SPECIAL
IPC12
COLLECTION

Object-Based Comparison of Data-Driven and Physics-Driven Satellite Estimates of Extreme Rainfall

ZHE LI,[a] DANIEL B. WRIGHT,[a] SARA Q. ZHANG,[b,c] DALIA B. KIRSCHBAUM,[b] AND SAMANTHA H. HARTKE[a]

[a] Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison, Wisconsin
[b] NASA Goddard Space Flight Center, Greenbelt, Maryland
[c] Science Applications International Corporation, McLean, Virginia

ABSTRACT: The Global Precipitation Measurement (GPM) constellation of spaceborne sensors provides a variety of direct and indirect measurements of precipitation processes. Such observations can be employed to derive spatially and temporally consistent gridded precipitation estimates either via data-driven retrieval algorithms or by assimilation into physically based numerical weather models. We compare the data-driven Integrated Multisatellite Retrievals for GPM (IMERG) and the assimilation-enabled NASA-Unified Weather Research and Forecasting (NU-WRF) model against Stage IV reference precipitation for four major extreme rainfall events in the southeastern United States using an object-based analysis framework that decomposes gridded precipitation fields into storm objects. As an alternative to conventional ''grid-by-grid analysis,'' the object-based approach provides a promising way to diagnose spatial properties of storms, trace them through space and time, and connect their accuracy to storm types and input data sources. The evolution of two tropical cyclones are generally captured by IMERG and NU-WRF, while the less organized spatial patterns of two mesoscale convective systems pose challenges for both. NU-WRF rain rates are generally more accurate, while IMERG better captures storm location and shape. Both show higher skill in detecting large, intense storms compared to smaller, weaker storms. IMERG's accuracy depends on the input microwave and infrared data sources; NU-WRF does not appear to exhibit this dependence. Findings highlight that an object-oriented view can provide deeper insights into satellite precipitation performance and that the satellite precipitation community should further explore the potential for ''hybrid'' data-driven and physics-driven estimates in order to make optimal usage of satellite observations.

KEYWORDS: Extreme events; Precipitation; Remote sensing; Satellite observations; Numerical weather prediction/forecasting

1. Introduction

Satellite-based remote sensing has provided unprecedented opportunities to monitor the Earth system (Lettenmaier et al. 2015; Wood et al. 2011). Particular emphasis has been placed on the estimation of precipitation using satellites (Skofronick-Jackson et al. 2018), due to its key role in weather and climate and in rainfall-driven hazards such as floods and landslides (e.g., Kirschbaum et al. 2017; Wright 2018). The most recent example is the Global Precipitation Measurement (GPM) joint mission from the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA).

Many applications require gridded precipitation estimates, often in near–real time, and call for high resolution and accurate estimation of extreme rainfall rates—both of which have posed hurdles to uptake by potential end-users (Maggioni et al. 2016a). The resolution and accuracy depend in part on the available observations from the various space-borne sensors such as the GPM ''constellation'' (Skofronick-Jackson et al. 2017). These multisensor observations must therefore be converted into precipitation rates and interpolated onto a consistent

spatial and temporal grid. The ''workhorse'' satellite instruments for precipitation estimates are passive microwave (PMW) radiometers, which observe along a satellite's ''swath,'' the relatively narrow band over Earth sampled by the onboard sensor as a satellite moves along its orbit. Infrared (IR) observations from geostationary satellites are also commonly used in the creation of gridded precipitation estimates (Joyce et al. 2001).

In this study, we posit that there are two fundamental approaches to satellite precipitation retrieval and interpolation onto spatially and temporally regular grids. The first we call the ''data-driven'' approach: precipitation estimates are derived from PMW/IR radiances using some manner of a priori database or data-driven algorithm. Prominent examples include the Goddard profiling algorithm (GPROF; Kummerow et al. 2001, 2015), the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN) family of products (Ashouri et al. 2015; Hsu et al. 1997), and ''cloud morphing''-based techniques such as the CPC morphing technique (CMORPH; Joyce et al. 2004; Xie et al. 2017), JAXA's Global Satellite Mapping of Precipitation (GsMAP; Kubota et al. 2007), and NASA's Integrated Multisatellite Retrievals for GPM (IMERG; Huffman et al. 2018). GPROF applies a large a priori database of coincident PMW brightness

temperature (TB) and precipitation estimates, from which a weighted combination of entries is selected to estimate the precipitation rate for any given PMW observation. PERSIANN relates precipitation estimates with gridded IR cloud-top TB observations via an artificial neural network (ANN) model, and its parameters are continuously adapted from sparsely sampled PMW observations. Cloud morphing uses motion vectors, usually derived from consecutive IR images, to spatially and temporally interpolate between PMW swaths. Recent data-driven datasets such as IMERG, which combines GPROF, CMOPRH's cloud morphing, and PERSIANN's ANN scheme, are more accurate and have higher resolution than their predecessors (Hou et al. 2014; Skofronick-Jackson et al. 2017).

The second method for obtaining gridded precipitation estimates from satellite remote sensing is via a numerical weather prediction (NWP) model. Precipitation estimates are produced by the model's dynamical equations and parameterizations, which are constrained through the assimilation of satellite radiances (Benjamin et al. 2019). Hence, we refer to this as the "physics-based" approach. A number of datasets, particularly reanalyses such as the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2017) from NASA and ERA5 (Hersbach et al. 2018) from the European Centre for Medium-Range Weather Forecasts assimilate PMW TBs to get gridded estimates of a wide range of atmospheric fields, including precipitation. A growing body of work has argued that NWP model simulations can, at least under specific conditions, yield precipitation estimates of comparable or better accuracy than data-driven satellite precipitation datasets (e.g., Lee et al. 2017; Nikolopoulos et al. 2015; X. Zhang et al. 2013, 2016; J. Zhang 2018; X. Zhang 2018; Lundquist et al. 2019). It is becoming increasingly feasible to run satellite-assimilating NWP model at "convection-permitting" resolutions (<4 km; Prein et al. 2015) for regional domains. One such model is NASA-Unified Weather Research and Forecasting (NU-WRF) with the Ensemble Data Assimilation System (EDAS), which is designed specifically for regional weather simulations and data assimilation at "satellite-resolved scales" (Peters-Lidard et al. 2015; S. Q. Zhang et al. 2013).

We argue that advances of modern high-resolution gridded satellite precipitation data demand alternative evaluation techniques to those commonly employed in the past: while high resolution offers potential benefits, it also poses challenges. For example, higher resolution leads to a greater likelihood of compounding errors in precipitation intensity and spatial location, which conventional grid-by-grid metrics fail to distinguish (Gilleland et al. 2009). These metrics include summary statistics at grid scale for detection skills (e.g., contingency tables) and for detected-rain errors (e.g., bias, root-mean-square error, mean absolute error, etc.), but they tend to result in so-called "double penalty" since they "punish" the estimate twice: once for missing observed rainfall at the correct location and again for falsely placing it elsewhere (Rossa et al. 2008).

This study presents an intercomparison of high-resolution gridded precipitation estimates from IMERG and NU-WRF for four extreme rainfall events in the southeastern United States. Their relative performances in reproducing key aspects of extreme precipitation, particularly storm intensity, location

and geometry are assessed. We perform an "object-based" evaluation, which provides a stronger basis in characterizing the spatial features related errors than more commonly used "grid-by-grid" metrics. Object-based evaluation methods are gaining popularity in the NWP forecasting community (e.g., Dorninger et al. 2018; Gilleland et al. 2010), but have received less attention in satellite precipitation evaluation (AghaKouchak et al. 2011; Demaria et al. 2011; Li et al. 2016). We also attempt to link the accuracy of IMERG and NU-WRF estimates to the actual satellite-borne sensors used at any particular time and location, again using object-based methods.

Section 2 describes the study region, datasets and selected storm events. Methodology follows in section 3. Section 4 presents the results; associated discussion follows in section 5. We close with a summary and conclusions in section 6.

2. Study region and data

a. Study area and case study storms

The study region is centered around the domain of the Integrated Precipitation and Hydrology Experiment (Barros et al. 2014), and spans the physiographic gradients from the Atlantic coast to the Blue Ridge Mountains in the southeastern United States (Fig. 1). This region is characterized by complex terrain along with a variety of extreme weather systems capable of producing floods and landslides (Barros et al. 2014; Mahoney et al. 2016; Moore et al. 2015; Schumacher and Johnson 2006), including tropical cyclones (TCs) and mesoscale convective systems (MCSs). Four heavy rainfall events during the period 2016–18 were selected for analysis: event 1 (6–9 October 2016) and event 4 (13–18 September 2018) were Hurricanes Matthew and Florence, respectively, while event 2 (22–25 April 2017) and event 3 (21–25 May 2017) were MCSs. These constitute four of the heaviest rainfall-producing storm systems for the region in the period since the launch of GPM in 2014. Key characteristics of the four storms can be found in Table 1.

b. IMERG gridded precipitation estimates

IMERG provides precipitation estimates every 30 min on a 0.1° grid with quasi-global coverage (60°S–60°N), and consists of three types of products: Early, Late (hereafter IMERG-L), and Final. We focus on IMERG-L, which includes some additional satellite observations not used in Early. IMERG-Final uses a rain gauge bias correction and is thus less relevant for our focus on the properties of near-real-time satellite-only precipitation estimates. We use IMERG version 05B (Huffman et al. 2018), which can be accessed at https://disc.gsfc.nasa.gov/.

IMERG contains multiple data fields in addition to the "best guess" precipitation estimates. These include a PMW source field which states which (if any) PMW sensor was used to create the estimates; PMW- and IR-only precipitation estimates, and an IR weight (IRW) which determines the extent to which IR information was used in a best guess estimate (Huffman et al. 2018). Five PMW instruments contributed to IMERG-L over this region at one or more times during the four case study storm events: the GPM Microwave Imager (GMI), the Advanced Microwave Scanning Radiometer 2 (AMSR2), the Special Sensor Microwave Imager/Sounder
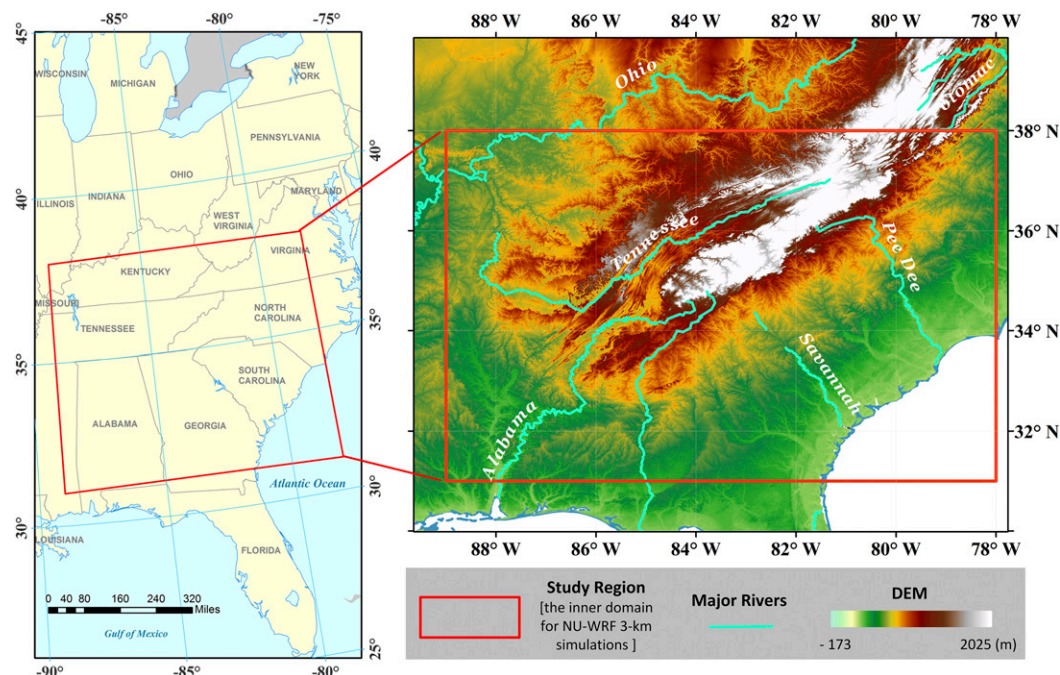
Fig. 1. The location, topography, and geographic boundaries of the study area.

(SSMIS), the Microwave Humidity Sounder (MHS), and the Advanced Technology Microwave Sounder (ATMS). More details on the role of PMW and IR auxiliary variables in IMERG algorithms can be found in Tan et al. (2016).

c. NU-WRF gridded precipitation estimates

NU-WRF combines the dynamical core of the Advanced Research WRF (ARW; Skamarock et al. 2008) with a collection schemes such as the Goddard Cumulus Ensemble (GCE; Tao et al. 2014) for cloud radiation–microphysics parameterization, the Land Information System (LIS; Peters-Lidard et al. 2007) for the land surface spinup fields, and EDAS (Zupanski et al. 2011) for data assimilation. NU-WRF EDAS assimilates precipitation-sensitive radiances using an all-sky radiative transfer simulator (Matsui et al. 2014) and a maximum likelihood ensemble filter (MLEF) to produce a 32-member ensemble used to update the state-dependent background error covariance (Zupanski et al. 2011). Recent studies have shown that NU-WRF performs well in simulating precipitation intensity and duration (Lee et al. 2017), while this EDAS module improves the estimates of both precipitation intensity and spatial pattern (S. Q. Zhang et al. 2013, 2017).

In this study, NU-WRF used 55 vertical levels (up to 50 hPa) and a 3-km inner horizontal grid (Fig. 1), nested within a 9-km horizontal grid (not shown). The Thompson microphysics scheme (Thompson et al. 2008) was used to provide microphysical simulation of clouds that are connected to satellite observation operators in radiance data assimilation, and Noah land surface model was used in atmospheric and land coupled simulation as well as within LIS spinup process. Boundary forcing came from the Global Forecast System (Whitaker et al. 2008). Hourly accumulated rainfall fields (currently NU-WRF EDAS does not facilitate output temporal resolutions finer than hourly) are generated at 3-km spatial resolution over the inner simulation domain.

Control variables in the data assimilation module EDAS include wind, temperature, surface pressure, water vapor, and five hydrometeors (the mixing ratios of cloud water, rain, ice, snow, and graupel). In summary, the observations for EDAS analysis cycles include in situ conventional data (radiosonde, pilot, wind profiler, and GPS integrated precipitable water data), clear-sky satellite radiances from the Advanced Microwave Sounding Unit A (AMSU-A), along

TABLE 1. Characteristics for the four storms evaluated in this study. Descriptions are archived and briefed from the panels of Storm Summaries and Mesoscale Precipitation Discussions at https://www.wpc.ncep.noaa.gov/.

	Event 1	Event 2	Event 3	Event 4
Year	2016	2017	2017	2018
Start time	2100 UTC 6 Oct	2100 UTC 22 Apr	0300 UTC 21 May	2100 UTC 13 Sep
End time	0300 UTC 9 Oct	0300 UTC 25 Apr	0300 UTC 25 May	1800 UTC 18 Sep
Duration (h)	54	54	96	117
Storm type	TC	MCS	MCS	TC
Description	Hurricane Matthew	Convective heavy rainfall	Convective heavy rainfall	Hurricane Florence

with precipitation-sensitive radiances from GMI, SSMIS, and AMSR2. Currently, data assimilation cycles of EDAS consists of an ensemble model simulation and an analysis at each 3-h interval. Conventional and satellite observations obtained around the analysis time pass the quality control and an online bias correction (Chambon et al. 2014) and enter into the optimization solver. For the assimilation of NU-WRF in this study, satellite swath observations such as PMW data within ±30 min of the analysis time is considered for assimilation.

d. Stage IV multisensor gridded precipitation

Stage IV multisensor precipitation is used in this study to evaluate the precipitation estimates from IMERG-L and NU-WRF. The Stage IV national mosaic merges weather radar and rain gauge measurements at 4 km hourly over the contiguous United States (Lin 2011), and has been widely used to validate satellite precipitation products and NWP simulation results (e.g., Beck et al. 2019; Lee et al. 2017; Nelson et al. 2016).

3. Methodology

a. Space–time resampling

IMERG-L, NU-WRF, and Stage IV vary in their spatial coverage, grid size, and temporal resolution. Thirty-minute IMERG-L precipitation estimates were temporally aggregated to obtain hourly totals. Hourly NU-WRF and Stage IV fields were first interpolated onto a regular 0.01° grid using nearest neighbor interpolation (Amidror 2002). The resulting finescale fields were then aggregated onto the 0.1° IMERG grid by block averaging.

b. Object-based characterization of rainstorms

We applied the object-based identification and characterization approach from the "SpatialVx" R package (Gilleland 2019) to precipitation fields from IMERG-L, NU-WRF, and Stage IV. This consisted of four steps:

Step 1: Smoothing. A two-dimensional smoothing kernel (Gilleland 2013) was applied to the accumulated hourly precipitation field (Fig. 2a) to obtain a contiguous rainy area. We used a circular disk kernel with a convolving radius of five grid lengths (i.e., 0.5° or roughly 50 km), which is slightly larger than the recommended minimum value of four grid lengths in Davis et al. (2006).

Step 2: Thresholding. Once precipitation fields were smoothed, pixels that exceeded a given threshold were identified, thus defining object boundaries. A threshold of 5 mm h^{-1} was applied in this study. Binary "masks" of distinct precipitation objects were then created (Fig. 2b). Only objects covering at least 50 grid cells (~1% of the study region, i.e., 5000 km^2) were included in further analyses.

Step 3: Characterization. By convolving the masks with their original precipitation fields (Figs. 2a,b), the precipitation distribution within each object was obtained and two groups of object properties were calculated: geometric properties—area, centroid, orientation, major and minor axis lengths, and aspect ratio, and rainfall intensity metrics—the 25th and 90th percentile (denoted hereafter as P25 and P90) of the precipitation distribution

within each object (Fig. 2c). More details on these object properties can be found in Davis et al. (2006).

Step 4: Matching. Precipitation objects were "matched" (we attempted to determine whether an object captured in the satellite precipitation estimates and Stage IV) in order to evaluate the object detection and estimation skills of IMERG-L and NU-WRF (Fig. 2d). Following after Davis et al. (2006), a match was counted if the separation distance (i.e., the Euclidean distance) between two objects' centroids was shorter than the sum of their sizes (size is the square root of the area of precipitation object).

Sensitivity of this object-based characterization to convolving radius, rainfall thresholds and matching rules are further discussed in section 5.

c. Evaluation metrics

We first applied two conventional methods to evaluate the relative performance of IMERG-L and NU-WRF: storm total accumulations and scatterplots of pixel-scale hourly precipitation estimates. For the latter, we quantified the performance in terms of three widely used evaluation metrics, including the relative bias (RB), root-mean-square error (RMSE), and Pearson linear correlation coefficient (CORR), which have been commonly defined and adopted in a large number of literatures (e.g., Tang et al. 2016; Tan et al. 2018).

The relative performance of IMERG-L and NU-WRF was further explored with an object-based view by comparing the properties of their identified precipitation objects. The capabilities of IMERG-L and NU-WRF to reproduce Stage IV objects' location, shape, and intensity (P25 and P90) were first compared without matching. Next, the detection skills of IMERG-L and NU-WRF were quantified in terms of object-oriented metrics using matched objects. These metrics include the object-based probability of detection (POD$_{obj}$), false alarm ratio (FAR$_{obj}$), and critical success index (CSI$_{obj}$), which are calculated as follows (Davis et al. 2006):

$$POD_{obj} = \frac{N_{matched}}{N_{matched} + N_{misses}}, \qquad (1)$$

$$FAR_{obj} = \frac{N_{false\_alarms}}{N_{matched} + N_{false\_alarms}}, \qquad (2)$$

$$CSI_{obj} = \frac{N_{matched}}{N_{matched} + N_{misses} + N_{false\_alarms}}, \qquad (3)$$

where $N_{matched}$ denotes the number of pairs of matched Stage IV and IMERG-L/NU-WRF objects, N_{misses} is the number of missed objects (i.e., Stage IV identified objects that are not captured by IMERG-L/NU-WRF), and N_{false_alarms} is the number of false alarms (i.e., IMERG-L/NU-WRF objects that are not observed by Stage IV).

The relative performance of IMERG-L and NU-WRF in characterizing matched precipitation objects' location, shape, and intensity can be evaluated according to a number of additional metrics such as the separation distance, intersection ratio (the ratio of intersected area divided by the area of smaller
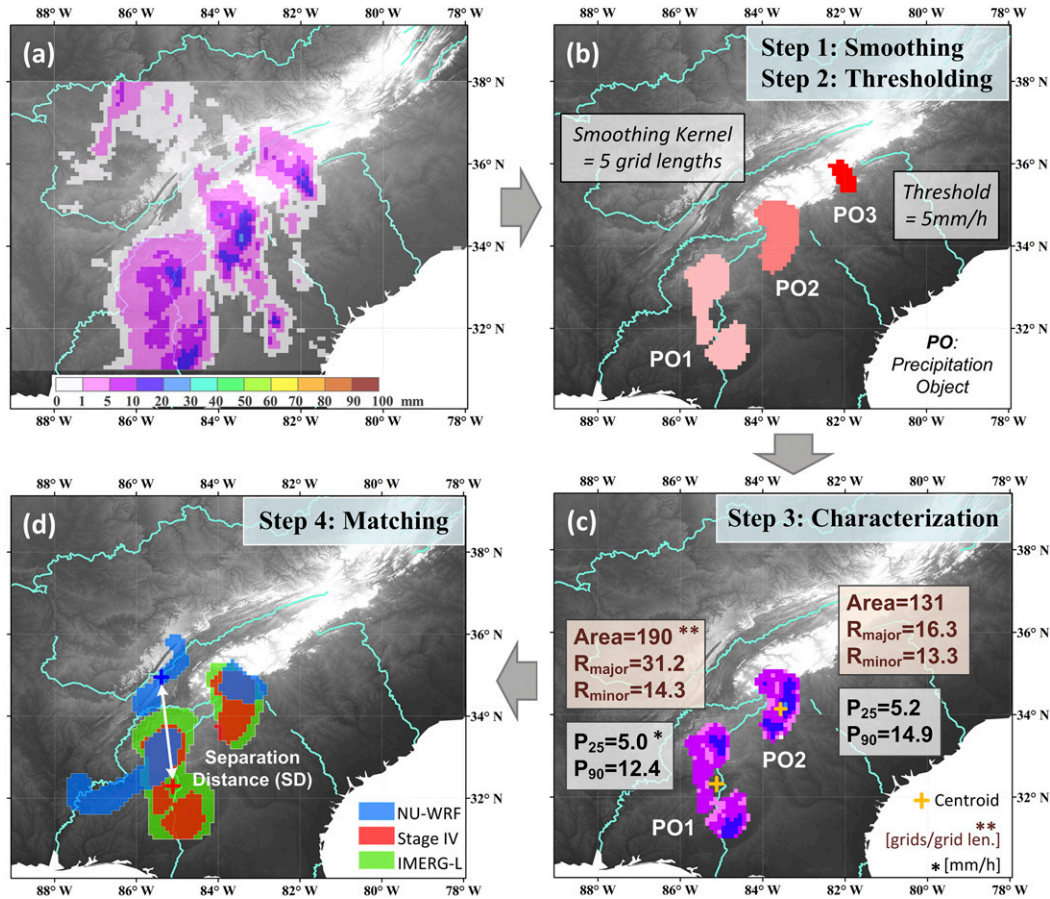
FIG. 2. Illustration of the object-based identification and characterization approach: (a) original precipitation field observed by Stage IV at 0700–0800 UTC 21 May 2017; (b) identified objects after smoothing and thresholding; (c) object-based characterization of storm properties; and (d) matching of precipitation objects from IMERG-L and NU-WRF with those from Stage IV.

matched object), and P25 and P90 estimation relative bias of matched objects:

$$RB_{P25} = \frac{P25_{est} - P25_{obs}}{P25_{obs}}, \qquad (4)$$

$$RB_{P90} = \frac{P90_{est} - P90_{obs}}{P90_{obs}}. \qquad (5)$$

$P25_{obs}$ ($P90_{obs}$) is Stage IV observed object's intensity, while $P25_{est}$ ($P90_{est}$) represents the intensity of matched precipitation object from IMERG-L or NU-WRF.

d. Categorization for conditional analysis

Precipitation objects were also grouped into various categories in terms of their properties to compare the conditional performance of IMERG-L and NU-WRF. Two characteristics, area and P90, were considered to classify the objects and to explore the detection skills of IMERG-L and NU-WRF as a function of object size and intensity.

We also examined PMW and IR inputs to IMERG-L and NU-WRF, which can impact both data- and model-based estimates' performance (Tan et al. 2016; S. Q. Zhang et al. 2013).

As mentioned above, there are embedded PMW source and IRW data fields in each 30-min IMERG file. These were overlapped on previously identified hourly IMERG-L objects to obtain half-hour object-based PMW and IRW masks. Based on every two consecutive half-hour PMW masks, identifiers for each 0.1° pixel of the mask were determined: no PMW (further classified as described below), one PMW (with sensor-specific identifier AMSR2, SSMIS, MHS, GMI, ATMS), and two consecutive PMWs (with identifier TCPMW). Object-based hourly PMW identifiers were then obtained by finding the majority identifiers across all the pixels of a mask. Objects without PMW input were categorized into two groups with further consideration of object-mean IRW: those derived purely from cloud morphing ("Morph only"; with an object-mean IRW of 0%), and those using a hybrid of morphing and IR-based estimates ("Morph + IR"; with an object-mean IRW larger than 0%), which can be further subdivided by IRW level; object-mean IRW was calculated by averaging the IRW of all pixels from the two consecutive half-hour IRW masks.

NU-WRF EDAS assimilates PMW data from multiple sensors but does not record the sensor identifier of each observation at a particular footprint after the online quality

control process. We used the following procedures to make an educated guess of the PMW measurements that may have been assimilated, and then base subsequent analysis on the assumption that such data passed quality checks. First, Level-1C intercalibrated orbital TB data from GMI, SSMIS (carried aboard *F16*, *F17*, and *F18*), and AMSR2, which passed over the study region during the four storms were collected (from https://disc.gsfc.nasa.gov/), including their overpass time stamps. Since NU-WRF EDAS only accepts PMW sensor-specific TB swaths within ± 30 min of the start time of 3-h simulation initialized by the assimilation at that time, we did not consider orbital TB data not in that time window and assign PMW identifiers (NO PMW, GMI, F16, F17, F18, AMSR2; MS is for multiple sensors) to each assimilation loop of NU-WRF. Since the assimilated PMW data constrain initial model state variables (e.g., hydrometeors) and thus simulated surface rainfall implicitly over the modeling domain (Zupanski et al. 2011), we assumed that swath-based data assimilation impacts all the hourly objects, and thus the same PMW identifier is assigned to any NU-WRF hourly object within each 3-h loop regardless of the object's location and shape.

4. Results

a. Conventional analysis of gridded precipitation fields

Visual comparison of storm total maps is a typical way of comparing precipitation estimates. For TC events (events 1 and 4), visual inspection of storm total rainfall accumulation suggests that NU-WRF outperforms IMERG-L in terms of precipitation magnitude and spatial distribution, especially for the heaviest parts within storm total rainfall fields, though it still underestimates the maximum accumulated precipitation amount (Fig. 3). IMERG-L tends to seriously underestimate the rainfall in the two TCs and fails to retrieve the heavy rain cores of storm total fields. The MCS storms (events 2 and 3) show a different picture: NU-WRF generally overestimates rainfall totals and shows a number of localized peaks, while IMERG-L captures the general spatial pattern and provides better estimates of rainfall magnitude overall.

This visual inspection highlights that the relative performances of IMERG-L and NU-WRF appears to depend on rainfall regime, albeit based on a small sample of four storms. Their relative skills mainly manifest in the highly varied representations of spatial rainfall structures, especially localized extreme rainfall accumulations.

Scatterplots and associated summary statistics such as bias, error (RMSE, mean absolute error, etc.), and correlations are common methods for comparing precipitation datasets. These methods might be limited, however, due to their lack of regard to spatial information (Gilleland et al. 2009). As shown in Fig. 4, scatterplots and evaluation metrics of pixel-scale hourly precipitation estimates from IMERG-L and NU-WRF provide some insights into their error characteristics, but also draw attention to their shortcomings. IMERG-L shows higher correlation and lower random error, while NU-WRF is less biased, with the exception of event 3. The results of Fig. 4 seem not

consistent, however, with Fig. 3: both NU-WRF and IMERG-L typically underestimate grid-scale precipitation according to the relative bias values, contrasting with the aforementioned NU-WRF's overestimation and IMERG-L's general agreement of storm total precipitation for the MCSs. This may be due to the fact that Fig. 4 only includes cases in which both Stage IV and IMERG or NU-WRF estimates are greater than 0.1 mm h^{-1}. Scatterplots can illustrate hit bias (Tian et al. 2009), but not detection error, which may play a critical role in heavy rainfall estimation for specific weather systems such as the MCSs shown in Fig. 3.

b. Characterization and comparison of precipitation objects

Maps of hourly precipitation object properties from IMERG-L, NU-WRF, and Stage IV for TC and MCS events (Figs. 5 and 6, respectively) highlight features not easily discerned from Figs. 3 and 4. Objects show larger size and much simpler spatial patterns for the TCs than for the MCSs, and both IMERG-L and NU-WRF are relatively accurate in terms of capturing the spatial evolution of the TCs. IMERG-L fails to retrieve the heavy rain cores of storm total fields in both TC events (Fig. 3): for Hurricane Matthew (event 1; Figs. 5a,c), IMERG-L identifies precipitation objects with approximately correct size and location but underestimates P90 along this hurricane's path; for Hurricane Florence (event 4; Figs. 5b,d), it yields too few precipitation objects and underestimates rainfall intensity through the entire event. NU-WRF, on the other hand, provides better estimates in terms of the number and intensity of precipitation objects during both TC events, though with a noticeable eastward displacement error in the case of Hurricane Matthew (Fig. 5a).

In contrast with TCs, the spatial pattern of MCS events tends to be more complicated (Fig. 6). Detection errors appear to dominate the performance of both IMERG-L and NU-WRF in MCS storms: IMERG-L misses several precipitation objects in events 2 and 3, including large objects near the coast (Figs. 6c,d), while NU-WRF generates several spurious high-intensity precipitation objects in the western (southwestern) portion of the study area during event 2 (event 3) and fails to detect many precipitation objects in the mountainous areas (Figs. 6a,b).

Boxplots of object geometric properties from IMERG-L and NU-WRF (left column in Fig. 7) show that major and minor axis lengths are relatively unbiased during events 1 and 3. Errors in object geometry plays greater roles in events 2 and 4: IMERG-L and NU-WRF both underestimate the precipitation object size in event 2; in event 4, NU-WRF overestimates the object sizes whereas IMERG-L underestimates. Since the relative length of a boxplot's whiskers indicates the skewness of a sample, Fig. 7 (left column) also suggests that IMERG-L and NU-WRF generally capture the overall skewness of precipitation objects' size distribution. In event 1 (negatively skewed), the size distribution is concentrated on large precipitation objects, which are more likely to be accurately estimated with both approaches; when the size distribution is shifted toward smaller objects in the positively skewed scenarios (events 2, 3, and 4), IMERG-L and NU-WRF tend to have larger uncertainty and discrepancy in resolving the shape of rainstorms.
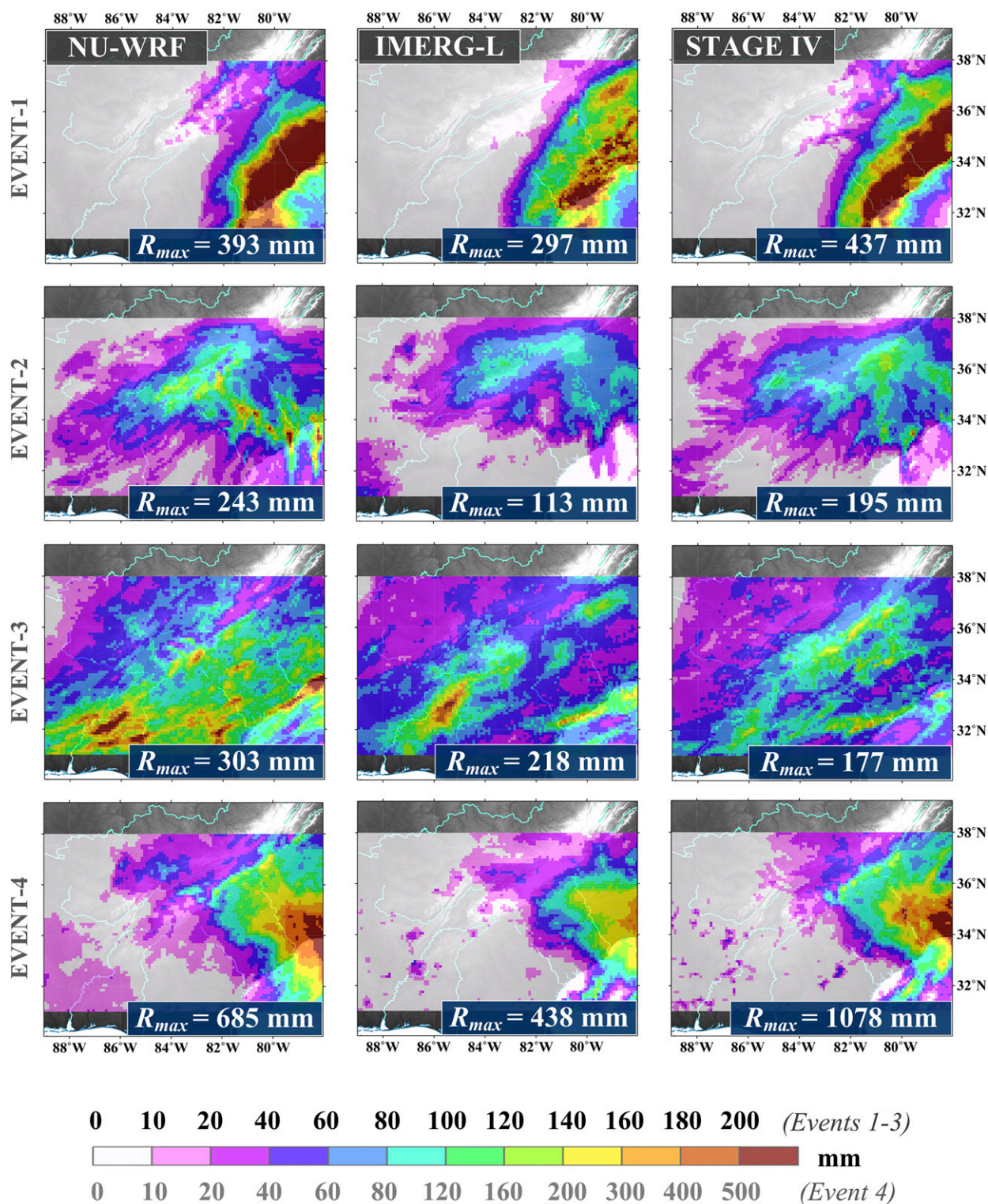
FIG. 3. Storm total accumulations from NU-WRF, IMERG-L, and Stage IV during the four storm events; R_{max} represents the maximum accumulated precipitation amount.
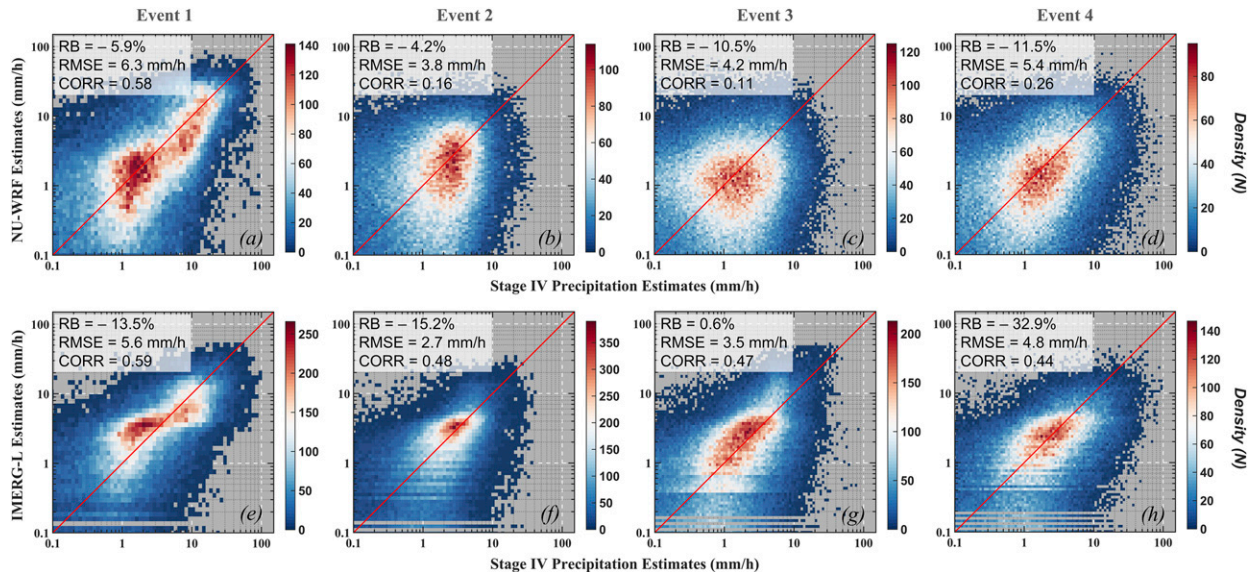
FIG. 4. Density scatterplots of gridded hourly precipitation estimates from (a)–(d) NU-WRF and (e)–(h) IMERG-L vs Stage IV during the four events. Only the gridded data which are simultaneously estimated to be greater than 0.1 mm h^{-1} by Stage IV and NU-WRF (IMERG-L) have been considered. RB and RMSE denote relative bias and root-mean-square error, respectively, while CORR denotes Pearson correlation coefficient.

Boxplots for P25 and P90 shed additional light on the relative skill of IMERG-L and NU-WRF with respect to rainfall intensity within the objects (center and right columns in Fig. 7). IMERG-L tended to accurately capture P25 and P90 rain rates for MCS events (events 2 and 3). NU-WRF, on the other hand, overestimated P90 and underestimated P25 in the MCS events. In the TC events (events 1 and 4), NU-WRF performs as well or better than IMERG-L. Both IMERG-L and NU-WRF slightly overestimate P25 and underestimate P90 in Hurricane Florence.

Figure 8 compares the object-based detection skills of IMERG-L and NU-WRF during the four events after matching, in terms of object-based POD, FAR and CSI (defined in section 3c). NU-WRF presents POD and CSI roughly comparable with IMERG-L, but a larger FAR for all events. FAR is much higher for both IMERG-L and NU-WRF under the MCS scenarios (events 2 and 3), confirming that both datasets struggle with properly depicting localized convection.

Figure 9 summarizes the properties of matched precipitation objects from IMERG-L and NU-WRF such as the separation distance, intersection ratio, and P25 and P90 estimation relative bias. Separation distance for NU-WRF matched objects tends to range from 0.6° to 1.0° (6–10 grid lengths), while it is mostly below 0.4° (4 grid lengths) for IMERG-L (Figs. 9a–d). Meanwhile, the general distribution of separation distance for NU-WRF is more dispersed than IMERG-L in most cases. IMERG-L objects typically have a larger intersection ratio (median of 0.5–0.8) than NU-WRF (median less than 0.5 except for event 1), especially for MCS rainstorm events (Figs. 9e–h). The histograms of intercentroid separation distance together with intersection ratio confirm that IMERG-L outperforms NU-WRF in locating storms and capturing their spatial coverage. Relative estimation bias of P25 and P90 for

matched objects from NU-WRF are generally low for TC events, while IMERG-L presents a relatively better performance in estimating MCS storms (Figs. 9i–p). Generally, NU-WRF underestimates low rainfall intensity (P25) in MCS storms, while IMERG-L underestimates high-intensity rain rates (P90) in TC events.

We examined the dependence of object-based POD on the size and P90 of Stage IV objects—in other words, the likelihood of detection of a "real" precipitation object as a function of object size and intensity. It was found that POD depends strongly on the size of precipitation object for both IMERG-L and NU-WRF (Fig. 10a), indicating that large precipitation objects are more likely to be captured by both approaches. A similar relationship exists between POD and P90 in NU-WRF, but not in IMERG-L (Fig. 10b): detection skill of NU-WRF increases with 90th percentile rainfall intensity. This is generally true in IMERG-L when rainfall intensity is lower than 18 mm h^{-1}, but then it shows a dramatic decline in detection ability due to IMERG's serious underestimates of extreme storms like Hurricane Florence.

c. Impact of PMW and IR data

As we connect their accuracy back to input data source (IRW, or a specific PMW sensor), the object-oriented evaluation reveals varied dependencies of IMERG-L and NU-WRF on IR and PMW data source. In the case of IMERG-L, objects with PMW generally present a higher hit ratio than those based on morphing purely or morphing combined with IR estimates ("Morph only" and "Morph + IR," respectively, in Fig. 11a). Hourly objects derived with two consecutive half-hour PMW inputs ("TCPMW" in Fig. 11a) have the highest hit ratio of 81% but are relatively uncommon (7% of observations). MHS and SSMIS also show strong performance though both are also characterized by a notably high FAR ratio of 25% and 16% for
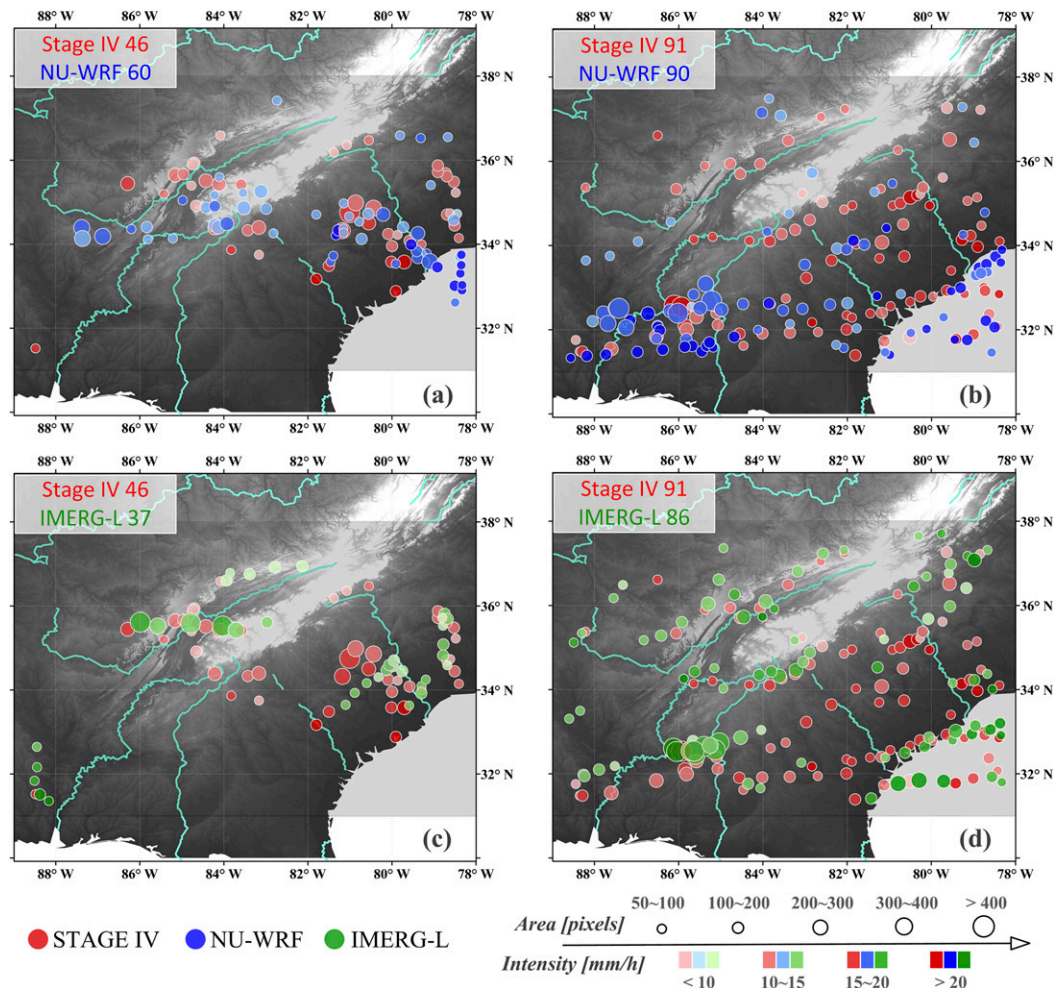
FIG. 5. Location of the centroids of hourly precipitation objects compared with Stage IV for (a),(b) NU-WRF and (c),(d) IMERG-L during TC rainstorm events: (left) event 1 (i.e., Hurricane Matthew) and (right) event 4 (i.e., Hurricane Florence). Circle size and color shading denotes the area and precipitation intensity (represented by P90) of hourly objects, respectively. The numbers of precipitation objects are shown in inset boxes.

MHS and SSMIS, respectively. This high FAR might be caused by the coarse resolutions of MHS and SSMIS scanning foot-prints (Tan et al. 2016). ASMR2 shows the lowest hit ratio of 50% among all PMW sources and a detection skill (with a false ratio of 13% and miss ratio of 38%) similar to combined morphing and IR.

NU-WRF's detection skills show a more complicated relation to PMW (Fig. 11b). Note that what appears as SSMIS in Fig. 11a appears as *F16*, *F17*, and *F18* in Fig. 11b, since the SSMIS sensor flies aboard those three satellites. Generally, NU-WRF presents a higher FAR (also shown in Fig. 8) compared to IMERG-L, especially as *F16*, *F17*, *F18*, and MS (note that MS category might also have incorporated swaths from *F16/F17/F18*), suggesting that assimilated PMW data have benefits in reducing false alarms. This means that NU-WRF objects that incorporate SSMIS tend to have larger FAR ratios than GMI and AMSR2. Unlike IMERG-L, NU-WRF estimates with GMI and AMSR2 present the best detection skills, with a high hit ratio (66% and 67%, respectively) and minor false alarms (9% and 6%, respectively). With the exception of GMI and AMSR2, the benefits of PMW assimilation in NU-WRF seem limited since observations without PMW inputs exhibit a rel-atively high hit ratio of 56% and a relatively small false alarm ratio of 22%, better than *F16*, *F17*, and *F18*.

Similar to the detection statistics above, PMW and IRW data sources have a greater impact on precipitation estimation bias in IMERG-L than NU-WRF (Fig. 12). Underestimation of P90 for IMERG-L objects is eliminated when PMW data are available (Fig. 12a). This does not extend to object area (Fig. 12b), however, most PMW-impacted categories (except GMI) present an obvious positive bias in characterizing object size, while groups without PMW input tend to be less biased. It is also noteworthy that SSMIS has a large bias in estimating object size and rainfall rate. NU-WRF matched precipitation objects exhibit no obvious dependence on PMW in charac-terizing detected storms (Figs. 12c,d).
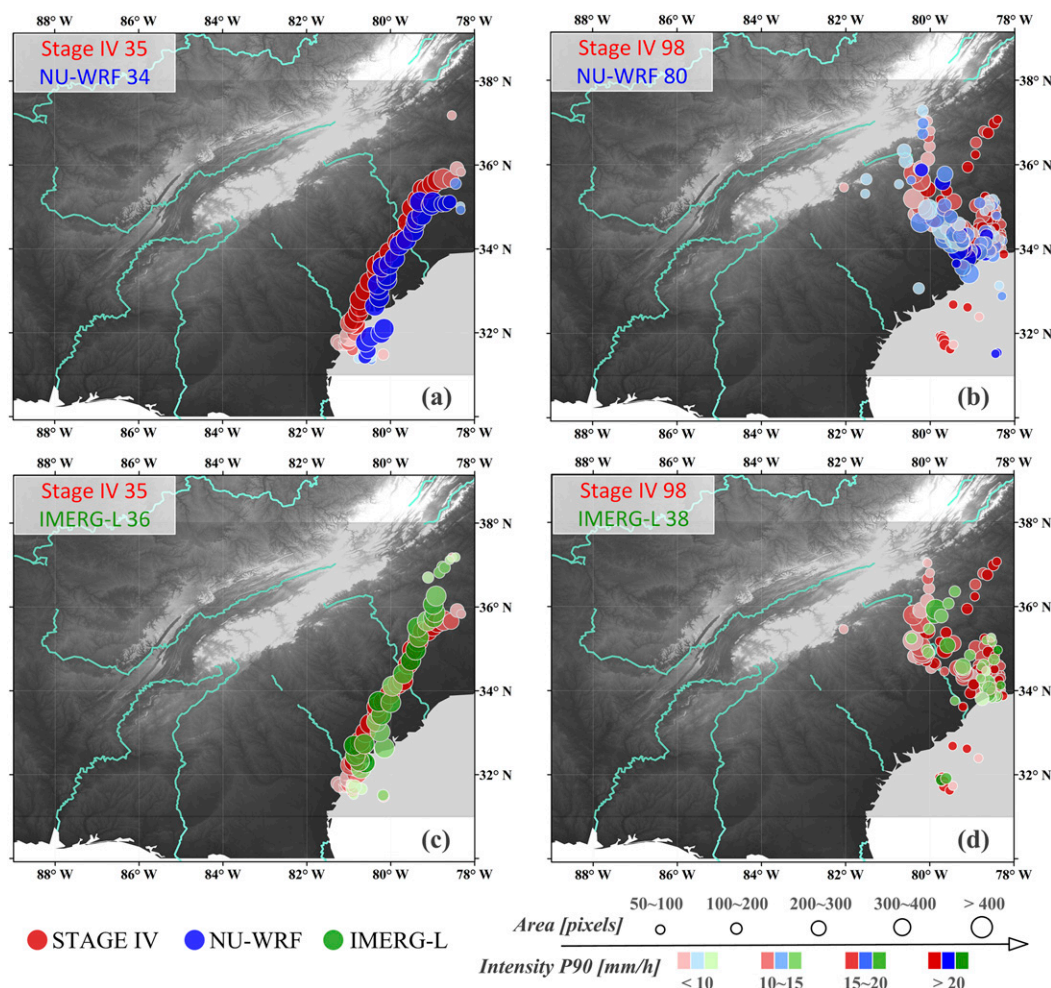
FIG. 6. As in Fig. 5, but for MCS rainstorm events: (a),(c) event 2 and (b),(d) event 3.

As a data-driven approach to estimate precipitation, IMERG relies on input data sources as discussed above. Due to the retrieval algorithms developed for IMERG and its predecessors like TMPA, CMORPH, and PERSIANN, the estimated precipitation spatial pattern can be traced back to the signals contained in input PMW overpasses and IR images. To better understand the relations between PMW–IR inputs and IMERG rainfall patterns, Fig. 13 compares IMERG-L with Stage IV over two consecutive hours in event 3, along with several 30-min input data fields of IRW, IR-only estimated precipitation, PMW sources, and PMW-only estimated precipitation. Compared to Stage IV, IMERG-L overestimates the rainy area in the first hour but reasonably captures its spatial pattern in the next hour. This "update" in estimated spatial pattern was due to the introduction of a new PMW observation into IMERG: the first hourly IMERG-L was greatly impacted by the IR (Fig. 13e; IRW of 39%), as can be seen by comparing Figs. 13c and 13f. In the subsequent hour, there is an AMSR2 overpass (Fig. 13g), and PMW-only estimates (Fig. 13h) dominate the retrieved spatial pattern in the following hourly IMERG-L (Fig. 13d). This highlights that

IMERG estimated precipitation accuracy is highly sensor-specific at fine temporal and spatial scales.

5. Discussion

The object-based evaluation in this study highlights that NU-WRF is capable of capturing extreme rainfall rates, but that these can be subject to substantial displacement errors (i.e., the right rainfall in the wrong place). While the prevalence of displacement errors in NWP have been well documented (Ebert and McBride 2000; Gilleland et al. 2010; Dorninger et al. 2018), its importance in the context of applications is likely growing as NWP pushes toward convection-permitting resolutions (<4 km; Prein et al. 2015). The finescale convection that can be simulated at high resolutions can in principle "unlock" forecasting of small-scale localized natural hazards such as flash floods and landslides. However, we found displacement errors of precipitation objects in NU-WRF on the order of 50–100 km, which could easily mean the difference between a successful and a botched localized forecast. Our analysis suggests that assimilation of satellite radiance data
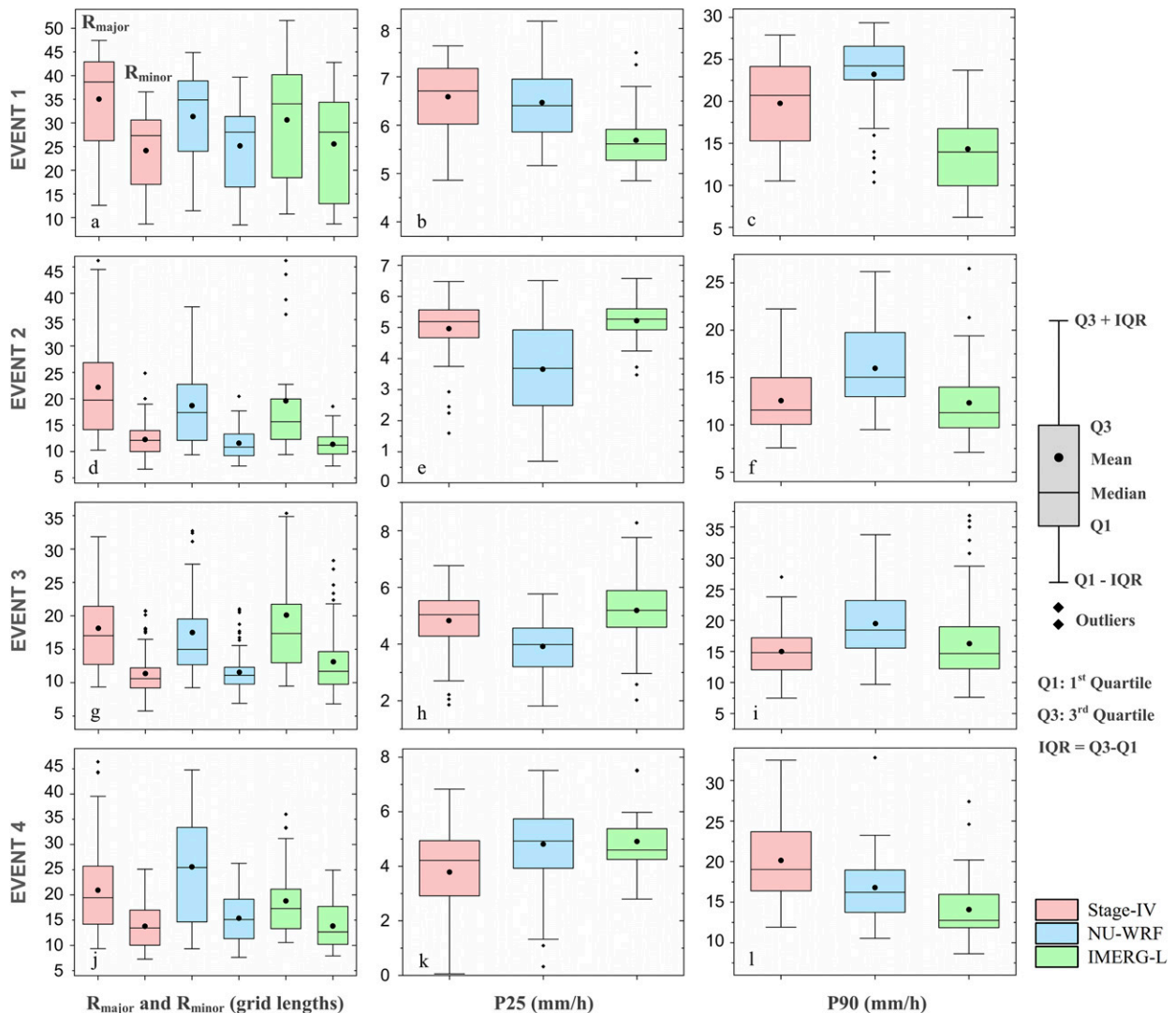
FIG. 7. Boxplots of (left) the lengths of major and minor axes (denoted as R_{major} and R_{minor}), (center) P25, and (right) P90 for Stage IV, IMERG-L, and NU-WRF identified precipitation objects during the four events.

may not be sufficient to constrain these displacement errors for local-scale forecasting.

As demonstrated in this study, data-driven approaches hold potential to capture the spatial features (e.g., the location and shape; Fig. 9) of rainstorms. An alternative to direct assimilation into NWP model, therefore, could be to directly leverage the skills of data-driven satellite precipitation estimation algorithms such as IMERG to "locate" storms. These locations could then be combined with rain rate estimates from convection-permitting NWP. Even relatively simple methods to implement NWP-based correction of satellite precipitation estimates can result in significant improvement of heavy rainfall estimation (X. Zhang et al. 2013; Nikolopoulos et al. 2015).

The object-based evaluation method used in this study can potentially help "bridge" satellite precipitation estimates and input PMW and IR sources. As the data-driven approach, IMERG's performance is found to rely more on the input

PMW and IR information (Figs. 11 and 12) than NU-WRF estimates. As demonstrated in Fig. 13, retrieved spatial patterns in IMERG can be traced back to the availability of PMW swaths and the weight of incorporated IR. This could move beyond the pixel scale investigation as in Tan et al. (2016), and further to facilitate a complete understanding of uncertainty sources of IMERG by considering extra constraints from the swathes and footprints (e.g., "effective resolutions"; Guilloteau et al. 2017) of PMW sensors. We plan to explore this idea and employ this object-based evaluation framework regionally using the recently available long-term IMERG records.

Historically, the satellite precipitation community has placed substantial emphasis on how to distinguish pixel-scale systematic and random error (e.g., Maggioni et al. 2016b; Tian et al. 2013; Wright et al. 2017). Our results suggest that the reality is more complex and that object-based evaluation can
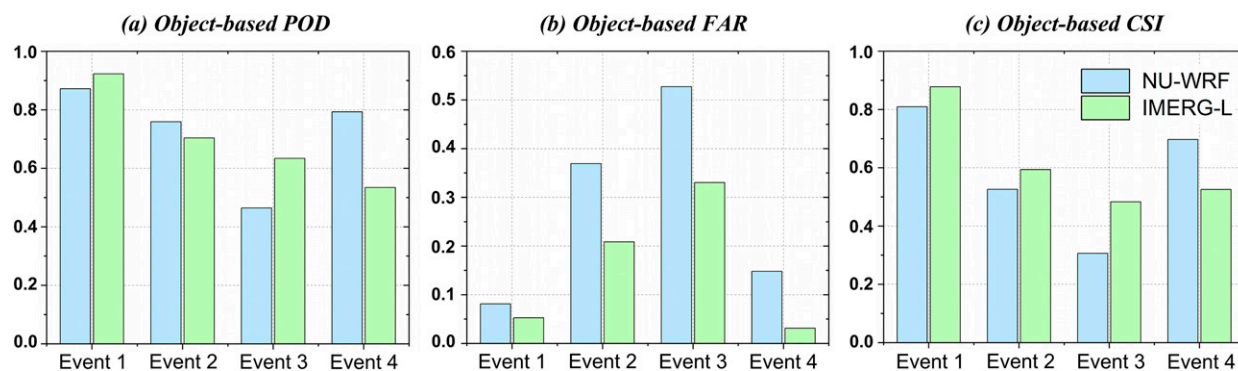
FIG. 8. Object-based detection skills: (a) POD, (b) FAR, and (c) CSI for IMERG-L and NU-WRF in the four events.

provide additional insights, which can help to disentangle errors in location and precipitation intensity. Comparing Fig. 4 to Figs. 5 and 6, for example, suggests that random errors in NU-WRF, particularly for the MCS events, are driven by displacement errors, reflecting the "double penalty" described in section 1.

This object identification and matching approach depends on the selected convolving radius, rainfall thresholds and

matching rules, as mentioned in section 3b. The choice of a convolving-disk radius mainly impacts on the number of objects, which varies inversely with the convolving radius. In this study, the radius of five grid lengths (i.e., 0.5°) was chosen by visual inspection since it appeared to include most localized rainstorm elements. Rainfall thresholds will influence the number, size and shape of identified precipitation objects, and thus will determine the scale at which heavy rainfall can be
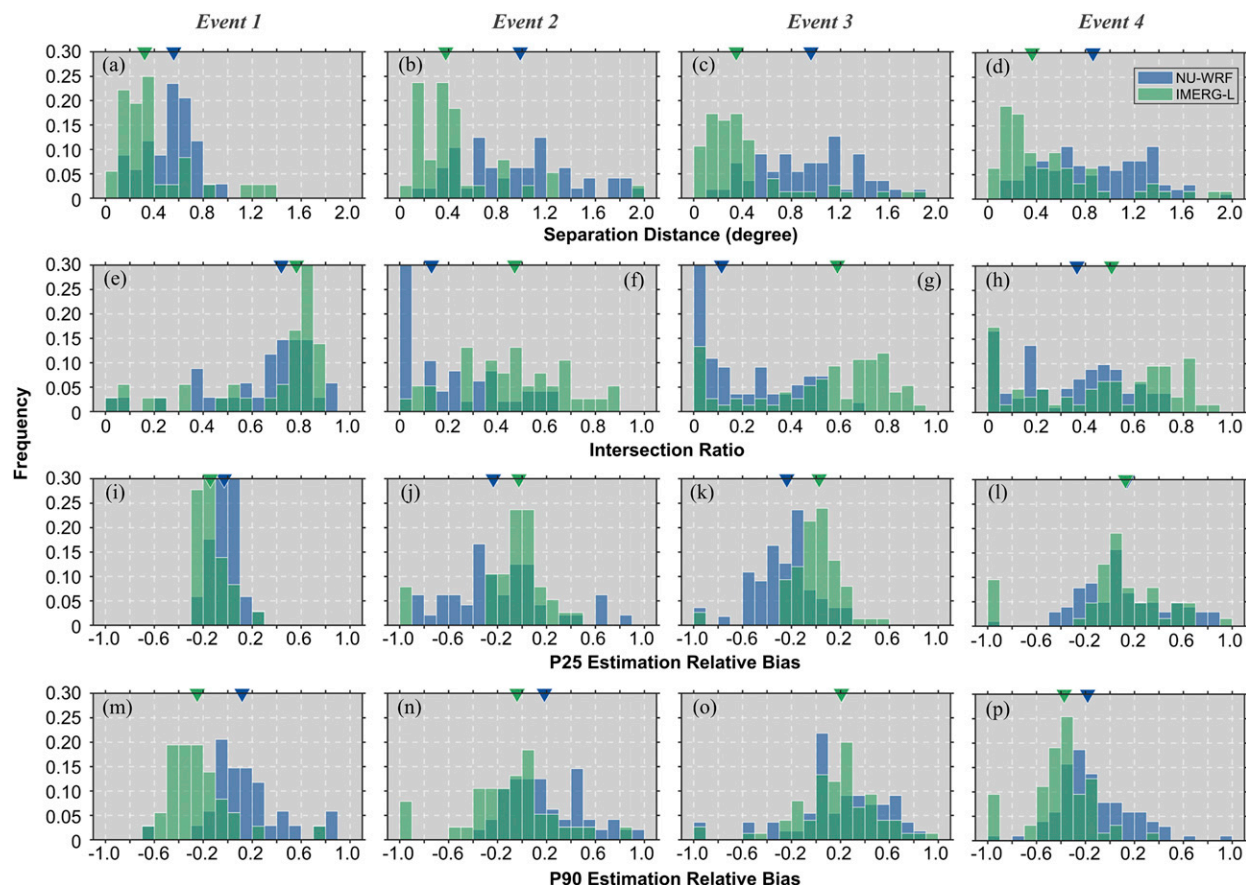

FIG. 9. The comparison of histograms for (a) separation distance (°), (b) intersection ratio, (c) P25 estimation relative bias, and (d) P90 estimation relative bias, of matched precipitation objects identified by IMERG-L and NU-WRF in the four events. Top triangles show the median value of the distribution.
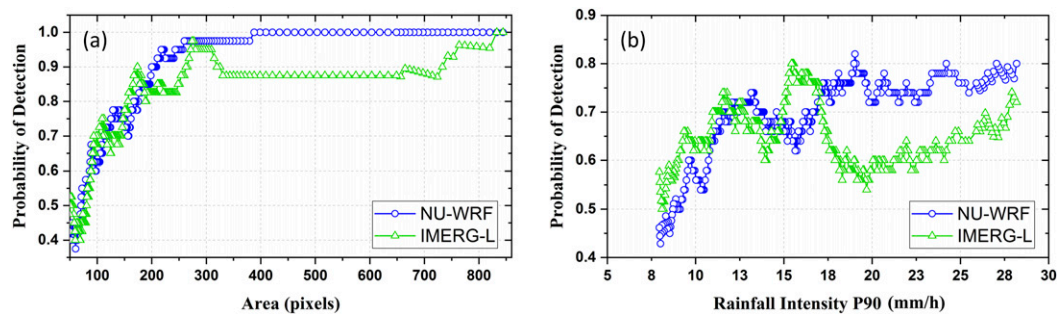
FIG. 10. The detection skill of NU-WRF and IMERG-L varies as a function of (a) the size and (b) P90 of precipitation objects that identified by Stage IV. A 20-neighbor moving average filter was performed to smooth the curves.

characterized. Therefore, thresholds should be tuned according to the dominant extreme weather systems (Dixon and Wiener 1993), precipitation climatology (Chang et al. 2016), and particular application purposes (Li et al. 2014; Morin et al. 2006). In addition to the $5 \, \text{mm h}^{-1}$ threshold applied in this study, we briefly examined object-based metrics stemming from other thresholds (e.g., 1, 2, 3, and $4 \, \text{mm h}^{-1}$). The results (not shown) suggest that our major findings still hold and that the relative skills of IMERG-L and NU-WRF in characterizing storms remain unchanged. However, due to the multiscale nature of extreme rainfall, a multi-threshold object-based analysis framework should be considered in the future to adequately explore performance under different heavy rainfall mechanisms. We selected the least restrictive matching rule as discussed in Davis et al. (2006), in recognition of the potential for large displacement errors in NU-WRF.

It should be also worth noting that the Stage IV data used as ground reference in this study contains their own errors. This is particularly true in mountainous regions such as our Southern Appalachians study area (Fig. 1) which pose specific challenges for radar measurement (Erlingis et al. 2018). Nelson et al. (2016) showed that Stage IV tends to overestimate light to moderate rainfall while slightly underestimated heavy rainfall around the study area in the summer and fall. This conditional bias probably suggests that Stage IV-based P90 (right column in Fig. 7) may be underestimated slightly, but its impact on the number and shape of identified objects is likely limited due to the relatively high threshold used. Prat and Nelson (2015) showed that Stage IV's detection skill decreases for some rainfall extremes due to the low-level orographic enhancement that cannot be detected by operational radars (Barros and Arulraj 2020). This implies that Stage IV may have missed some localized storms in the mountainous area (i.e., the north in Fig. 5), though it is likely that the larger storm elements, both in terms of spatial extent and magnitude, were detected.
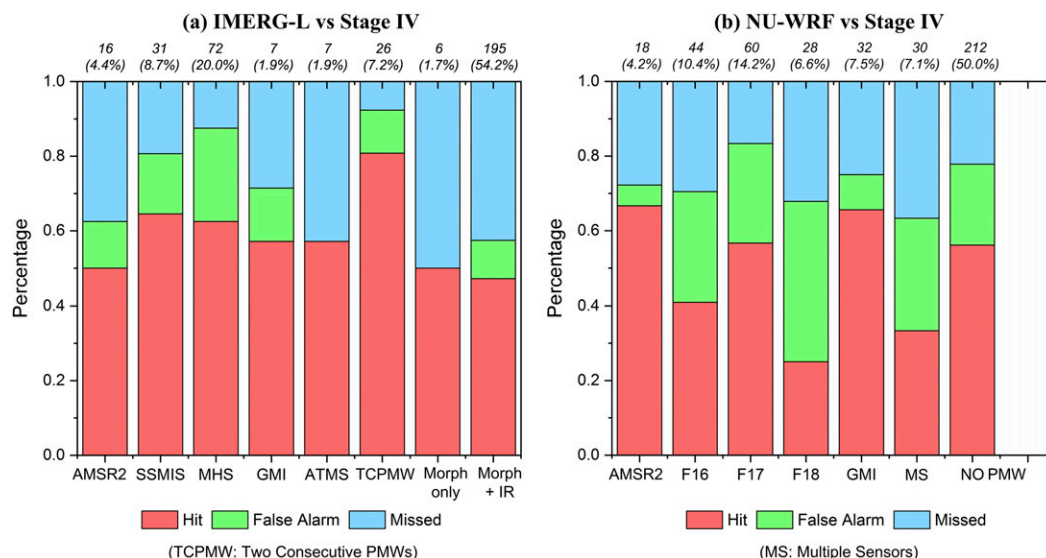

FIG. 11. Statistics of the hit (matched), missed, and false alarm objects conditioned on data sources for IMERG-L and NU-WRF. The number on the top of each bar indicates the total number of objects for each data source category and its percentage.
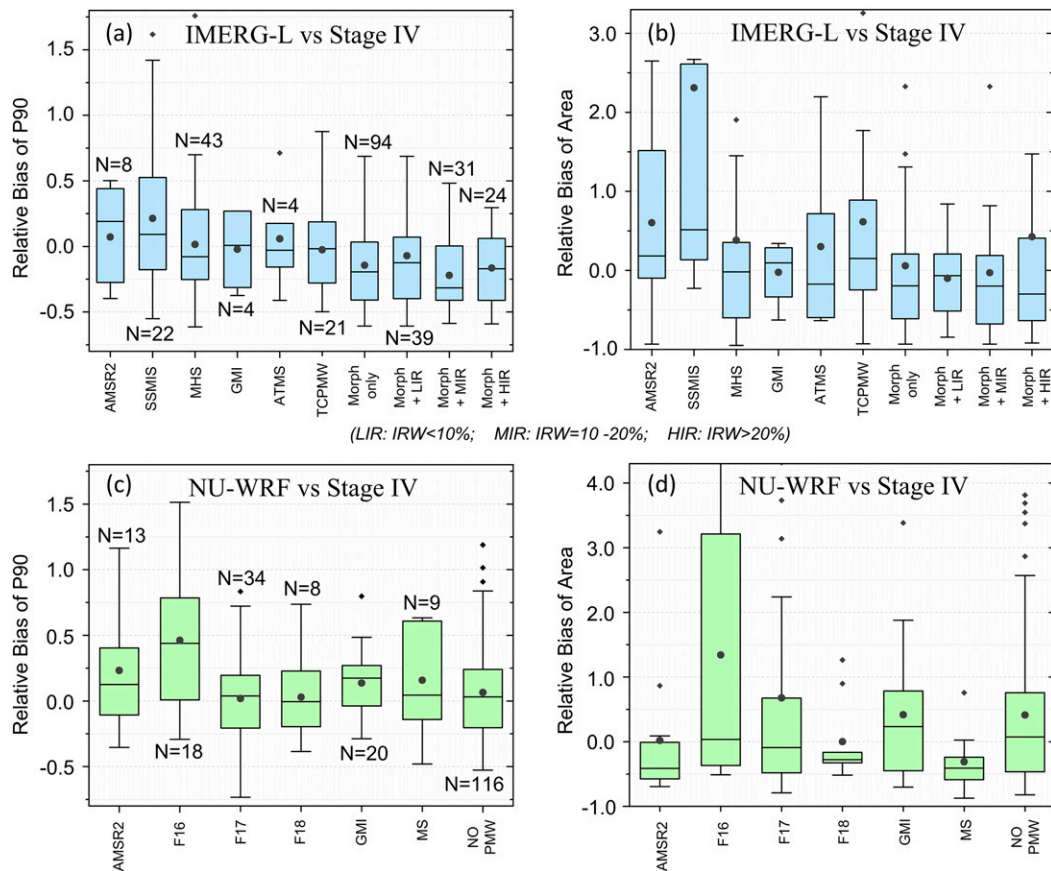
FIG. 12. Boxplots of the relative estimation bias of (a),(c) P90 and (b),(d) area in matched precipitation objects for IMERG-L and NU-WRF, categorized by different PMW sources and IRW. Number N indicates the sample size of matched objects in each category.

6. Summary and conclusions

In this study, we compare gridded precipitation estimates in four extreme storms from the "late" version of IMERG multisatellite merged dataset and from NU-WRF, a numerical weather prediction (NWP) model that assimilates satellite radiances. We argue that the two approaches represent two fundamental ways of producing gridded precipitation estimates from multisatellite observations. We refer to IMERG as data-driven, as it is built upon a variety of retrieval algorithms that relate multiple satellite-observed radiances to precipitation rates, and refer to the output from NU-WRF as physically based, since it assimilates the radiances into the model's microphysics scheme and dynamical equations to simulate precipitation.

Both satellite multisensor datasets such as IMERG and NWP are being pushed to increasingly high spatial and temporal resolutions. These are certainly welcome advances, in part since they unlock new potential applications such as localized monitoring and forecasting of flash floods and landslides. Nevertheless, this presents challenges for the "conventional" satellite precipitation analysis, since even small errors in the locations or sizes of convective cells can mask otherwise good performance of a particular precipitation

dataset. In this regard, we argue that it is necessary to shift the satellite precipitation evaluation paradigm toward object-based approaches, especially since future precipitation estimates are expected to continue to increase in spatial and temporal resolution.

As an example, in this study we move beyond the conventional grid-by-grid manner that has been mainly adopted in existing works to evaluate satellite precipitation datasets, to compare the relative performance of IMERG and NU-WRF using an object-based analysis framework. This framework enables the decomposition of gridded precipitation fields into separated storm objects, and provides a way to diagnose spatial feature-related errors (such as displacement errors), trace them across space and time, and connect their accuracy to input data sources and storm types. Major findings are summarized as follows:

1) Both IMERG and NU-WRF can generally capture the spatial patterns of storm total rainfall in the four rainstorms. Notwithstanding the small sample size of only four events, performance depends on storm type: NU-WRF tends to typically overestimate precipitation during MCSs, while IMERG seriously underestimates peak precipitation rates during TCs.
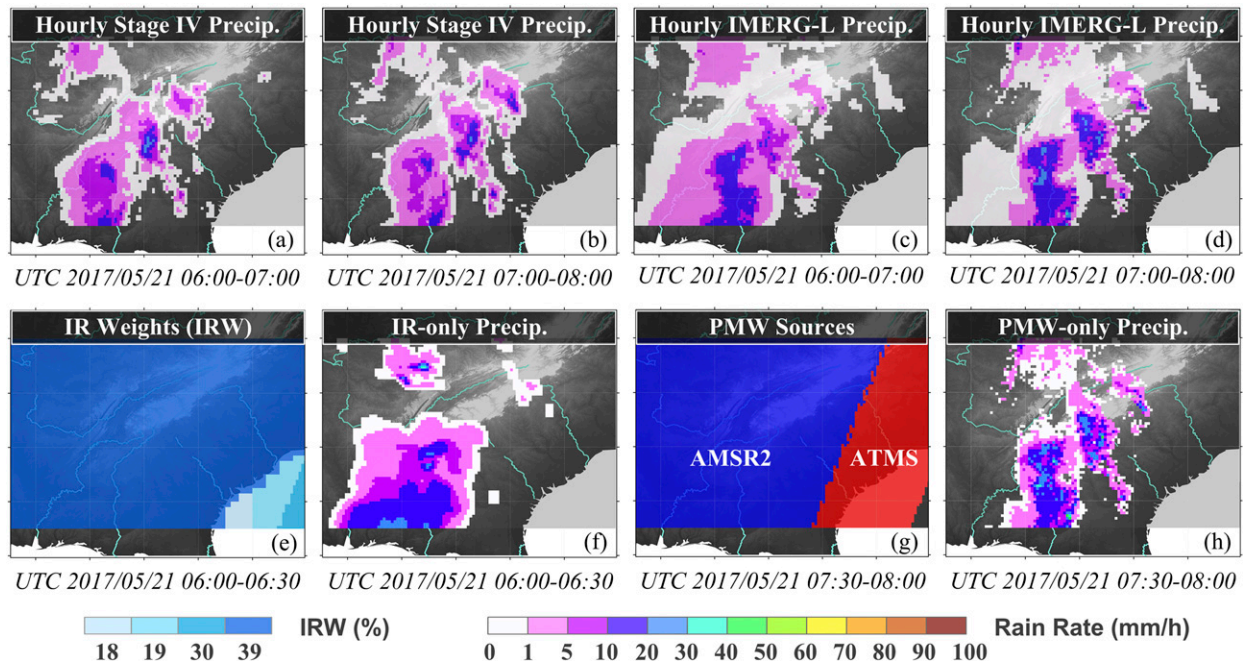
FIG. 13. Comparison of hourly precipitation estimates from (a),(b) Stage IV and (c),(d) IMERG-L for 0600–0800 UTC 21 May 2017, with (e) the fields of IRW and (f) IR-only precipitation estimates in the first hour, and (g) PMW sources and (h) PMW-only precipitation estimates in the second hour.

2) By decomposing gridded precipitation fields into separate storm objects, we find that the coherent spatial pattern of TC events is generally retrieved by NU-WRF in terms of the number and intensity of objects albeit with a noticeable displacement error, while IMERG is characterized by underestimation of TC rainfall intensity. The less coherent and highly localized precipitation structure of the MCS events poses challenges to both IMERG and NU-WRF in accurately capturing the location and intensity of storms, with displacement errors and false alarm precipitation objects being particularly prevalent in NU-WRF results.

3) There is no general bias for IMERG and NU-WRF in estimating precipitation object shape and area. NU-WRF tends to underestimate the 25th percentile but overestimate the 90th percentile intensity of storm objects; and in contrast, IMERG can generally capture 25th percentile intensity but underestimates 90th percentile intensity, especially in TC events, suggesting IMERG tends to flatten the distribution of rain rates within the storm objects.

4) NU-WRF shows a similar detection skill as IMERG in terms of object-based POD and CSI. However, NU-WRF generates many false alarm precipitation objects during the MCS events. Once the estimated objects are matched with Stage IV observations, IMERG demonstrates better skill in locating the storms (with shorter separation distances) and in capturing their spatial extents (with larger intersection ratios), while NU-WRF offers better estimates of high intensity (in terms of 90th percentile intensity) within rainy areas.

5) Precipitation object detection skill of IMERG and NU-WRF depends on storm size, and decreases substantially for objects smaller than 200 pixels (approximately $20\,000\,\text{km}^2$). Detection skill of both estimates improves with increasing rainfall intensity, but IMERG shows a dramatic decline as rainfall becomes larger than $18\,\text{mm}\,\text{h}^{-1}$, suggesting the potential limitations of IMERG in capturing the high-intensity rain rates in the storms.

6) As a data-driven method, IMERG's performance shows a stronger dependence on PMW data sources regarding both detection skill and estimation bias. IMERG precipitation objects based purely on PMW data have the highest hit ratio (i.e., are most likely to be detected), while objects without any PMW data show the lowest hit ratio, along with obvious underestimation of 90th percentile intensity of matched objects. SSMIS-impacted IMERG objects present large biases in size and rainfall rate. NU-WRF matched precipitation objects show no obvious dependence on PMW data sources in characterizing detected storms.

Finally, it should be noted that the physics-based versus data-driven distinction used in this study is becoming increasingly blurred—the GPROF retrieval algorithm in IMERG, for example, applies a physical model to establish elements of its priori "lookup database" for different radiometer observations (Kummerow et al. 2015) while the most recent version of IMERG (version 06B; not used in this study) uses atmospheric variables from relatively coarse-resolution numerical models to upgrade its cloud morphing scheme (Huffman et al. 2019). Nevertheless, the complementary performance of data-driven IMERG and physics-driven NU-WRF revealed by our object-based analysis, along with other studies using different

methods (e.g., Nikolopoulos et al. 2015; X. Zhang et al. 2013, 2016; J. Zhang 2018; X. Zhang 2018), suggest that even closer union of the two approaches holds promise for the future of satellite precipitation estimation.

REFERENCES

AghaKouchak, A., N. Nasrollahi, J. Li, B. Imam, and S. Sorooshian, 2011: Geometrical characterization of precipitation patterns. *J. Hydrometeor.*, **12**, 274–285, https://doi.org/10.1175/2010JHM1298.1.

Amidror, I., 2002: Scattered data interpolation methods for electronic imaging systems: A survey. *J. Electron. Imaging*, **11**, 157, https://doi.org/10.1117/1.1455013.

Ashouri, H., K.-L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Amer. Meteor. Soc.*, **96**, 69–83, https://doi.org/10.1175/BAMS-D-13-00068.1.

Barros, A. P., and M. Arulraj, 2020: Remote sensing of orographic precipitation. *Satellite Precipitation Measurement*, V. Levizzani et al., Eds., Advances in Global Change Research, Vol. 69, Springer International Publishing, 559–582.

——, and Coauthors, 2014: NASA GPM—Ground validation: Integrated precipitation and hydrology experiment 2014 science plan. NASA Tech. Rep., 64 pp., https://doi.org/10.7924/G8CC0XMR.

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

Benjamin, S. G., J. M. Brown, G. Brunet, P. Lynch, K. Saito, and T. W. Schlatter, 2019: 100 years of progress in forecasting and NWP applications. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial, Meteor. Monogr.*, No. 59, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0020.1.

Chambon, P., S. Q. Zhang, A. Y. Hou, M. Zupanski, and S. Cheung, 2014: Assessing the impact of pre-GPM microwave precipitation observations in the Goddard WRF ensemble data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **140**, 1219–1235, https://doi.org/10.1002/qj.2215.

Chang, W., M. L. Stein, J. Wang, V. R. Kotamarthi, and E. J. Moyer, 2016: Changes in spatiotemporal precipitation patterns in changing climate conditions. *J. Climate*, **29**, 8355–8376, https://doi.org/10.1175/JCLI-D-15-0844.1.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

Demaria, E. M. C., D. A. Rodriguez, E. E. Ebert, P. Salio, F. Su, and J. B. Valdes, 2011: Evaluation of mesoscale convective systems in South America using multiple satellite products and an object-based approach. *J. Geophys. Res.*, **116**, D08103, https://doi.org/10.1029/2010JD015157.

Dixon, M., and G. Wiener, 1993: TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology. *J. Atmos. Oceanic Technol.*, **10**, 785–797, https://doi.org/10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2.

Dorninger, M., E. Gilleland, B. Casati, M. P. Mittermaier, E. E. Ebert, B. G. Brown, and L. J. Wilson, 2018: The setup of the MesoVICT project. *Bull. Amer. Meteor. Soc.*, **99**, 1887–1906, https://doi.org/10.1175/BAMS-D-17-0164.1.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, **239**, 179–202, https://doi.org/10.1016/S0022-1694(00)00343-7.

Erlingis, J. M., J. J. Gourley, P.-E. Kirstetter, E. N. Anagnostou, J. Kalogiros, M. N. Anagnostou, and W. Petersen, 2018: Evaluation of operational and experimental precipitation algorithms and microphysical insights during IPHEx. *J. Hydrometeor.*, **19**, 113–125, https://doi.org/10.1175/JHM-D-17-0080.1.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Gilleland, E., 2013: Two-dimensional kernel smoothing: Using the R package smoothie. NCAR Tech. Note NCAR/TN-502+STR, 24 pp., http://doi.org/10.5065/D61834G2.

——, 2019: SpatialVx: Spatial forecast verification, version 0.7. R package, https://cran.r-project.org/package=SpatialVx.

——, D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430, https://doi.org/10.1175/2009WAF2222269.1.

——, D. A. Ahijevych, B. G. Brown, and E. E. Ebert, 2010: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.*, **91**, 1365–1376, https://doi.org/10.1175/2010BAMS2819.1.

Guilloteau, C., E. Foufoula-Georgiou, and C. D. Kummerow, 2017: Global multiscale evaluation of satellite passive microwave retrieval of precipitation during the TRMM and GPM eras: Effective resolution and regional diagnostics for future algorithm development. *J. Hydrometeor.*, **18**, 3051–3070, https://doi.org/10.1175/JHM-D-17-0087.1.

Hersbach, H., and Coauthors, 2018: Operational global reanalysis: Progress, future directions and synergies with NWP. ERA Rep. Series 27, 63 pp., https://doi.org/10.21957/tkic6g3wm.

Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1.

Hsu, K., X. Gao, S. Sorooshian, and H. V. Gupta, 1997: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *J. Appl. Meteor.*, **36**, 1176–1190, https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.

Huffman, G. J., and Coauthors, 2018: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 5.2, 35 pp., https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V5.2_0.pdf.

——, and Coauthors, 2019: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 6, 34 pp., https://gpm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf.

Joyce, R., J. Janowiak, and G. Huffman, 2001: Latitudinally and seasonally dependent zenith-angle corrections for geostationary satellite IR brightness temperatures. *J. Appl. Meteor.*, **40**, 689–703, https://doi.org/10.1175/1520-0450(2001)040<0689:LASDZA>2.0.CO;2.

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

Kirschbaum, D. B., and Coauthors, 2017: NASA's remotely sensed precipitation: A reservoir for applications users. *Bull. Amer. Meteor. Soc.*, **98**, 1169–1184, https://doi.org/10.1175/BAMS-D-15-00296.1.

Kubota, T., and Coauthors, 2007: Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation. *IEEE Trans. Geosci. Remote Sens.*, **45**, 2259–2275, https://doi.org/10.1109/TGRS.2007.895337.

Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, and G. Petty, 2001: The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor.*, **40**, 1801–1820, https://doi.org/10.1175/1520-0450(2001)040<1801:TEOTGP>2.0.CO;2.

Kummerow, C. D., D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic, 2015: The evolution of the Goddard profiling algorithm to a fully parametric scheme. *J. Atmos. Oceanic Technol.*, **32**, 2265–2280, https://doi.org/10.1175/JTECH-D-15-0039.1.

Lee, H., D. E. Waliser, R. Ferraro, T. Iguchi, C. D. Peters-Lidard, B. Tian, P. C. Loikith, and D. B. Wright, 2017: Evaluating hourly rainfall characteristics over the U.S. Great Plains in dynamically downscaled climate model simulations using NASA-Unified WRF. *J. Geophys. Res. Atmos.*, **122**, 7371–7384, https://doi.org/10.1002/2017JD026564.

Lettenmaier, D. P., D. Alsdorf, J. Dozier, G. J. Huffman, M. Pan, and E. F. Wood, 2015: Inroads of remote sensing into hydrologic science during the WRR era. *Water Resour. Res.*, **51**, 7309–7342, https://doi.org/10.1002/2015WR017616.

Li, J., K.-L. Hsu, A. AghaKouchak, and S. Sorooshian, 2016: Object-based assessment of satellite precipitation products. *Remote Sens.*, **8**, 547, https://doi.org/10.3390/rs8070547.

Li, Z., D. Yang, Y. Hong, J. Zhang, and Y. Qi, 2014: Characterizing spatiotemporal variations of hourly rainfall by gauge and radar in the mountainous three gorges region. *J. Appl. Meteor. Climatol.*, **53**, 873–889, https://doi.org/10.1175/JAMC-D-13-0277.1.

Lin, Y., 2011: GCIP/EOP surface: Precipitation NCEP/EMC 4KM gridded data (GRIB) stage IV data. Version 1.0, UCAR/NCAR EOL, accessed 1 October 2018, https://doi.org/10.5065/d6pg1qdd.

Lundquist, J., M. Hughes, E. Gutmann, and S. Kapnick, 2019: Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. *Bull. Amer. Meteor. Soc.*, **100**, 2473–2490, https://doi.org/10.1175/BAMS-D-19-0001.1.

Maggioni, V., P. C. Meyers, and M. D. Robinson, 2016a: A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. *J. Hydrometeor.*, **17**, 1101–1117, https://doi.org/10.1175/JHM-D-15-0190.1.

——, M. R. P. Sapiano, and R. F. Adler, 2016b: Estimating uncertainties in high-resolution satellite precipitation products: Systematic or random error? *J. Hydrometeor.*, **17**, 1119–1129, https://doi.org/10.1175/JHM-D-15-0094.1.

Mahoney, K., and Coauthors, 2016: Understanding the role of atmospheric rivers in heavy precipitation in the southeast United States. *Mon. Wea. Rev.*, **144**, 1617–1632, https://doi.org/10.1175/MWR-D-15-0279.1.

Matsui, T., and Coauthors, 2014: Introducing multisensor satellite radiance-based evaluation for regional Earth System modeling. *J. Geophys. Res. Atmos.*, **119**, 8450–8475, https://doi.org/10.1002/2013JD021424.

Moore, B. J., K. M. Mahoney, E. M. Sukovich, R. Cifelli, and T. M. Hamill, 2015: Climatology and environmental characteristics of extreme precipitation events in the southeastern United States. *Mon. Wea. Rev.*, **143**, 718–741, https://doi.org/10.1175/MWR-D-14-00065.1.

Morin, E., D. C. Goodrich, R. A. Maddox, X. Gao, H. V. Gupta, and S. Sorooshian, 2006: Spatial patterns in thunderstorm rainfall events and their coupling with watershed hydrological response. *Adv. Water Resour.*, **29**, 843–860, https://doi.org/10.1016/j.advwatres.2005.07.014.

Nelson, B. R., O. P. Prat, D.-J. Seo, and E. Habib, 2016: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons. *Wea. Forecasting*, **31**, 371–394, https://doi.org/10.1175/WAF-D-14-00112.1.

Nikolopoulos, E. I., N. S. Bartsotas, E. N. Anagnostou, and G. Kallos, 2015: Using high-resolution numerical weather Forecasts to improve remotely sensed rainfall estimates: The case of the 2013 Colorado flash flood. *J. Hydrometeor.*, **16**, 1742–1751, https://doi.org/10.1175/JHM-D-14-0207.1.

Peters-Lidard, C. D., and Coauthors, 2007: High-performance Earth system modeling with NASA/GSFC's land information system. *Innovation Syst. Software Eng.*, **3**, 157–165, https://doi.org/10.1007/s11334-007-0028-x.

——, and Coauthors, 2015: Integrated modeling of aerosol, cloud, precipitation and land processes at satellite-resolved scales. *Environ. Modell. Software*, **67**, 149–159, https://doi.org/10.1016/j.envsoft.2015.01.007.

Prat, O. P., and B. R. Nelson, 2015: Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002–2012). *Hydrol. Earth Syst. Sci.*, **19**, 2037–2056, https://doi.org/10.5194/hess-19-2037-2015.

Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, **53**, 323–361, https://doi.org/10.1002/2014RG000475.

Rossa, A., P. Nurmi, and E. Ebert, 2008: Overview of methods for the verification of quantitative precipitation forecasts. *Precipitation: Advances in Measurement, Estimation and Prediction*, S. Michaelides, Ed., Springer, 419–452.

Schumacher, R. S., and R. H. Johnson, 2006: Characteristics of U.S. extreme rain events during 1999–2003. *Wea. Forecasting.*, **21**, 69–85, https://doi.org/10.1175/WAF900.1.

Skamarock, W., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., http://doi.org/10.5065/D68S4MVH.

Skofronick-Jackson, G., and Coauthors, 2017: The Global Precipitation Measurement (GPM) mission for science and society. *Bull. Amer. Meteor. Soc.*, **98**, 1679–1695, https://doi.org/10.1175/BAMS-D-15-00306.1.

——, D. Kirschbaum, W. Petersen, G. Huffman, C. Kidd, E. Stocker, 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.

Tan, J., W. A. Petersen, and A. Tokay, 2016: A novel approach to identify sources of errors in IMERG for GPM ground validation. *J. Hydrometeor.*, **17**, 2477–2491, https://doi.org/10.1175/JHM-D-16-0079.1.

——, ——, G. Kirchengast, D. C. Goodrich, and D. B. Wolff, 2018: Evaluation of global precipitation measurement rainfall estimates against three dense gauge networks. *J. Hydrometeor.*, **19**, 517–532, https://doi.org/10.1175/JHM-D-17-0174.1.

Tang, G., Y. Ma, D. Long, L. Zhong, and Y. Hong, 2016: Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *J. Hydrol.*, **533**, 152–167, https://doi.org/10.1016/j.jhydrol.2015.12.008.

Tao, W.-K., and Coauthors, 2014: The Goddard Cumulus Ensemble model (GCE): Improvements and applications for studying precipitation processes. *Atmos. Res.*, **143**, 392–424, https://doi.org/10.1016/j.atmosres.2014.03.005.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

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.

——, G. J. Huffman, R. F. Adler, L. Tang, M. Sapiano, V. Maggioni, and H. Wu, 2013: Modeling errors in daily precipitation measurements: Additive or multiplicative?. *Geophys. Res. Lett.*, **40**, 2060–2065, https://doi.org/10.1002/grl.50320.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463–482, https://doi.org/10.1175/2007MWR2018.1.

Wood, E. F., and Coauthors, 2011: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resour. Res.*, **47**, W05301, https://doi.org/10.1029/2010WR010090.

Wright, D. B., 2018: Rainfall information for global flood modeling. *Global Flood Hazard: Applications in Modeling, Mapping, and Forecasting, Geophys. Monogr.*, Vol. 233, Amer. Geophys. Union, 17–42, https://doi.org/10.1002/9781119217886.ch2.

——, D. B. Kirschbaum, and S. Yatheendradas, 2017: Satellite precipitation characterization, error modeling, and error correction using censored shifted gamma distributions. *J. Hydrometeor.*, **18**, 2801–2815, https://doi.org/10.1175/JHM-D-17-0060.1.

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.

Zhang, J., L.-F. Lin, and R. L. Bras, 2018: Evaluation of the quality of precipitation products: A case study using WRF and IMERG data over the central United States. *J. Hydrometeor.*, **19**, 2007–2020, https://doi.org/10.1175/JHM-D-18-0153.1.

Zhang, S. Q., M. Zupanski, A. Y. Hou, X. Lin, and S. H. Cheung, 2013: Assimilation of precipitation-affected radiances in a cloud-resolving WRF ensemble data assimilation system. *Mon. Wea. Rev.*, **141**, 754–772, https://doi.org/10.1175/MWR-D-12-00055.1.

——, T. Matsui, S. Cheung, M. Zupanski, and C. Peters-Lidard, 2017: Impact of assimilated precipitation-sensitive radiances on the NU-WRF simulation of the West African monsoon. *Mon. Wea. Rev.*, **145**, 3881–3900, https://doi.org/10.1175/MWR-D-16-0389.1.

Zhang, X., E. N. Anagnostou, M. Frediani, S. Solomos, and G. Kallos, 2013: Using NWP simulations in satellite rainfall estimation of heavy precipitation events over mountainous areas. *J. Hydrometeor.*, **14**, 1844–1858, https://doi.org/10.1175/JHM-D-12-0174.1.

——, ——, and H. Vergara, 2016: Hydrologic evaluation of NWP-adjusted CMORPH estimates of hurricane-induced precipitation in the southern Appalachians. *J. Hydrometeor.*, **17**, 1087–1099, https://doi.org/10.1175/JHM-D-15-0088.1.

——, E. Anagnostou, and C. Schwartz, 2018: NWP-based adjustment of IMERG precipitation for flood-inducing complex terrain storms: Evaluation over CONUS. *Remote Sens.*, **10**, 642, https://doi.org/10.3390/rs10040642.

Zupanski, D., S. Q. Zhang, M. Zupanski, A. Y. Hou, and S. H. Cheung, 2011: A prototype WRF-based ensemble data assimilation system for dynamically downscaling satellite precipitation observations. *J. Hydrometeor.*, **12**, 118–134, https://doi.org/10.1175/2010JHM1271.1.