APCP	Surface
	Downward shortwave radiation flux	DSWRF	Surface
	Downward longwave radiation flux	DLWRF	Surface
	Pressure	PRES	Surface
	Air temperature	TMP	Shelter height (2 m)
	Specific humidity	SPFH	Shelter height (2 m)
	Zonal (U) wind	UGRD	Anemometer height (10 m)
	Meridional (V) wind	VGRD	Anemometer height (10 m)

constraining precipitation via climatology is an effective means of maintaining temporal and spatial consistency, as long as it does not do away with actual extreme events and other real-world statistical anomalies.

## 2 | CHARACTERISTICS OF THE AORC-CONUS 1.1 DATASET

### 2.1 | Spatial coverage and raster geometry

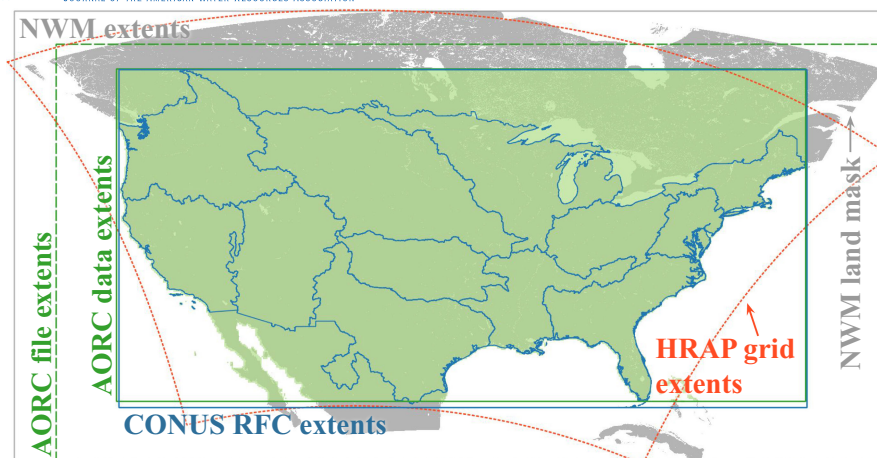
The AORC-CONUS dataset is generated on a geographic (longitude/latitude) World Geodetic System 1984 (WGS 84) coordinate system, at 30-arc-second (roughly 800m) spatial resolution. The spatial coverage of the AORC-CONUS domain, illustrated in [Figure 1](#), is a bounding box from roughly –125° E to –67° E longitude, and from roughly 25° N to 53° N latitude. While these coordinates do not fully encompass the National Water Model's operational CONUS domain, they do accommodate calibration within inland U.S. basins where model calibration is generally practiced, including the bounds of all CONUS RFC regions apart from the easternmost portion of the Northeast RFC region in New Brunswick, the southernmost portion of the West Gulf RFC region in Mexico, and the Florida Keys. Within its usable extents, the AORC dataset does not contain missing values in any inland location, including over large water bodies such as the Great Lakes. Its data/no-data mask, which excludes only offshore areas of the Atlantic Ocean, the Pacific Ocean, and the Gulf of Mexico, is largely consistent with that of forcing grids from NLDAS-2, an essential input data source for the AORC. Note that the full extents of the AORC grid as defined in its data files (labeled “AORC file extents” in [Figure 1](#)) exceed those cited above (labeled “AORC data extents” in [Figure 1](#)); those outermost rows and columns of data grids are filled with missing values, and are the remnant of an early set of required AORC extents that have since been adjusted inward.

### 2.2 | Essential climatologies

Because temperature and precipitation exert a dominant influence on hydrologic processes relative to other hydrometeorological elements (listed in [Table 2](#)), the AORC-CONUS 1.1 relies on gridded monthly 30-year climatologies of mean precipitation, mean daily minimum temperature, and mean daily maximum temperature to constrain the behavior of those elements. Prior to the development of the AORC, no single climatology grid in existence provided these quantities over the entire AORC-CONUS spatial domain; therefore, several products were combined to fulfill the need for a complete climatology.

The 1981–2010, 800-m CONUS PRISM dataset (Daly et al., 2008) provides the necessary grids over most of the AORC-CONUS domain. For coverage outside the CONUS, a regional 1981–2010 PRISM climatology covering the Canadian portion of the Columbia River basin (Pacific Climate Impacts Consortium, University of Victoria; available at <https://www.pacificclimate.org/data/prism-climatology-and-monthly-timeseries>; last accessed March 2023) is used, and remaining portions of southern Canada are filled with the 5 km NOAA Monthly U.S. Climate Gridded Dataset (NClimGrid-Monthly; Vose et al., 2014a, 2014b). The NClimGrid-Monthly provides coverage over northern Mexico but we found that it suffers from poor accuracy there; consequently, an independent climatology for the region was developed at OWP using





**FIGURE 1** AORC spatial domain. This figure depicts bounding rectangles of the National Water Model (NWM) CONUS domain, AORC grids, and CONUS River Forecast Center (RFC) extents, in geographic (longitude/latitude) coordinates. The area shaded in green is the AORC data/no-data mask. The “NWM land mask” is included to illustrate the distortion between the Lambert conformal NWM grid and geographic coordinates, and to provide a familiar visual reference for the bounding rectangles. Finally, the inverse-projected extents of the polar stereographic Hydrologic Rainfall Analysis Project (HRAP; Reed & Maidment, 1999) grid are shown as an orange dotted line. A substantial portion of AORC processing is performed on this grid.

1981–2010 station climatic values from Mexico's Servicio Meteorológico Nacional (SMN). These were interpolated to 30-arc-second resolution using the ANUSPLIN 4.4 thin-plate spline package (Hutchinson & Xu, 2013), with required gridded terrain heights provided by the Shuttle Radar Topography Mission (SRTM; Farr et al., 2007; NASA, 2013). The resulting precipitation climatology grids show appreciable improvement in terms of absolute accuracy relative to the NClimGrid-Monthly dataset, and to both the WorldClim v1.4 1960–1990 climatology (Hijmans et al., 2005) and the Uniatmos 1902–2011 climatology (Fernández Eguarte et al., 2012). Precipitation root-mean-squared error (RMSE) is reduced by 25%–50% relative to the Uniatmos climatology for most months at stations excluded from the spline analysis. For temperature, RMSE is usually  $<1.5^{\circ}\text{C}$  for both Uniatmos and OWP climatologies.

To reduce discontinuities at the boundaries between sources and to preserve information from the PRISM datasets over adjacent non-PRISM areas, NClimGrid-Monthly values within 40 km of PRISM cells were blended with nearby PRISM data via weighted average. Finally, the PRISM and NClimGrid-Monthly datasets are both undefined over large inland water bodies, such as the Great Lakes, Lake Winnipeg, and the Great Salt Lake. Values for these areas were estimated from 30-year (1981–2010) monthly mean values from NLDAS-2, which were then modified to agree with the original climatologies at water boundaries by treating adjacent PRISM values as boundary values and performing statistical relaxation using NLDAS-2 as the initial guess.

The combined climatology produced through these efforts is hereafter referred to as the PRISM/NCEI/OWP climatology.

## 2.3 | Methodology for non-precipitation elements

While the subsequent portions of this study focus on AORC precipitation, in this section the sources and methods used to generate non-precipitation elements of the AORC are described.

### 2.3.1 | Air temperature

Apart from precipitation, 2-m air temperature is widely regarded as the physical element that most significantly influences runoff processes, via its approximately determinative relationships with potential evapotranspiration, precipitation phase, and snowmelt (McCabe & Wolock, 2011). Furthermore, other near-surface hydrometeorological fields in the AORC are adjusted for consistency with temperature; therefore, we begin with that element.

For the years 1979–2015, AORC-CONUS air temperature grids are generated by merging two datasets. The first consists of 2-m temperatures from hourly NLDAS-2 land surface forcing fields, delivered at  $0.125^{\circ}$  ( $\sim 12\text{-km}$ ) spatial resolution and 1-h temporal resolution, having been interpolated from a native 40-km/3-h resolution (Cosgrove et al., 2003). The second are Livneh et al. (2015) minimum and maximum temperatures, which have a  $\frac{1}{2}^{\circ}$  ( $\sim 16\text{-km}$ ) spatial resolution and are available at both daily and monthly frequencies (both frequencies contribute

to AORC-CONUS temperatures). While the original Livneh et al. (2015) data were available only through 2013, the dataset was extended after publication, in 2016, to cover the years through 2015 using the same monthly bias/downscaling ratios generated for the 1981–2010 period. This dataset is hence referred to as LIV16. In the most general terms, the merging process downscales hourly NLDAS-2 temperatures using a method designed to maximize their consistency with LIV16 daily temperature extrema, the latter having been climatologically downscaled using the difference between the PRISM/NCEI/OWP climatology of monthly temperature extrema and the corresponding averages from 30 years (1981–2010) of monthly LIV16 temperature extrema.

In 2016, AORC-CONUS temperatures begin a transition to hourly temperature grids derived from the 2.5-km/1-h UnRestricted Mesoscale Analysis (URMA; Pondeva et al., 2015). URMA is generally considered an authoritative gridded analysis for model forecast verification within the NWS. For 2016 and 2017, URMA temperatures are climatologically downscaled and adjusted by the differences between monthly 2010–2015 URMA means and the corresponding 2010–2015 AORC means. Beginning in 2018, AORC-CONUS 1.1 temperatures are nearest-neighbor sampled directly from URMA grids, with no additional measures to downscale them from 2.5- to 1-km resolution.

### 2.3.2 | Pressure and specific humidity

From 1979 to 2017, terrain-based adjustment of hourly surface pressure from the input data source is performed following the calculation of 2-m temperatures in such a way as to preserve hydrostatic balance between the near-surface hydrometeorological state of the input source and its AORC counterpart, based on the difference between their surface elevations. For specific humidity, atmospheric demand for water vapor is preserved by maintaining the 2-m relative humidity of the input source, given the previously calculated AORC temperature and pressure. This approach is identical to that used for downscaling retrospective forcing fields in the original NLDAS project (Cosgrove et al., 2003). The input source for surface pressure and 2-m specific humidity is NLDAS-2 for 1979–2015, followed by URMA thereafter, with Global Data Assimilation System (GDAS) cycles of the Global Forecast System (GFS) atmospheric model (Kleist et al., 2009) providing data outside the bounds of the URMA domain. From 2018 onward, no terrain-based adjustment of URMA pressure and specific humidity is performed because temperature, pressure, and humidity are all nearest-neighbor sampled, unmodified, from URMA grids.

### 2.3.3 | Downward longwave radiation

From 1979 to 2015, hourly downward surface longwave radiation fluxes (DLWRF) from NLDAS-2 are adjusted for consistency with the AORC temperature and emissivity (a function of pressure, specific humidity, and temperature). As is the case for pressure and specific humidity, the procedure followed is identical to that described in Cosgrove et al. (2003). After 2015, URMA is the primary source of temperature, pressure, and specific humidity in the AORC. Nevertheless, adjustment of NLDAS-2 DLWRF using those temperatures and derived emissivities continues for 2016 and 2017, justified by the climatological similarity between NLDAS-2 and URMA temperatures during that time. Beginning in 2018, however, no such reasoning applies, and NLDAS-2 DLWRF grids are bilinearly interpolated to the AORC grid without further modification.

### 2.3.4 | Downward shortwave radiation

For downward surface shortwave radiation, no downscaling process apart from simple bilinear interpolation is applied to input data sources. NLDAS-2 is used from 1979 onward, with GDAS-derived grids available beginning in 2018 as a contingency for missing NLDAS-2 data.

### 2.3.5 | U and V wind

Hourly U and V wind grids at the 10-m anemometer height are bilinearly interpolated from NLDAS-2 from 1979 to 2017. From 2018 onward, the input grids are provided by URMA and by GDAS when and where URMA is unavailable.

## 2.4 | Methodology for precipitation

The general strategy for generating hourly gridded precipitation for the AORC, described with additional detail in the subsections below, begins with high-accuracy and low-bias monthly gridded precipitation estimates. These are conservatively disaggregated, first into daily (herein defined as the 24-h accumulation from 12 UTC to 12 UTC) and finally into hourly estimates, using datasets that are less constrained, and on

their own more prone to error and bias, but provide the desired time step sizes. For the first 31 years, from 1979 to 2009 (to which this study refers as the “historical period”), this approach is explicitly followed. During an interim 6-year period from 2010 to 2015, daily estimates are subdivided into hourly amounts as during the historical period, but total monthly precipitation is not directly adjusted to match a gridded analysis; instead, it is constrained for climatological consistency. Finally, beginning in 2016, daily estimates are subdivided into hourly amounts, but those daily estimates are unfettered, with neither a monthly precipitation analysis nor a climatology to constrain them.

Temporal and spatial manipulation of observation data to produce gridded datasets can suppress extreme events present in the underlying observation data (Pierce et al., 2021). More generally, the statistical behavior of those gridded results may differ significantly from that of observations, in particular by possessing a higher frequency of low-intensity precipitation and a lower frequency of heavy precipitation (Ensor & Robeson, 2008). These impacts may be reduced in the AORC through the use of observation-based daily and hourly datasets to subdivide monthly precipitation from 1979 to 2015; however, these disaggregation methods, and the gridded datasets used in their implementation, themselves have the potential to misrepresent precipitation extremes and further disrupt the intensity-duration-frequency characteristics of the resulting dataset. A comparison of precipitation intensity distributions from the AORC with those of observations is outside the scope of this study, but users of the data are advised to carefully consider these matters.

The remainder of this section describes in more detail the methods used to generate monthly and associated daily precipitation amounts for the AORC for the three eras described above, followed by a discussion of disaggregation of daily precipitation into hourly amounts.

## 2.4.1 | Monthly and daily precipitation

The availability and accuracy of different gridded input datasets have evolved significantly over the years covered by the period of interest for the AORC (1979–present), as listed in Table 3. Consequently, the methods for combining them into a minimally biased and consistent product follow that evolution. Note that all processing of daily precipitation is performed on the 4.7625-km HRAP grid after converting input data sources, if necessary, to that coordinate system via budget regridding/reprojection, with downscaling to the AORC grid to be performed later. The budget regridding/reprojection method used in AORC processing calculates each output grid value as the simple average over a 5×5 raster that subdivides the (HRAP) output cell into a 0.9525-km resolution subgrid, with bilinear interpolation of the input grid performed at the center of each of those 25 subgrid cells. This approach, provided by the IPOLATES software library via the wgrib2 utility (National Centers for Environmental Prediction, Version 0.2.0.4, February 2016), approximates more precise conservative regridding methods such as that described by Jones (1999).

**TABLE 3** Summary of input datasets for monthly and daily (i.e., the 24-h accumulation ending at 12 UTC) AORC precipitation.

Dataset	Monthly or daily grids	Spatial resolution	Dates applied to AORC	Notes
LIV16	Monthly	$\frac{1}{2}^{\circ}$ (~16 km)	Feb. 1979–Dec. 2015	Strong constraint on AORC precipitation from 1979 to 2009; climatological constraint from 2010 to 2015
NLDAS-2	Daily	$\frac{1}{2}^{\circ}$ (~12 km)	Feb. 1979–present	Primary daily data source from 1979 to 2001; mainly used for OCONUS after 2001; used only for gap-filling after 2015
Stage IV	Daily	4.7625 km	Jan. 2002–present	Primary daily CONUS data source beginning in January 2002
CPC	Daily	$\frac{1}{2}^{\circ}$ (~48 km)	Jan. 2016–present	Primary daily OCONUS and secondary daily CONUS data source beginning in January 2016
CaPA	Daily	10 km	May 2012–present	Primary source of daily precipitation for the Canadian portion of the lower Great Lakes region from May 2012 till present

Note: LIV16= Livneh et al. (2015) dataset, extended through 2015; NLDAS-2=North American Land Data Assimilation System project, phase 2 (Xia et al., 2009); Stage IV=precipitation grids (Lin & Mitchell, 2005; Du, 2011) mosaiced from U.S. River Forecast Center (RFC) quantitative precipitation estimates (QPE); CPC=NOAA Climate Prediction Center Global Unified Gauge-Based Analysis of Daily Precipitation (Xie et al., 2007); CaPA=Government of Canada's Regional Deterministic Precipitation Analysis (RDPA) based on the Canadian Precipitation Analysis (CaPA) system (Fortin et al., 2018).

#### 2.4.1.1 | Historical period (1979–2009)

Precipitation processing for this period begins with LIV16 monthly precipitation grids covering the full 31-year period. For the first 23 years (February 1979 through December 2001), disaggregation from monthly LIV16 to daily amounts is based on NLDAS-2 daily (12 UTC to 12 UTC) precipitation accumulations. Beginning in January 2002, Stage IV daily precipitation grids (Du, 2011; Lin & Mitchell, 2005) mosaiced from U.S. RFC quantitative precipitation estimates (QPE) are used where available over the CONUS; consequently, the influence of NLDAS-2 on daily precipitation within the bounds of CONUS RFC regions is diminished after 2001. However, given the varied nature and priorities of RFC operations, the Stage IV mosaic occasionally has regions of missing data, typically associated with non-receipt of QPE from one or more RFCs. When this circumstance arises, NLDAS-2 precipitation is used to fill areas of missing Stage IV data. In areas outside the coverage of LIV16 precipitation grids, NLDAS-2 daily precipitation grids are used with no monthly constraint.

#### 2.4.1.2 | Interim period (2010–2015)

The PRISM/NCEI/OWP climatology described above in Section 2.2 covers the period 1981–2010, and is therefore no longer directly applicable after 2010; however, the LIV16 dataset does cover the years 2010–2015, as discussed in Section 2.3. To generate AORC precipitation grids for this period, monthly LIV16 gridded precipitation accumulations are not explicitly disaggregated into daily amounts as they are for the years 1979–2009. As is done for the latter portion of the historical period, daily precipitation estimates are built from Stage IV grids, with NLDAS-2 used outside the coverage of Stage IV grids. However, instead of adjusting these grids so their monthly accumulations match that of the LIV16 monthly product (as was done for 1979–2009), they are bias-adjusted such that the 6-year mean for each grid location for each calendar month is equal to the corresponding 6-year mean from the LIV16 dataset. Because the LIV16 dataset for 2010–2015 is generated using a spatial distribution and bias adjustment based on the 1981–2010 PRISM climatology, this approach effectively does the same for the AORC data for those years. Finally, internal evaluations revealed AORC-CONUS version 1.0 precipitation to possess high bias in the portion of southern Canada near the lower Great Lakes beginning in 2012 (see Gronewold et al., 2018), and replacement of NLDAS-2 with the Government of Canada's 10-km Regional Deterministic Precipitation Analysis (RDPA) based on the Canadian Precipitation Analysis (CaPA) system (Fortin et al., 2018) was retrospectively applied in this area beginning in May 2012 as part of the AORC-CONUS 1.1 effort (see Table 1). The area of interest, which includes lakes Huron, Erie, and Ontario, is defined by a polygon, bounded on the west by Lake Huron, on the north by a path between Sault Ste. Marie and Montreal, Canada, and on the south and east by the U.S. shores of the Great Lakes and the U.S./Canada border.

#### 2.4.1.3 | Near-real-time operations (post-2015)

The LIV16 dataset is unavailable after 2015. This circumstance has thus far prevented AORC precipitation from undergoing any bias adjustment for dates beginning in January 2016, and no disaggregation of monthly precipitation analysis is performed during this time either. Instead, daily precipitation amounts are generated by combining, in descending order of preference, Stage IV 24-h precipitation, the 0.5-degree CPC Global Unified Gauge-Based Analysis of Daily Precipitation (Xie et al., 2007), and daily CaPA grids, using the highest priority product available for each grid location. As was done from May 2012 through December 2015, CaPA daily precipitation supplants other precipitation estimates (primarily the CPC analysis) over the lower Great Lakes region in Canada. In spite of the lack of bias adjustment on precipitation analyses during this period—which includes near-real-time updates—improvements in QPE generation such as the key contribution of Multi-Radar, Multi-Sensor (MRMS) precipitation products (Zhang et al., 2016) have improved the skill of Stage IV QPE since early 2015. Even so, the lack of bias adjustment does alter the nature of AORC-CONUS 1.1 precipitation beginning in 2016, as the sections to follow illustrate.

### 2.4.2 | Hourly precipitation

Disaggregation of daily precipitation accumulations into 1-h amounts is based on gridded hourly precipitation from a variety of input sources, these varying significantly in their availability over the period of interest. These are listed in Table 4, and the timeline for which each is used, including occasional data gaps, is illustrated in Figure 2. Note that total daily accumulation is never altered by the disaggregation process.

For daily precipitation estimates from February 1979 through December 2015, all available hourly input sources are combined using a weighted average. The associated weighting grids for each day are generated so as to maximize the Nash–Sutcliffe Efficiency (NSE), calculated over regional subgrids, between aggregated 1-h precipitation and the associated daily precipitation estimate. In general, for any grid location, the weighted average is based on the best-performing (highest NSE) source receiving a weight of 1.0 and all others receiving weights of zero, but a smoothing process blends the weights to prevent discontinuous results.

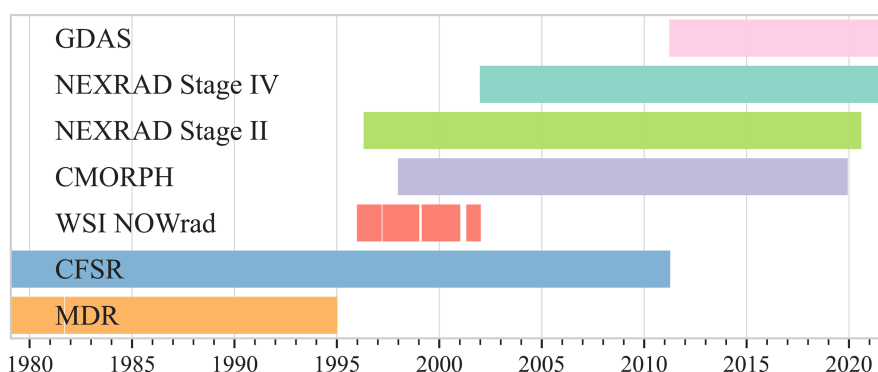
Beginning in 2016, a simpler method is used. Disaggregation at each HRAP grid location is performed using one of the following, in order of preference depending on what is available: NEXRAD Stage IV, NEXRAD Stage II, CMORPH, or GDAS.

As was the case with daily precipitation processing, the above steps are performed on the HRAP grid after budget regridding of input data sources to that coordinate system. Downscaling to AORC grid coordinates using the PRISM/NCEI/OWP climatology, and bias adjustment for consistency with LIV16 (monthly accumulations for 1979–2009, monthly climatologies for 2010–2015) is the final step in processing.

**TABLE 4** Input data sources used to disaggregate 24-h precipitation into 1-h accumulations in AORC-CONUS 1.1.

Hourly precipitation dataset	Spatial resolution	Dates applied to AORC
MDR	47 km	Feb. 1979–Dec. 1994
CFSR	38 km	Feb. 1979–Mar. 2011
WSI-NOWrad	4.7625 km	Jan. 1996–Dec. 2001
CMORPH	8 km	Jan. 1998–Nov. 2019
NEXRAD Stage II	4.7625 km	May 1996–Jul. 2020
NEXRAD Stage IV	4.7625 km	Jan. 2002–present
GDAS	$\frac{1}{4}^{\circ}$ (~24 km)	Apr. 2011–present

Note: MDR=Manually Digitized Radar (Miller & Kitzmiller, 2017); CFSR=Climate Forecast System Reanalysis (Saha et al., 2010); WSI-NOWrad=Weather Services International National Operational Weather Radar (Zhang et al., 2017); CMORPH=Climate Prediction Center Morphing technique (Xie et al., 2017, 2019); NEXRAD=Next Generation Weather Radar (Lin & Mitchell, 2005); GDAS=Global Data Assimilation System cycles of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) model (Kleist et al., 2009).

**FIGURE 2** Availability of gridded input data sources for disaggregating 24-h precipitation into 1-h accumulations in AORC-CONUS 1.1, for February 1979–December 2021.

### 3 | PRECIPITATION EVALUATION

The following evaluation of AORC-CONUS 1.1 precipitation is focused on daily (24 h) precipitation amounts ending at 12 UTC. This approach allows for the use of a large, well-established database of observations, but it does not consider the question of how successfully AORC processing procedures subdivide daily precipitation into hourly amounts. A discussion of intense hourly precipitation occurrences in the AORC follows the evaluation results.

#### 3.1 | Observation data

The observations with which AORC-CONUS 1.1 precipitation and its gridded input data sources are compared in this study are provided by the Global Historical Climatology Network–Daily (GHCN-D) database, version 3.28 (Menne, Durre, Korzeniewski, et al., 2012; Menne, Durre, Vose, et al., 2012). The subset of the database used in this study includes all observing locations in the U.S., Canada, and Mexico (US/CA/MX hereafter), excluding the following:

- Reports for which a local time of observation of 12:00 or later is indicated (reports for which no local time is indicated are not excluded; see the following discussion);
- “Global Summary of the Day” reports (source flag “S”), which are derived from hourly synoptic reports and may differ significantly, particularly in the 24-h period they represent, from daily precipitation measured at or near 12 UTC (Menne, Durre, Korzeniewski, et al., 2012; Menne, Durre, Vose, et al., 2012);
- Reports flagged as having failed any quality assurance check (Durre et al., 2010).

The first two criteria ensure that only a small proportion of the remaining observations were collected more than 2–3 h from 12 UTC. A more detailed discussion of the reasoning behind this assertion, and of the timing of GHCN-D precipitation reports in general, is provided in the “Observation Times of GHCN-Daily Precipitation Reports” section of the [Appendix](#).

### 3.2 | Evaluation methodology

For each evaluation covering a particular time period and spatial domain, all qualifying GHCN-D precipitation observations are selected from the US/CA/MX database. Gridded precipitation estimates under investigation are then aggregated, if necessary, into 24-h amounts ending at 12 UTC. These are nearest-neighbor sampled at station coordinates, then paired with corresponding precipitation observations at those dates and locations. Stations must report on at least 50% of the days covering an evaluation period to be included in that evaluation.

To supplement the GHCN-D quality flags, two additional checks on precipitation reports are performed during each evaluation. First, any report of >10 inches of daily precipitation must be accompanied by at least nine other observations of >10 inches on the same day, without regard for the locations of these observations, to be included. Second, any report of either zero or >1 inch of daily precipitation must be corroborated by observations in its vicinity, if available. Otherwise, such a report is flagged as a local oddity and excluded from evaluation. Specifically, a report of zero is flagged if more than 95% of the reporting stations in its neighborhood observed >1 inch, and 100% of its nearest neighbors (defined as the nearest 5% of reporting stations in the neighborhood) observed >1 inch. A report of >1 inch is flagged if more than 95% of the reporting stations in its neighborhood observed  $\leq 0.1$  inch, and 100% of its nearest neighbors observed  $\leq 0.1$  inch. Reports at stations located at elevations more than 2 standard deviations from their mean neighborhood elevation (based on the elevations of neighboring stations) are exempt from this check. The neighborhood size for this check is a 100-km radius around any reporting station. This distance effectively balances practicability with the likelihoods of type I and type II errors, and a typical decorrelation distance for daily precipitation is on the order of 2–3 times this radius (van Leth et al., 2021).

Once observations and accompanying sampled precipitation estimates are collected and prepared, a variety of bias and skill metrics are calculated both across the spatial domain and the time domain. For metrics calculated across the spatial domain (which produce time series), data are collected over aggregation periods of 28 days. Metrics for a 28-day aggregation period are calculated using all the data in that sample population and attributed to the midpoint of the 28-day range. This approach produces more stable metrics for time series than what is obtained by calculating an independent metric for each day. Note that observations in any such sample population and their gridded, sampled counterparts are never themselves summed into 28-day accumulations; each daily pair of observed and gridded data at any given location is a discrete member of the sample, and any single observing station may provide up to 28 such pairs.

For metrics calculated across the time domain, a simpler approach applies: all the daily pairs of observed and gridded precipitation attributed to a particular station participate in the computation of its bias and skill metrics.

Among all metrics examined, this study relies largely on the unrestricted accumulation (multiplicative) bias (UAMB; the ratio of total analysis precipitation to total observed precipitation) to present our evaluation results. In some circumstances, it is beneficial to categorize events on a contingency table and treat bias and error as exclusive categories; for example, to calculate bias based only on hits (cases in which both gridded and observed daily precipitation exceed some threshold amount). It may also be worthwhile to examine the statistical qualities of individual values of multiplicative bias (MB; the ratio of analysis precipitation to observed precipitation on a station-by-station, day-by-day basis). By contrast, UAMB is simply the sum of all gridded samples in a population divided by the sum of all corresponding observed values, without regard for whether any particular gridded value would be judged a categorical error by any standard. Among a variety of metrics, UAMB has been found to best illustrate the overall accuracy of the gridded precipitation products used in the AORC, as well as that of the AORC itself. It provides a practical measure of the aggregate effect of bias and error across a given temporal and/or spatial domain. The use of UAMB is largely consistent with a multiplicative error model, which has been shown to better represent errors in estimates of daily precipitation than an additive model (Tian et al., 2013) and is often used to characterize precipitation uncertainty in hydrologic modeling applications (Kavetski et al., 2006; McMillan et al., 2011).

An additional examination of deviation and bias in AORC-CONUS 1.1 precipitation, provided in the “Error Statistics of AORC-CONUS 1.1 Precipitation” section of the [Appendix](#), demonstrates that multiplicative bias (MB) exhibits a distinctly log-normal tendency. In this study, we address this tendency by using logarithmic coordinates for all UAMB (vertical) axes in time-series plots, constructing color ramps that are logarithmically symmetric about a bias value of 1.0, and using geometric statistics as defined by Koopmans et al. (1964) whenever possible to quantify the behavior of both UAMB and MB.

In addition to AORC-CONUS 1.1 precipitation grids, 24-h Stage IV and NLDAS-2 precipitation estimates are also evaluated here, as these are the primary building blocks used to generate daily AORC precipitation. Daily LIV16 precipitation grids, while available, are not considered in this study. They do not contribute to AORC precipitation grids, as they are non-synoptic, representing accumulation periods ending at local midnight. Finally, it is worthy of note that all NWM versions prior to 2.1 were calibrated using NLDAS-2 forcings; therefore, the behavior of that dataset may provide clues as to how the calibration of those NWM versions may have differed from later versions calibrated with the AORC.

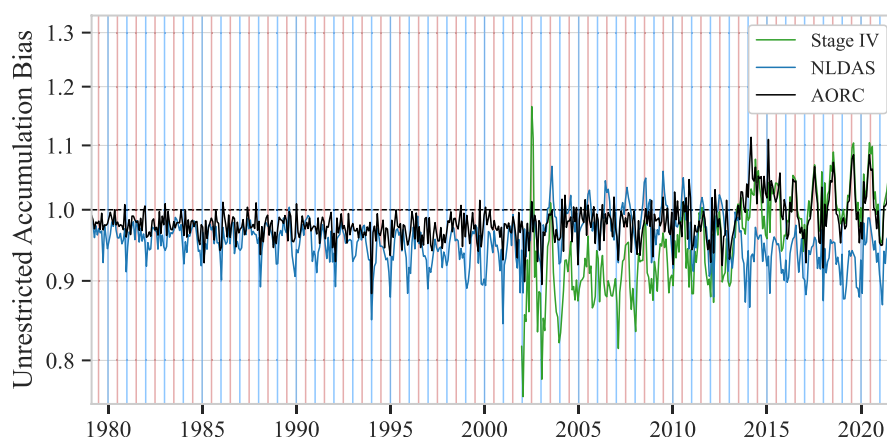


### 3.3 | Evaluation results

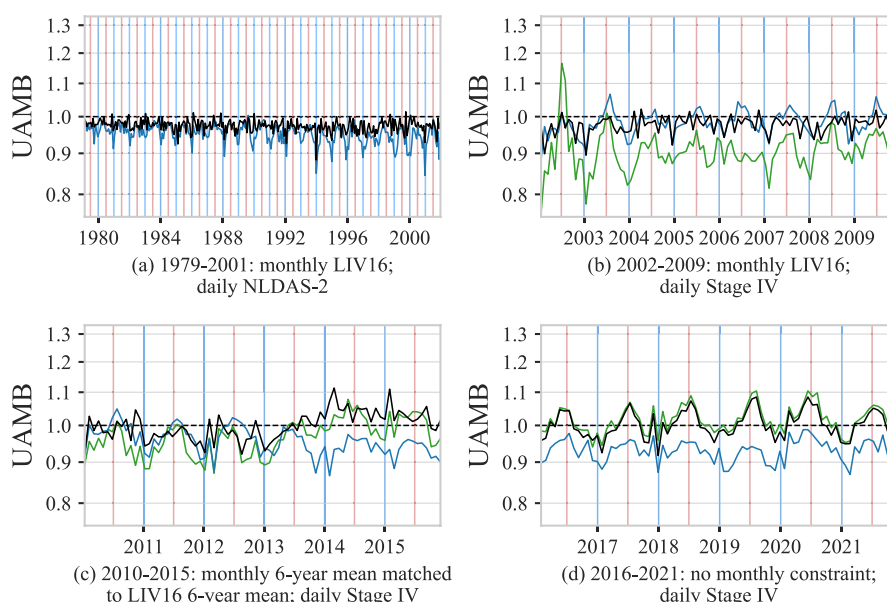
#### 3.3.1 | Overall bias behavior of the AORC

Figure 3 shows the UAMB of AORC, NLDAS-2, and Stage IV (beginning in 2002) precipitation calculated across the full CONUS domain from 1979 through 2021 using 28-day aggregation periods. This result demonstrates that the AORC shows little overall bias and generally less bias than its counterparts, with a minimum of 0.88, a maximum of 1.11, and a median value of 0.98. However, these values are calculated for the full CONUS domain, using all observations gathered over 28-day aggregation periods, and such an aggregation can conceal significant local and regional biases.

Figure 4 presents the time series of Figure 3 separated into each of the four sub-periods within which input data sources and methods were consistent. The latter part of the historical period, from roughly 2004 to 2009 (Figure 4b), in which Stage IV QPE mosaics contribute to daily precipitation estimates, shows a slight improvement in AORC bias over earlier periods. The first 2 years in which AORC uses daily



**FIGURE 3** Unrestricted accumulation bias (UAMB) of 24-h AORC-CONUS 1.1, NLDAS-2, and Stage IV precipitation compared with Global Historical Climatology Network-Daily (GHCN-D) observations from February 1979 through December 2021. Vertical blue and red lines indicate January 1 and July 1, respectively. The vertical axis is logarithmic.



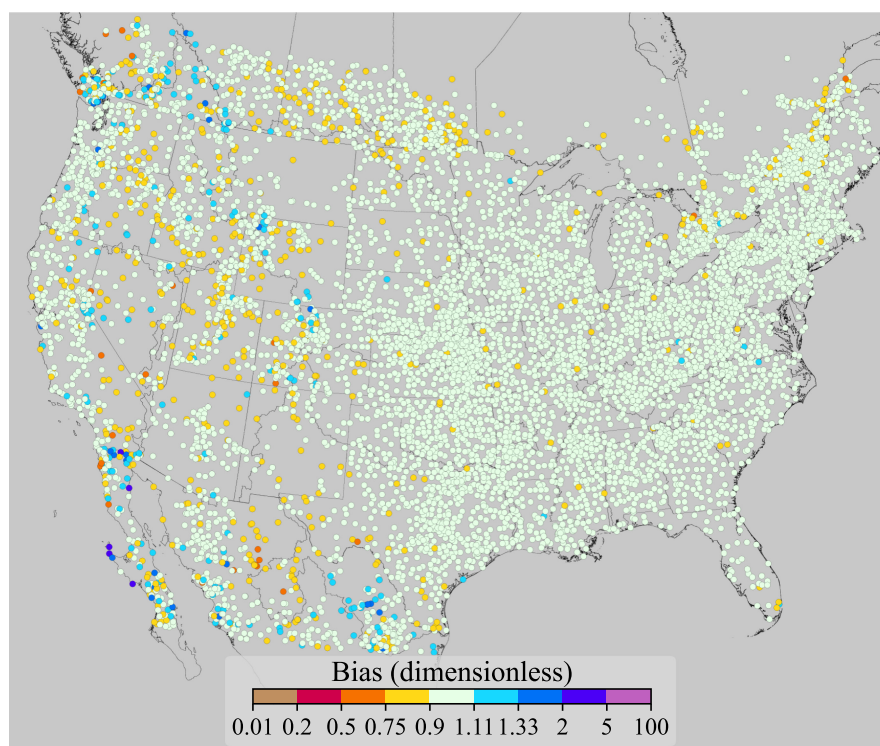
**FIGURE 4** UAMB time series for 24-h AORC-CONUS 1.1, NLDAS-2, and Stage IV precipitation compared with GHCN-D observations over the CONUS: (a) February 1979 through December 2001, (b) January 2002 through December 2009, (c) January 2010 through December 2015, and (d) January 2016 through December 2021. Vertical blue and red lines indicate January 1 and July 1, respectively. Individual captions describe the primary input datasets used during this period, including any constraints imposed on monthly precipitation totals. A legend for line colors is provided in Figure 3.



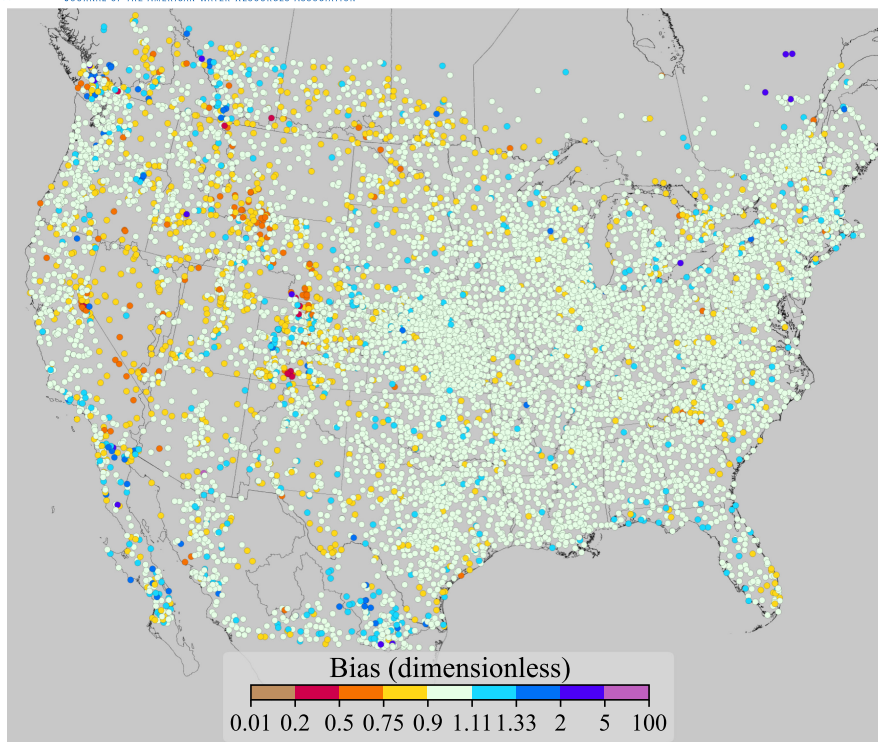
Stage IV QPE (2002–2003) show less improvement, likely due to the instability and seasonal bias extremes seen in the Stage IV during that time. Over the interim period from 2010 to 2015 (Figure 4c), the average precipitation for each calendar month is constrained by the 1981–2010 climatology, but monthly precipitation is not explicitly matched to LIV16 analysis. Consequently, more interannual variability is present during this time, with a mostly dry-biased period from early 2011 through mid-2013, followed by a mostly wet-biased period from mid-2013 through late 2015. Beginning in 2016, AORC precipitation is not constrained by climatology. The potential bias that this change might introduce is mitigated by improvement in the overall performance of the Stage IV QPE mosaics in late 2013 to early 2014, likely due to the operational deployment of MRMS QPE at that time. Operational MRMS products were rapidly integrated into QPE generation processes at many RFCs, and a sustained improvement in the bias performance of Stage IV QPE followed. However, AORC precipitation bias is also tightly correlated with that of the Stage IV beginning in 2016, and this correlation includes a clear annual variation in bias, with winter minima and summer maxima. This signal persists to varying degrees throughout all 20 years of the Stage IV QPE record examined in this study, but prior to 2016 the control asserted by LIV16 monthly precipitation and associated climatologies largely prevents the associated biases from impacting the AORC.

Examination of UAMB calculated across time at specific station locations reveals more about the performance of the AORC. In Figure 5, which shows UAMB at GHCN-D observing stations from 1979 to 2001, some bias extremes are apparent, particularly in Mexico and on coastlines, but with the exception of some dryness across the Western U.S. and Canadian Plains, most bias values are close to 1.0. The geometric mean (GM) of the UAMB values shown is 0.97, and the geometric coefficient of variation (GCV) is 7.8%. In Figure 6 (2002–2009), more extremes in bias, particularly dryness in the Western U.S., appear to have been introduced into the AORC. The influence of eastern portions of the domain, which contain more observing locations and suffer from less bias, prevents these dry areas from significantly affecting the time series of Figures 3 and 4. Still, the GM for this period is 0.98 and GCV is 11.5%, the latter value revealing an increase in spatial scatter over the results in Figure 5.

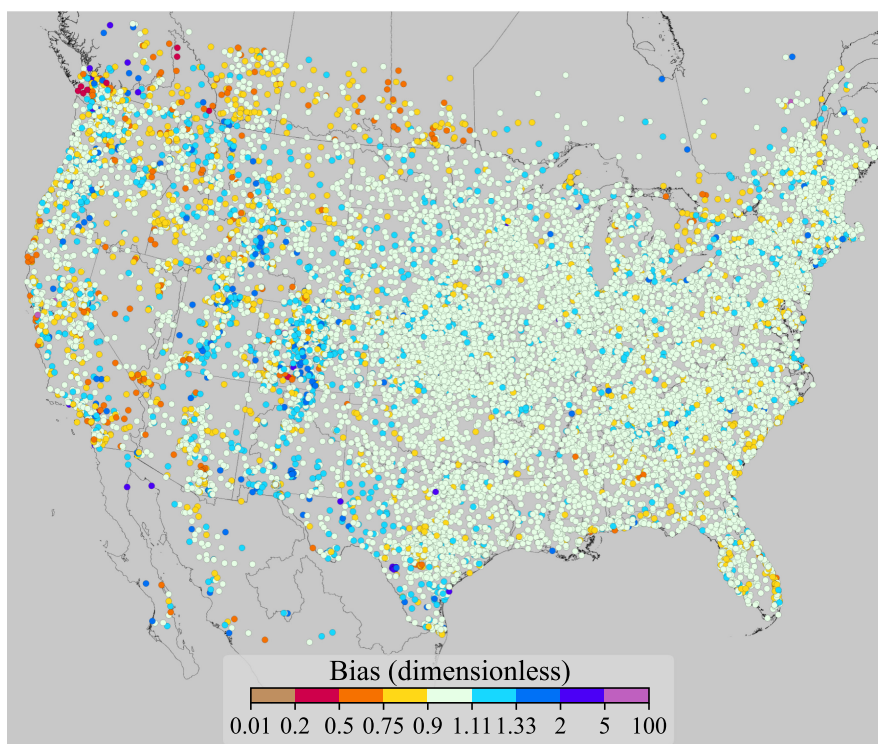
For the period from 2010 to 2015 (Figure 7), less systematic dryness in the Western U.S. has been replaced with greater overall variability in bias between both low (dry) and high (wet) results across the domain. The GM and GCV for this map are 1.00 and 13.2%. Finally, lacking climatological constraints, AORC precipitation during the period from 2016 to 2021 (Figure 8) suffers from far more significant systematic bias in several portions of the domain, including dry bias across all Western RFC regions, and still more increase in scatter relative to earlier periods: the GM and GCV in UAMB across the domain for the 2010–2015 map are 1.01% and 16.7%.



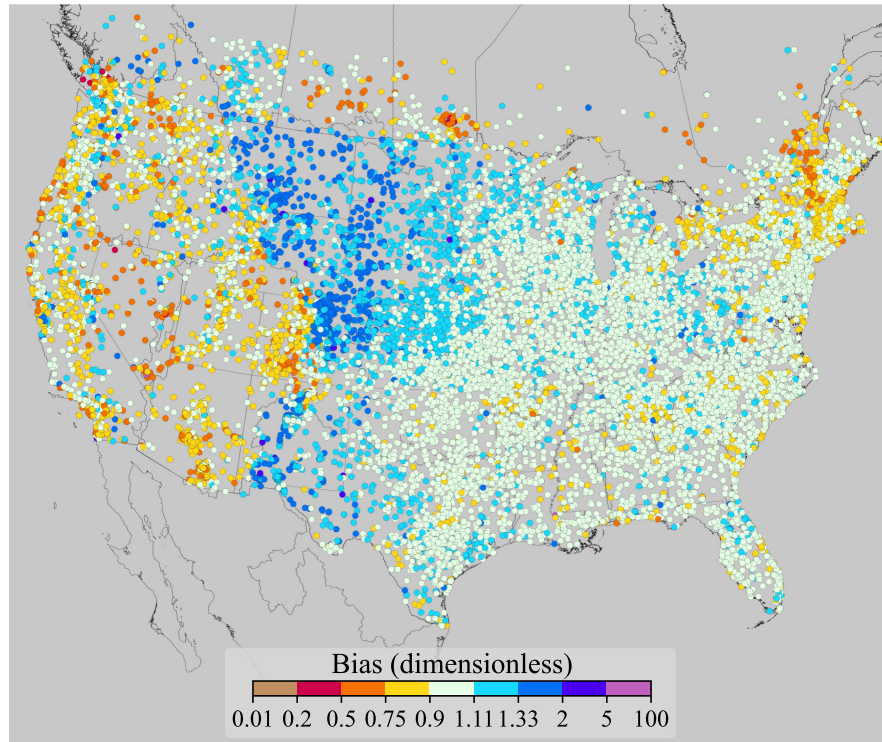
**FIGURE 5** UAMB of 24-h AORC-CONUS 1.1 precipitation compared with GHCN-D observations from February 1979 through December 2001.



**FIGURE 6** UAMB of 24-h AORC-CONUS 1.1 precipitation compared with GHCN-D observations from January 2002 through December 2009.



**FIGURE 7** UAMB of 24-h AORC-CONUS 1.1 precipitation compared with GHCN-D observations from January 2010 through December 2015.



**FIGURE 8** UAMB of 24-h AORC-CONUS 1.1 precipitation compared with GHCN-D observations from January 2016 through December 2021.

### 3.3.2 | Other evaluation metrics

The UAMB captures much of the error and bias behavior of the AORC and related datasets; however, additional insights can be gained from examining other statistical metrics, particularly those which quantify categorical skill and variability in bias.

First, we present a discussion of categorical skill in the AORC, which delivers an event-by-event analysis of AORC precipitation that the UAMB, given its aggregate nature, does not provide. For this analysis, a precipitation event is defined as any observation or grid value that exceeds 0.1 inch in a day (constrained to 12 UTC to 12 UTC as is the case throughout this study). We limit this discussion to two basic, well-known metrics:

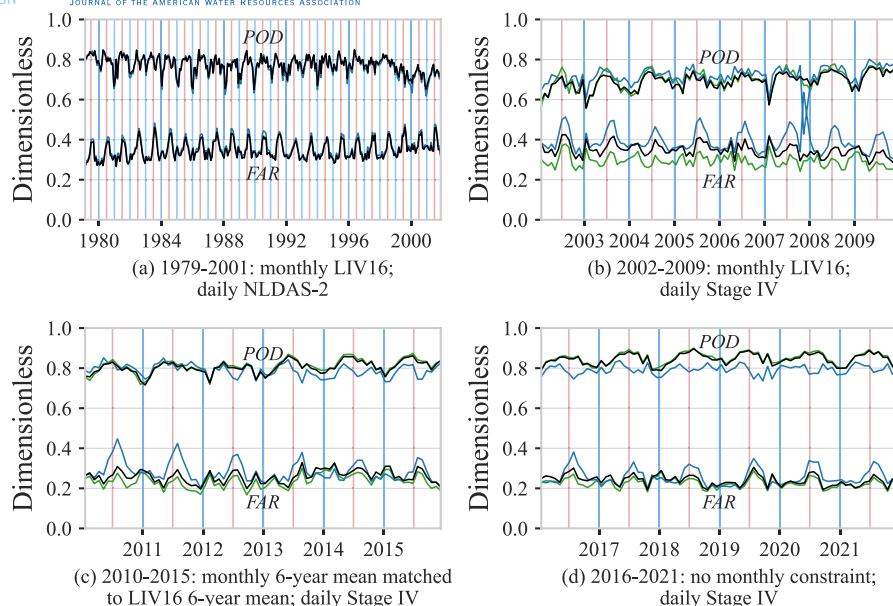
1. Probability of Detection (POD) =  $\text{hits} \div (\text{hits} + \text{misses})$
2. False Alarm Ratio (FAR) =  $\text{false alarms} \div (\text{hits} + \text{false alarms})$

Figure 9 depicts time series of POD and FAR for each of the four sub-periods previously examined. The AORC and its component counterparts exhibit a combination of biannual and annual variation in errors, in particular showing many local maxima in FAR near the summer solstice and local minima in POD near the winter solstice. These errors are particularly noticeable in the NLDAS-2 results, which impact the AORC across its full spatial domain through 2001, and continue to affect most areas outside the bounds of CONUS RFC regions through 2015, as discussed in Section 2.4.

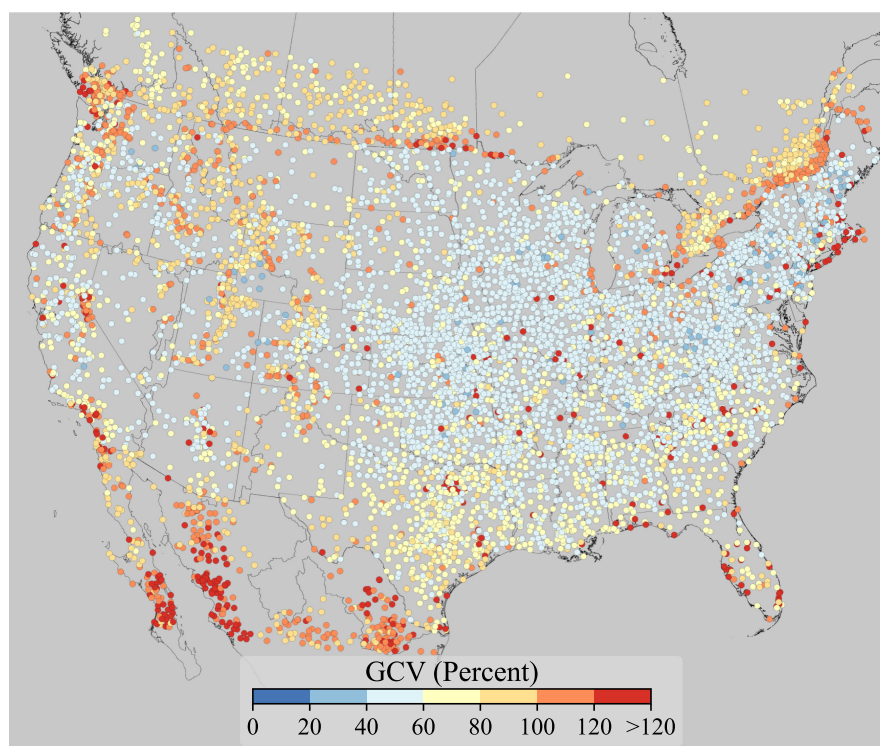
A note on the peculiarity in the POD and FAR in late 2007 (Figure 9b): this oddity occurs because for most of November and December 2007, NLDAS-2 precipitation coverage inside U.S. boundaries appears to be lagged by 4 days; for example, NLDAS-2 precipitation over the CONUS for December 14 is based on data from December 10. To our knowledge, this issue has not been reported elsewhere, and does not occur at any other time in the NLDAS-2 forcing data record.

Another characteristic of AORC behavior not illustrated by the UAMB is the spread in bias from location to location, and event to event. For this study, we measured the spread in bias using the GCV in multiplicative bias (GCVMB). Figure 10 shows the GCVMB in AORC-CONUS 1.1 precipitation for the entire evaluation period. Results for each of the four sub-periods (not shown) are similar, though the magnitudes of GCVMB outside the CONUS are largest for the 2002–2009 period. This bias spread is likely attributable to anomalies and unrealistic gradients in NLDAS-2 precipitation outside U.S. boundaries; these have previously been reported by Gronewold et al. (2018). To illustrate this proposition, the UAMB for NLDAS-2 precipitation for each of the four eras examined in this study is displayed in Figure 11. Of particular note is the high (wet) bias in the lower Great Lakes region during 2010–2015, which is generally low (dry) during other eras. This anomaly is actually



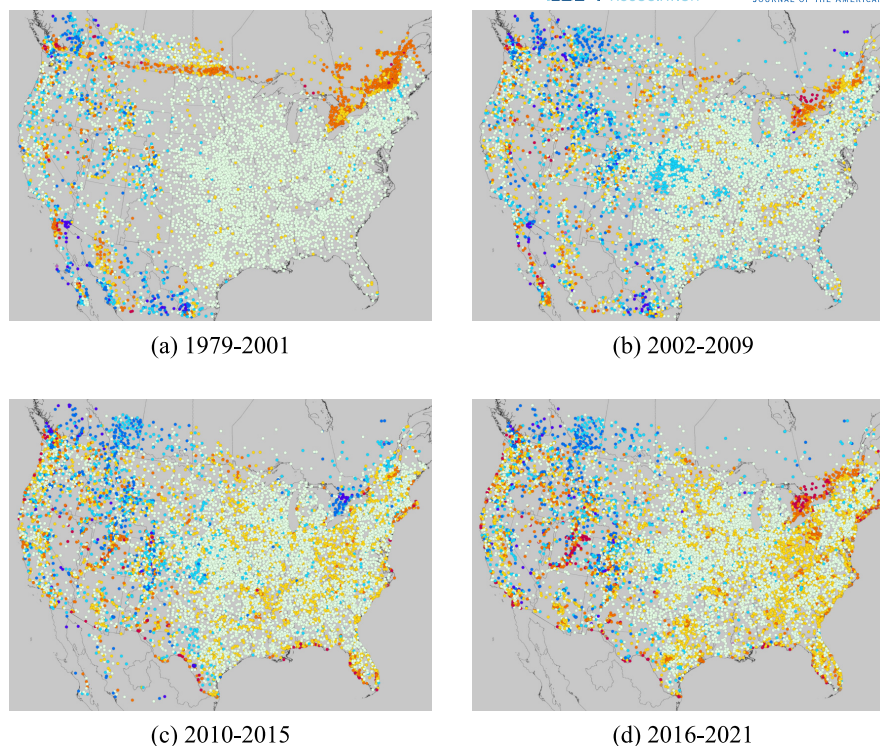


**FIGURE 9** Probability of Detection (POD) and False Alarm Ratio (FAR) time series for 24-h AORC-CONUS 1.1, NLDAS-2, and Stage IV precipitation compared with GHCN-D observations over the CONUS: (a) February 1979 through December 2001, (b) January 2002 through December 2009, (c) January 2010 through December 2015, and (d) January 2016 through December 2021. Vertical blue and red lines indicate January 1 and July 1, respectively. Individual captions describe the primary input datasets used during this period, including any constraints imposed on monthly precipitation totals. A legend for line colors is provided in [Figure 3](#).



**FIGURE 10** Geometric coefficient of variation (GCV) in multiplicative bias in AORC-CONUS 1.1 precipitation compared with GHCN-D observations from February 1979 through December 2021.

confined to the period from May 2012 through July 2013, but anomalous low bias also occurs in the region for brief periods in early 2012 and early 2013. As [Figures 5–8](#) demonstrate, these issues have been addressed in AORC-CONUS 1.1 through monthly gridded bias adjustments through 2015, and via prioritizing CaPA over NLDAS-2 precipitation over the lower Great Lakes beginning in May 2012, but [Figure 10](#) reveals that traces of the inaccuracies in NLDAS-2 along and outside U.S. boundaries remain, and appear as bias variability in the AORC precipitation



**FIGURE 11** UAMB of NLDAS-2 precipitation compared with GHCN-D observations from (a) February 1979 through January 2001, (b) January 2002 through December 2009, (c) January 2010 through December 2015, (d) January 2016 through December 2021. A color ramp legend is provided in [Figures 5–8](#).

data. Finally, regional variations in the UAMB of NLDAS-2 precipitation within the CONUS are not unlike regional anomalies reported by Mo et al. (2012).

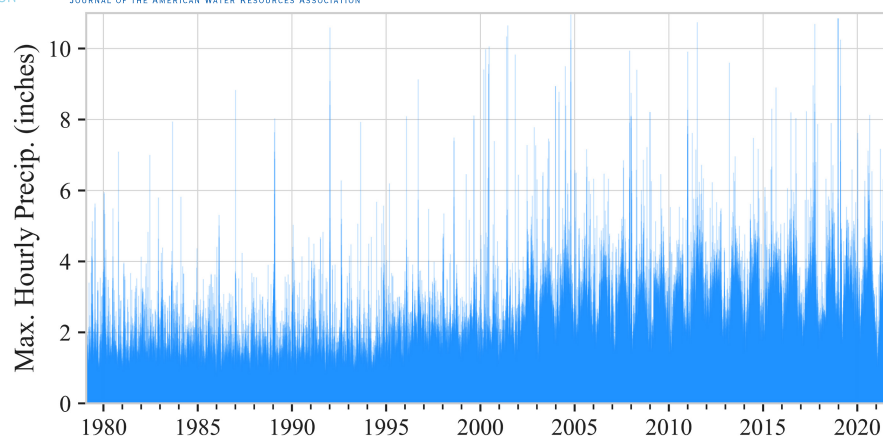
The release of climatological constraints in the AORC beginning in January 2016 exposes the analysis to bias and error from that point forward. [Figures 4d](#) and [8](#) illustrate the change in precipitation bias across the AORC-CONUS spatial domain beginning in 2016. In the “Precipitation Evaluation for Two RFC Regions” section of the [Appendix](#), we illustrate how this and other changes from one era to another are manifested on a more regional basis, focusing on two CONUS RFC regions.

### 3.4 | Unresolved heavy hourly precipitation amounts

During the development of AORC-CONUS 1.1, several high-priority fixes to the AORC component elements were implemented, including the removal of the very highest, and most fundamentally unrealistic, hourly precipitation values (see [Table 1](#)). Specifically, for all hours during which any AORC precipitation grid value exceeded 280 mm/h (11 in./h), the entire AORC-CONUS grid was replaced with NLDAS-2 precipitation. Note that currently, “world record exceedance” tests for hourly precipitation generally reference an observation of 12 inches of hourly rainfall in Holt, Missouri, on June 22, 1947 (Cerveny et al., 2007).

The removal of precipitation exceeding 11 in./h leaves open the possibility that unrealistic hourly precipitation amounts not reaching that threshold remain unresolved in AORC-CONUS 1.1. Indeed, as [Figure 12](#) illustrates, if one identifies the maximum value from each hourly AORC-CONUS 1.1 precipitation grid, it is clear that extreme precipitation amounts still occur frequently in the dataset. There appear to exist three distinct eras visible in the figure: the period prior to 1995, the period from 1995 to 2001, and the post-2002 period. It is worth noting, in terms of hourly data used to disaggregate daily precipitation totals, a few key milestones (see [Table 4](#); [Figure 2](#)) that are likely related to the changes distinguishing these eras:

- December 1994: the inclusion of Manually Digitized Radar (MDR; Miller & Kitzmiller, 2017) reflectivity data ends;
- January 1996: the inclusion of Weather Services International National Operational Weather Radar (WSI-NOWrad; Zhang et al., 2017) reflectivity grids begins;
- May 1996: the inclusion of NEXRAD Stage II (Lin & Mitchell, 2005) precipitation grids begins;
- January 2002: the inclusion of RFC-generated Stage IV precipitation estimates begins.



**FIGURE 12** Maximum hourly AORC-CONUS 1.1 precipitation from 1979 through 2021.

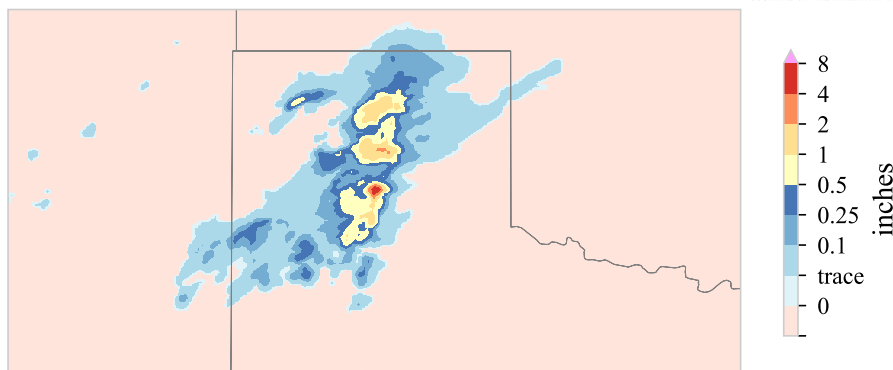
**TABLE 5** Inventory of occurrences (hours) of maximum AORC-CONUS 1.1 h precipitation in various ranges, for February 1979 through December 2021 (376,200h).

Range (inches)	Number of hours	Subtotal	Number (%) consistent with adjacent hours
>11.0	0	0	-
(10.0, 11.0]	13	13	3 (23%)
(9.0, 10.0]	14	27	2 (14%)
(8.0, 9.0]	28	55	9 (32%)
(7.0, 8.0]	34	89	8 (24%)
(6.0, 7.0]	66	155	11 (17%)
(5.0, 6.0]	189	344	50 (26%)
(4.0, 5.0]	851	1195	322 (38%)
(3.0, 4.0]	5038	6233	2303 (46%)
(2.0, 3.0]	32,402	38,635	18,409 (57%)
(1.0, 2.0]	96,306	134,941	59,367 (62%)
(0.0, 1.0]	241,259	376,200	Not inventoried

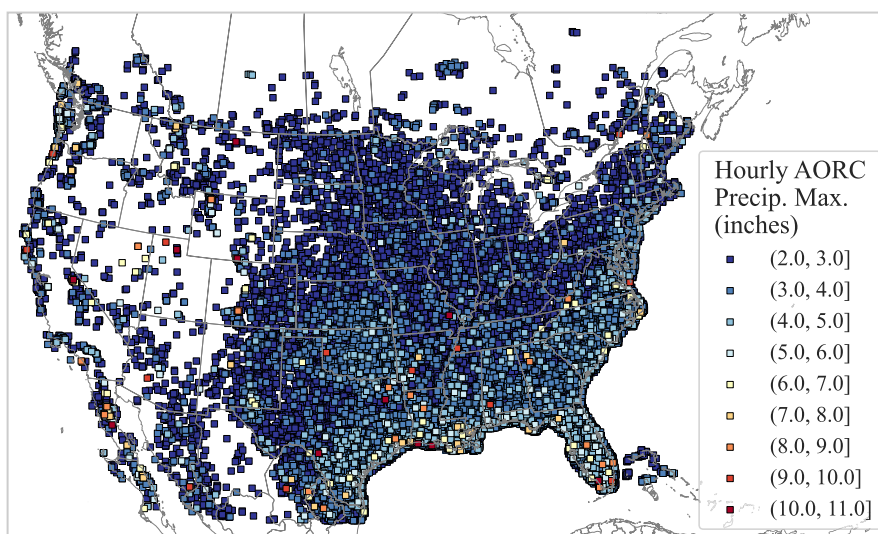
An inventory of maximum hourly precipitation in various ranges (using units of inches) for all 376,200 h (February 1979 through December 2021) of the AORC-CONUS 1.1 dataset considered for this study is provided in [Table 5](#). Precipitation rates exceeding 50 mm/h (about 2 in./h) are conventionally categorized as “violent” (Met Office, 2012); therefore, these results indicate 38,635 discrete 1-h occurrences of violent rain over the 43-year period (10% of the hours in the dataset, averaging to almost 2.5 occurrences per day), and 134,941 h (36% of the hours in the dataset) during which heavy-to-violent rain (>1 in./h) occurred in at least one location on the AORC grid.

Next, we examine the question of whether heavy-to-violent hourly precipitation amounts are sustained or isolated in time. Short durations are expected when heavy-to-violent precipitation occurs, but we are nonetheless interested to know to what extent the anomalies identified here are localized in time. For a preliminary answer, we examined grid locations within a 75 km radius of each maximum hourly gridded value during the two adjacent hours of AORC precipitation, and identified the number of cases in which the grid for either or both of the adjacent hours includes a maximum value in this regional neighborhood within 25% of the reference maximum. These results are listed in the fourth column of [Table 5](#). They indicate that the likelihood of a heavy-to-violent hourly maximum precipitation rate being sustained for more than 1 h in the same neighborhood is 62% for the (1.0, 2.0] in./h range, 57% for (2.0, 3.0] in./h, and decreases substantially as intensity increases beyond 3 in./h. This result suggests that a subset of the intense precipitation identified in this study is sustained for at least 2 h, but that most hourly maxima above 3–4 in./h are isolated in time.

[Figure 13](#) provides one example of intense precipitation, showing a maximum precipitation amount of 6.6 in./h in the Texas Panhandle on September 20, 2019, for the hour ending at 00 UTC. This result, and the precipitation in the surrounding area, is consistent with the Stage IV precipitation estimate for the same hour. The thunderstorms that produced this precipitation were reported by the NWS Weather Forecast Office (WFO) in Lubbock, TX (Lubbock WFO, 2019), to have generated “very heavy” rain, but the extreme amount seen here appears to be associated with



**FIGURE 13** Hourly AORC-CONUS 1.1 precipitation in Texas for the hour ending at 00 UTC on September 20, 2019. A maximum hourly precipitation amount of 6.6 inches appears in the center of the map.



**FIGURE 14** All 38,635 maximum hourly precipitation values exceeding 2 in./h in the AORC dataset. Points were drawn in ascending order of precipitation intensity; therefore, higher values often obscure lower values at similar locations.

a single ground report of 6.21 inches near Claude, Texas, for the hour ending at 00 UTC, and an associated report of 9.54 inches for the 24 h ending at 12 UTC on September 20, 2019, from the same observer. Both reports are outliers relative to other surface observations in the vicinity, where the highest 24-h accumulations reported were from 2 to 3 inches. They were, however, described by the WFO as valid reports.

Typically, when erroneous extreme precipitation intensities appear in the AORC-CONUS 1.1 dataset, they are the result of a minimal quality assurance process, which does not seek to identify spurious values present in input data sources or emerging from data processing steps. Figure 14 shows the locations and values of all hourly maxima from AORC-CONUS precipitation grids that exceed 2 in./h. This collection represents a subset of such events, since it includes at most one grid location (that of the maximum) for each hour, but given that the dataset we have examined covers 376,200h, it provides a large sample, illustrating the spatial distribution of intense precipitation events throughout the AORC dataset. Note that the 6.6 in./h value from September 20, 2019, described in Figure 13, is visible over the Texas Panhandle. Only 11 of the 13 maxima above 10 in./h are visible in the figure because three occur at exactly the same grid location in northern Utah over a 12-h period around December 24, 2018.

Most occurrences of extreme precipitation are seen in areas frequently affected by tropical cyclones and severe convective storms. The figure suggests that while some extreme events present in the AORC-CONUS 1.1 are depicted accurately, it is likely that a significant subset of them are the result of insufficient quality assurance in AORC data processing.

In an effort intended to simulate a prototypical quality assurance test, each of the hourly maxima exceeding 2 in./h summarized in Table 5 was compared, if possible, with a database of hourly precipitation observations from surface stations covering the years 2005–2021. To determine whether the maxima were supported by observations, all observations within a 75 km radius of the hourly maximum's grid location and attributed to times within 90 min of the same date and time were collected. Whenever at least five such neighboring observations were found, those reporting precipitation within 25% of the maximum AORC grid value were judged as corroborating reports.



**TABLE 6** Inventory of corroborating observations of extreme hourly AORC precipitation maxima in various ranges, for January 2005 through December 2021.

Range (inches)	# Hours	Subtotal	# with $\geq 5$ neighbors	Subtotal	# Corroborated by 1 neighbor	# Corroborated by $>1$ neighbor
>10.0	7	7	6	6	3	0
(9.0, 10.0]	5	12	4	10	0	0
(8.0, 9.0]	18	30	17	27	0	0
(7.0, 8.0]	19	49	18	45	0	0
(6.0, 7.0]	51	100	47	92	2	0
(5.0, 6.0]	141	241	132	224	3	0
(4.0, 5.0]	706	947	650	874	41	9
(3.0, 4.0]	4226	5173	3918	4792	342	82
(2.0, 3.0]	26,280	31,453	24,580	29,372	3050	477

The results of this effort are shown in Table 6. No hourly maximum above 5 in./h was corroborated by more than one neighboring observation, and only 8 out of 224 maxima considered in this range (the top six rows of Table 6) had even one corroborating neighbor. Overall, of the 29,372 cases in which at least five neighboring reports were found within 75 km of hourly maxima greater than 2 in./h, 4009 (14%) were corroborated by at least one observation, and 568 (2%) were corroborated by two or more observations.

While this finding suggests that extreme hourly precipitation in the AORC should be viewed with caution, intense precipitation is often extremely localized, and decorrelation distances for hourly precipitation observations typically range from 25 to 50 km (Kirsetter et al., 2010). Therefore, the above findings do not necessarily indicate that the majority of the intense precipitation events in the AORC dataset are erroneous; they do, however, suggest a lack of evidence of their validity. This preliminary study illustrates that effective quality assurance to eliminate spurious depictions of intense precipitation is particularly challenging, but also necessary, given the importance of extreme precipitation and the consequent need to avoid its erroneous inclusion or removal.

## 4 | SUMMARY AND CONCLUSION

The practice of hydrology depends on historical input forcing data for model parameter calibration and validation, and operational hydrology in the U.S. calls for a high-resolution, multi-decade, CONUS-scale record of hydrometeorological elements to fill this role. The AORC was developed for this purpose, and following the calibration of NWM 2.1 with AORC-CONUS 1.0, several patches were applied to address issues uncovered during that process. The result of those efforts, AORC-CONUS 1.1, is described in this study. It provides gridded hydrometeorological elements from February 1979 to the present, with near-real-time updates lagged by 9 days. Spatial resolution is 800 m, temporal resolution is 1 h, and coverage matches that of NLDAS-2 forcings, including the CONUS and portions of southern Canada and Mexico, with no missing data for any onshore grid location for any hour, including large inland water bodies.

AORC-CONUS 1.1 draws on numerous input datasets to generate its grids; its most significant data sources are NLDAS-2, LIV16, NEXRAD Stage IV, URMA, and the PRISM/NCEI/OWP climatology of temperature and precipitation developed specifically for the AORC. A description of how different data elements are processed for the AORC makes up a portion of this study, with particular attention given to temperature and precipitation inputs.

An evaluation of daily (12 UTC to 12 UTC) AORC-CONUS 1.1 precipitation shows that the AORC performs well relative to related datasets such as NEXRAD Stage IV and NLDAS-2, in general exhibiting less bias than either. The constraint imposed by the monthly LIV16 analysis from 1979 to 2015 appears particularly effective in limiting precipitation bias, and with the lifting of that constraint at the end of 2015, significant regional and seasonal biases are introduced into the product. Practitioners who wish to use the AORC for hydrologic model calibration should be attentive to this change, since it does represent a disruption in the temporal consistency of the bias behavior of AORC precipitation, such consistency an important scientific requirement for hydrologic calibration datasets. Ongoing bias adjustment, even at the potential expense of timeliness, should be a priority for future versions or alternatives to the AORC.

Finally, we identified a large number of extreme precipitation events occurring in the AORC throughout the 43 years of hourly data examined, even though world-record-exceeding hourly precipitation was eliminated from AORC-CONUS 1.1. A search for observations corroborating these events did not find them in most cases, and this finding calls for effective quality assurance measures in the generation of any future analysis. We are nevertheless mindful that actual extreme weather events described by input data streams must not be flagged or eliminated by overly aggressive quality control.



## AUTHOR CONTRIBUTIONS

**Greg Fall:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; validation; visualization; writing – original draft. **David Kitzmiller:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; validation; writing – review and editing. **Sandra Pavlovic:** Formal analysis; investigation; methodology; project administration; validation. **Ziya Zhang:** Formal analysis; investigation; methodology; software; validation; visualization; writing – review and editing. **Nathan Patrick:** Data curation; methodology; software; validation. **Michael St. Laurent:** Investigation; methodology; software; validation. **Carl Trypaluk:** Investigation; methodology; software; validation. **Wanru Wu:** Investigation; software; validation; visualization; writing – review and editing. **Dennis Miller:** Investigation; validation; writing – review and editing.

## ACKNOWLEDGMENTS

PRISM datasets covering the CONUS are available from the PRISM Climate Group, Oregon State University, at <https://prism.oregonstate.edu> (1981–2010 datasets are available from the PRISM Climate Group upon request); last accessed March 2023. PRISM data covering western Canada for 1981–2010 are available from the Pacific Climate Impacts Consortium (PCIC), University of Victoria, at <https://www.pacificclimate.org/data/prism-climatology-and-monthly-timeseries>; last accessed March 2023. WorldClim 1.4 historical climate data are available at <https://www.worldclim.org/data/v1.4/worldclim14.html>; last accessed March 2023. The Uniatmos 1902–2011 climatology is available from Universidad Nacional Autónoma de México (UNAM), Instituto de Ciencias de la Atmósfera y Cambio Climático, at <https://uniatmos.atmosfera.unam.mx/atlas/>; last accessed February 2023. The September 2016 update to the Livneh meteorological dataset (LIV16), including daily minimum and maximum temperatures and monthly precipitation, is available at <ftp://livnehpublicstorage.colorado.edu/public/Livneh.2016.Dataset>; last accessed July 2022. Archived CPC Global Unified Gauge-Based Analysis of Daily Precipitation data are available from the NOAA Physical Sciences Laboratory, Boulder, Colorado, USA, at <https://psl.noaa.gov>; last accessed March 2023. Climate Forecast System Reanalysis (CFSR) data produced by the National Centers for Environmental Prediction (NCEP) are available at <https://www.ncei.noaa.gov/products/weather-climate-models/climate-forecast-system>; last accessed October 2016.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

## DATA AVAILABILITY STATEMENT

Detailed information on obtaining most of the data discussed in this study is provided in the *Acknowledgements* and, when appropriate, via citations in the *Literature Cited*. Some of the datasets used to develop AORC-CONUS 1.1 are not to our knowledge publicly distributed (though several may remain in production as non-archived, near-real-time products). These include the following:

- CaPA daily precipitation
- URMA hourly hydrometeorological elements
- GDAS 6-h hydrometeorological elements
- MDR hourly precipitation
- WSI-NOWRad hourly precipitation
- NEXRAD Stage II hourly precipitation

The portions of the above datasets used to develop AORC-CONUS 1.1, and the AORC-CONUS 1.1 data itself, may be obtained upon reasonable request by contacting the corresponding author.

## ORCID

Greg Fall  <https://orcid.org/0000-0002-5059-4703>

## REFERENCES

- Anderson, E.A. 2002. *Calibration of Conceptual Hydrologic Models for Use in River Forecasting*. Silver Spring, MD: Office of Hydrologic Development, U.S. National Weather Service. [https://www.weather.gov/media/owp/oh/hrl/docs/1\\_Anderson\\_CalbManual.pdf](https://www.weather.gov/media/owp/oh/hrl/docs/1_Anderson_CalbManual.pdf).
- Boulmaiz, T., M. Guermoui, and H. Boutaghe. 2020. "Impact of Training Data Size on the LSTM Performances for Rainfall-Runoff-Modeling." *Modeling Earth Systems and Environment* 6: 2153–64. <https://doi.org/10.1007/s40808-020-00830-w>.
- Cerveny, R.S., J. Lawrimore, R. Edwards, and C. Landsea. 2007. "Extreme Weather Records." *The Bulletin of the American Meteorological Society* 88(6): 853–60. <https://doi.org/10.1175/BAMS-88-6-853>.
- Cosgrove, B.A., D. Lohmann, K.E. Mitchell, P.R. Houser, E.F. Wood, John C. Schaake, A. Robock, et al. 2003. "Real-Time and Retrospective Forcing in the North American Land Data Assimilation System (NLDAS) Project." *Journal of Geophysical Research-Atmospheres* 108(D22): 8842. <https://doi.org/10.1029/2002JD003118>.

- Daly, C., M. Halbleib, J.I. Smith, W.P. Gibson, M.K. Doggett, George H. Taylor, J. Curtis, and P.A. Pasteris. 2008. "Physiographically-Sensitive Mapping of Temperature and Precipitation across the Conterminous United States." *International Journal of Climatology* 28: 2031–64. <https://doi.org/10.1002/joc.1688>.
- Du, J. 2011. "NCEP/EMC 4 Km Gridded Binary (GRIB) Stage IV Data." Version 1.0. UCAR/NCAR Earth Observing Laboratory (EOL). <https://doi.org/10.5065/D6PG1QDD>.
- Durre, I., A. Arguez, C.J. Schreck, III, M.F. Squires, and R.S. Vose. 2022. "Daily High-Resolution Temperature and Precipitation Fields for the Contiguous United States from 1951 to Present." *The Journal of Atmospheric and Oceanic Technology* 39(12): 1837–55. <https://doi.org/10.1175/JTECH-D-22-0024.1>.
- Durre, I., M.J. Menne, B.E. Gleason, T.G. Houston, and R.S. Vose. 2010. "Comprehensive Automated Quality Assurance of Daily Surface Observations." *The Journal of Applied Meteorology and Climatology* 49: 1615–33. <https://doi.org/10.1175/2010JAMC2375.1>.
- Ensor, L.A., and S.M. Robeson. 2008. "Statistical Characteristics of Daily Precipitation: Comparisons of Gridded and Point Datasets." *The Journal of Applied Meteorology and Climatology* 47(9): 2468–76. <https://doi.org/10.1175/2008JAMC1757.1>.
- Farr, T.G., P.A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, et al. 2007. "The Shuttle Radar Topography Mission." *Reviews of Geophysics* 45: RG2004. <https://doi.org/10.1029/2005RG000183>.
- Fernández Eguarte, A., R. Romero Centeno, and J. Zavala Hidalgo. 2012. *Atlas Climático de México y Áreas Adyacentes*, Vol 1. Mexico: Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México. <http://atlasclimatico.unam.mx/ACM/>.
- Fortin, V., G. Roy, T. Stadnyk, K. Koenig, N. Gasset, and A. Mahidjiba. 2018. "Ten Years of Science Based on the Canadian Precipitation Analysis: A CaPA System Overview and Literature Review." *Atmosphere-Ocean* 56(3): 178–96. <https://doi.org/10.1080/07055900.2018.1474728>.
- Gronewold, A.D., V. Fortin, R. Caldwell, and J. Noel. 2018. "Resolving Hydrometeorological Data Discontinuities along an International Border." *The Bulletin of the American Meteorological Society* 99(5): 899–910. <https://doi.org/10.1175/BAMS-D-16-0060.1>.
- Gupta, H.V., K.J. Beven, and T. Wagener. 2005. "Model Calibration and Uncertainty Estimation." In *Encyclopedia of Hydrological Sciences*, edited by M.G. Anderson, 2015–31. West Sussex, England: John Wiley & Sons. <https://doi.org/10.1002/0470848944.hsa138>.
- Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." *The International Journal of Climatology* 25: 1965–78. <https://doi.org/10.1002/joc.1276>.
- Hutchinson, M., and T. Xu. 2013. *ANUSPLIN Version 4.4 User Guide*. Canberra: The Australian National University, Fenner School of Environment and Society. <http://fennerschool.anu.edu.au/files/anusplin44.pdf>.
- Jones, P.W. 1999. "First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates." *The Monthly Weather Review* 127(9): 2204–10. [https://doi.org/10.1175/1520-0493\(1999\)127<2204:FASOCR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2).
- Kavetski, D., G. Kuczera, and S.W. Franks. 2006. "Bayesian Analysis of Input Uncertainty in Hydrological Modeling: 1. Theory." *Water Resources Research* 42(3): 4368. <https://doi.org/10.1029/2005WR004368>.
- Khakbaz, B., B. Imam, K. Hsu, and S. Sorooshian. 2012. "From Lumped to Distributed Via Semi-Distributed: Calibration Strategies for Semi-Distributed Hydrologic Models." *Journal of Hydrology* 418–419: 61–77. <https://doi.org/10.1016/j.jhydrol.2009.02.021>.
- Kirseter, P.-E., G. Delrieu, B. Boudevillian, and C. Obled. 2010. "Toward an Error Model for Radar Quantitative Precipitation Estimation in the Cévennes-Vivarais Region, France." *Journal of Hydrology* 394(1–2): 28–41. <https://doi.org/10.1016/j.jhydrol.2010.01.009>.
- Kleist, D.T., D.F. Parrish, J.C. Derber, R. Treadon, W.-S. Wu, and S. Lord. 2009. "Introduction of the GSI into the NCEP Global Data Assimilation System." *Weather and Forecasting* 24(6): 1691–705. <https://doi.org/10.1175/2009WAF2222201.1>.
- Koopmans, L.H., D.B. Owen, and J.I. Rosenblatt. 1964. "Confidence Intervals for the Coefficient of Variation for the Normal and Log Normal Distributions." *Biometrika* 51(1–2): 25–32. <https://doi.org/10.1093/biomet/51.1-2.25>.
- Lahmers, T.M., P. Hazenberg, H. Gupta, C. Castro, D. Gochis, A. Dugger, D. Yates, L. Read, L. Karsten, and Y.-H. Wang. 2021. "Evaluation of NOAA National Water Model Parameter Calibration in Semiarid Environments Prone to Channel Infiltration." *The Journal of Hydrometeorology* 22: 2939–69. <https://doi.org/10.1175/JHM-D-20-0198.1>.
- Lin, Y., and K.E. Mitchell. 2005. "The NCEP Stage II/IV Hourly Precipitation Analyses: Development and Applications." Preprints, 19th Conference on Hydrology, American Meteorological Society, San Diego, CA, January 9–13, 2005. Paper 1.2. <https://www.emc.ncep.noaa.gov/mmb/SREF/pcpan/l/refs/stage2-4.19hydro.pdf>.
- Livneh, B., T. Bohn, D. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, Daniel R. Cayan, and L. Brekke. 2015. "A Spatially Comprehensive, Hydrometeorological Data Set for Mexico, the U.S., and Southern Canada 1950–2013." *Scientific Data* 2: 150042. <https://doi.org/10.1038/sdata.2015.42>.
- Lubbock WFO. 2019. *Mid-September 2019 Drops Rain on the High Plains*. Lubbock, TX: National Weather Service Forecast Office. <https://www.weather.gov/lub/events-2019-20190921-rain>.
- McCabe, G.J., and D.M. Wolock. 2011. "Independent Effects of Temperature and Precipitation on Modeled Runoff in the Conterminous United States." *Water Resources Research* 47(11): 10630. <https://doi.org/10.1029/2011WR010630>.
- McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods. 2011. "Rainfall Uncertainty in Hydrological Modelling: An Evaluation of Multiplicative Error Models." *Journal of Hydrology* 400(1–2): 83–94. <https://doi.org/10.1016/j.jhydrol.2011.01.026>.
- Menne, M.J., I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, et al. 2012. "Global Historical Climatology Network - Daily (GHCN-Daily)." Version 3.28. NOAA National Climatic Data Center. <https://doi.org/10.7289/V5D21VHZ>.
- 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." *The Journal of Atmospheric and Oceanic Technology* 29: 897–910. <https://doi.org/10.1175/JTECH-D-11-00103.1>.
- Met Office. 2012. *Fact Sheet No. 3 – Water in the Atmosphere*. UK: National Meteorological Library and Archive. Met Office. [https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/library-and-archive/library/publications/factsheets/factsheet\\_3-water-in-the-atmosphere-v02.pdf](https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/library-and-archive/library/publications/factsheets/factsheet_3-water-in-the-atmosphere-v02.pdf).
- Miller, D., and D. Kitzmiller. 2017. "Application of Pre-NEXRAD Reflectivity Data to Hourly Precipitation Analyses." *Journal of Hydrologic Engineering* 22: 5. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001310](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001310).
- Mizukami, N., and M.B. Smith. 2012. "Analysis of Inconsistencies in Multi-Year Gridded Quantitative Precipitation Estimate over Complex Terrain and its Impact on Hydrologic Modeling." *Journal of Hydrology* 428–429: 129–41. <https://doi.org/10.1016/j.jhydrol.2012.01.030>.
- Mo, K.C., L.-C. Chen, S. Shukla, T.J. Bohn, and D.P. Lettenmaier. 2012. "Uncertainties in North American Land Data Assimilation Systems over the Contiguous United States." *The Journal of Hydrometeorology* 13(3): 996–1009. <https://doi.org/10.1175/JHM-D-11-0132.1>.



- NASA (National Aeronautics and Space Administration). 2013. *Shuttle Radar Topography Mission (SRTM) Global*, Version 3.0. Distributed by Open Topography. <https://doi.org/10.5069/G9445JDF>.
- Newman, A.J., M.P. Clark, K. Sampson, A. Wood, L.E. Hay, A. Bock, R.J. Viger, et al. 2015. "Development of a Large-Sample Watershed-Scale Hydrometeorological Data Set for the Contiguous USA: Data Set Characteristics and Assessment of Regional Variability in Hydrologic Model Performance." *Hydrology and Earth System Sciences* 19: 209–23. <https://doi.org/10.5194/hess-19-209-2015>.
- Oudin, L., C. Perrin, T. Mathevet, V. Andréassian, and C. Michel. 2006. "Impact of Biased and Randomly Corrupted Inputs on the Efficiency and the Parameters of Watershed Models." *Journal of Hydrology* 320(1–2): 62–83. <https://doi.org/10.1016/j.jhydrol.2005.07.016>.
- Pierce, D.W., L. Su, D.R. Cayan, M.D. Risser, B. Livneh, and D.P. Lettenmaier. 2021. "An Extreme-Preserving Long-Term Gridded Daily Precipitation Dataset for the Conterminous United States." *The Journal of Hydrometeorology* 22(7): 1883–95. <https://doi.org/10.1175/JHM-D-20-0212.1>.
- Pondeca, M., S. Levine, J. Carley, Y. Lin, Y. Zhu, J. Purser, J. McQueen, et al. 2015. "Ongoing Improvements to the Real Time Mesoscale Analysis (RTMA) and Unrestricted Mesoscale Analysis (URMA) and NCEP/EMC." World Climate Research Organization. [http://www.wcrp-climate.org/WGNE/BlueBook/2015/individual-articles/01\\_Pondeca\\_Manuel\\_et\\_al\\_RTMA.pdf](http://www.wcrp-climate.org/WGNE/BlueBook/2015/individual-articles/01_Pondeca_Manuel_et_al_RTMA.pdf).
- Reed, S., and D.R. Maidment. 1999. "Coordinate Transformations for Using NEXRAD Data in GIS-Based Hydrologic Modeling." *Journal of Hydrologic Engineering* 4(2): 174–82. [https://doi.org/10.1061/\(ASCE\)1084-0699\(1999\)4:2\(174\)](https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(174)).
- Refsgaard, J.C. 1997. "Parameterisation, Calibration and Validation of Distributed Hydrologic Models." *Journal of Hydrology* 198(1–4): 69–97. [https://doi.org/10.1016/S0022-1694\(96\)03329-X](https://doi.org/10.1016/S0022-1694(96)03329-X).
- 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 Resources Research* 46(5): 8328. <https://doi.org/10.1029/2009WR008328>.
- Saha, S., S. Moorthi, H. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, et al. 2010. "The NCEP Climate Forecast System Reanalysis." *The Bulletin of the American Meteorological Society* 97: 1015–58. <https://doi.org/10.1175/2010BAMS3001.1>.
- Shen, H., B.A. Tolson, and J. Mai. 2022. "Time to Update the Split-Sample Approach in Hydrological Model Calibration." *Water Resources Research* 58(3): 31523. <https://doi.org/10.1029/2021WR031523>.
- Smith, M.B., D.P. Laurine, V.I. Koren, S.M. Reed, and Z. Zhang. 2003. "Hydrologic Model Calibration in the National Weather Service." In *Calibration of Watershed Models*, edited by Q. Duan, H.V. Gupta, S. Sorooshian, A.N. Rousseau, and R. Turcotte. Washington, DC: American Geophysical Union. <https://doi.org/10.1029/WS006>.
- Tian, Y., G.J. Huffman, R.F. Adler, L. Tang, M. Sapiiano, V. Maggioni, and H. Wu. 2013. "Modeling Errors in Daily Precipitation Measurements: Additive or Multiplicative?" *Geophysical Research Letters* 40: 1–6. <https://doi.org/10.1002/grl.50320>.
- van Leth, T.C., H. Leijnse, A. Overeem, and R. Uijlenhoet. 2021. "Rainfall Spatiotemporal Correlation and Intermittency Structure from Micro- $\gamma$  to Meso- $\beta$  Scale in The Netherlands." *The Journal of Hydrometeorology* 22(8): 2227–40. <https://doi.org/10.1175/JHM-D-20-0311.1>.
- Vose, R.S., S. Applequist, M. Squires, I. Durre, M.J. Menne, C.N. Williams, Jr., C. Fenimore, K. Gleason, and D. Arndt. 2014a. "Improved Historical Temperature and Precipitation Time Series for US Climate Divisions." *The Journal of Applied Meteorology and Climatology* 53(5): 1232–51. <https://doi.org/10.1175/JAMC-D-13-0248.1>.
- Vose, R.S., S. Applequist, M. Squires, I. Durre, M.J. Menne, Claude N. Williams, Jr., Chris Fenimore, Karin Gleason, and D. Arndt. 2014b. "NOAA Monthly U.S. Climate Gridded Data set (NclimGrid-Monthly)." Version 1. NOAA National Centers for Environmental Information. <https://doi.org/10.7289/V5SX6B56>.
- Wang, J., L. Li, J.J. Gourley, S.I. Khan, K.K. Yilmaz, R.F. Adler, F.S. Policelli, et al. 2011. "The Coupled Routing and Excess Storage (CREST) Distributed Hydrological Model." *Hydrological Sciences Journal* 56(1): 84–98. <https://doi.org/10.1080/02626667.2010.543087>.
- Wang, J., L. Zhuo, D. Han, Y. Liu, and M.A. Rico-Ramirez. 2023. "Hydrological Model Adaptability to Rainfall Inputs of Varied Quality." *Water Resources Research* 59(5): 32484. <https://doi.org/10.1029/2022WR032484>.
- Whittaker, G., R. Confesor, Jr., M. Di Luzio, and J.G. Arnold. 2010. "Detection of Overparameterization and Overfitting in an Automatic Calibration of SWAT." *Transactions of the ASABE* 53(5): 1487–99. <https://doi.org/10.13031/2013.34909>.
- Xia, Y., M. Ek, H. Wei, and J. Meng. 2012. "Comparative Analysis of Relationships between NLDAS-2 Forcings and Model Outputs." *Hydrological Processes* 26: 467–74. <https://doi.org/10.1002/hyp.8240>.
- Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, et al. 2009. "NLDAS Primary Forcing Data L4 Hourly 0.125  $\times$  0.125 Degree V002." In *NASA/GSFC/HSL*, edited by David Mocko. Greenbelt, MD: Goddard Earth Sciences Data and Information Services Center (GES DISC). <https://doi.org/10.5067/6J5LHHOZHNN4>.
- Xie, P., R. Joyce, S. Wu, S. Yoo, Y. Yarosh, F. Sun, and R. Lin. 2017. "Reprocessed, Bias-Corrected CMORPH Global High-Resolution Precipitation Estimates from 1998." *The Journal of Hydrometeorology* 18: 1617–41. <https://doi.org/10.1175/JHM-D-16-0168.1>.
- Xie, P., S. Joyce, S.-H. Wu, Y. Yoo, F. Sun Yarosh, and R. Lin. 2019. "NOAA Climate Data Record (CDR) of CPC Morphing Technique (CMORPH) High Resolution Global Precipitation Estimates." Version 1. NOAA National Centers for Environmental Information (NCEI). <https://doi.org/10.25921/w9va-q159>.
- Xie, P., A. Yatagai, M. Chen, T. Hayasaka, Y. Fukushima, C. Liu, and S. Yang. 2007. "A Gauge-Based Analysis of Daily Precipitation over East Asia." *The Journal of Hydrometeorology* 8: 607–26. <https://doi.org/10.1175/JHM583.1>.
- Zhang, J., K. Howard, C. Langston, B. Kaney, Y. Qi, L. Tang, H. Grams, et al. 2016. "Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities." *The Bulletin of the American Meteorological Society* 97: 621–38. <https://doi.org/10.1175/BAMS-D-14-00174.1>.
- Zhang, Y., D. Kitzmiller, D.-J. Seo, D. Kim, and R. Cifelli. 2017. "Creation of Multisensor Precipitation Products from WSI NOWrad Reflectivity Data." *Journal of Hydrologic Engineering* 22: E4015001. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001216](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001216).

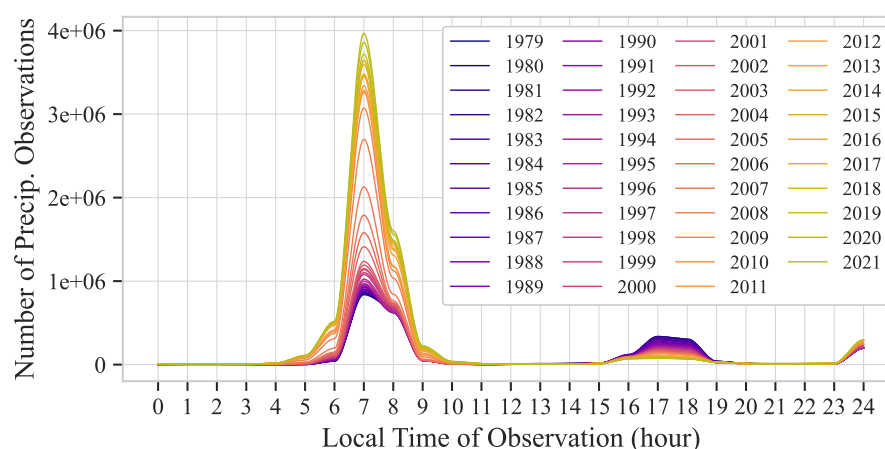
**How to cite this article:** Fall, Greg, David Kitzmiller, Sandra Pavlovic, Ziya Zhang, Nathan Patrick, Michael St. Laurent, Carl Trypaluk, Wanru Wu and Dennis Miller. 2023. "The Office of Water Prediction's Analysis of Record For Calibration, Version 1.1: Dataset Description and Precipitation Evaluation." *JAWRA Journal of the American Water Resources Association* 59 (6): 1246–1272. <https://doi.org/10.1111/1752-1688.13143>.

## APPENDIX

## OBSERVATION TIMES OF GHCN-DAILY PRECIPITATION REPORTS

The evaluation of gridded precipitation presented in this study compares daily (24 h) precipitation amounts ending at 12 UTC, summed from hourly grids, with corresponding observations. The GHCN-Daily database that provides the observations for these comparisons does not constrain its data to 12 UTC, and in fact some data providers end their daily 24-h period at local midnight (Menne, Durre, Vose, et al., 2012), which for the AORC-CONUS domain ranges from 04 UTC to 08 UTC. Other data providers instruct observers to report at 07:00 local time, which ranges from 11 UTC to 15 UTC for the AORC-CONUS domain.

Despite some ambiguity in the timing of GHCN-D precipitation reports, most—specifically, 69% of those in US/CA/MX locations from 1979 to 2021—are accompanied by the local time of day (HH:MM) of observation (LTO), that is, the ending time of the measured 24-h accumulation. For calendar year 2021, the LTO is included with 91% of US/CA/MX GHCN-D precipitation observations. For earlier years, this proportion is smaller (e.g., 56% in 1979), but it increases steeply from 2003 to 2013 with the addition of the Community Collaborative Rain, Hail, and Snow (CoCoRaHS) network to the GHCN-D database. When available LTOs are binned over 24 h for each calendar year, the resulting histograms, shown in Figure A1 for 1979–2021, are distinctly bimodal, with a primary peak from 05:00 to 09:00, a secondary peak from 16:00 to 18:00, and very small observation counts outside those periods. From 1979 to 2021, the primary peak grows larger and the secondary grows smaller. It is therefore reasonable to assume that by excluding all observations reporting an LTO of 12:00 or later while also including observations reporting no LTO, a large majority of the observations used in this study will have been collected within 2–3 h of 12 UTC. We expect that only a small subset of reports not reporting an LTO occurred far from 12 UTC, and these are not expected to significantly degrade the results reported here. Finally, the proper interpretation of an LTO value of 24:00 is ambiguous and currently unknown to us, but since such observations are unused in this evaluation, the question is not relevant to this study. Additional discussion and analysis of GHCN-D observation times is provided by Pierce et al. (2021) and Durre et al. (2022).



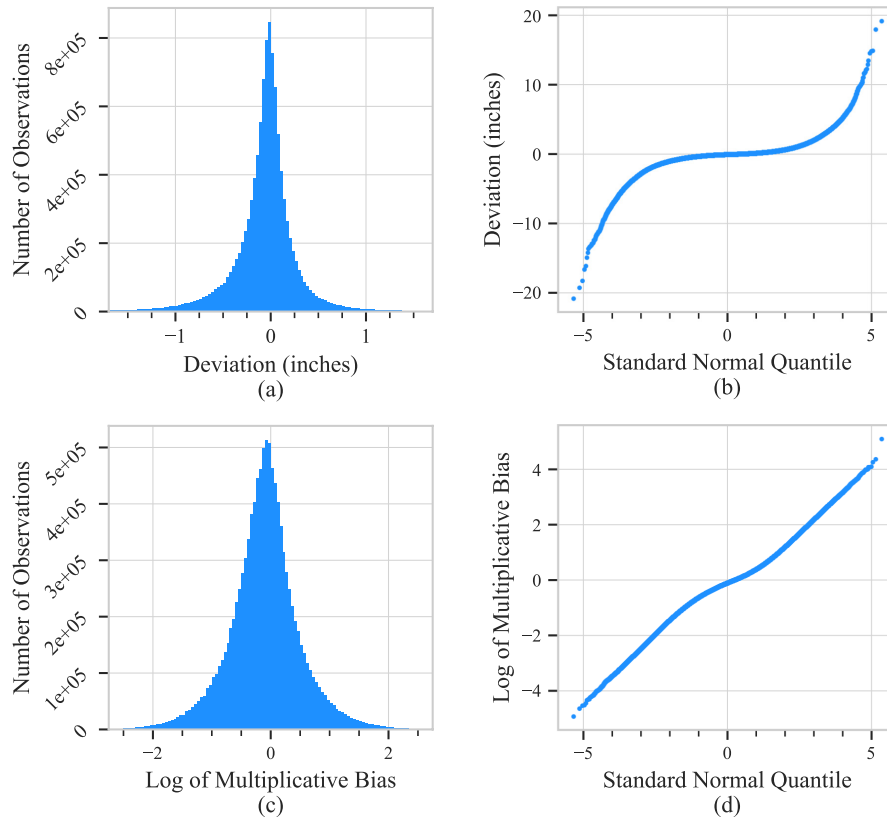
**FIGURE A1** Histograms of local time of observation (LTO) for precipitation observations from the Global Historical Climatology Network-Daily (GHCN-D) database, version 3.28 (Menne, Durre, Korzeniewski, et al., 2012; Menne, Durre, Vose, et al., 2012), over the U.S., Canada, and Mexico, for each year from 1979 to 2021. Observations attributed to a given hour may have occurred at any time during the subsequent 60 min, but as most LTO values are provided with minutes recorded as “00,” observation counts in this figure are attributed to the start of the hour, and plotted as such, with the “minutes” portion of the LTO disregarded. A cubic Hermite spline has been applied to improve readability.

## ERROR STATISTICS OF AORC-CONUS 1.1 PRECIPITATION

The evaluation of AORC-CONUS 1.1 precipitation and its constituent counterparts presented in this study preferentially employs statistics based on multiplicative bias (MB, the ratio of analysis precipitation to observed precipitation), rather than deviation (the difference between analysis precipitation and observed precipitation), to characterize error and bias in the dataset. Here, the statistical behavior of these two quantities is compared, with analysis precipitation represented by daily (24-h accumulation ending at 12 UTC) AORC-CONUS 1.1 precipitation, nearest-neighbor sampled at GHCN-D station locations, and observed precipitation represented by the corresponding 24-h GHCN-D precipitation observations at those stations.

Figure A2 illustrates both the non-normal nature of deviation and the tendency of MB toward log-normality. Deviation (Figure A2a) has only a slight negative (left) skew, but its high positive excess kurtosis differs significantly from that of a sample from a normal distribution. This characteristic is further illustrated in the Q–Q plot of deviation (Figure A2b). The natural log of MB, or log-bias (Figure A2c), shows little skewness and a small degree of excess positive kurtosis. The latter characteristic is visible in the slight horizontal flattening of the log-bias Q–Q plot





**FIGURE A2** Daily AORC-CONUS 1.1 precipitation compared with GHCN-D observations from February 1979 through December 2021. For this figure, deviation (the difference between AORC and observed daily precipitation) and multiplicative bias (MB; the ratio of AORC to observed daily precipitation) are calculated only for *hits*, that is, cases in which both estimated and observed precipitation accumulations exceed 0.1 inch. The resulting sample size is 10,850,233. (a) Histogram of AORC deviation, with skewness =  $-1.90$  and excess kurtosis =  $36.21$ . (b) Q-Q plot of deviation. (c) Histogram of the natural log of MB (log-bias), with skewness =  $-0.04$  and excess kurtosis =  $1.91$ . (d) Q-Q plot of log-bias.

**TABLE A1** UAMB time series geometric mean (GM) and geometric coefficient of variation (GCV) for different AORC eras and domains, for Stage IV daily QPE and AORC.

Domain	CONUS				MBRFC				NERFC			
	Stage IV		AORC		Stage IV		AORC		Stage IV		AORC	
Product	GM	GCV (%)	GM	GCV (%)	GM	GCV (%)	GM	GCV (%)	GM	GCV (%)	GM	GCV (%)
UAMB												
1979–2001	-	-	0.98	1.8	-	-	0.98	5.0	-	-	0.97	3.9
2002–2009	0.91	5.9	0.98	2.4	0.79	33.9	0.93	16.3	0.93	8.0	0.99	4.9
2010–2015	0.97	4.8	1.00	3.7	1.16	10.4	1.03	8.5	0.91	5.8	1.00	4.3
2016–2021	1.00	3.1	1.00	2.9	1.17	7.5	1.16	6.9	0.92	4.5	0.91	4.6

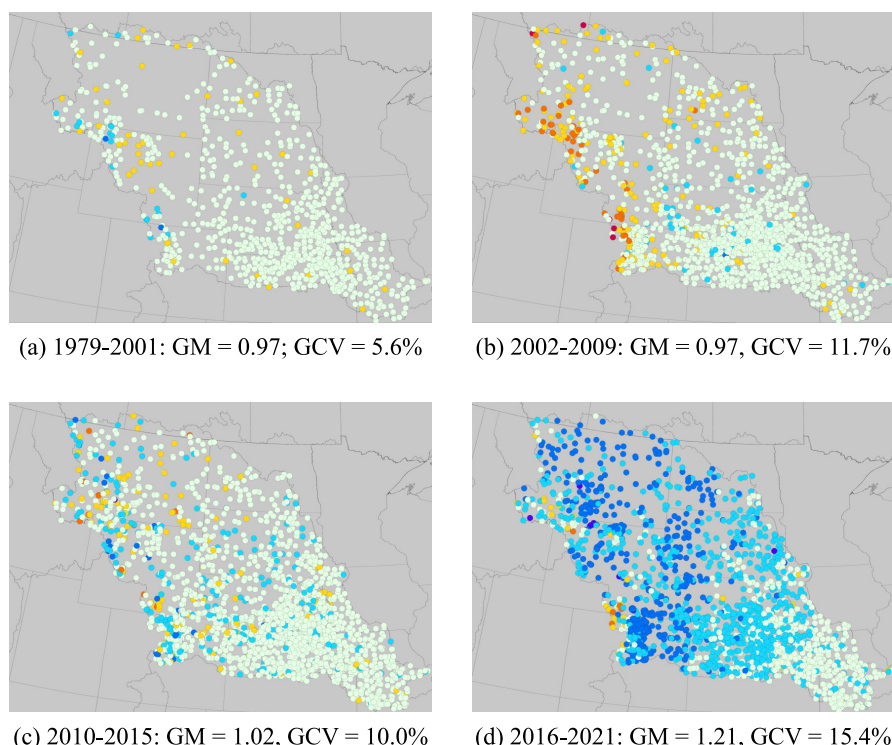
Note: GM values less than 0.95 are shown in red to indicate low (dry) average bias, and values greater than 1.05 are shown in blue to indicate high (wet) average bias.

near zero (Figure A2d), but the otherwise strong and nearly linear correlation between the log-bias and corresponding values of the standard normal quantile function indicates that the MB values closely resemble a sample from a log-normal distribution.

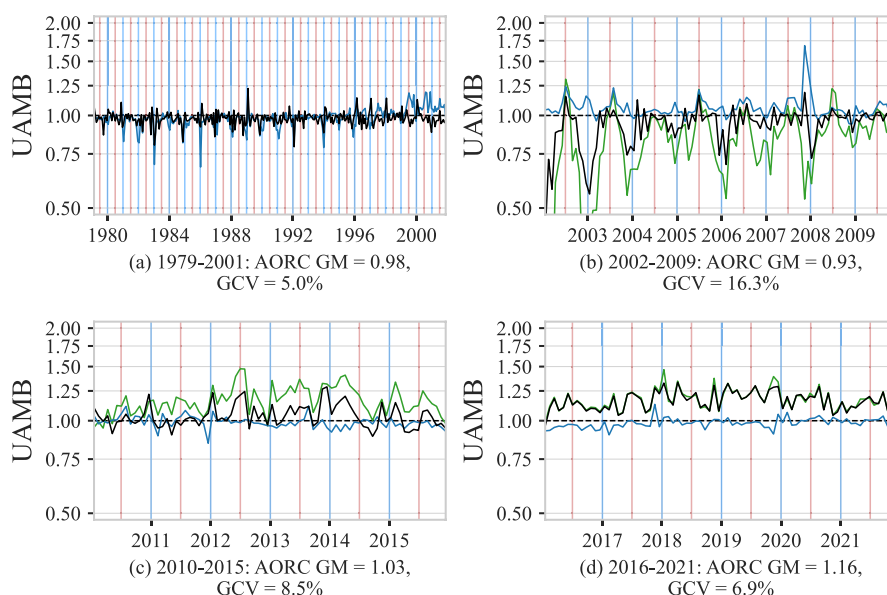
## PRECIPITATION EVALUATION FOR TWO RFC REGIONS

The AORC-CONUS 1.1 precipitation is not subject to any bias correction or climatological constraint after 2015; consequently, beginning in January 2016, the dataset becomes more vulnerable to bias in its constituent datasets. Here, we discuss the impact of this vulnerability in two different CONUS RFC regions: the Missouri Basin RFC (MBRFC) and the Northeast RFC (NERFC). Table A1 shows UAMB time-series geometric means (GM) and geometric coefficients of variation (GCV) for the CONUS, MBRFC, and NERFC domains for each generation of

AORC precipitation methodology, with the historical period (1979–2009) subdivided into eras before and after the introduction of Stage IV daily QPE, as elsewhere in this study. The table illustrates the resistance of AORC to bias in the Stage IV over all three regions prior to 2016, and this resistance breaking down in the RFC region results from 2016 to 2021.



**FIGURE A3** Unrestricted accumulation bias (UAMB) at individual observing stations of 24-h AORC-CONUS 1.1 precipitation, compared with GHCN-D observations over the Missouri Basin RFC (MBRFC) region: (a) February 1979 through December 2001, (b) January 2002 through December 2009, (c) January 2010 through December 2015, and (d) January 2016 through December 2021. The geometric mean (GM) and coefficient of variation (GCV) for the displayed UAMB values are provided in the caption below each subfigure. A color ramp legend is provided in [Figures 5–8](#).



**FIGURE A4** Unrestricted accumulation bias (UAMB) time series for 24-h AORC, NLDAS-2, and Stage IV precipitation compared with GHCN-D observations over the Missouri Basin RFC (MBRFC) region: (a) February 1979 through December 2001, (b) January 2002 through December 2009, (c) January 2010 through December 2015, and (d) January 2016 through December 2021. Vertical blue and red lines indicate January 1 and July 1, respectively. The GM and GCV for the AORC UAMB time series are provided in the caption below each subfigure. A legend for line colors is provided in [Figure 3](#).

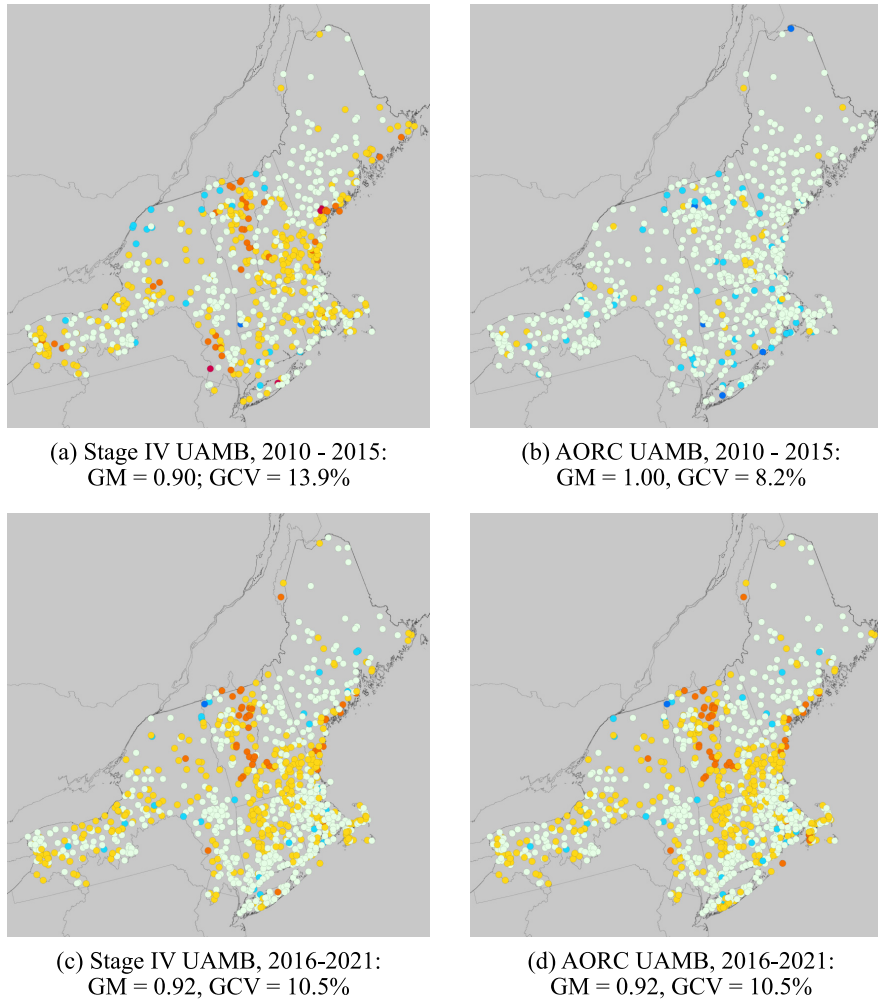


### Missouri Basin RFC region

Figure A3 shows UAMB in AORC precipitation at station locations, as was done for the CONUS in Figures 5–8, but limited in this case to the MBRFC region. Figure A4 provides the corresponding UAMB time series. From 1979 to 2001, AORC precipitation is consistently unbiased across the region (Figures A3a and A4a). With the introduction of Stage IV daily QPE in 2002, dry biases arise in the mountainous Western portions of the domain (Figure A3b), and an annual variation in UAMB with winter minima and summer maxima appears and remains until about 2008 (Figure A4b). The climatological constraint imposed by the LIV16 data reduces the amplitude of this variation for AORC versus Stage IV, but the geometric time-averaged bias across the domain for the AORC is still somewhat low (0.93) from 2002 to 2009 (Table A1; Figure A4b)—though the GM bias for the points in Figure A3b is slightly higher at 0.97. A wet bias in Stage IV QPE (GM=1.16 in Table A1) over the MBRFC region emerges during 2010 (Figure A4c) and increases and persists thereafter (Figure A4d). This bias arises from a merging of different datasets in the Multisensor Precipitation Estimation (MPE) process at MBRFC, which serves the needs of operational hydrology for that region but impacts the accuracy of the Stage IV mosaic. From 2010 to 2015, this bias is, as has been noted previously, constrained by the process of maintaining climatological similarity between LIV16 and AORC for those years, but a slight wet bias does appear in the AORC over portions of the region during that period (Figure A3c), with a GM of 1.03 for the time series in Figure A4c, 1.02 for the points Figure A3c. Beginning in 2016, there is virtually no difference between AORC and Stage IV (Figure A4d), resulting in a distinct wet bias in AORC across the region, particularly its western portion (Figure A3d).

### Northeast RFC region

Numerous occurrences of low (dry) bias in AORC precipitation are evident in the Northeastern U.S. for the 2016–2021 period in Figure 8. In Figure A5, UAMB values at individual observing stations for 2010–2015 (Figure A5a,b) and 2016–2021 (Figure A5c,d) are shown for Stage IV



**FIGURE A5** Unrestricted accumulation bias (UAMB) at individual observing stations of 24-h Stage IV and AORC-CONUS 1.1 precipitation, compared with GHCN-D observations over the Northeast RFC (NERFC) region: (a) Stage IV, 2010–2015; (b) AORC, 2010–2015; (c) Stage IV, 2016–2021; and (d) AORC, 2016–2021. The GM and GCV for the displayed UAMB values are provided in the caption below each subfigure. A color ramp legend is provided in Figures 5–8.



QPE (Figure A5a,c) and the AORC (Figure A5b,d). As Table A1 indicates, and Figure A5a,c illustrates, the UAMB for Stage IV, calculated across the region, is low, with a GM value below 0.95, from 2002 onward. After 2015, as has been shown previously for the CONUS and MBRFC regions, the lack of climatological control over bias permits the bias in daily Stage IV QPE to impact the AORC from 2016 onward (Figure A5d). The GM value of UAMB for 2010–2015 (Figure A5b) is 1.00, but it decreases to 0.92 for 2016–2021 (Figure A5d).