A NASA-Air Force Precipitation Analysis for Near-Real-Time Operations

Eric M. Kemp,[a,b] Jerry W. Wegiel,[a,c] Sujay V. Kumar,[a] James V. Geiger,[a,d]
David M. Mocko,[a,c] Jossy P. Jacob,[a,b] and Christa D. Peters-Lidard[e]

[a] *Hydrological Sciences Laboratory, NASA GSFC, Greenbelt, Maryland*

[b] *Science Systems and Applications, Inc, Lanham, Maryland*

[c] *Science Applications International Corporation, Reston, Virginia*

[d] *Science Data Processing Branch, NASA GSFC, Greenbelt, Maryland*

[e] *Sciences and Exploration Directorate, NASA GSFC, Greenbelt, Maryland*

Corresponding author: Eric M. Kemp, eric.kemp@nasa.gov

Revised submission to *Journal of Hydrometeorology*

ABSTRACT

This article describes a new precipitation analysis algorithm developed by NASA for time-sensitive operations at the United States Air Force. Implemented as part of the Land Information System -- a land modeling and data assimilation software framework -- this NASA-Air Force Precipitation Analysis (NAFPA) combines numerical weather prediction model outputs with rain gauge measurements and satellite estimates to produce global, gridded 3-hr accumulated precipitation fields at approximately 10-km resolution. Input observations are subjected to quality control checks before being used by the Bratseth analysis algorithm that converges to Optimal Interpolation. NAFPA assimilates up to 3.5 million observations without artificial data thinning or selection. To evaluate this new approach, a multi-year reanalysis is generated and intercompared with eight alternative precipitation products across the contiguous United States, Africa, and the monsoon region of eastern Asia. NAFPA yields superior accuracy and correlation over low-latency (up to 14-hour) alternatives (numerical weather prediction and satellite retrievals), and often outperforms high-latency (up to 3.5-month) products, although the details for the latter vary by region and product. The development of NAFPA offers a high-quality, near-real-time product for use in meteorological, land surface, and hydrological research and applications.

SIGNIFICANCE STATEMENT

Precipitation is a key input to land modeling systems due to effects on soil moisture and other parts of the hydrologic cycle. It is also of interest to government decision makers due to impacts on human activities. Here we present a new precipitation analysis based on available near-real-time data. By running the program for prior years and comparing with alternative products, we demonstrate that our analysis provides better accuracy and usually less bias than near-real-time satellite data alone, and better accuracy and correlation than data provided by numerical weather models. Our analysis is also competitive with other products created months after the fact, justifying confidence in using our analysis in near-real-time operations.

1. Introduction

As part of its responsibilities for providing timely meteorological and land surface information across the globe, the United States Air Force (USAF) operates the NASA Land Information System (LIS; Kumar et al. 2006; Peters-Lidard et al. 2007; Wegiel et al. 2020), a land modeling and data assimilation (DA) software framework. Included in this analysis system is the specification of 3-hourly precipitation, based on data available in near-real-time

(NRT) to the USAF 557th Weather Wing (557 WW). Precipitation is a key input to land surface models, with strong impacts on near-surface soil moisture (Gottschalck et al. 2005; Guo et al. 2006; Wei et al. 2008; Wang and Zeng 2011), runoff (Fekete et al. 2004; Wang and Zeng 2011; Qi et al. 2020), and snowpack (Pan et al. 2003; Wang et al. 2020). In addition, the analyzed precipitation can be of direct use to USAF customers in the military, intelligence, and foreign agriculture sectors for a variety of applications (crop yield, transportation disruptions, hydroelectric power generation, etc.). It is therefore important to periodically update the analysis approach to keep abreast of technological developments in observing systems, satellite retrievals, and DA algorithms. Operational security requirements lead to keeping analysis generation "in house," minimizing dependencies that could become unavailable if external communications are disrupted. A further challenge is to provide a NRT product with comparable quality to reanalysis and high-latency (up to 3.5-month) alternatives. This paper documents recent and successful changes to that effect.

When LIS first became operational at the USAF in 2009, 3-hourly precipitation forcing was provided by the legacy USAF Agricultural Meteorology (AGRMET) software (Moore et al. 1991; Gayno and Wegiel 2000; Air Force Weather Agency 2002), which was converted into a coupled component of LIS (Eylander et al. 2005). While AGRMET benefits from using rain gauges, several different satellite retrievals, numerical weather prediction (NWP), and climatology, the overall algorithm has several drawbacks. For example, the preliminary value of each analysis grid box is directly inserted from a single "best" datum in that box, based on the assumed quality of available data sources and their distances to the center of the box. As a result, most observations in data-rich regions are rejected as redundant, and the data selection logic is convoluted. Further, for those analysis boxes without rain gauges, values are further adjusted by interpolating from nearby gauge-associated boxes using a modified Barnes (1964) scheme. This generates a "ballooning effect" of precipitation extrema (Tian et al. 2009).

When designing a new algorithm, multiple alternative NRT precipitation products were examined. The 0.1° NASA Integrated Multisatellite Retrievals for Global Precipitation Measurement (IMERG) products (Tan et al. 2019; Huffman et al. 2020a) provide semi-global 30-min analyses from time-interpolated passive microwave (PMW) and infrared (IR) satellite estimates, with two NRT feeds (Early Run and Late Run) available with latencies of 4- to 14-hrs. The older 8-km NOAA CPC Morphing Technique (CMORPH; Joyce et al. 2004; Xie et al. 2017) also produces semi-global, 30-min satellite-based analyses with a latency of 2-hrs,

and bias-corrected analyses once a day. The NCEP Multi-Radar Multi-Sensor system (Zhang et al. 2016) combines radar, gauge, NWP, and climatology from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994, 2008) to produce hourly, 0.01° analyses over the contiguous United States (CONUS), with a latency of around 90 minutes. The 10-km regional Canadian Precipitation Analysis (CaPA; Lespinas et al. 2021; Fortin et al. 2018) uses a form of Optimal Interpolation (OI; Daley 1993) to blend rain gauges, IMERG (beginning in mid-2021), and radar estimates with NWP, with 6-hr accumulations analyzed four times a day. The 0.5° NOAA CPC Unified Gauge-Based Analysis of Global Daily Precipitation (CPCU) uses modified OI developed by Xie et al. (2007) and Chen et al. (2008) to blend 24-hr gauge reports over global land areas once per day. The 0.05° Climate Hazards Group InfraRed Precipitation with Stations Version 2 (CHIRPSv2; Funk et al. 2015) estimates daily overland precipitation from gauge reports and calibrated geostationary IR data, with NOAA Climate Forecast System (CFS; Saha et al. 2014) data used for temporal disaggregation. The CHIRPSv2 Preliminary product is generated two days after each pentad,[1] while the Final product is generated the third week of the following month. None of these alternatives combine NWP, gauge, and satellite data on a global scale at 3-hr intervals, and many have longer latencies than acceptable for USAF operations[2].

NASA adopted the philosophy of approximate optimal blending of data, based on estimated error covariances of all the input. The new algorithm, the NASA-Air Force

---

[1] There are six pentads per month: Five 5-day time periods, plus one pentad for the remaining 3- to 6-days of the month (Funk et al. 2015).

[2]While this manuscript was under revision, we learned of a NRT version of the Multi-Source Weighted-Ensemble Precipitation product (MSWEP; http://www.gloh2o.org/mswep/). Based on earlier reanalysis versions of MSWEP (Beck et al. 2017, 2019a, b), this combines multiple gauge, NWP, satellite, reanalysis, and climatology data together into 3-hr accumulations, with a reported latency of ~3 hours.

Precipitation Analysis (NAFPA), also takes advantage of parallelism using distributed-memory computing clusters to assimilate several million observations across the globe using over one thousand processors. All observations that are statistically "close enough" to affect an analysis point (based on error covariances) are allowed to do so, provided they first pass quality control (QC) checks. Adding and removing data sources is more straightforward, and NAFPA handles varying observation density much more smoothly.

This manuscript describes the development and technical details of NAFPA. Daily summed NAFPA values are evaluated over the period of 2012-2019 (2012-2015 in Asia) against alternative precipitation products, using a suite of quantitative metrics. The superior quality of NAFPA over low-latency alternatives, and often against high-latency reanalysis data, is demonstrated over different regions of the world. In addition, the evaluation also identifies areas for future improvement. We believe the development of this product presents a significant contribution to the NWP and operational land and hydrological communities.

Section 2 describes the current data sources for NAFPA, QC tests, DA algorithm, error covariance models, and choice of analysis variable. Section 3 describes the multi-year reanalysis produced with NAFPA, as well as the alternative precipitation products used for intercomparisons. Section 4 presents those intercomparisons. Section 5 presents overall conclusions and future work.

2. Algorithm Description

NAFPA is heavily inspired by CaPA (Fortin et al. 2018). Like CaPA, NAFPA blends a short-term NWP field with rain gauge observations based on the estimated error characteristics of the input data. However, there are several major differences from the CaPA approach. First, NAFPA interpolates data across the entire globe with a resolution of around 10-km (equidistant cylindrical grid, with cells dimensioned 0.140625° longitude by 0.093750° latitude). Second, satellite rainfall estimates from up to four different sources can be assimilated in NAFPA. Third, radar rainfall estimates are excluded due to limited global coverage. Fourth, the OI local approximation (artificially limiting the number of observations affecting a particular analysis point) is avoided by using an iterative algorithm that avoids direct matrix inversion (Bratseth 1986). Last, NAFPA produces 3-hr accumulations instead of 6-hr totals.

a. Data sources

The background first-guess for NAFPA usually comes from the USAF Global Air-Land Weather Exploitation Model (GALWEM; Stoffler 2017), an implementation of the UK Met Office Unified Model (Brown et al. 2012). Currently, GALWEM is run four times a day by the 557 WW on a global, equidistant cylindrical grid at approximately 17-km resolution (grid cells dimensioned 0.234375° longitude by 0.156250° latitude). 3-hr accumulated precipitation forecasts are used, preferably from 9-hr and 12-hr GALWEM forecasts (to allow precipitation to "spin-up" in the model) with older/longer-range forecasts substituted if necessary. If GALWEM forecasts are not available, NAFPA will fall back to using 0.5° NOAA Global Forecast System (GFS) precipitation. In practice, GFS is typically only used for producing re-analyses prior to mid-2017, after which GALWEM data become available. Since the USAF GFS archive only includes 3-hr and 6-hr forecast files, NAFPA will attempt to use GFS 3-hr and 6-hr forecasts before searching for older/longer-range forecasts for the background.

557 WW receives rain gauge reports from several sources, including the WMO Global Telecommunications System (GTS). These reports are primarily from fixed locations and include Binary Universal Form of Representation of Meteorological Data (BUFR) reports, aviation routine weather reports (METARs), surface synoptic observations (SYNOPs), and aviation special weather reports (SPECIs). Some sea station weather reports (SHIPs) and surface observations from mobile land stations (SYNOP MOBL) reports are also received. Depending on the source, the reports contain 6-hr, 24-hr, and/or "miscellaneous" accumulations. Fig. 1 shows the locations of NAFPA grid points containing archived rain gauge reports, and their reporting frequencies, for the period 2012-2019. Note the frequent, high-resolution reports in parts of Europe, parts of the Middle East, Japan, and southern Russia; the more sporadic but dense coverage in CONUS and eastern Europe; and the sporadic, coarse coverage in Africa, South America, and western China. These reporting frequencies are affected by coding and decoding errors, data transmission speed and reliability, and deadlines for fetching reports for the NAFPA analysis (typically below two hours after the last analysis valid time). We also notice some spread in latitude/longitude coordinates for reports in Canada, which are due to differences in the received reports.

After fetching available reports, NAFPA performs intercomparisons to establish a more-consistent time series of accumulations. Special logic is invoked to handle some unique regional reporting standards or patterns:

- Indian stations only report accumulations valid from 0300 UTC;

- Sri Lankan stations only issue reports with a duration of 3 hrs;
- Many reports from former Soviet Union nations do not indicate duration, and are assumed to have a duration of 12 hrs;
- Many South American stations do not report at 0600 UTC, and issue 1200 UTC "zero precipitation" reports without indicating duration; these are assumed to mean "no precipitation in last 12 hrs."

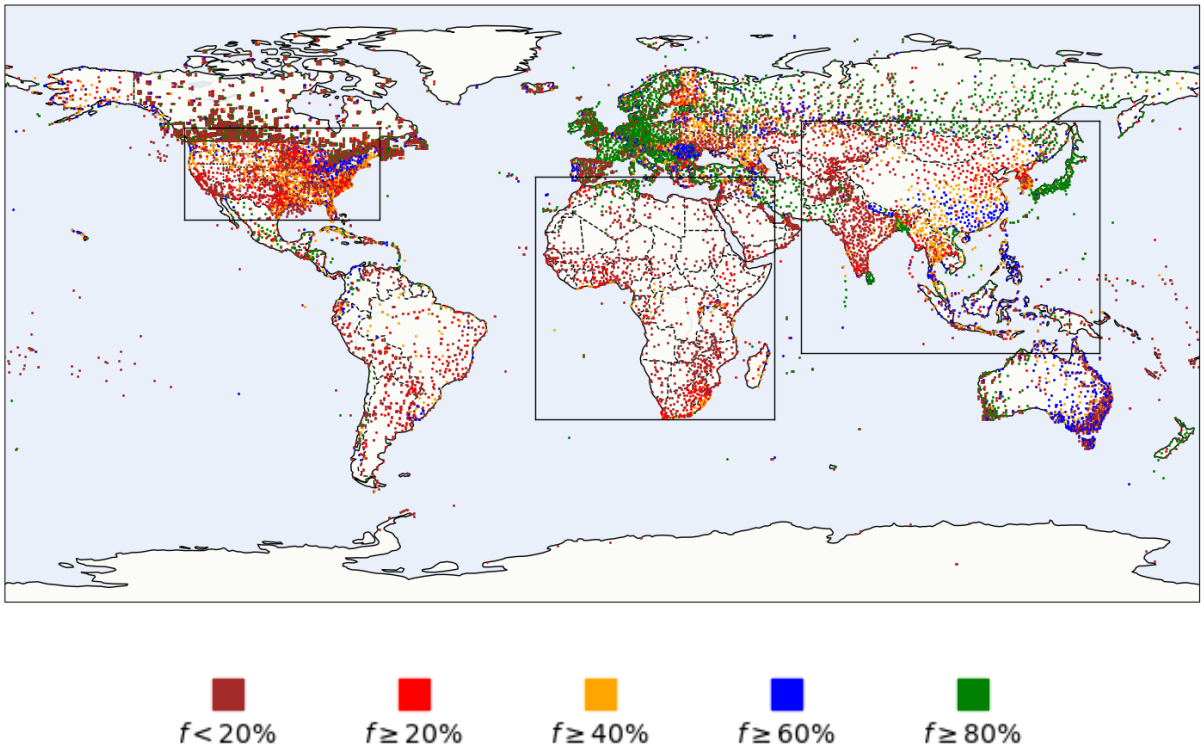


Fig. 1. Locations and frequencies *f* of analysis grid boxes containing at least one rain gauge report per day for the period of 2012-2019. Left box indicates CONUS domain, excluding water points and non-U.S. land points. Center box indicates Africa domain, excluding water points. Right box indicates Monsoon Asia domain, excluding water points.

As NAFPA runs, the revised 12-hr gauge reports are read in. (6-hr reports are substituted if the NAFPA run is six simulation hours away from completion.) Once read in, the gauge reports are disaggregated into 3-hr totals by comparing to interpolated background values. If the interpolated background has no precipitation for the entire gauge reporting period, the gauge values are split equally across the 3-hr time periods. Otherwise, the gauge totals are split according to the fractions of total background precipitation that occur in each 3-hr period. This procedure to disaggregate gauge totals using other temporally finer data sources, while preserving the gauge total amounts, is similar to that used in the NOAA North American Land Data Assimilation System (NLDAS; Cosgrove et al. 2003).

NAFPA supports Version 6 of IMERG (Tan et al. 2019; Huffman et al. 2020a). Although all three IMERG data feeds are supported in LIS (the NRT Early Run and Late Run, and the research-grade Final Run), presently the Early Run (IMERG-ER) is targeted in NAFPA, due to its low latency (~4 hrs). 30-min calibrated rain rates are read from each IMERG file and converted to accumulations, if the IMERG algorithm diagnosed 100% liquid precipitation. [See Huffman et al. (2020a), pp 18-19 for details on the precipitation phase algorithm.] Sequential 30-min accumulations are then summed to 3-hr totals. If any 30-min data are missing or rejected at a given point, the corresponding 3-hr total is set to missing.

Three additional legacy sources of satellite estimates are supported. First, 3-hr PMW estimates are available from the Special Sensor Microwave Imager/Sounder (SSMIS) instruments (Kunkee et al. 2008) aboard recent Defense Meteorological Satellite Program spacecraft. Second, 3-hr IR-based estimates (Vicente et al. 1998) from geostationary weather satellites (GEOPRECIP) may be used. Last, bias-corrected estimates from a USAF variant of CMORPH (Xie et al. 2017) are available. All three products are presently generated by the 557 WW and used in USAF operations as of mid-March 2022, but these will be replaced in the near-future by the NASA IMERG-ER product. This is partially due to the USAF desire to retire the older products, and partially due to limitations of each product [SSMIS has limited spatial coverage and the instruments have likely limited remaining lifetimes; IR-based satellite retrievals are generally inferior to PMW (Joyce et al. 2004); and bias-corrected CMORPH is only available once a day].

b. Quality control

A series of QC checks are performed on the input observations, borrowing on the algorithms presented by Mahfouf et al. (2007), Lopez (2013), and Lespinas et al. (2015), as well as leveraging logic implemented in AGRMET. At present, the checks are administered sequentially, with observations failing a particular check excluded from further consideration and use by LIS. A schematic of the QC workflow is given in Fig. 2. Below we describe the tests in some detail.

Since NAFPA is intended for land-based applications, most observations over water are considered less important. Furthermore, the large number of satellite estimates can be a serious challenge to store in memory. Accordingly, NAFPA implements a simple *WaterQC* check where observation locations are compared to the land/water mask used by LIS. All observations that are located over a water grid box are flagged and rejected.

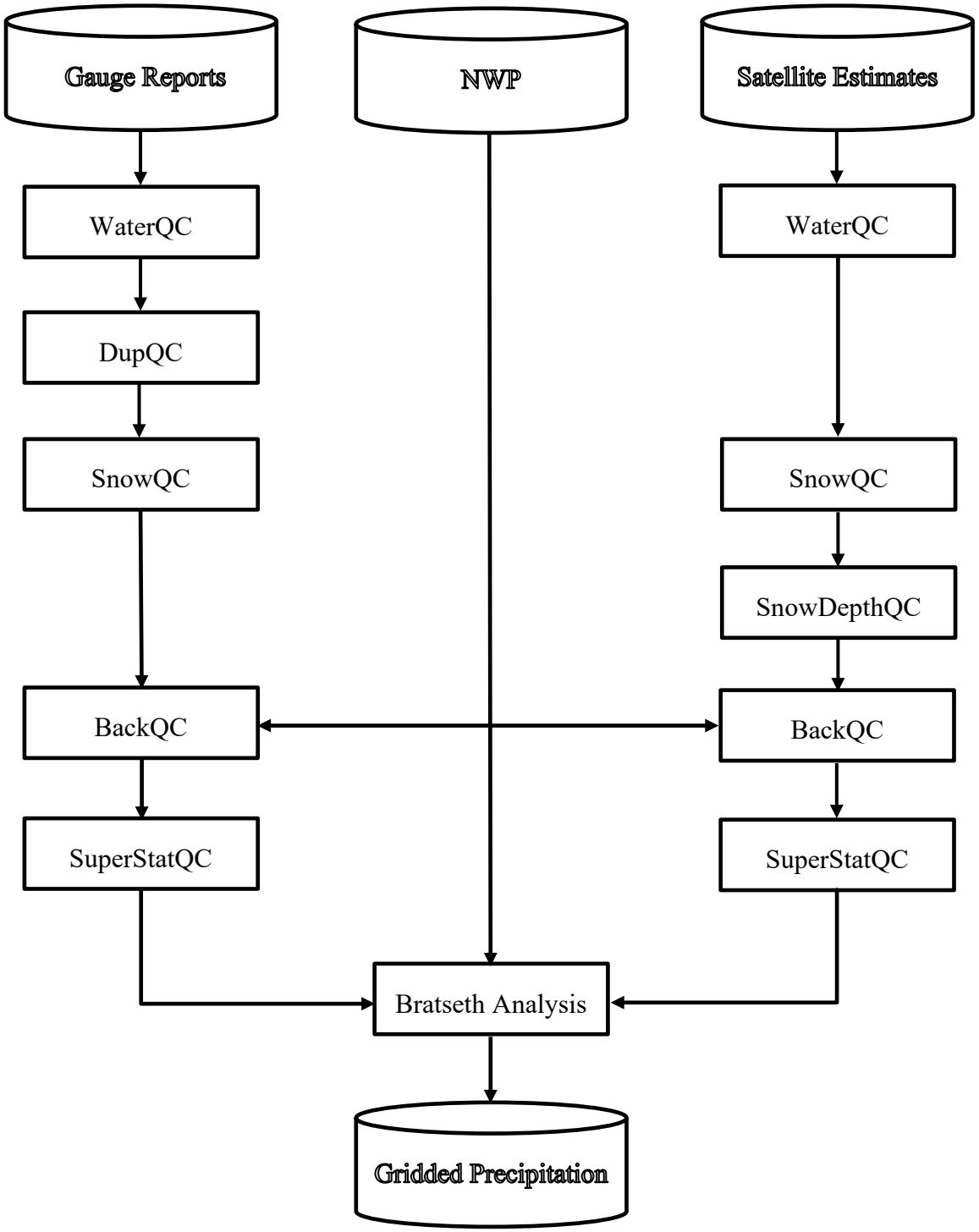


Fig. 2. Workflow chart showing application of QC checks on input data before execution of Bratseth analysis. See section 2b for descriptions of each test.

The possibility exists that duplicate rain gauge reports will be received. Accordingly, a *DupQC* check is performed on the 3-hrly disaggregated gauge totals. This consists of several sub-checks:

- If reports from the same station and time are received with multiple latitude/longitude coordinates (due to precision differences between data feeds), the first observation is

arbitrarily saved and the rest rejected. [This was implemented as a "hot-fix" in operations.]

- If only two reports exist for the same location and time but have different values, the reports are rejected if the squared difference in values exceeds the observation error variance (see section 2d below). Otherwise, the two values are averaged together to form a single "super-observation." This follows Mahfouf et al. (2007).
- If multiple reports from the same station and time are received but the values are identical, all but the first report are rejected.
- Otherwise, all the reports are rejected unless only a single report is received.[3]

Measuring active snowfall is a challenge for gauges and satellite retrievals. It is well known that gauges suffer from undercatch bias, but the details of the bias are affected by gauge type and shielding (Goodison et al. 1998; Rasmussen et al. 2012). In the satellite realm, the SSMIS, GEOPRECIP, and CMORPH products are rainfall-only. IMERG does provide a snowfall estimate based on the Global Precipitation Measurement (GPM) *Core Observatory*; but the retrievals show significant differences from *CloudSat* spaceborne radar measurements (Stephens et al. 2002), and there are difficulties in calibrating against surface observations and surface radars (Skofronick-Jackson et al. 2018). For simplicity, NAFPA attempts to filter out observed snowfall. The *SnowQC* test will reject a precipitation observation (gauge or satellite) if the near-surface temperature is less than 2°C (following Lopez 2013). IR- and PMW-based satellite retrievals can also have significant errors if data are taken over a snowpack—for example, dry snow, glacial snow, and precipitation all act as

[3] We acknowledge some flaws in the current *DupQC* approach: arbitrarily picking the first report when the latitude/longitude coordinates vary does not ensure the report is correct; and rejecting reports with more than two unique precipitation values may throw away useful information. In the future we plan to simplify *DupQC*, only rejecting exact duplicate reports and relying on *SuperstatQC* (described at the end of section 2b) to form "super-observations" after rejecting outliers.

microwave scatterers, making it difficult to isolate precipitation scattering effects (Grody 1991). To address this, the *SnowDepthQC* test will reject a satellite estimate if the local snow depth modelled by LIS exceeds 0 kg m^{-2}.

As with many other DA systems, NAFPA will attempt to detect gross observation errors by comparing observations to the interpolated background field. The *BackQC* test follows Lopez (2013), and an observation is rejected if $|x_b - y_o| > 4(\sigma_o^2 + \sigma_b^2)^{1/2}$. Here x_b is the background value, y_o is the observed value, and σ_o^2 and σ_b^2 are the observation and background error variances, respectively.

Lastly, a *SuperstatQC* test is invoked to compare observations of the same type located in the same LIS grid box. Here the arithmetic mean $\overline{y}_o$ of the observations is calculated, and each individual observation y_o is compared to it. An observation will be rejected if $\left|\overline{y}_o - y_o\right| < 3\sigma_o^2[m/(m-1)]^{1/2}$, where *m* is the number of compared observations. Those observations that pass the test are used to calculate a new "super-observation." This follows Lespinas et al. (2015).

c. The Bratseth scheme

The analysis algorithm is related to OI (Gandin 1965). Define a domain with *n* analysis points and containing *p* observations. Assume the input data are unbiased. The OI analysis equation can then be written as:

$$\mathbf{x}_a = \mathbf{x}_b + \widetilde{\mathbf{B}}(\mathbf{B} + \mathbf{R})^{-1}(\mathbf{y}_o - \mathbf{y}_b) \quad (1)$$

Here $\mathbf{x}$ is a *n*-length vector of values at analysis points; $\mathbf{y}$ is a *p*-length vector of values at the observation points; and subscripts *a*, *b*, and *o* indicate analysis, background, and observed values, respectively. (Below we refer to $\mathbf{y}_o - \mathbf{y}_b$ as the *observation increments*.) $\widetilde{\mathbf{B}}$ is a $n \times p$ background error covariance matrix with covariances between observation and analysis points. $\mathbf{B}$ is a $p \times p$ background error covariance matrix, and $\mathbf{R}$ is a $p \times p$ observation error covariance matrix; both are symmetric, positive-definite, and contain error covariances between observation points.

The computational cost for directly solving (1) can be quite high due to matrix inversion. In practice, OI implementations often reduce the problem size by calculating $(\mathbf{B} + \mathbf{R})^{-1}$ for individual analysis points and using an artificially small number of observations for each case (e.g., Fortin et al. 2015; Soci et al. 2016). Alternatively, (1) is sometimes solved for overlapping subset regions which are then averaged together (e.g., Lorenc 1981; Goerss and

Phoebus 1992). Either way, truly global analyses using all observations are rarely constructed.

Bratseth (1986) proposed an alternative which avoids direct matrix inversion, which we describe here following Kalnay (2003). If matrix $\mathbf{A}$ is symmetric and positive-definite, and has eigenvalues μ in the range $0 \leq \mu \leq 2$, then the inverse $\mathbf{A}^{-1}$ can be approximated by iterating a geometric series:

$$\mathbf{A}^{-1} \approx \sum_{k=0}^{v} (\mathbf{I} - \mathbf{A})^k$$

where $\mathbf{I}$ is the identity matrix with the same dimensions as $\mathbf{A}$. Likewise, a matrix-vector product $\mathbf{A}^{-1}\mathbf{v}$ can be approximated:

$$\mathbf{A}^{-1}\mathbf{v} \approx \sum_{k=0}^{v} (\mathbf{I} - \mathbf{A})^k \mathbf{v}$$

where $\mathbf{v}$ is a vector. This approach can be leveraged for the OI problem by breaking (1) into two equations (Kalnay 2003, pp 165-166):

$$\mathbf{x}_a = \mathbf{x}_b + \widetilde{\mathbf{B}}\mathbf{M}^{-1}\mathbf{d} \tag{2}$$

$$\mathbf{d} = \mathbf{M}(\mathbf{B} + \mathbf{R})^{-1}(\mathbf{y}_o - \mathbf{y}_b) \tag{3}$$

where $\mathbf{M}$ is a $p \times p$ diagonal matrix selected to shrink the eigenvalues of the product $(\mathbf{B} + \mathbf{R})\mathbf{M}^{-1}$.

Bratseth (1986) showed that if the diagonal values of $\mathbf{M}$ are defined as:

$$(\mathbf{M})_{j,j} = \sum_{k=1}^{p} \left| (\mathbf{B})_{j,k} + (\mathbf{R})_{j,k} \right| \tag{4}$$

where the subscripts are row and column indices, then the eigenvalues of $(\mathbf{B} + \mathbf{R})\mathbf{M}^{-1}$ will lie within the range $0 < \mu \leq 1$. Thus, (3) can be approximated by iterating the series:

$$\mathbf{d} = \sum_{k=0}^{v} [\mathbf{I} - (\mathbf{B} + \mathbf{R})\mathbf{M}^{-1}]^k (\mathbf{y}_o - \mathbf{y}_b) \tag{5}$$

From the above, we can approximate OI in four steps:

1. Use equation (4) to calculate the diagonal matrix $\mathbf{M}$ from the column values of $\mathbf{B}$ and $\mathbf{R}$.
2. Calculate the inverted diagonal matrix $\mathbf{M}^{-1}$.
3. Iterate equation (5) until $\mathbf{d}$ converges, or until an iteration limit is reached (currently 5).
4. Plug $\mathbf{d}$ into equation (2) and solve for $\mathbf{x}_a$.

This algorithm requires significantly less memory than classic OI and can be parallelized and distributed across multiple processors. The error variances (diagonals of $\mathbf{B}$ and $\mathbf{R}$) can be set by look-up tables, and other values can be calculated on-the-fly (e.g., covariances calculated from a best-fit function; see next subsection). This makes it computationally feasible to use all quality-controlled observations globally, provided the observations are statistically "close enough" to affect an analysis point. We are still left with the task of identifying such observations in an efficient manner. In our implementation, an array of observation lists is constructed, with each list specifying the observations located within a unique analysis grid box. The row and column indices of the analysis grid are then used to search lists of neighbors relative to a particular location. This allows us to skip covariance calculations where the distances are obviously too large to matter (as explained in the next subsection).

d. Error covariances

The covariance matrices are parameterized to save memory, as is common in DA systems. Presently we assume the background errors are uncorrelated with observation errors (ignoring the use of the background field to disaggregate gauge data to 3-hr totals; see section 2a above). We also assume no correlated errors between observations, except when comparing two estimates from the same satellite-based product. The background error covariances are modelled as Gaussian:

$$\left(\widetilde{\mathbf{B}}\right)_{i,j} = \sigma_b^2 \cdot \exp\left(-d_{i,j}^2 / L_b^2\right) \tag{8}$$

$$\left(\mathbf{B}\right)_{i,j} = \sigma_b^2 \cdot \exp\left(-d_{i,j}^2 / L_b^2\right) \tag{9}$$

Here σ_b^2 is the background error variance, $d_{i,j}$ is the great circle distance between points i and j [calculated using a special case of the Vincenty (1975) equation, assuming a spherical Earth; see https://en.wikipedia.org/wiki/Great-circle_distance], and L_b is the background error correlation length.

The observation error covariances are parameterized as:

$$(\mathbf{R})_{j,j} = \sigma_o^2 \tag{10}$$

and either:

$$(\mathbf{R})_{i,j} = \sigma_o^2 \cdot \exp\left(-d_{i,j}^2 / L_o^2\right), i \neq j \tag{11}$$

or:

$$(\mathbf{R})_{i,j} = 0, i \neq j \tag{12}$$

Here σ_o^2 is the observation error variance, and L_o is the observation error correlation length. (11) is only used when comparing two estimates of the same satellite-based product (e.g., IMERG); otherwise, (12) is used. Note that σ_o^2 and (for satellite data) L_o are uniquely set for each observation type.

With the error correlations parameterized by the Gaussian function, it is straightforward to specify a radius of influence (the limit where observations can be considered too far to affect an analysis calculation and can be safely excluded). We define the scale length L_{max}, which is the maximum of L_b and the L_o specified for each observation type. We then set the radius of influence as $2L_{max}$, noting that the resulting Gaussian correlation is no more than ~0.02.

The error variances and correlation length scales are currently based on a global semivariogram analysis conducted with at least one month of input data[4]. A semivariogram γ (Matheron 1963; Cressie 2015) is defined here as one-half of the variance of differences in observation increments separated by great circle distance:

$$\gamma = \frac{1}{2}\langle\left(z_i - z_j\right)^2\rangle$$

[4] We note here that using a single global fit for the error covariances is a major simplification of the true errors in the data. The actual statistics will likely vary regionally based on the local weather conditions. Improving this error model is a focus of future work.

where z is the observation increment at point i or j, and the angle brackets indicate the arithmetic mean. The semivariogram procedure is as follows:

- Apply the *WaterQC*, *SnowQC*, and *DupQC* checks to the gauge reports.
- Interpolate 3-hr background precipitation to quality-controlled rain gauges.
- Calculate mean observation increments at each gauge location.
- Reject gauges with absolute mean observation increments exceeding 10 kg m^{-2}.
- Using the remaining data, calculate the semivariogram values in 10-km bins for great circle distances d up to 500 km (Fortin et al. 2015).
- Fit a Gaussian semivariogram function to the empirical results:

$$\gamma(d) = \sigma_{o,g}^2 + \sigma_b^2 \cdot [1 - \exp(-d^2/L_b^2)], d > 0 \quad (13)$$

  where $\sigma_{o,g}^2$ is the (NWP-relative) observation error variance for gauges. The best-fit is performed with the Levenberg-Marquardt non-linear least-squares algorithm (Levenberg 1944; Marquardt 1963), implemented in the MINPACK library (Moré et al. 1980; https://netlib.org/minpack), and called from the *scipy.optimize.curve_fit* function in the SciPy library (Virtanen et al. 2020; https://docs.scipy.org/doc/scipy-1.5.2/reference/optimize.html).

- For each satellite product:
  - Apply the *WaterQC*, *SnowQC*, and *SnowDepthQC* checks to the satellite estimates.
  - Match each gauge report (not screened out above) to the nearest satellite retrieval within 25 km of the gauge. Calculate new semivariogram values in 10-km bins up to 500 km, with the satellite values taking the place of the NWP background.
  - Fit another Gaussian function to the binned semivariogram values:

$$\gamma(d) = \sigma_{o,g*}^2 + \sigma_{o,s*}^2 \cdot \left[1 - \exp\left(-d^2/L_{o,s}^2\right)\right], d > 0, \quad (14)$$

    with $\sigma_{o,g*}^2$ representing the new *satellite-relative* gauge error variance estimate, $\sigma_{o,s*}^2$ representing the *gauge-relative* satellite error variance estimate, and $L_{o,s}^2$ representing the error correlation length of the satellite product.

  - Finally, rescale the satellite error variance estimate for consistency with the NWP background field. Here we assume $\sigma_{o,s}^2 = \sigma_{o,g}^2\left(\sigma_{o,s*}^2/\sigma_{o,g*}^2\right)$, where $\sigma_{o,s}^2$ is the *NWP-relative* estimated satellite error variance estimated from (13). This preserves the ratio of gauge-to-satellite error variances estimated from (14).

406 *e. Choice of analysis variable*

407 When developing NAFPA we originally used a Box-Cox cubic transformed analysis
408 variable following Fortin et al. (2015). The Box-Cox transform (Box and Cox 1964; Wilks
409 1995, pp 37-41) is:

410 $$Y^* = \left(Y^\lambda - 1\right)/\lambda, \;\; \lambda \neq 0,$$

411 $$Y^* = \log(Y), \quad \lambda = 0, \tag{15}$$

412 where Y and Y^* are the original and transformed variable, respectively, and λ is a tunable
413 transformation parameter which we set to 1/3 (Fortin et al. 2015). The intent was to transform
414 the underlying probability distribution function (PDF) of precipitation to a quasi-Gaussian
415 distribution. This would allow the Bratseth analysis value (and the OI answer it
416 approximates) to be interpreted as a maximum likelihood estimate (MLE; Daley 1991;
417 Lorenc 1986) as well as a minimum variance estimate (MVE; Daley 1991; Kalnay 2003).
418 After generating the analysis, it was necessary to apply a second transformation to convert
419 the analyzed value back into physical precipitation:

420 $$Y = (\lambda Y^* + 1)^\lambda, \quad \lambda \neq 0,$$

421 $$Y = \exp(Y^*), \quad \lambda = 0, \tag{16}$$

422 with λ again set to 1/3. Unfortunately, while transformations (15) and (16) preserve the
423 quantiles of underlying PDF (Wilks 1995, p. 38), they do *not* preserve the PDF mode or
424 mean. Instead, the mode and mean of the back-transformed PDF may shift towards different
425 quantiles due to skewness and kurtosis. This subtle detail is important, because the MLE is
426 defined as the mode of the PDF while the MVE is the mean (Lorenc 1986). If we want either
427 interpretation to be preserved after the back-transformation, we must apply a correction to
428 (16).

429 MVE correction equations are provided by Fortin et al. (2015, their equation 10) and for
430 the lognormal case by Cohn (1997, his equation 5.61), but both require estimating the
431 analysis error variance in transformed space. This calculation is not performed by the
432 Bratseth scheme because it requires explicit matrix inversion [the $\widetilde{\mathbf{B}}(\mathbf{B}+\mathbf{R})^{-1}$ term in (1)
433 above], which the algorithm avoids. Given this predicament, we originally used (15) and (16)
434 as-is, without corrections. Subsequent evaluations showed a significant dry bias was
435 introduced, which offset low mean square error scores estimated from gauge reports (Kemp
436 et al. 2020). We recently removed transformations (15) and (16), and instead analyze physical

precipitation directly; the results in this paper were generated without this transformation. Thus, NAFPA should be interpreted as an MVE analysis.

3. Evaluation Approach

To support related USAF LIS development work, NAFPA was used to generate a multi-year reanalysis from November 2007-July 2020, avoiding the variable transform discussed in section 2e to prevent a dry bias. This reanalysis used operational data (GALWEM, GFS, and rain gauges) archived in NRT by the USAF, as well as IMERG-ER V06B data. The error covariances were set using global semivariogram fitting from 2 Feb – 7 Mar 2020, and kept frozen through the reanalysis. This configuration is a proxy for a future operational configuration of NAFPA (planned for mid-2022), with the IMERG-ER data feed replacing SSMIS, GEOPRECIP, and CMORPH. Archived 0.5° GFS data are used as the background field prior to October 2017, after which 17-km GALWEM data are used when available. For a graphical depiction of the analysis grid points containing rain gauge reports and their reporting frequencies, see Fig. 1.

NAFPA was then intercompared with eight other precipitation products, which are summarized in Table 1 and detailed here:

Product	Latency	Spatial Resolution	Temporal Resolution	Data Sources
GDAS	7 hours	34-km before 14 Jan 2015, 13-km afterward	3-hr	N
IMERG-LR V06B	14 hours	0.1°	30-min	S
NLDAS-2*	4 days	0.125°	1-hr	C, G
CHIRPSv2-Final*	3rd week of following month	0.05°	1-day	C, G, S
MERRA-2	3rd week of following month	0.5° × 0.625°	1-hr	G, N
ERA5	3 months	31-km	3-hr	N
IMERG-FR V06B	3.5 months	0.1°	30-min	G, S
APHRODITE-2 V1901*	Completed Reanalysis	0.25°	1-day	G

Table 1. Summary of precipitation products intercompared with NAFPA in this paper, listed in order of increasing latency. APHRODITE-2 data are available from 2012-2015 during the evaluation period, while other products are fully available from 2012-2019. Products marked with [*] are used as references

for metric calculations. Data sources are abbreviated as C (climatology), G (gauges), N (NWP), and S (satellite estimates).

- Three low-latency (up to 4-day) products: (i) the NOAA Global Data Assimilation System (GDAS) model run produced every 6 hours by NCEP; (ii) the IMERG Late Run (IMERG-LR) generated by the NASA Precipitation Processing System (Huffman et al. 2020b); and (iii) precipitation forcing from the NOAA NLDAS Phase 2 (NLDAS-2; Xia et al. 2012; https://ldas.gsfc.nasa.gov/nldas/v2/forcing).[5]
- Four high-latency (up to 3.5-month) products: (i) the CHIRPSv2-Final product (Funk et al. 2015; https://chc.ucsb.edu/data/chirps); (ii) bias-corrected precipitation from the NASA Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2; Reichle et al. 2017; https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/), which is constrained by the NOAA CPCU analysis (Xie et al. 2007; Chen et al. 2008) and the NOAA CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997); (iii) the quality-assured ECMWF Reanalysis Version 5 (ERA5; Hersbach et al. 2020; https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5); and (iv) the NASA IMERG-Final Run (IMERG-FR) research product (Huffman et al. 2020b; https://gpm.nasa.gov/data/imerg).
- One completed reanalysis: Version 1901 of the Asian Precipitation-Highly Resolved Observational Data Integration Towards Evaluation of Water Resources, Phase 2 (APHRODITE-2; Yatagai et al. 2012, 2020; http://aphrodite.st.hirosaki-u.ac.jp/).

These alternative products vary in choice of input data. One relies on quality-controlled gauge reports (APHRODITE-2), and another adjusts gauge data with climatological bias-correction (NLDAS-2); one is satellite-based with NWP used for quasi-Lagrangian interpolation (IMERG-LR); two are NWP adjusted by atmospheric DA (GDAS and ERA5), including assimilation of rain rate data (Hersbach et al. 2020); and others blend gauges, satellite-rain rate retrievals, and/or NWP estimates. Additional data are sometimes used for

[5] We also evaluated the IMERG-ER data and found very similar results to the IMERG-LR, so IMERG-ER results will not be presented here.

482 temporal disaggregation (CHIRPSv2-Final, MERRA-2, NLDAS-2), or for QC of gauge
483 reports (APHRODITE-2).

484 We focus on three regional domains (see Fig. 1), each with a particular reference product
485 for calculating verification metrics:

486 • Within the CONUS, we use NLDAS-2. Here daily precipitation values are derived from
487 high-quality/high-density gauge data remapped to 0.125° resolution using OI (Chen et al.
488 2008), with monthly bias-correction towards PRISM climatology (Daly et al. 1994,
489 2008). The daily totals are disaggregated to hourly values using (in order of availability):
490 4-km Doppler radar-based Stage II analyses (Fulton et al. 1998; Baldwin and Mitchell
491 1997); 8-km CMORPH satellite estimates (Joyce et al. 2004); the legacy 2° × 2.5° NOAA
492 CPC Hourly Precipitation Data gauge analysis (Higgins et al. 1996); and NWP values
493 from the NOAA Regional Climate Data Assimilation System (Mesinger et al. 2006). This
494 allows us to calculate 24-hourly metrics, which we generate for 2012-2019.
495 • For Africa and adjoining land areas, we use the 0.05° CHIRPSv2-Final daily product.
496 This product is designed to provide quality rainfall estimates in regions where surface
497 observations are sparse (Funk et al. 2015); and is used by the U.S. Foreign Agriculture
498 Service/International Production Assessment Division (https://ipad.fas.usda.gov/) and by
499 the U.S. Agency for International Development for drought monitoring (McNally et al.
500 2017, 2019; Funk et al. 2019; Arsenault et al. 2020). For this region we calculate 24-hr
501 metrics for 2012-2019.
502 • Finally, in eastern Asia, we use the 0.25° daily APHRODITE-2 V1901 Monsoon Asia
503 reanalysis. This product is based on historical gauge records from multiple collaborators,
504 and the quality-controlled observation count exceeds the normal GTS report volume by
505 2.3 to 4.5 times (Yatagai et al. 2012). V1901 has also been designed to better capture
506 climatological extremes, and to account for different reporting times between countries
507 (Yatagai et al. 2020). Here we evaluate 24-hr totals from 2012-2015 (APHRODITE-2 is
508 not available past 2015).

509 Our selection of these three regions was based on the difficulties of choosing acceptable
510 reference datasets. In the past (Kemp et al. 2020) we used the Global Historical Climate
511 Network-Daily (Menne et al. 2012) but found many cases where all the tested precipitation
512 analyses had large differences for particular observations (suggesting poor observation
513 quality). Africa also suffers from poor gauge coverage, leading to a desire to use a gridded

514 analysis as a reference. The analyses from the Global Precipitation Climatology Centre
515 (GPCC; Becker et al. 2013; Schneider et al. 2016;
516 https://www.dwd.de/EN/ourservices/gpcc/gpcc.html), the Global Precipitation Climatology
517 Project (Huffman et al. 2009; Adler et al. 2018; http://gpcp.umd.edu/), and CMAP (Xie and
518 Arkin 1997) are much coarser in temporal resolution (pentad or monthly) and/or spatial
519 resolution (0.5°, 1.0°, or 2.5°) than NAFPA. Also, CPCU (Xie et al. 2007; Chen et al. 2008)
520 is unreliable in parts of Africa due to poor gauge coverage.[6] In the end, we opted to choose
521 several references (NLDAS-2, CHIRPSv2-Final, and APHRODITE-2) which we felt had
522 good quality in their respective regions.

523 For each regional domain, rainfall totals are interpolated to land points in the local
524 reference analysis grids (0.125°, 0.05°, and 0.25° resolution, for CONUS, Africa, and
525 Monsoon Asia, respectively). [Note that the coarser resolutions increasingly average/smooth
526 observed values together, potentially damping the reference accumulations used in the
527 evaluation. Also, the sample sizes for the arithmetic, domain-wide means decrease with
528 coarser resolution, increasing the width of confidence intervals.] In the case of CONUS, non-
529 U.S. land points are also excluded. When downscaling to a finer resolution, precipitation
530 values are remapped using budget interpolation (Accadia et al. 2003), while upscaling to
531 coarser resolution is performed using linear averaging. At each evaluation point, we create a
532 time series of n paired values (values of interpolated and reference rainfall, denoted y_k and o_k
533 for $k = 1, 2, \ldots, n$). From this series we calculate four metrics, which are summarized on
534 Table 2. Here we note that mean error (ME) measures the apparent bias; root mean square
535 error (RMSE) reflects the lack of apparent accuracy; and the Pearson product-moment
536 coefficient (R) measures linear correlation (Wilks 1995). Finally, for each domain, the
537 gridded metrics are combined into mean scores for easier comparison.

538

[6] We also considered MSWEP (Beck et al. 2017, 2019a, b), but these data were not available to us when we started the evaluation.

Metric	Equation	Interpretation	Perfect Score
Mean Error (ME)	$\mathrm{ME} = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)$	Bias (positive is wet, negative is dry)	0 kg m^{-2}
Root Mean Squared Error (RMSE)	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2}$	Lack of accuracy	0 kg m^{-2}
Pearson product-moment coefficient (R)	$R = \frac{\mathrm{cov}(y, o)}{\sigma_y \sigma_o}$	Linear correlation (0 indicates independence)	1

Table 2. Summary of evaluation metrics. Here $\boldsymbol{n}$ is the number of paired values of the interpolated product $\boldsymbol{y_k}$ and reference $\boldsymbol{o_k}$ over time series $\boldsymbol{k} = \boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{n}$; $\mathbf{cov}(\boldsymbol{y}, \boldsymbol{o})$ is the covariance of the interpolated and reference products, while $\boldsymbol{\sigma_y}$ and $\boldsymbol{\sigma_o}$ are the standard deviations of the interpolated and reference products. From Wilks (1995).

4. Results

a. Domain-averaged results

Domain-averaged 24-hr metrics for CONUS are given in Fig. 3, for Africa in Fig. 4, and for the Monsoon Asia region in Fig. 5. Note that CHIRPSv2-Final is excluded from the Monsoon Asia intercomparison since it does not cover the full domain. For convenience, a summary "score card" is given in Fig. 6 to aid interpretation.

Over CONUS (Fig. 3), NAFPA outperforms the low-latency GDAS and IMERG-LR for all four metrics; and outperforms ERA5 and CHIRPSv2-Final in all metrics except bias. However, for bias, NAFPA is too wet compared to ERA5, CHIRPSv2-Final, MERRA-2, and IMERG-FR. Note that the last three alternative products all include bias-correction, which is absent in NAFPA. MERRA-2 also outperforms NAFPA in accuracy and correlation.

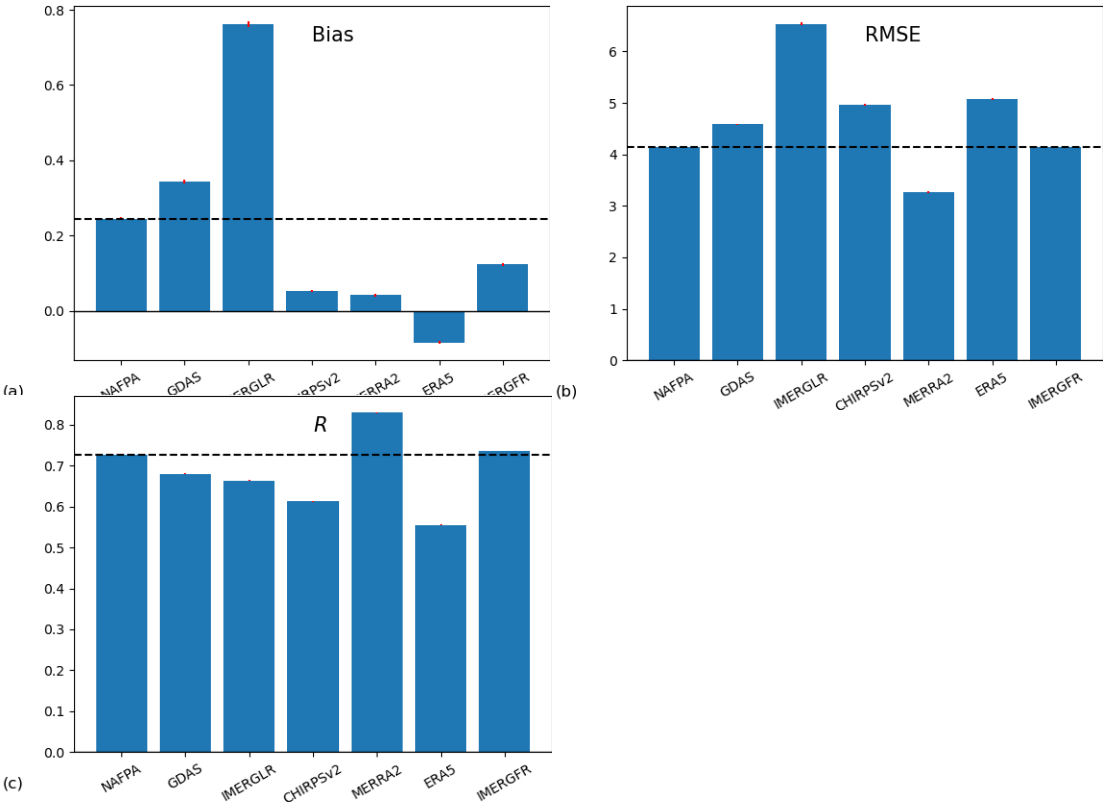


Fig. 3. Arithmetic means of 24-hr precipitation metrics calculated at each NLDAS-2 analysis land point within CONUS for 2012-2019. (a) ME (apparent bias) in kg m^{-2}; (b) RMSE (departure from apparent accuracy) in kg m^{-2}; and (c) *R* (linear correlation). Dashed line indicates metric score from NAFPA. Alternatives to NAFPA are listed left-to-right in order of latency. Red lines at top of bars indicate 95% confidence intervals of domain means.

Over Africa (Fig. 4), we again see NAFPA outperform GDAS in all three metrics. NAFPA also outperforms the low-latency IMERG-LR in ME and RMSE (less wet bias, better accuracy). Further, the higher-latency ERA5 and MERRA-2 are outperformed by NAFPA for all metrics except bias (ME), while the IMERG-FR has roughly the same apparent accuracy (RMSE). Regarding bias, we again see NAFPA is too wet compared to ERA5, MERRA-2, and IMERG-FR. Finally, NAFPA gives inferior *R* scores than both the low-latency IMERG-LR and high-latency IMERG-FR.

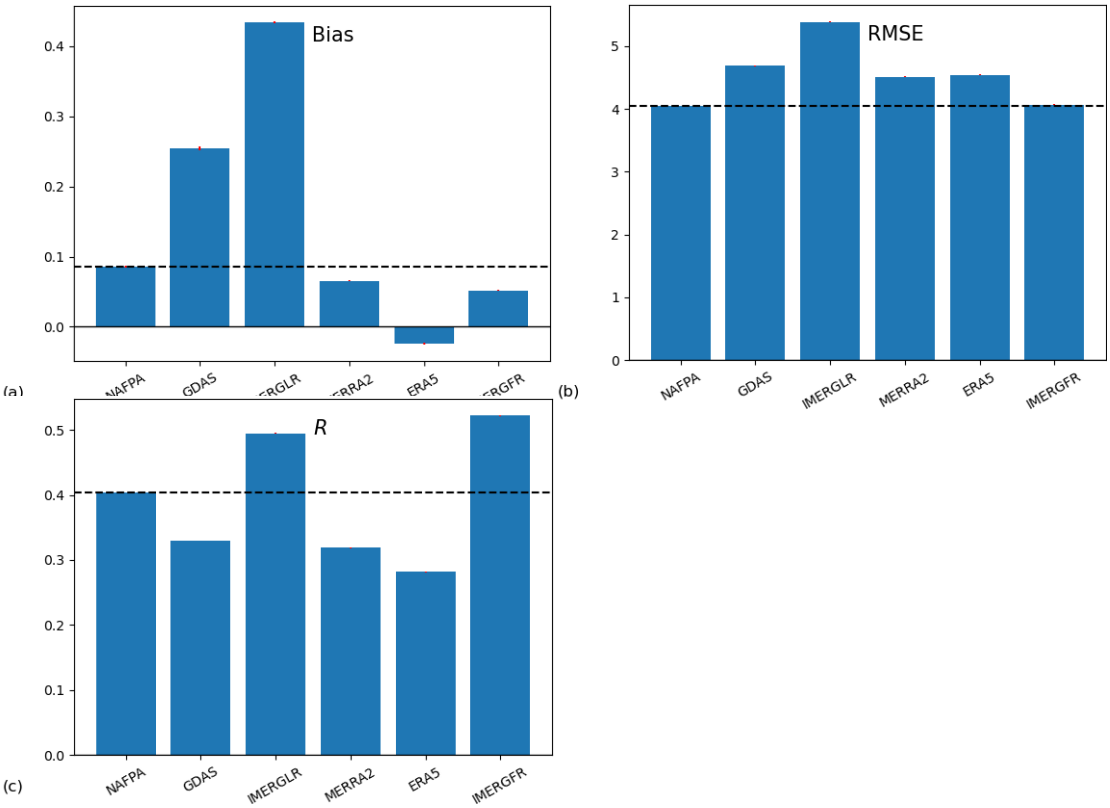
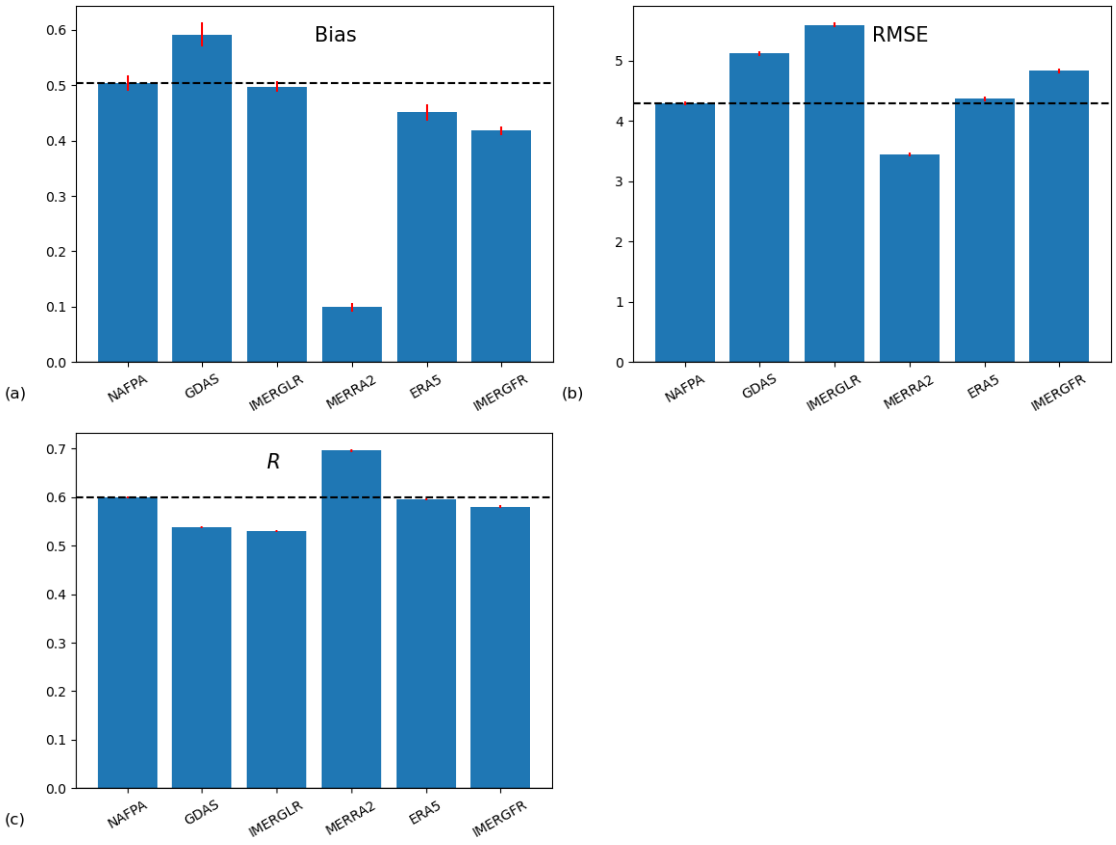


Fig. 4. Arithmetic means of 24-hr precipitation metrics calculated at each CHIRPSv2-Final analysis land point in the Africa domain for 2012-2019. (a) ME (apparent bias) in kg m^{-2}; (b) RMSE (departure from apparent accuracy) in kg m^{-2}; and (c) *R* (linear correlation). Dashed line indicates metric score from NAFPA. Alternatives to NAFPA are listed left-to-right in order of latency. Red lines at top of bars indicate 95% confidence intervals of domain means.

Over Monsoon Asia (Fig. 5), NAFPA again outperforms GDAS in all three quality measures. It also outperforms IMERG-LR in RMSE (accuracy) and *R* (correlation), while yielding the same domain-averaged ME (bias). Notably, NAFPA also outperforms the high-latency IMERG-FR in RMSE and *R*; and outperforms ERA5 in correlation with slightly better accuracy. However, NAFPA is again too wet compared to ERA5, IMERG-FR, and MERRA-2. Finally, MERRA-2 outperforms NAFPA in accuracy and correlation.

When intercomparing the three domain-averaged results, we reach several conclusions. First, NAFPA consistently outperforms the low-latency IMERG-LR in RMSE, and yields either the same or superior bias scores (IMERG-LR is often too wet). Second, NAFPA consistently outperforms the low-latency GDAS for all metrics; and consistently outperforms the high-latency ERA5 for daily *R* and RMSE. Third, NAFPA also shows either the same or superior RMSE scores than the IMERG-FR (with a 3.5-month latency). *This demonstrates the value of merging gauge reports, NWP, and satellite retrievals*, *which improves upon using NRT satellite data alone* (IMERG-LR) *or NWP with atmospheric DA* (GDAS, ERA5),

589 *and which can improve upon products generated months after the fact* (ERA5, IMERG-FR).
590 This last point is further supported by NAFPA outperforming CHIRPSv2-Final in accuracy
591 and correlation.

592


593

594 Fig. 5. Arithmetic means of 24-hr precipitation metrics calculated at each APHRODITE-2 analysis
595 land point in the Monsoon Asia domain for 2012-2015. (a) ME (apparent bias) in kg m^{-2}; (b) RMSE
596 (departure from apparent accuracy) in kg m^{-2}; and (c) *R* (linear correlation). Dashed line indicates metric
597 score from NAFPA. Alternatives to NAFPA are listed left-to-right in order of latency. Red lines at top of
598 bars indicate 95% confidence intervals of domain means.

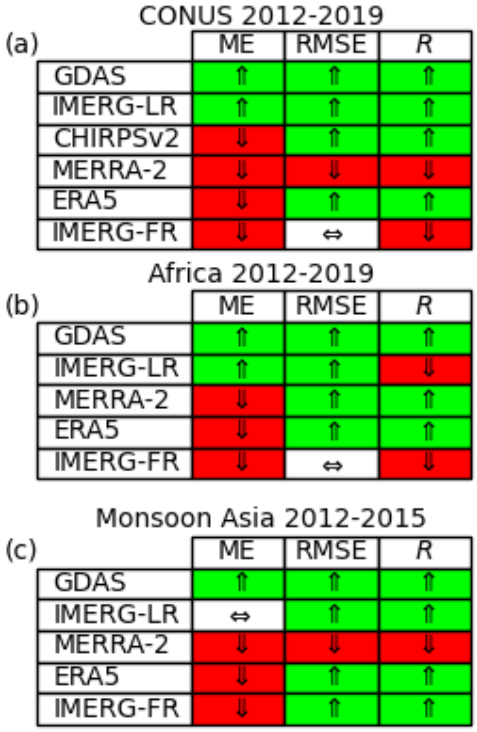


Fig. 6. Summary of 24-hr domain averaged results. Green-highlighted up arrows, red-highlighted down arrows, and side-to-side arrows indicate superior, inferior, and equivalent performance by NAFPA over the product identified in the leftmost column. (a) Results for CONUS 2012-2019 using NLDAS-2 as reference. (b) Results for Africa 2012-2019 using CHIRPSv2-Final as reference. (c) Results for Monsoon Asia 2012-2015 using APHRODITE-2 as reference.

Some limitations of the NAFPA product are also evidenced through our analysis. NAFPA always shows a wet bias compared to ERA5, MERRA-2, and IMERG-FR; and shows a wet bias compared to CHIRPSv2-Final in CONUS. It is not clear why ERA5 is drier than NAFPA, though we note that the underlying atmospheric DA system does make use of rain rate data (Hersbach et al. 2020). Meanwhile, MERRA-2, IMERG-FR, and CHIRPSv2-Final all use bias-correction in their precipitation estimates. MERRA-2 bias-corrects toward CMAP analyses over Africa, and toward CPCU analyses elsewhere for latitudes (lat) of $|lat| \leq 62.5°$, with linear tapering between $42.5° \leq |lat| \leq 62.5°$ (Reichle et al. 2017). IMERG-FR corrects toward Global Precipitation Climate Centre monthly gauge analyses (Huffman et al. 2020a), while CHIRPSv2-Final uses IR-based estimates calibrated towards climatology and local gauges (Funk et al. 2015). It appears adding a bias-correction step could therefore improve NAFPA. We also note that, for regions where MERRA-2 is bias-corrected toward CPCU (CONUS and Monsoon Asia, both with relatively high gauge counts), MERRA-2 outperforms NAFPA in accuracy and correlation. In contrast, where MERRA-2 is bias-corrected toward CMAP (Africa, with 2.5° spatial resolution and pentad temporal resolution), the RMSE and *R* scores are inferior to NAFPA. This suggests the resolution of the bias-

correction data (CPCU versus CMAP) can impact (and even degrade) analysis quality metrics beyond just bias itself.

Finally, we note again that over Africa the IMERG products outperform NAFPA in correlation. Since NAFPA assimilates IMERG-ER data, and since gauge data are generally sparce in Africa (Fig. 1), we attribute this limitation to the NWP background field. We also note that the *BackQC* test (see section 2b above) will remove observations that differ "too much" from the background field, on the assumption that such differences indicate a gross error in the observation. Thus, it may be worth relaxing this test in the future, after adding a bias-correction step.

In the next three subsections we investigate regional (sub-domain) differences in the evaluation metrics.

b. Gridded results over CONUS (NLDAS-2 reference)

Fig. 7 shows plots of 24-hr ME (bias) at each evaluation point, over CONUS for 2012-2019, for GDAS, IMERG-LR, CHIRPSv2-Final, MERRA-2, ERA5, IMERG-FR, and NAFPA. IMERG-LR shows widespread and strong wet bias in the eastern two-thirds of CONUS, but dry bias in high elevations in the western domain. GDAS shows less wet bias than IMERG-LR; but is still too wet across most of the CONUS including in the Mountain West. In contrast, NAFPA shows superior (less) bias magnitude, though there are still wet bias area across the CONUS. ERA5 shows some regions of dry bias across the eastern CONUS and in high western mountain ranges, with little bias elsewhere. The MERRA-2 and IMERG-FR products outperform NAFPA in bias. Notably, IMERG-FR has much less wet bias than IMERG-LR, although dry bias persists in parts of the western domain. Finally, CHIRPSv2-Final shows a slight wet bias in the eastern U.S. and a slight dry bias in parts of the Pacific Northwest, both superior to NAFPA.

Fig. 8 shows the differences in RMSE scores for each evaluation point for GDAS, IMERG-LR, CHIRPSv2-Final, MERRA-2, ERA5, and IMERG-FR, where the RMSEs from NAFPA are subtracted from those of each alternative product. It is apparent that NAFPA increases accuracy over IMERG-LR across most of the CONUS, with little added skill in parts of the Rocky Mountains and some degradation in South Carolina and Texas. NAFPA also improves over GDAS and ERA5 across almost all of CONUS; and over CHIRPSv2-Final outside of the central CONUS. For IMERG-FR, the pattern is very different: NAFPA

has inferior accuracy in the central U.S. and along the Gulf Coast and southeast U.S., but has superior accuracy along the West Coast, in the northeast U.S., in the Appalachian Mountains, and in parts of Michigan. As for MERRA-2, NAFPA is generally inferior except in very small local regions, especially along the West Coast.

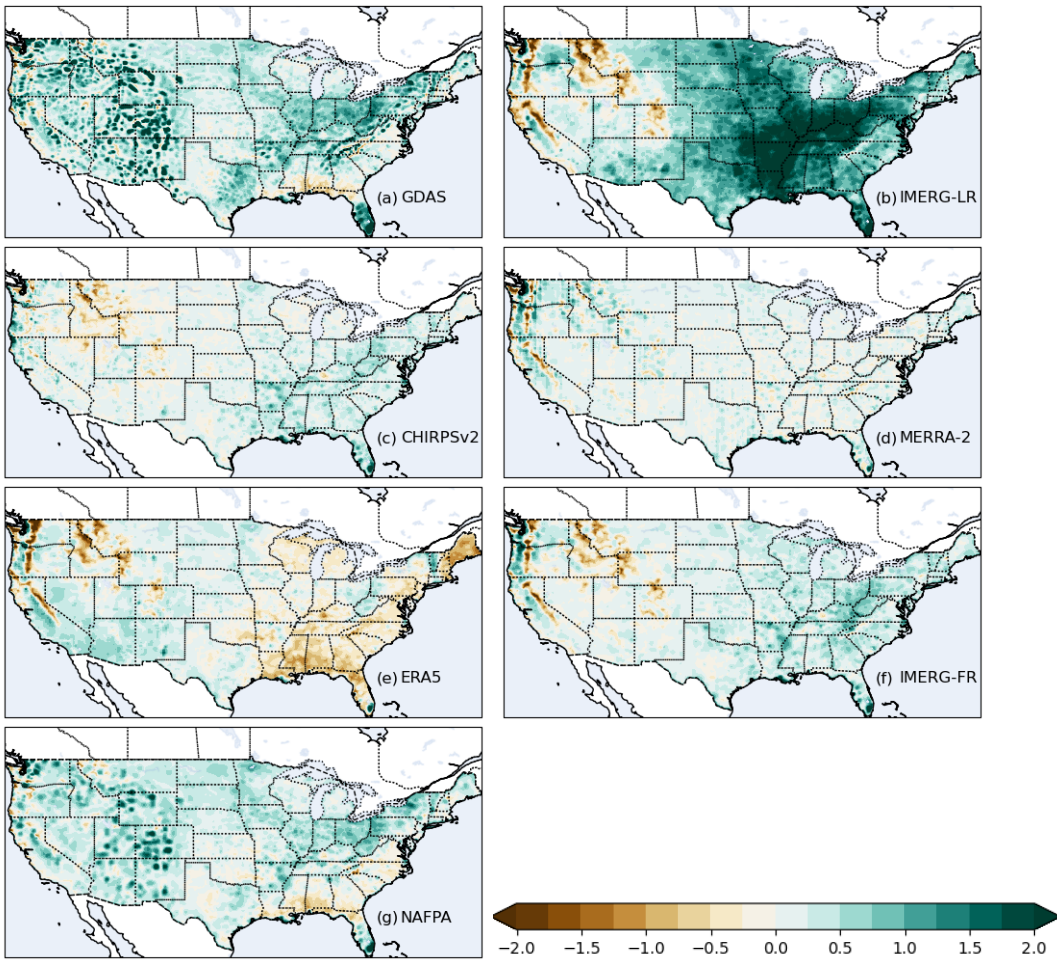


Fig 7. 24-hr ME (in kg m^{-2}) at each evaluation point in CONUS for 2012-2019, using NLDAS-2 as reference for calculation. Positive (negative) values indicate wet (dry) apparent bias. Results are for (a) GDAS; (b) IMERG-LR; (c) CHIRPSv2-Final; (d) MERRA-2; (e) ERA5; (f) IMERG-FR; and (g) NAFPA.

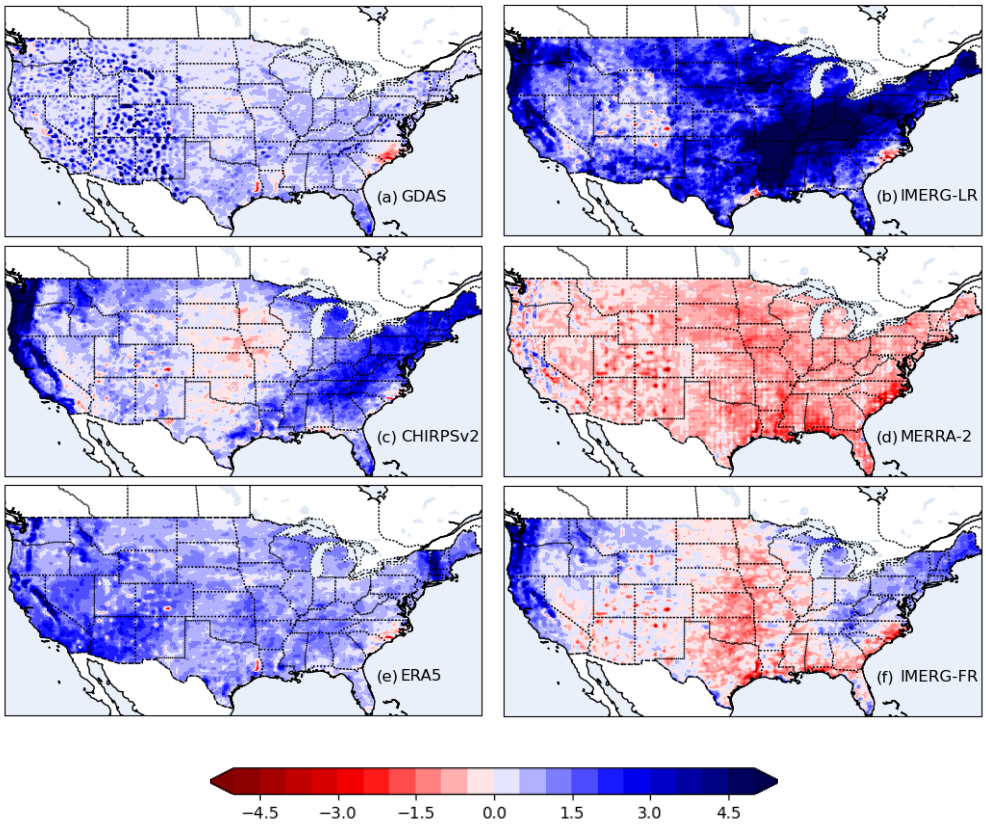


Fig. 8. Differences in 24-hr RMSE (in kg m^{-2}) in CONUS domain, with NAFPA RMSE subtracted from RMSE of other products. RMSE scores calculated for 2012-2019, using NLDAS-2 as reference. Positive (negative) values indicate superior (inferior) NAFPA accuracy. Results are for (a) GDAS; (b) IMERG-LR; (c) CHIRPSv2-Final; (d) MERRA-2; (e) ERA5; and (f) IMERG-FR.

Fig. 9 shows the differences in *R* at each evaluation point for the same precipitation products. (Note that here the subtraction occurs *from* the NAFPA value, since larger *R* indicates better performance.) NAFPA shows superior correlations to the IMERG products in the northern and western Intermountain West, and in the northeast CONUS. However, IMERG (especially the IMERG-FR) has superior daily correlations in the central U.S. and along the coastal southeast CONUS. MERRA-2 is generally superior over all the entire domain, while CHIRPSv2-Final is generally inferior except in the central U.S. GDAS and ERA5 have inferior daily correlation scores than NAFPA across the entire domain, and ERA5 has especially poor scores in the western CONUS.

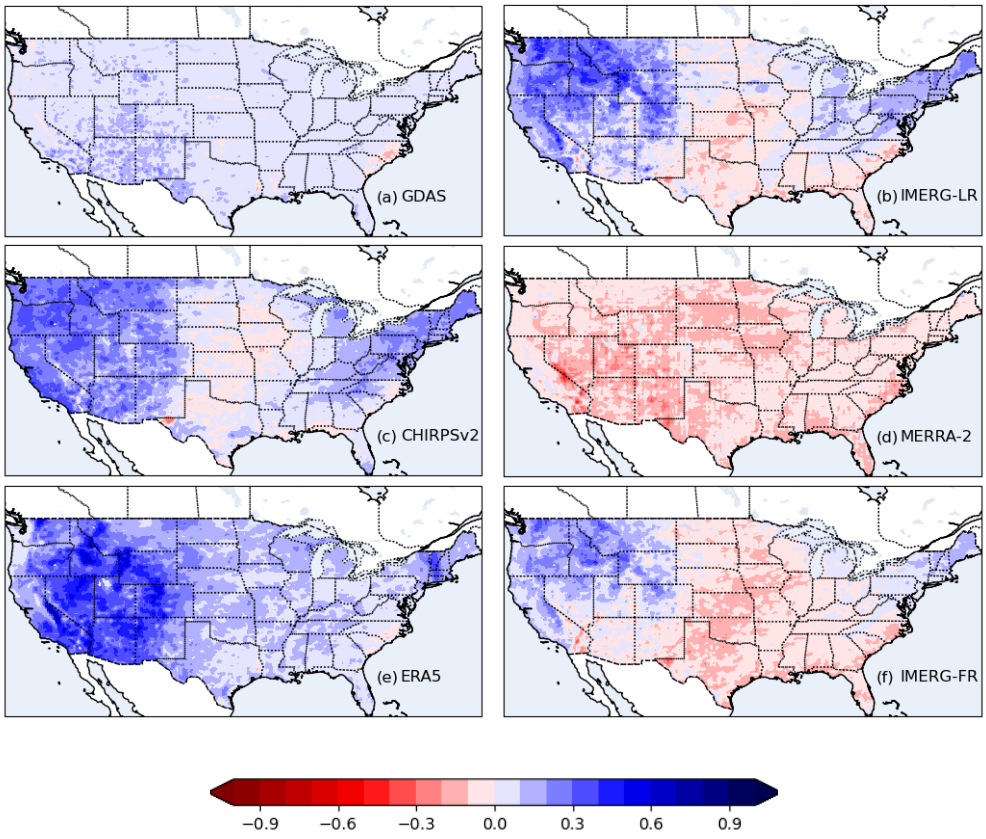


Fig. 9. Differences in 24-hr *R* in CONUS domain, with *R* from each product subtracted from NAFPA. *R* scores calculated for 2012-2019, using NLDAS-2 as reference. Positive (negative) values indicate superior (inferior) NAFPA linear correlation. Results for (a) GDAS; (b) IMERG-LR; (c) CHIRPSv2-Final; (d) MERRA-2; (e) ERA5; and (f) IMERG-FR.

c. Gridded results over Africa (CHIRPSv2-Final reference)

Fig. 10 shows the apparent daily bias for GDAS, IMERG-LR, MERRA-2, ERA5, IMERG-FR, and NAFPA over the African region from 2012-2019. Although the domain-average ME is low for NAFPA, the figure shows significant wet bias across most of east Africa and Madagascar, and along coastal central Africa; while dry bias exists across portions of central and west Africa south of the Sahara Desert. Similar patterns exist for GDAS (with stronger biases) and ERA5 (with weaker biases). In contrast, a wet bias is indicated for the NRT IMERG-LR across almost all the evaluation domain. This bias is strongly damped in the IMERG-FR version. Finally, MERRA-2 displays wet bias across most of west and east Africa, and some dry bias in central Africa.

Fig. 11 shows the RMSE differences for GDAS, IMERG-LR, MERRA-2, ERA5, and IMERG-FR, where NAFPA RMSE scores are subtracted from those of the other products. NAFPA shows apparent superior accuracy compared to IMERG-LR, except in parts of eastern Africa, Madagascar, and the area of Liberia, Sierra Leone, and Guinea. When compared to IMERG-FR, the apparent superior accuracy of NAFPA is restricted to the

Congo Basin. Compared to GDAS and MERRA-2, NAFPA appears superior across most of the Africa south of the Sahara Desert. NAFPA is also superior to ERA5 across most of Africa, excluding the western Sahara and parts of Zambia and Angola. Fig. 12 shows differences in daily *R*. Daily correlations from both IMERG runs are generally superior to NAFPA south of the Sahara Desert, while daily correlations for GDAS, MERRA-2, and ERA5 are inferior to NAFPA across almost the entire domain.

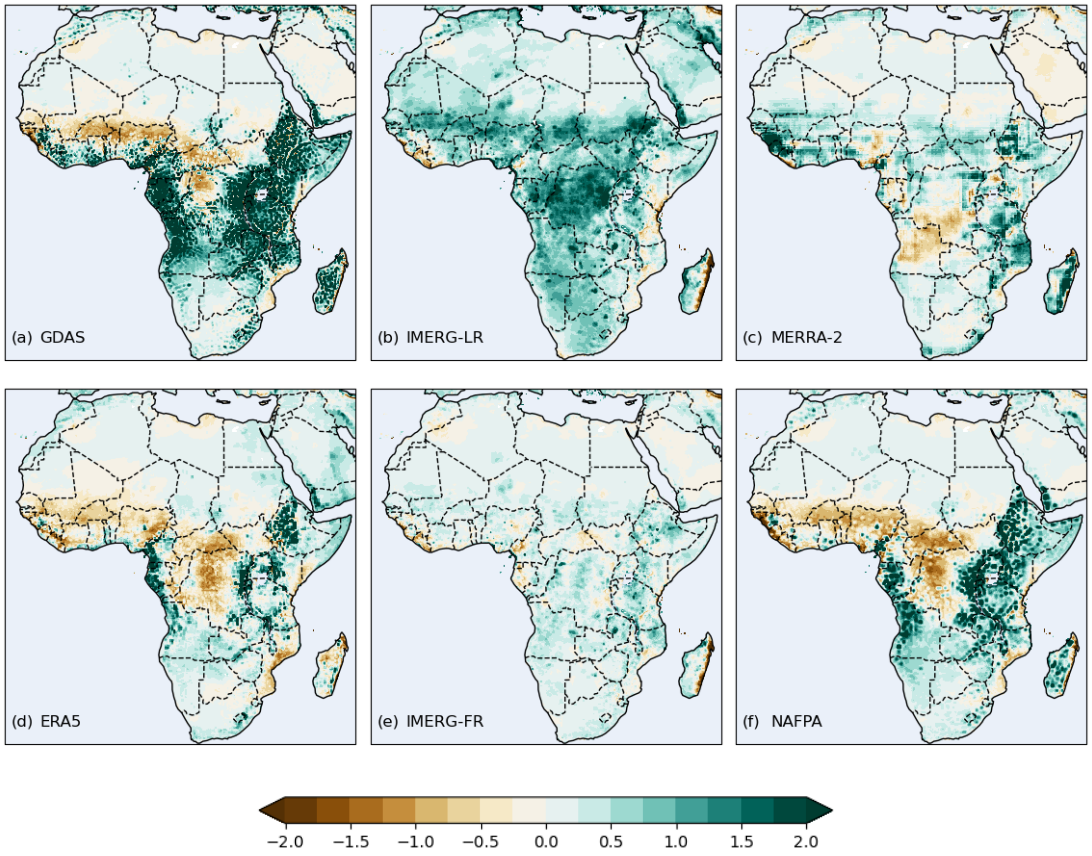


Fig. 10. 24-hr ME (in kg m^{-2}) at each evaluation point in Africa domain for 2012-2019, using CHIRPSv2-Final as reference for ME calculation. Positive (negative) values indicate wet (dry) apparent bias. Results for (a) GDAS; (b) IMERG-LR; (c) MERRA-2; (d) ERA5; (e) IMERG-FR; and (f) NAFPA.

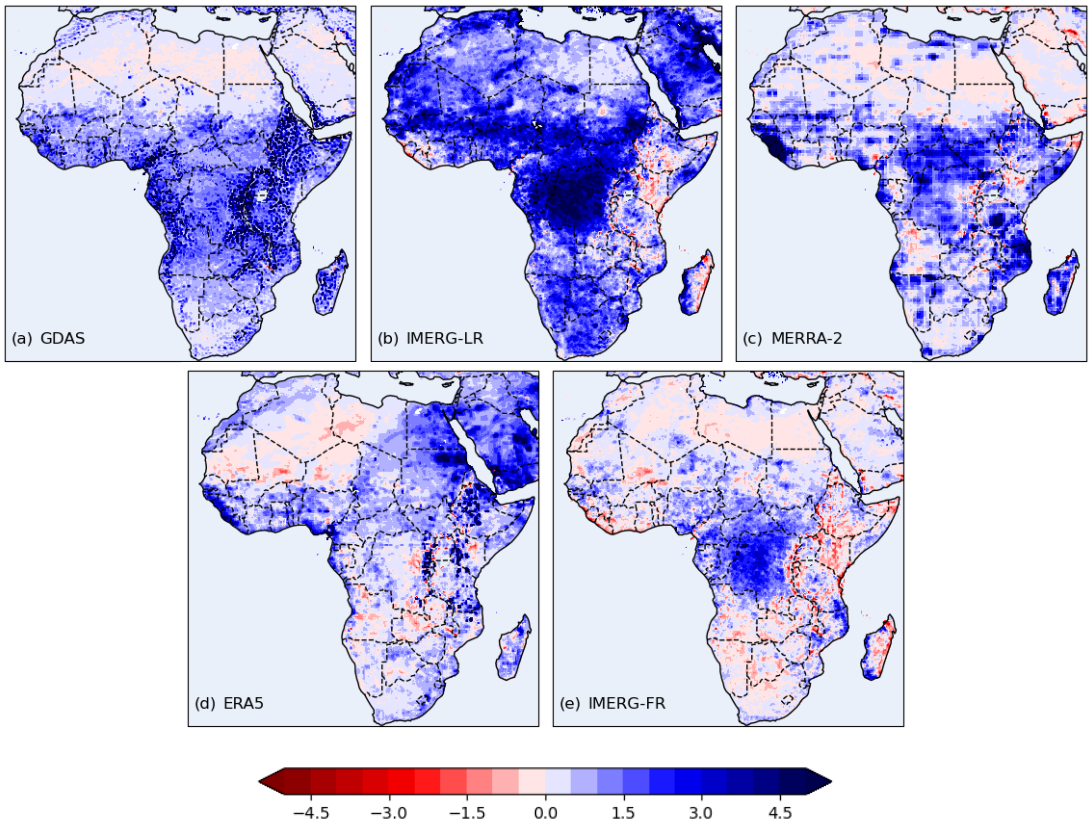


Fig. 11. Differences in 24-hr RMSE (in kg m^{-2}) in Africa domain, with RMSE from NAFPA subtracted from other products. RMSE scores calculated for 2012-2019, using CHIRPSv2-Final as reference. Positive (negative) values indicate superior (inferior) NAFPA accuracy. Results for (a) GDAS; (b) IMERG-LR; (c) MERRA-2; (d) ERA5; and (e) IMERG-FR.

d. Gridded results over Monsoon Asia (APHRODITE-2 reference)

Fig. 13 shows the 24-hr ME maps for GDAS, IMERG-LR, MERRA-2, ERA5, IMERG-FR, and NAFPA, over the Monsoon Asia domain for 2012-2015 (the end of the APHRODITE-2 dataset). All products have a dry bias in parts of Pakistan and Afghanistan, but elsewhere there are significant differences in the bias pattern. GDAS and NAFPA tend to be too wet in southeast Asia, Bangladesh, Nepal, Kyrgyzstan, and northwest of Mongolia. IMERG-LR has a wet bias in India, Bangladesh, Myanmar, Thailand, and Cambodia. MERRA-2 has damped wet bias in India, Thailand, and Cambodia, but dry bias in Vietnam, Laos, and parts of Myanmar. ERA5 looks similar to GDAS and NAFPA, except for damped bias northwest of Mongolia and in southern Russia. Finally, IMERG-FR has wet bias in Bangladesh, coastal Myanmar, Thailand, and Cambodia.

Plots of RMSE difference (Fig. 14) indicate that NAFPA is generally more accurate than both IMERG products (especially IMERG-LR) in eastern China, southeast Asia, most of India, and central Kazakhstan. NAFPA is also superior to GDAS across most of the domain. However, MERRA-2 is superior to NAFPA across most of southern China, parts of southern

Asia, and southern India. Regional comparisons with ERA5 are mixed, with NAFPA showing better accuracy in eastern China, Vietnam, and parts of India; and ERA5 looking superior in parts of the Tibetan Plateau. Correlation difference plots (Fig. 15) show NAFPA having superior *R* compared to the IMERG products across most of mainland Asia, excluding India and parts of southeast Asia; however, MERRA-2 correlations appear superior to NAFPA over almost all the domain. NAFPA has superior correlation scores than GDAS domain-wide, while superior correlation over ERA5 is restricted to the border regions of western China, northern India, and northeast Pakistan and Afghanistan.

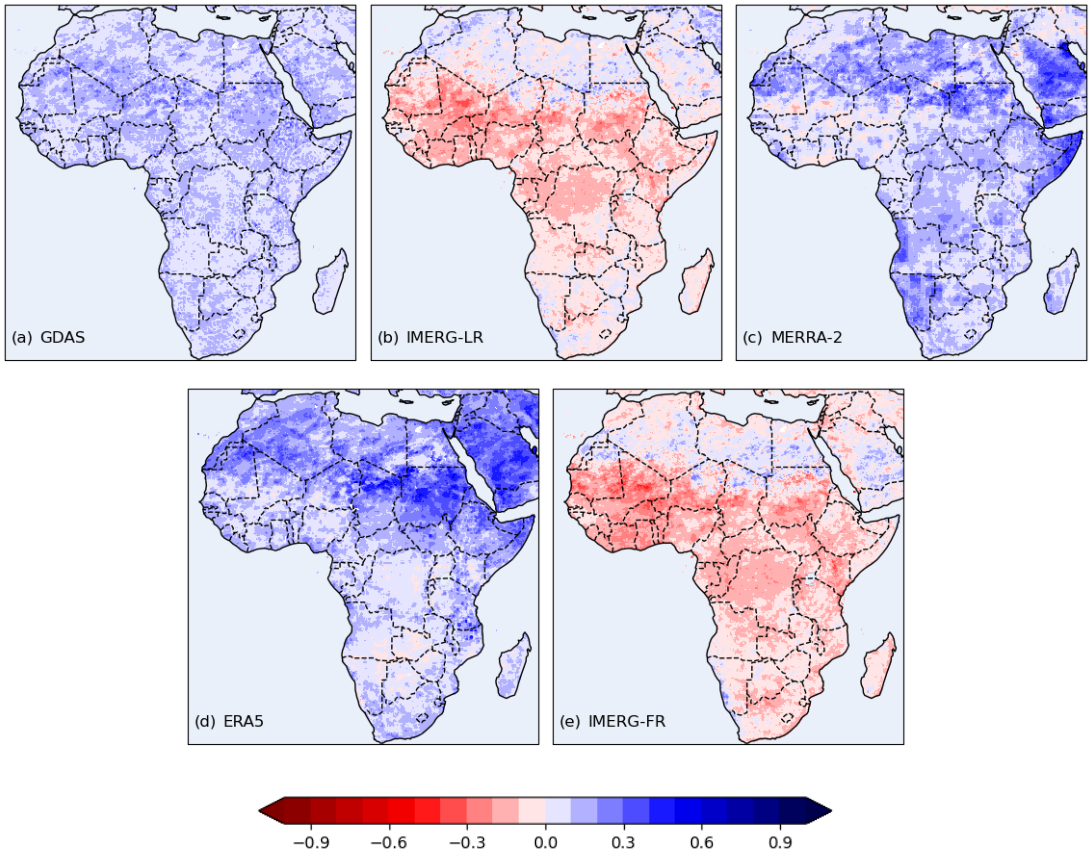


Fig. 12. Differences in 24-hr *R* in Africa domain, with *R* from other products subtracted from NAFPA. *R* scores calculated for 2012-2019, using CHIRPSv2-Final as reference. Positive (negative) values indicate superior (inferior) NAFPA linear correlation. Results for (a) GDAS; (b) IMERG-LR; (c) MERRA-2; (d) ERA5; and (e) IMERG-FR.

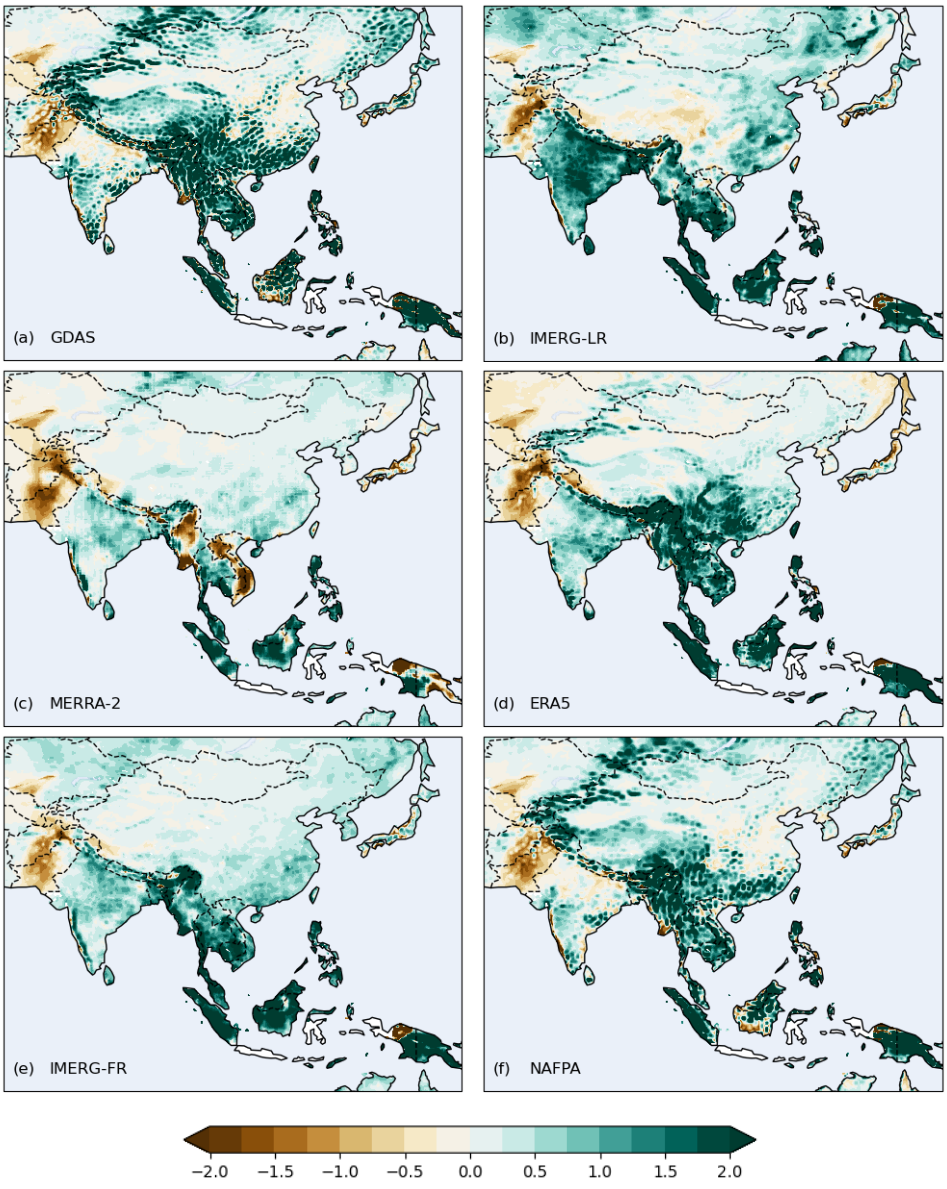


Fig. 13. 24-hr ME (in kg m^{-2}) in Monsoon Asia domain for 2012-2015, using APHRODITE-2 as reference for ME calculation. Positive (negative) values indicate wet (dry) apparent bias. Results for (a) GDAS; (b) IMERG-LR; (c) MERRA-2; (d) ERA5; (e) IMERG-FR; and (f) NAFPA.

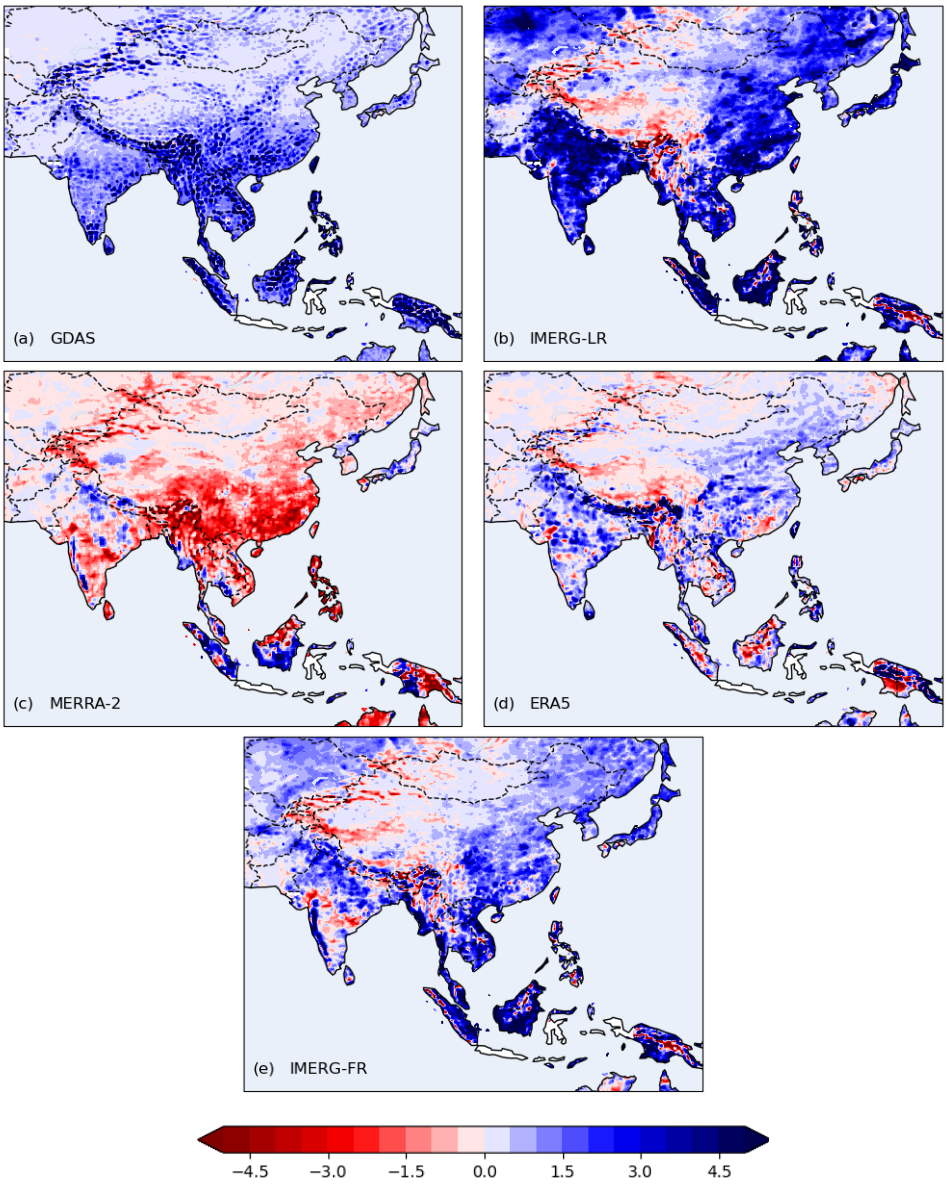


Fig. 14. Differences in 24-hr RMSE (in kg m^{-2}) in Monsoon Asia domain, with RMSE from NAFPA subtracted from other products. RMSE scores calculated for 2012-2015, using APHRODITE-2 as reference. Positive (negative) values indicate superior (inferior) NAFPA accuracy. Results for (a) GDAS; (b) IMERG-LR; (c) MERRA-2; (d) ERA5; and (e) IMERG-FR.

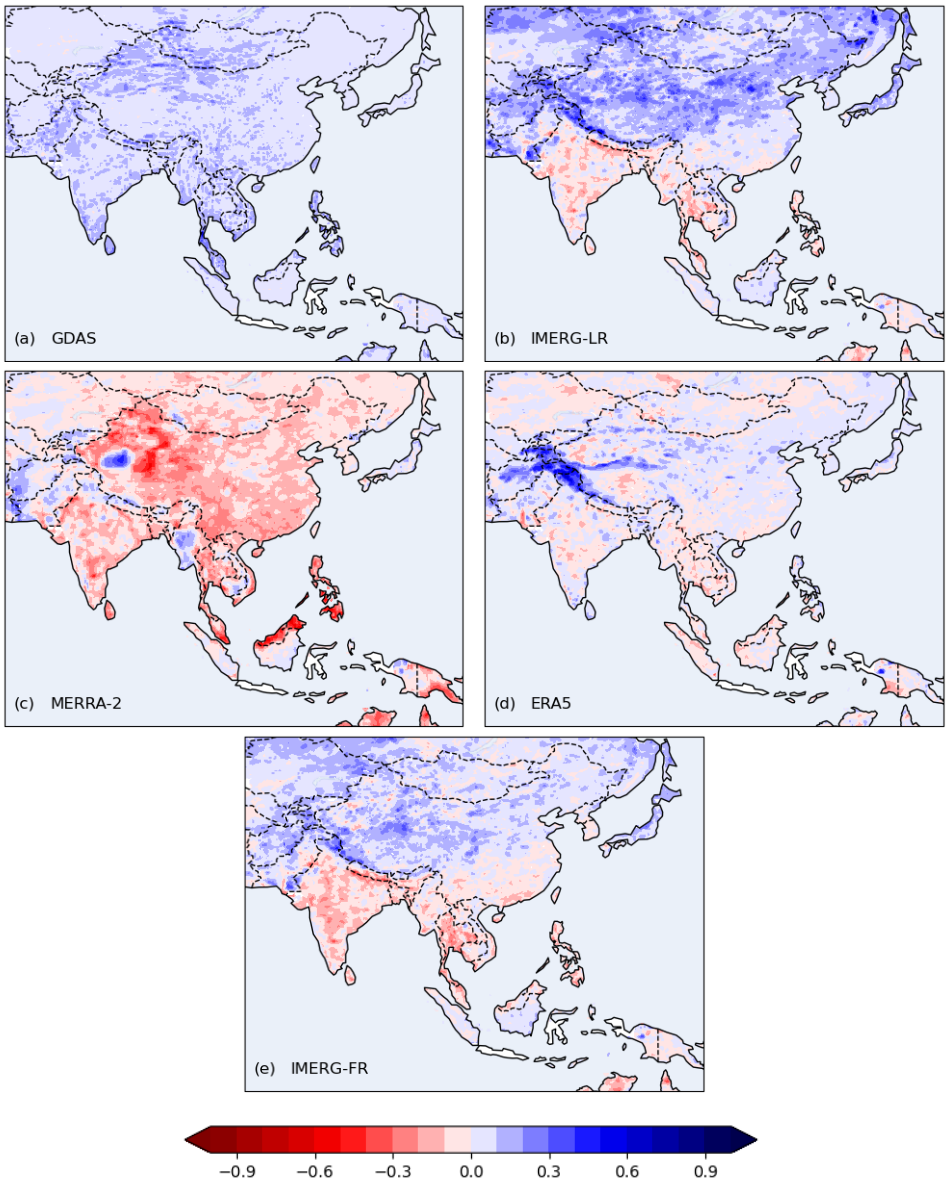


Fig. 15. Differences in 24-hr *R* in Monsoon Asia domain, with *R* from other products subtracted from NAFPA. *R* score calculated for 2012-2015, using APHRODITE-2 as reference. Positive (negative) values indicate superior (inferior) NAFPA linear correlation. Results for (a) GDAS; (b) IMERG-LR; (c) MERRA-2; (d) ERA5; and (e) IMERG-FR.

e. Discussion

The above intercomparisons point to generally superior results for NAFPA compared to alternative low-latency products (GDAS, IMERG-LR). Perhaps these results should be expected, since GDAS is an NWP system with atmospheric DA, while IMERG-LR is satellite-only (Huffman et al. 2020a). In addition, good results from merging gauges, satellite, and NWP data have earlier been reported for NRT regional domains (CaPA; Lespinas et al. 2015), and for global reanalyses (MSWEP; Beck et al. 2017, 2019a, b). On the other hand, the use of frozen global error covariances (see sections 2d and 3) could act to hinder good analyses from the source data, since the local error characteristics are averaged out. In any

784 case, it appears the basic approach of assimilating gauge reports and satellite retrievals into a
785 global NWP first guess is a successful one for NRT operations. *Furthermore, NAFPA is*
786 *competitive with products with a much-higher latency.* NAFPA usually outperforms ERA5
787 except for bias (ERA5 is significantly drier in all domains), and in some locations NAFPA
788 outperforms the IMERG-FR in accuracy. Over CONUS, NAFPA has better accuracy and
789 correlation than CHIRPSv2-Final. Similar conclusions are made in Africa for MERRA-2,
790 likely because MERRA-2 is NWP data bias-corrected toward very coarse (2.5°, pentad)
791 CMAP data in that region (Reichle et al. 2017).

792 As discussed in section 2e, the current NAFPA analysis removes a significant dry bias
793 that was originally present when using an analysis variable transformation (Kemp et al.
794 2020), and the results in this paper reflect that change. However, the current evaluation points
795 to further room for improvement. The alternative products that incorporate formal bias-
796 correction (CHIRPSv2-Final, MERRA-2, and IMERG-FR) all generally have similar-to-
797 superior bias scores. Furthermore, the low-latency IMERG-LR data clearly shows significant
798 wet bias in many regions. The presence of this bias can be problematic for NAFPA, since the
799 Bratseth scheme (and OI, and many other DA schemes) is derived with the assumption that
800 input data have no systematic errors. In general, this bias cannot be simply filtered out by
801 adjusting the error covariances, since those terms do not explicitly account for bias (Dee and
802 da Silva 1998).

803 ## 5. Summary and Future Directions

804 The NASA-Air Force Precipitation Analysis has been developed to improve precipitation
805 forcing of an operational land data assimilation system run by the USAF, as well as to
806 improve precipitation information for USAF customers. This new, 3-hourly product blends
807 NWP with rain gauge reports and satellite estimates, including the NASA NRT IMERG-ER
808 product. NAFPA uses an iterative algorithm (Bratseth 1986) that converges to OI (Gandin
809 1965) without direct matrix inversion. The software is parallelized to use distributed memory,
810 with over one thousand processors used on NASA and Department of Defense
811 supercomputers. NAFPA has been in operations at 557 WW since November 2019, with an
812 upgrade in October 2020 to fix an analysis variable dry bias. A further upgrade is anticipated
813 in mid-2022 to assimilate IMERG-ER operationally, to support the upcoming IMERG V07
814 data feed, and to automatically update the error covariances at 0000 UTC using the prior 28
815 days of data.

NAFPA is used to generate a multi-year global reanalysis (November 2007 to July 2020) at approximately 10-km resolution. Intercomparisons are made to alternative products in three different regions, with NLDAS-2 (2012-2019), CHIRPSv2-Final (2012-2019), and APHRODITE-2 V1901 (2012-2015) selected as reference analyses in CONUS, Africa, and Monsoon Asia, respectively. *Results indicate that NAFPA performs better than alternate NRT products, and in many instances has superior results than high-latency analyses produced up to 3.5 months after the fact.* More specifically, NAFPA provides better accuracy and less bias than low-latency IMERG-LR alone, and in some cases has better accuracy than IMERG-FR depending on the region. NAFPA also generally outperforms NRT NWP fields from GDAS; and for daily accuracy and correlation, NAFPA outperforms ERA5 (all domains), CHIRPSv2-Final in CONUS, and MERRA-2 in Africa.

There are, however, avenues for further improvement. Those alternative products with bias-correction (CHIRPSv2-Final, MERRA-2, IMERG-FR) generally have better domain-averaged bias scores than NAFPA. Also, the NRT IMERG data (represented here by IMERG-LR) usually have a significant wet bias that is not explicitly accounted for by NAFPA, while daily NAFPA correlations in Africa are inferior to IMERG. Considerations will be made in future versions to add a bias correction, likely by comparing with gridded monthly precipitation climatology (e.g., Karger et al. 2021). We will also explore implementing regionally varying error covariances, as these likely vary with local weather conditions (e.g., mid-latitude synoptic weather systems vs smaller-scale convection).

The regional results in CONUS also point to superior *R* scores for satellite data (IMERG-LR, CHIRPSv2-Final, IMERG-FR) in convective favored regions, such as the central CONUS and along the Gulf Coast (Liang and Fritsch 1997; Cecil et al. 2015; Ashley et al. 2019; Cheeks et al. 2020). We also note superior *R* scores for IMERG across most of Africa, including equatorial Africa which is a "hot spot" for convection (Liang and Fritsch 1997; Cecil et al. 2015). It is possible that errors in the cumulus parameterization used for the NWP background mar the resulting NAFPA analysis. It may therefore be worth disabling the *BackQC* test for satellite data, which would increase assimilation of IMERG when a mismatch occurs with the background field. Note that CaPA does not use a *BackQC* test for regional IMERG assimilation (Boluwade et al. 2018; Lespinas et al. 2021). But there are also risks with such a change, as the IMERG products in central Africa also have high RMSE scores with little bias.

It is worth noting that most of the evaluated precipitation products (except for MERRA-2) have wet biases across most of the Monsoon Asia region, while the biases in CONUS and Africa are more mixed. It is possible this is due to the relative coarse resolution of the APHRODITE-2 reference analysis (0.25°), compared to NLDAS-2 (0.125°) and CHIRPSv2-Final (0.05°), leading to more averaging/smoothing of small-scale precipitation events and lower totals in the reference field. It is also notable that IMERG-FR has somewhat more bias than MERRA-2, perhaps due to bias-correction with coarser resolution data (monthly/1.0° data from GPCC versus daily/0.5° data from CPCU).

As mentioned in section 2b, the present *DupQC* has some flaws – namely, improperly accepting the first gauge report if multiple reports are found with varying latitude/longitude coordinates, and all reports for a station are rejected if more than two unique values are found. We plan to revise *DupQC* to be more precise, only rejecting exact duplicate reports but leaving the remainder to face additional quality controls. In this new paradigm, *SuperstatQC* will be left with the responsibility of creating "super-observations" after checking for outliers. Additional QC tests are also envisioned, including a temporal test to reject gauge reports with a history of questionable data. Examples of such tests are documented in Chen et al. (2008b) and Lespinas et al. (2015).

Longer term, a switch from the Bratseth (1986) scheme to a more recent "dual formulation" variational approach (Courtier 1997) could be considered, where statistical estimation remains defined in observation space (e.g., Cohn et al. 1998; Daley and Barker 2001). Two candidates are the Restricted Preconditioned Conjugate Gradient technique (Gratton and Tshimanga 2009) and the Restricted B-preconditioned Lanczos method (Gürol et al. 2014). Both algorithms are known to converge at the same rate as the "primal formulation" variational method performed in model space. Such an approach may converge to the optimal solution faster than the Bratseth scheme, and thus yield critical wall clock savings.

Acknowledgments.

This work was financially supported by NASA GSFC under the Hydrospheric, Biospheric, and Geophysics Support contract (80GSFC20C004). Computing resources were provided by the Department of Defense High Performance Computing Modernization Program and the NASA Center for Climate Simulation. 557 WW is recognized for archiving operational input data for our NAFPA reanalyses. GDAS data were archived and provided by

H. Beaudoing, J. Jacob, and K. Arsenault of the NASA Global Land Data Assimilation System team at GSFC. Key third-party software packages used in this work include Cartopy, EcCodes, ESMF, HDF5, Matplotlib, MINPACK, Ncview, NCO, NetCDF4, NumPy, Python, and SciPy.

Data Availability Statement.

IMERG-ER, IMERG-LR, IMERG-FR, MERRA-2, and NLDAS-2 data used in this work are openly available from the NASA Goddard Earth Sciences Data and Information Services Center at https://doi.org/10.5067/GPM/IMERG/3B-HH-E/06, https://doi.org/10.5067/GPM/IMERG/3B-HH-L/06, https://doi.org/10.5067/GPM/IMERG/3B-HH/06, https://doi.org/10.5067/RKPHT8KC1Y1T, and https://doi.org/10.5067/6J5LHHOHZHN4. ERA5 precipitation data are openly available from the Copernicus Climate Change Service at https://doi.org/10.24381/cds.e2161bac. GDAS precipitation data on Gaussian grids are openly available from the NASA GLDAS team at https://portal.nccs.nasa.gov/lisdata_pub/data/MET_FORCING/GDAS/. CHIRPSv2-Final data are openly available from the Climate Hazards Center, University of California Santa Barbara at https://chc.ucsb.edu/data/chirps/. APHRODITE-2 data is available from the University of Hirosaki only to registered, non-commercial users at http://aphrodite.st.hirosaki-u.ac.jp/conditions.html. Metrics and rain gauge data used to construct the figures in this paper are openly available at https://doi.org/10.5281/zenodo.5714443.

REFERENCES

Accadia, C., S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting*, **18**, 918-932, https://doi.org/10.1175/1520-0434(2003)018<0918:SOPFSS>2.0.CO;2.

Adler, R. F., and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new Version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, https://doi.org/10.3390/atmos9040138.

Air Force Weather Agency, 2002: Data format handbook for AGRMET, 17 pp., https://www2.mmm.ucar.edu/mm5/documents/DATA_FORMAT_HANDBOOK.pdf.

910 Arsenault, K. R., and Coauthors, 2020: The NASA hydrological forecast system for food and
911 water security applications. *Bull. Amer. Meteor. Soc.*, **101**, E1007-E1025,
912 https://doi.org/10.1175/BAMS-D-18-0264.1.

913 Ashley, W. S., A. M. Haberlie, and J. Strohn, 2019: A climatology of quasi-linear convective
914 systems and their hazards in the United States. *Wea. Forecasting*, **34**, 1605-1631,
915 https://doi.org/10.1175/WAF-D-19-0014.1.

916 Baldwin, M., and K. E. Mitchell, 1997: The NCEP hourly multi-sensor U.S. precipitation
917 analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology*,
918 Boston, MA, Amer. Meteor. Soc., 54-55.

919 Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis.
920 *J. Appl. Meteor.*, **3**, 396-409, https://doi.org/10.1175/1520-
921 0450(1964)003<0396:ATFMDI>2.0.CO;2.

922 Beck, H. E., and Coauthors, 2017: Global-scale evaluation of 22 precipitation datasets using
923 gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201-
924 6217, https://doi.org/10.5194/hess-21-6201-2017.

925 Beck, H. E., and Coauthors, 2019a: Daily evaluation of 26 precipitation datasets using Stage-
926 IV gauge radar data for the CONUS. *Hydrol. Earth Syst. Sci*., **23**, 207-224,
927 https://doi.org/10.5194/hess-23-207-2019.

928 Beck, H. E., and Coauthors, 2019b: MSWEP V2 global 3-hourly 0.1° precipitation –
929 Methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, **100**, 473-500,
930 https://doi.org/10.1175/BAMS-D-17-0138.1.

931 Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M.
932 Ziese, 2013: A description of the global land-surface precipitation data products of the
933 Global Precipitation Climatology Centre with sample applications including
934 centennial (trend) analysis from 1901-present. *Earth Syst. Sci. Data*, **5**, 71,
935 https://doi.org/10.5194/essd-5-71-2013.

936 Boluwade, A., T. Stadnyk, V. Fortin, and G. Roy, 2018: Assimilation of precipitation
937 estimates from the Integrated Multisatellite Retrievals for GPM (IMERG, Early Run)
938 in the Canadian Precipitation Analysis (CaPA). *J. Hydrol. Regional Studies*, **14**, 10-
939 22, https://doi.org/10.1016/j.ejrh.2017.10.005.

940 Box, G. E. P., and D. R. Cox, 1964: An analysis of transformation. *J. Roy. Stat. Soc.*, **26B**,
941 211-243, https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

942 Bratseth, A. M., 1986: Statistical interpolation by means of successive corrections. *Tellus*,
943 **38A**, 439-447, https://doi.org/10.3402/tellusa.v38i5.11730.

944 Brown, A., S. Milton, M. Cullen, B. Golding, J. Mitchell, and A. Shelly, 2012: Unified
945 modeling and prediction of weather and climate: A 25-year journey. *Bull. Amer.*
946 *Meteor. Soc*., **93**, 1865-1877, https://doi.org/10.1175/BAMS-D-12-00018.1.

947 Cecil, D. J., D. E. Buechler, and R. J. Blakeslee, 2015: TRMM LIS climatology of
948 thunderstorm occurrence and conditional flash rates. *J. Climate*, **28**, 6536-6547,
949 https://doi.org/10.1175/JCLI-D-15-0124.1.

950 Cheeks, S. M., S. Fueglistaler, and S. T. Garner, 2020: A satellite-based climatology of
951 central and southeastern U.S. mesoscale convective systems. *Mon. Wea. Rev.*, **148**,
952 2607-2621, https://doi.org/10.1175/MWR-D-20-0027.1.

953 Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak,
954 2008a: Assessing objective techniques for gauge-based analyses of global daily
955 precipitation. *J. Geophys. Res. Atmos.*, **113**, D04110,
956 https://doi.org/10.1029/2007JD009132.

957 Chen, M., and Coauthors, 2008b: Quality control of daily precipitation reports at
958 NOAA/CPC. *12th Conf. on IOAS-AOLS*, New Orleans, LA, Amer. Meteor. Soc., 3.3,
959 6pp.

960 Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257-288,
961 https://doi.org/10.2151/jmsj1965.75.1B_257.

962 Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects
963 of data selection with the DAO Physical-Space Statistical Analysis System. *Mon.*
964 *Wea. Rev.*, **126**, 2913-2926, https://doi.org/10.1175/1520-
965 0493(1998)126<2913:ATEODS>2.0.CO;2.

966 Cosgrove, B. A., and Coauthors, 2003: Real-time and retrospective forcing in the North
967 American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res. Atmos.*,
968 **108**, 8842, https://doi.org/10.1029/2002JD003118.

969 Courtier, P., 1997: Dual formulation of four-dimensional variational assimilation. *Quart. J.*
970 *Roy. Meteor. Soc.*, **123**, 2449-2461, https://doi.org/10.1002/qj.49712354414.

971 Cressie, N. A. C., 2015: *Statistics for Spatial Data, Revised Edition*. John Wiley and Sons,
972 900 pp.

973 Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

974 Daley, R., and E. Barker, 2001: NAVDAS: Formulation and diagnostics. *Mon. Wea. Rev.*,
975 **129**, 869-883, https://doi.org/10.1175/1520-0493(2001)129<0869:NFAD>2.0.CO;2.

976 Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping
977 climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140-158,
978 https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

979 Daly, C. R., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis,
980 and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological
981 temperature and precipitation across the conterminous United States. *Int. J. Climatol.*,
982 **28**, 2031-2064, https://doi.org/10.1002/joc.1688.

983 Dee, D. P., and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias.
984 *Quart. J. Roy. Meteor. Soc.*, **124**, 269-295, https://doi.org/10.1002/qj.49712454512.

985 Eylander, J. B., C. D. Peters-Lidard, and S. V. Kumar, 2005: The AFWA next generation
986 land data assimilation system. Proceedings, *2005 Battlespace Atmospheric and Cloud*
987 *Impacts on Military Operations Conference*, Monterey CA, Naval Research
988 Laboratory, 5.01,
989 https://www.nrlmry.navy.mil/BACIMO/2005/Proceedings/5%20NWP/5.02%20NWP
990 %20Eylander%20Land%20Surface%20Assimilation%20Paper.pdf.

991 Fekete, B. M., C. J. Vörösmarty, J. O. Roads, and C. J. Willmott, 2004: Uncertainties in
992 precipitation and their impacts on runoff estimates. *J. Climate*, **17**, 294-304,
993 https://doi.org/10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2.

994 Fortin, V., G. Roy, N. Donaldson, and A. Mahidjiba, 2015: Assimilation of radar quantitative
995 precipitation estimations in the Canadian Precipitation Analysis (CaPA). *J. Hydrol.*,
996 **531**, 296-307, https://doi.org/10.1016/j.jhydrol.2015.08.003.

997 Fortin, V., G. Roy, T. Stadnyk, K. Koenig, N. Gasset, and A. Mahidjiba, 2018: Ten years of
998 science based on the Canadian Precipitation Analysis: A CaPA system overview and
999 literature review. *Atmos.-Ocean*, **56**, 178-196,
1000 https://doi.org/10.1080/07055900.2018.1474728.

1001 Fulton, R. A., J. P. Breidenbach, D.-J. Seo, D. A. Miller, and T. O'Bannon, 1998: The WSR-
1002 88D rainfall algorithm. *Wea. Forecasting*, **13**, 377-395, https://doi.org/10.1175/1520-
1003 0434(1998)013<0377:TWRA>2.0.CO;2.

1004 Funk, C., and Coauthors, 2015: The Climate Hazards Infrared Precipitation with Stations—A
1005 new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066,
1006 https://doi.org/10.1038/sdata.2015.66.

1007 Funk, C., and Coauthors, 2019: Recognizing the Famine Early Warning Systems Network:
1008 Over 30 years of drought early warning science advances and partnerships promoting
1009 global food security. *Bull. Amer. Meteor. Soc*, **100**, 1011-1027,
1010 https://doi.org/10.1175/BAMS-D-17-0233.1.

1011 Gandin, L. S., 1965: *Objective Analysis of Meteorological Fields*. Israel Program for
1012 Scientific Translations, 242 pp.

1013 Gayno, G. A., and J. Wegiel, 2000: Incorporating global real-time surface fields into MM5 at
1014 the Air Force Weather Agency. Preprints, *10th Penn State/NCAR MM5 Users'*
1015 *Workshop*, Boulder, CO, NCAR, 62-65,
1016 http://www2.mmm.ucar.edu/mm5/workshop/ws00/Gayno.doc.

1017 Goerss, J. S., and P. A. Phoebus, 1992: The Navy's operational atmospheric analysis. *Wea.*
1018 *Forecasting*, **7**, 232-249, https://doi.org/10.1175/1520-
1019 0434(1992)007<0232:TNOAA>2.0.CO;2.

1020 Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement
1021 intercomparison. WMO Rep. WMO TD-872, 212 pp,
1022 https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-67-solid-
1023 precip/WMOtd872.pdf.

1024 Gottschalck, J., J. Meng, M. Rodell, and P. Houser, 2005: Analysis of multiple precipitation
1025 products and preliminary assessment of their impact on global land data assimilation
1026 system land surface states. *J. Hydrometeor.*, **6**, 573-598,
1027 https://doi.org/10.1175/JHM437.1.

1028 Gratton, S., and J. Tshimanga, 2009: An observation-space formulation of variational
1029 assimilation using a restricted preconditioned conjugate gradient algorithm. *Quart. J.*
1030 *Roy. Meteor. Soc.*, **135**, 1573-1585, https://doi.org/10.1002/qj.477.

1031 Grody, N. C., 1991: Classification of snow cover and precipitation using the Special Sensor
1032 Microwave Imager. *J. Geophys. Res. Atmos.*, **96**, 7423-7435,
1033 https://doi.org/10.1029/91JD00045.

1034 Guo, Z., P. A. Dirmeyer, Z.-Z. Hu, X. Gao, and M. Zhao, 2009: Evaluation of the Second
1035 Global Wetness Project soil moisture simulations: 2. Sensitivity to external
1036 meteorological forcing. *J. Geophys. Res. Atmos.*, **111**, D22S03,
1037 https://doi.org/10.1029/2006JD007845.

1038 Gürol, S., A. T. Weaver, A. M. Moore, A. Piacentini, H. G. Arango, and S. Gratton, 2013: B-
1039 preconditioned minimization algorithms for variational data assimilation with the dual
1040 formulation. *Quart. J. Roy. Meteor. Soc.*, **140**, 539-556,
1041 https://doi.org/10.1002/qj.2150.

1042 Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*,
1043 **146**, 1999-2049, https://doi.org/10.1002/qj.3803.

1044 Higgins, R. W., J. E. Janowiak, and Y.-P. Yao, 1996: A gridded hourly precipitation data
1045 base for the United States (1963-1993). NCEP/Climate Prediction Center Atlas No. 1,
1046 accessed 3 May 2021,
1047 https://www.cpc.ncep.noaa.gov/research_papers/ncep_cpc_atlas/1/index.html.

1048 Huffman, G. J., R. F. Adler, D. Bolvin, and G. Gu, 2009: Improvements in the GPCP global
1049 precipitation record: GPCP Version 2.1. *Geophys. Res. Lett.*, **36**, L17808,
1050 https://doi.org/10.1029/2009GL040000.

1051 Huffman, G. J., and Coauthors, 2020a: NASA Global Precipitation Measurement (GPM)
1052 Integrated Multi-Satellite Retrievals for GPM (IMERG). NASA Algorithm
1053 Theoretical Basis Doc., Version 06, 35pp.,
1054 https://gpm.nasa.gov/sites/default/files/2020-05/IMERG_ATBD_V06.3.pdf.

1055 Huffman, G., D. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, and P. Xie, 2020b: Integrated
1056 Multi-satellitE Retrievals for GPM (IMERG), version 06B. NASA Precipitation
1057 Processing Center, accessed July 2020,
1058 https://jsimpsonhttps.pps.eosdis.nasa.gov/imerg/early/YYYYMM,
1059 https://jsimpsonhttps.pps.eosdis.nasa.gov/imerg/late/YYYYMM, and
1060 ftp://arthurhou.pps.eosdis.nasa.gov/pub/gpmdata/YYYY/MM/DD/imerg.

1061 Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that
1062 produces global precipitation estimates from passive microwave and infrared data at
1063 high spatial and temporal resolution. *J. Hydrometeor.*, **5**, 487-503,
1064 https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

1065 Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge
1066 University Press, 341 pp.

1067 Karger, D. N., A. M. Wilson, C. Mahony, N. E. Zimmerman, and W. Jetz, 2021: Global daily
1068 1km land surface precipitation based on cloud cover-informed downscaling. *Sci.*
1069 *Data*, **8**, 307, https://doi.org/10.1038/s41597-021-01084-6.

1070 Kemp, E. M., J. Wegiel, S. V. Kumar, J. Geiger, and C. Peters-Lidard, 2020: Evaluation of a
1071 new global precipitation analysis at the US Air Force 557th Weather Wing. *34th Conf.*
1072 *on Hydrology*, Boston, MA, Amer. Meteor. Soc., 13B.4,
1073 https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/368159.

1074 Kumar, S. V., and Coauthors, 2006: Land Information System: An interoperable framework
1075 for high resolution land surface modeling. *Environ. Modell. Software*, **21**, 1402-1415,
1076 https://doi.org/10.1016/j.envsoft.2005.07.004.

1077 Kunkee, D. B., G. A. Poe, D. J. Boucher, S. D. Swadley, Y. Hong, J. E. Wessel, and E. A.
1078 Uliana, 2008: Design and evaluation of the first special sensor microwave
1079 imager/sounder. *IEEE Trans. Geosci. Remote Sens.*, **46**, 863-883,
1080 https://doi.org/10.1109/TGRS.2008.917980.

1081 Laing, A. G., and J. M. Fritsch, 1997: The global population of mesoscale convective
1082 complexes. *Quart. J. Roy. Meteor. Soc.*, **123**, 389-405,
1083 https://doi.org/10.1002/qj.49712353807.

1084 Lespinas, F., V. Fortin, G. Roy, P. Rasmussen, and T. Stadnyk, 2015: Performance evaluation
1085 of the Canadian Precipitation Analysis (CaPA). *J. Hydrometeor.*, **16**, 2045-2064,
1086 https://doi.org/10.1175/JHM-D-14-0191.1.

1087 Lespinas, F., G. Roy, A. Mahidjiba, and V. Fortin, 2021: Implementation of the Regional
1088 Deterministic Precipitation Analysis System (CaPA-RDPA) version 5.0.0. Technical
1089 note, Environment and Climate Change Canada, 16pp,
1090 https://collaboration.cmc.ec.gc.ca/cmc/cmoi/product_guide/docs/tech_notes/technote_
1091 capa_rdpa-500_e.pdf.

1092 Levenberg, K., 1944: A method for the solution of certain problems in least squares. *Quart.*
1093 *Appl. Math.*, **2**, 164-168, https://doi.org/10.1090/qam/10666.

1094 Lopez, P., 2013: Experimental 4D-Var assimilation of SYNOP rain gauge data at ECMWF.
1095 *Mon. Wea. Rev.*, **141**, 1527-1544, https://doi.org/10.1175/MWR-D-12-00024.1.

1096 Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme.
1097 *Mon. Wea. Rev.*, **109**, 701-721, https://doi.org/10.1175/1520-
1098 0493(1981)109<0701:AGTDMS>2.0.CO;2.

1099 Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy.*
1100 *Meteor. Soc.*, **112,** 1177-1194, https://doi.org/10.1002/qj.49711247414.

1101 Mahfouf, J.-F., B. Brasnett, and S. Gagnon, 2007: A Canadian Precipitation Analysis (CaPA)
1102 project: Description and preliminary results. *Atmos.-Ocean*, **45**, 1-17,
1103 https://doi.org/10.3137/ao.450101.

1104 Marquardt, D. W., 1963: An algorithm for least-squares estimation of nonlinear parameters.
1105 *J. Soc. Ind. Appl. Math.*, **11**, 431-441, https://doi.org/10.1137/0111030.

1106 Matheron, G., 1963: Principles of geostatistics. *Econ. Geol.*, **58**, 1246-1266,
1107 https://doi.org/10.2113/gsecongeo.58.8.1246.

1108 McNally, A., and Coauthors, 2017: A land data assimilation system for sub-Saharan Africa
1109 food and water security applications. *Sci. Data*, **4**, 170012,
1110 https://doi.org/10.1038/sdata.2017.12.

1111 McNally, A., and Coauthors, 2019: Acute water-scarcity monitoring for Africa. *Water,* **11**,
1112 1968, https://doi.org/10.3390/w11101968.

1113 Menne, M. J., and Coauthors, 2012: Global Historical Climatology Network-Daily (GHCN-
1114 D), Version 3.26. NOAA NCEI, http://doi.org/10.7289/V5D21VHZ. [Accessed 10
1115 Aug 2019.]

1116 Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer.*
1117 *Meteor. Soc.*, **87**, 343-360, https://doi.org/10.1175/BAMS-87-3-343.

1118 Moore, B., S. Bertone, P. Rice, R. Neill, and K. Mitchell, 1991: A worldwide near-real time
1119 diagnostic agrometeorological model. Preprints, *20th Conf. on Agricultural and Forest*
1120 *Meteorology*, Salt Lake City UT, Amer. Meteor. Soc., 7-11.

Moré, J. J., B. S. Garbow, and K. E. Hillstrom, 1980: User guide for MINPACK-1. Rep. ANL-80-74, Argonne National Laboratory, Argonne IL, 261 pp, https://doi.org/10.2172/6997568.

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

Peters-Lidard, C. D., and Coauthors, 2007: High-performance Earth system modeling with NASA/GSFC's Land Information System. *Innov. Syst. Softw. Eng.*, **3**, 157-165, https://doi.org/10.1007/s11334-007-0028-x.

Qi, W., J. Liu, J. Xia, and D. Chen, 2020: Divergent sensitivity of surface water and energy variables to precipitation product uncertainty in the Tibetan Plateau. *J. Hydrol.*, **581**, 124338, https://doi.org/10.1016/j.jhydrol.2019.124338.

Rasmussen, R., and Coauthors, 2012: How well are we measuring snow? The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, **93**, 811-829, https://doi.org/10.1175/BAMS-D-11-00052.1.

Reichle, R. H., Q. Liu, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2017: Land surface precipitation in MERRA-2. *J. Climate*, **30**, 1643-1664, https://doi.org/10.1175/JCLI-D-16-0570.1.

Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185-2208, https://doi.org/10.1175/JCLI-D-12-00823.1.

Schneider, U., M. Ziese, A. Meyer-Christoffer, P. Finger, E. Rustemeier, and A. Becker, 2016: The new portfolio of global precipitation data products of the Global Precipitation Climatology Centre suitable to assess and quantify the global water cycle and resources. *Proc. Int. Assoc. Hydrol. Sci.*, **374**, 29-34, https://doi.org/10.5194/piahs-374-29-2016.

Skofronick-Jackson, G., D. Kirschbaum, W. Petersen, G. Huffman, C. Kidd, E. Stocher, and R. Kakar, 2018: The Global Precipitation Measurement (GPM) mission's scientific achievements and societal contributions: Reviewing four years of advanced rain and snow observations. *Quart. J. Roy. Meteor. Soc.*, **144**, 27-48, https://doi.org/10.1002/qj.3313.

Soci, C., E. Bazile, F. Besson, and T. Landelius, 2016: High-resolution precipitation re-analysis system for climatological purposes. *Tellus*, **68A**, 29879, https://doi.org/10.3402/tellusa.v68.29879.

Stephens, G. L., and Coauthors, 2002: The CloudSat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation. *Bull. Amer. Meteor. Soc.*, **83**, 1771-1790, https://doi.org/10.1175/BAMS-83-12-1771.

Stoffler, R. O., 2017: The USAF GALWEM: Improving military decision making, advancing national capability. *28th Conf. on Weather Analysis and Forecasting / 24th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 2B.2, https://ams.confex.com/ams/97Annual/webprogram/Paper313034.html.

Tan, J., G. J. Huffman, D. T. Bolvin, and E. J. Nelkin, 2019: IMERG V06: Changes to the morphing algorithm. *J. Atmos. Oceanic Technol.*, **32**, 2265-2280. https://doi.org/10.1175/JTECH-D-19-0114.1.

Tian, Y., and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates. *J. Geophys. Res.,* **114**, D24101, https://doi.org/10.1029/2009JD011949.

Vicente, G. A., R. A. Scofield, and W. P. Menzel, 1998: The operational GOES infrared rainfall estimation technique. *Bull. Amer. Meteor. Soc.*, **79**, 1883-1898, https://doi.org/10.1175/1520-0477(1998)079<1883:TOGIRE>2.0.CO;2.

Vincenty, T., 1975: Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review*, **23**, 88-93, https://doi.org/10.1179/sre.1975.23.176.88.

Virtanen, P., and Coauthors, 2020: SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**, 261-272, https://doi.org/10.1038/s41592-019-0686-2.

Wang, A., and X. Zeng, 2011: Sensitivities of terrestrial water cycle simulations to the variations of precipitation and air temperature in China. *J. Geophys. Res. Atmos.*, **116**, D02107, https://doi.org/10.1029/2010JD014659.

Wang, Y., and Coauthors, 2020: Sensitivity of snow simulations to different atmospheric forcing data sets in the land surface model CAS-LSM. *J. Geophys. Res. Atmos.*, **125,** e2019JD032001, https://doi.org/10.1029/2019JD032001.

Wegiel, J., and Coauthors, 2020: Data assimilation enhancements to Air Force Weather's Land Information System. *34th Conf. on Hydrology*, Boston, MA, Amer. Meteor. Soc., 366952, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/366952.

Wei, J., P. A. Dirmeyer, and Z. Guo, 2008: Sensitivities of soil wetness simulation to uncertainties in precipitation and radiation. *Geophys. Res. Lett.*, **35**, L15703, https://doi.org/10.1029/2008GL034494.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 464 pp.

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

Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539-2558, https://doi.org/10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.

Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeor.*, **8**, 607-626, https://doi.org/10.1175/JHM583.1.

Xie, P., R. Joyce, S. Wu, S.-H. Yoo, Y. Yarosh, F. Sun, and R. Lin, 2017: Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *J. Hydrometeor.*, **18**, 1617-1641, https://doi.org/10.1175/JHM-D-16-0168.1.

Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh, 2012: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401-1415, https://doi.org/10.1175/BAMS-D-11-00122.1.

Yatagai, A., M. Maeda, S. Khadgarai, M. Masuda, and P. Xie, 2020: End of the Day (EOD) judgment for daily rain-gauge data. *Atmosphere*, **11**, 772, https://doi.org/10.3390/atmos11080772.

1212 Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative
1213 precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**,
1214 621-638, https://doi.org/10.1175/BAMS-D-14-00174.1.
1215