Evaluation of Quantitative Precipitation Estimations through Hydrological Modeling in IFloodS River Basins

HUAN WU
School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, China, and Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland

ROBERT F. ADLER, YUDONG TIAN, AND GUOJUN GU
Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland

GEORGE J. HUFFMAN
NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 20 August 2015, in final form 3 November 2016)

ABSTRACT

A multiple-product-driven hydrologic modeling framework (MMF) is utilized for evaluation of quantitative precipitation estimation (QPE) products, motivated by improving the utility of satellite QPE in global flood modeling. This work addresses the challenge of objectively determining the relative value of various QPEs at river basin/subbasin scales. A reference precipitation dataset is created using a long-term water-balance approach with independent data sources. The intercomparison of nine QPEs and corresponding hydrologic simulations indicates that all products with long-term (2002–13) records have similar merits as over the short-term (April–June 2013) Iowa Flood Studies period. The model performance in calculated streamflow varies approximately linearly with precipitation bias, demonstrating that the model successfully translated the level of precipitation quality to streamflow quality with better streamflow simulations from QPEs with less bias. Phase 2 of the North American Land Data Assimilation System (NLDAS-2) has the best streamflow results for the Iowa–Cedar River basin, with daily and monthly Nash–Sutcliffe coefficients and mean annual bias of 0.81, 0.88, and −2.1%, respectively, for the long-term period. The evaluation also indicates that a further adjustment of NLDAS-2 to form the best precipitation estimation should consider spatial–temporal distribution of bias. The satellite-only products have lower performance (peak and timing) than other products, while simple bias adjustment can intermediately improve the quality of simulated streamflow. The TMPA research product (TMPA-RP; research-quality data) can generate results approaching those of the ground-based products with only monthly gauge-based adjustment to the TMPA real-time product (TMPA-RT; near-real-time data). It is further noted that the streamflow bias is strongly correlated to precipitation bias at various time scales, though other factors may play a role as well, especially on the daily time scale.

1. Introduction
   a. Background

Satellite-based precipitation products have been demonstrated to be valuable in land surface and hydrological modeling for various applications (e.g., Artan et al. 2007; Hong et al. 2007; Kidd and Levizzani 2011; Su et al. 2011; Bitew and Gebremichael 2011a; Kuligowski et al. 2013; Hou et al. 2014; Nikolopoulos et al. 2013).

Corresponding author e-mail: Huan Wu, huanwu@umd.edu

DOI: 10.1175/JHM-D-15-0149.1
© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
runoff-routing model to form the Dominant River
Tracing-Routing Integrated with VIC Environment
(DRIVE) model system. The DRIVE model serves as
the core of the GFMS (http://flood.umd.edu) driven by
the real-time TMPA satellite-based precipitation, rou-
tinely providing global flood information every 3 h at 1/8°
(or ~12 km) resolution. The GFMS has been available
to a wide range of users and has been providing essential
inputs in catastrophe response activities by various hu-
manitarian relief agencies such as the United Nations
World Food Programme (WFP) and International Fed-
eration of Red Cross and Red Crescent Societies.

A global evaluation of the aforementioned DRIVE
hydrologic model of the GFMS has been reported by
W2014, with generally good model performance in
both streamflow simulation and flood event de-
tection. The accuracy of the GFMS in flood pre-
diction is determined by model inputs, particularly
precipitation input, and the DRIVE model physics
and parameterization. However, accurate quantita-
tive precipitation estimation (QPE) remains very
challenging (Kidd and Levizzani 2011). All existing
methods for precipitation information estimation
have both strengths and weaknesses, with significant
uncertainties being reported even in ground-based
radar and gauge observations (e.g., Adam et al. 2006;
Clark and Slater 2006). In situ (gauge) observations
measure precipitation directly, while they have
weakness in spatial coverage and issues of undercatch
during windy weather and occasional loss of information
in extreme (flood) events (Ciach 2003; Refsgaard et al.
2006; Tian et al. 2007; Volkmann et al. 2010). Satellite-
based precipitation has the advantage in spatial
coverage while having limitations in terms of tem-
poral sampling, accuracy of retrieval algorithms, and
spatial resolution (Bell and Kundu 2000; Nijssen and
Lettenmaier 2004). Radar-based products generally have
better accuracy than satellite products, but they have is-
ues in terms of intercalibration to obtain homogeneous
quality for a large domain and the lack of consistent,
unbiased reflectivity–rain-rate (Z–R) relationships (Villarini
and Krajewski 2010).

The recent (28 February 2014) successful launch of the
Global Precipitation Measurement (GPM) Core Observa-
tory, a joint Earth-observing mission (as the follow-on of
TRMM) between NASA and the Japan Aerospace Ex-
ploration Agency (JAXA), provides a broader future for
real-time flood prediction through hydrological mod-
eling using higher-quality satellite-based precipitation
information (Hou et al. 2014; Kirschbaum et al. 2017).
As a new generation of global rainfall and snowfall
observations, the new GPM real-time multisatellite
precipitation product, that is, Integrated Multisatellite
Retrievals for GPM (IMERG), extends the measure-
ment of precipitation range from heavy to moderate
rain attained by TRMM to include light-intensity pre-
cipitation (i.e., <0.5 mm h⁻¹) and cold-season solid
precipitation with more accurate instantaneous pre-
cipitation estimates based on advanced radar and ra-
diometer systems (Hou et al. 2014; Huffman et al.
2015). The expected, significantly better quantification
of precipitation characteristics and more accurate
precipitation products with higher temporal–spatial
resolutions will benefit hydrological research and ap-
plication including the GFMS. This study has a com-
mon purpose with GPM’s Ground Validation (GV)
program with regard to pursuing better understanding
of the strengths and limitations of satellite pre-
cipitation products, in comparison with other available
precipitation estimates, in the context of improving
global hydrologic applications (http://gpm.nsstc.nasa.gov/
ifloods/).

t.

t.

b. Evaluation of QPE products through hydrologic
modeling

Evaluating QPE products (hereafter referred to as
QPEs) is an important part of achieving the best QPE
for climatic and hydrological applications. In some
locales, ground-based validation data can be utilized
to estimate QPE errors, although there is no absolute
truth at a river basin scale because of errors in ground-
based measurements and space and time sampling er-
rors. A physically based hydrological model simulating
the whole process from precipitation to streamflow
with a closed water budget at a river basin scale can be
a useful tool for evaluating precipitation products
(Boone et al. 2004), especially in locations where
quality surface precipitation information is lacking.
Previous studies on the evaluation of satellite pre-
cipitation products using hydrologic models (e.g.,
Artan et al. 2007; Bitew and Gebremichael 2011a,b;
2011; Su et al. 2011; Tong et al. 2014), including our
previous work at global scale (Wu et al. 2012a;
W2014), were mostly focused on evaluating the suit-
ability or utility of satellite precipitation for streamflow
simulation. These studies shared the same strategy
based on hydrological modeling with different QPE in-
puts from both satellite and ground observations. These
analyses were usually based on one or a multiple of the
following comparisons: 1) satellite precipitation versus
gauge data, 2) satellite precipitation–based streamflow
simulation against river gauge–observed streamflow,
and 3) satellite precipitation–based streamflow simulation
against a benchmark simulation usually using gauge
precipitation input.
Hydrologic models also have been involved in understanding precipitation uncertainty impacts on the hydrological applications (Fekete et al. 2004; Decharme and Douville 2006; Renard et al. 2011). For example, uncertainties in precipitation could be magnified in runoff and streamflow in terms of relative change (%) by model activities (e.g., Decharme and Douville 2006; Fekete et al. 2004; Bitew and Gebremichael 2011b).

Hence, the purpose of this study is to evaluate various QPE precipitation products over the Iowa Flood Studies (IFloodS) region with the DRIVE modeling system used in the GFMS. By doing the study over the conventionally instrumented IFloodS domain, we can gain evidence of how streamflow errors might be related to precipitation errors, which is very useful in applications over regions lacking adequate conventional precipitation information, as is the case for many GFMS applications. This study does not intend to conclude which products are superior to the others, either as individual products or as types of products (e.g., conventional vs satellite based), but rather intends to better understand the impact of precipitation errors on the hydrological errors at different space and time scales, at least in one location. The results can lead to a better use of current precipitation information and provide directions for QPE improvement and possibly even for further hydrological model improvement and application, including how to use improved real-time satellite precipitation retrievals in the GFMS.

In the rest of this paper, section 2 describes the study domain; section 3 describes methodology, that is, the evaluation framework and various precipitation products, with the DRIVE model and other datasets used for this study detailed in the appendices; section 4 presents the intercomparison of precipitation products over a long- and short-term period; sections 5 and 6 focus on the intercomparison and analysis of precipitation products and corresponding simulations from multiple perspectives and temporal scales; and sections 7 and 8 present conclusions and future work.

2. Study domain

IFloodS is the first of several hydrology-oriented field efforts of the GPM GV program and was conducted mainly in the Iowa–Cedar River basin (ICRB), located in Iowa (Fig. 1), during the spring (April–June) of 2013. The rivers within the ICRB drain from the northwest to the southeast, and the two major tributaries (the Iowa and Cedar Rivers) meet at Columbus Junction and then enter the Mississippi River at about 65 river miles downstream from Iowa City. A USGS river gauge (05465000) located about 13.9 river miles downstream of the confluence is used as the approximate outlet of the entire ICRB (32,381 km²) for this study. The study area has a climate with cold winters, hot summers, and wet springs. Both of the river basins have relatively simple land use/cover and surface morphology and are dominated by corn–soybean rotation agriculture during April–November, with fallow fields for the rest of the year.

The Iowa River is significantly regulated by the Coralville Reservoir and Dam, whose primary purpose is to reduce flooding along the Iowa River and Mississippi River. The secondary purpose of the dam is for low-flow augmentation by storing and releasing water during dry situations. The Coralville Dam, constructed in the 1950s by the U.S. Army Corps of Engineers, has demonstrated its flood control function in two notable major floods in 1993 and 2008 by reducing the flooding in downstream and allowing several days for flood preparation in downstream locations. The Cedar River, as the primary tributary of the ICRB, has a larger streamflow than the Iowa River at their confluence and is under a relatively more natural status, with only some low head dams on the river. Therefore, more detailed analyses are performed using the Cedar River basin with the location of the USGS gauge 05465000 as its outlet, resulting in an upstream drainage area of 20,057 km².

3. Methodology

The general purpose of this effort is to evaluate various QPE inputs over a well-instrumented region in the context of assessing the performances of the DRIVE model system used routinely, quasi globally in the GFMS. While the DRIVE model is described in detail in W2014, a brief description is provided in appendix A. Instead of using artificial precipitation inputs, we perform extensive hydrological simulations with the DRIVE model forced by a set of existing real precipitation products, including gauge observations and radar-based, satellite-based, and merged products. This multiple-product-driven hydrological modeling framework (MMF) is thus utilized to address the challenge of objectively determining the relative value of various QPEs at river basin/subbasin scales. The results will help us better understand the impact of the various QPEs and their associated errors on hydrological calculations. In this study, the MMF simulations are performed over two time periods: the long-term (2002–13) and the short-term IFloodS (from 1 April to 30 June 2013) periods. The long-term simulations provide a context (model performance, initial condition, etc.) for the short-term simulation and analysis, allowing analyses across temporal scales. The short-term period comparisons are performed at daily and event level.
A set of existing precipitation products is adopted in this study, consisting of TMPA research product and TMPA real-time product (TMPA-RP and TMPA-RT, respectively; Huffman et al. 2007), CMORPH and CMORPH gauge adjusted (CMORPH-adj; Joyce et al. 2004), NMQ/Q2 (or Q2; http://nmq.ou.edu; Zhang et al. 2011), stage IV (Lin and Mitchell 2005; Baldwin and Mitchell 1998), phase 2 of the North American Land Data Assimilation System (NLDAS-2; Mitchell et al. 2004; http://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing_download.php), and CPC Unified (CPC-U; Xie et al. 2007; Chen et al. 2008). An additional radar dataset produced by the Iowa Flood Center (IFC) over the IFloodS period is also used in this study. Table 1 provides a concise summary of these products, including their full names and original spatial and temporal resolutions for the versions used. To reduce the scale effects on the intercomparisons and to be consistent with the current GFMS, all precipitation inputs are prepared to feed the DRIVE model at $\frac{1}{8}^\circ$ resolution and a 3-h time step, while all other DRIVE model inputs are kept the same for each simulation (see appendix B). According to the time period of the datasets, seven of them with the common 12-yr period (2002–13) are used for the long-term simulations, while the two additional radar datasets (i.e., Q2 and IFC) are for the IFloodS short-term simulations and analyses. A 4-yr (from 1 January 1998 to 31 December 2001) DRIVE model simulation using NLDAS-2 is performed first to spin up the model to define the initial condition for all the other long-term simulations. The NLDAS-2-based simulation for 31 March 2013 provides the initial condition for the two radar-based short-term simulations.

The model performances are evaluated primarily through standard metrics including daily and monthly Nash–Sutcliffe coefficient (NSC; Nash and Sutcliffe 1970), correlation coefficients, and mean annual relative error (MARE), based on the simulated and observed time series of streamflow ($\text{m}^3\text{s}^{-1}$). Simulated time series (hydrograph) and water budget components are also
intercompared among simulations with additional independent observations. The multiyear reference precipitation is then derived for the MMF through a water budget analysis method using independent observed streamflow data and an evapotranspiration (ET) dataset (detailed in section 5a). Furthermore, the streamflow bias responding to precipitation bias is quantified at annual, seasonal, and daily scales, which will be useful for the quality evaluation for the GFMS real-time application in areas where precipitation data are lacking.

4. Precipitation intercomparison

a. Long-term precipitation

The gridded mean annual (Fig. 1) and seasonal (Fig. 2) precipitation estimations for the period of 2002–13 show very similar spatial patterns and magnitudes for NLDAS-2, stage IV, and CPC-U. There are also only minor differences in their mean annual precipitation averaged over the entire ICBR, with the values of 870, 868, and 888 mm for CPC-U, stage IV, and NLDAS-2, respectively. However, the satellite-based products show a wide range and have large deviations from the CPC-U (gauge only) estimations. For the mean annual precipitation of the entire ICBR, the CMORPH-adj has the lowest value of 764 mm, while TMPA-RT has the highest estimation of 1207 mm (Fig. 1). Larger differences among products appear in winter than in summer, particularly for the two real-time, satellite-only products. CMORPH has the highest summer estimate with 81% more than CPC-U, but the lowest winter estimate with 96% less than CPC-U (Fig. 2). The TMPA-RT has more precipitation than the gauge observations in both summer (57%) and winter (114%).

Table 1. Summary of precipitation datasets used. A common time period of 2002–13 was defined from all the products for the long-term simulations. A 3-month period (from 1 Apr to 30 Jun 2013) of simulations using Q2 and IFC radar were added for the short-term evaluation. The 4 years (1998–2001) of NLDAS-2 were used to spin up the DRIVE model to initiate all other scenarios.

<table>
<thead>
<tr>
<th>Product</th>
<th>Main sources</th>
<th>Temporal resolution</th>
<th>Spatial resolution</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMPA-RT</td>
<td>Satellite</td>
<td>3 h</td>
<td>0.25°</td>
<td>Huffman et al. (2007)</td>
</tr>
<tr>
<td>CMORPH</td>
<td>Satellite</td>
<td>3 h</td>
<td>0.25°</td>
<td>Joyce et al. (2004)</td>
</tr>
<tr>
<td>TMPA-RP</td>
<td>Satellite and gauge</td>
<td>3 h</td>
<td>0.25°</td>
<td>Huffman et al. (2007)</td>
</tr>
<tr>
<td>CMORPH-adj</td>
<td>Satellite and gauge</td>
<td>3 h</td>
<td>0.25°</td>
<td>ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/CRT/ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/CRT/CMORPH_V1.0/CRT/</td>
</tr>
<tr>
<td>NMQ/Q2</td>
<td>Radar</td>
<td>5 min</td>
<td>0.01°</td>
<td><a href="http://nmq.ou.edu">http://nmq.ou.edu</a>; Zhang et al. (2011)</td>
</tr>
<tr>
<td>Stage IV</td>
<td>Radar and gauge</td>
<td>1 h</td>
<td>4 km</td>
<td>Lin and Mitchell (2005); Baldwin and Mitchell (1998)</td>
</tr>
<tr>
<td>IFC</td>
<td>Radar</td>
<td>1 h</td>
<td>0.004°</td>
<td>—</td>
</tr>
<tr>
<td>NLDAS-2</td>
<td>Gauge, radar, and satellite</td>
<td>1 h</td>
<td>0.125°</td>
<td>Mitchell et al. (2004)</td>
</tr>
<tr>
<td>CPC-U</td>
<td>Gauge</td>
<td>1 day</td>
<td>0.25°</td>
<td>Xie et al. (2007); Chen et al. (2008)</td>
</tr>
</tbody>
</table>

b. The IFloodS period precipitation

A basinwide very wet spring occurred in 2013, when the IFloodS enhanced observations took place, as compared to the climatology estimations by all the products. The total precipitation estimates during the spring of 2013 for CPC-U, NLDAS-2, and TMPA-RT are 551, 546, and 490 mm, respectively (Fig. 3) for the entire Iowa–Cedar River basin, which are much higher than their climatological spring precipitation during the 12-yr record: 349, 356, and 389 mm, respectively. The IFC radar dataset provides the lowest estimate (465 mm), while Q2 has the most precipitation (684 mm). Similar to the long-term datasets, the closest estimations are...
obtained from NLDAS-2 (546 mm), stage IV (546 mm), and CPC-U (551 mm). Consistent with the long-term climatology, TMPA-RT underestimates the precipitation (490 mm) during the period compared to the conventional observations. The relative order of the precipitation amount among all the products is consistent across the river basin from upstream to downstream areas. Q2 provides consistently higher estimation than all other products for all subbasin drainage areas and the entire Iowa–Cedar River basin.

5. Annual and seasonal intercomparison from long-term simulations

a. The reference precipitation dataset and annual-water-budget-based evaluation

Because of the complex nature of precipitation processes, none of the existing precipitation products are perfect or superior to the others at all times or at all places. Instead of designating one of them to be the “truth” to directly evaluate others, a “reference” of mean annual precipitation is constructed from independent sources based on the same multiyear (2002–13) USGS streamflow observations and an ET product from the University of Montana (Zhang et al. 2010). Both the mean annual precipitation and ET are averaged over the upstream drainage area above each river gauge, while the mean annual runoff is calculated as the mean annual discharge divided by the river basin area. Assuming the changes in (soil and surface) water storage are negligible at the mean annual scale for a closed river basin system without significant streamflow diversions or impact of reservoirs (Adam et al. 2006), the mean annual precipitation can be estimated from observed streamflow and calculated ET. The “reference value” of mean annual precipitation (mm) is defined as the sum of the observed mean annual runoff and calculated ET. The upstream basin-area averaged mean annual precipitation is derived from each precipitation product for every gauge location and then compared with the “reference” precipitation to define the precipitation bias in each product. Table 2 shows the 12-yr mean annual water budgets from both observations and the model simulations for the ICRB according to USGS gauge 0546550 with a drainage area of 32,381 km². A mean ET of 627 mm is obtained from Zhang et al. (2010) with a satellite remote-sensing–based ET algorithm to estimate canopy transpiration and soil evaporation using a modified Penman–Monteith approach and open water evaporation using a Priestley–Taylor approach. The observed mean runoff is 298 mm, hence resulting in a reference mean annual precipitation of 925 mm over the entire ICRB.

The uncertainty in the “reference” has to be well understood in this evaluating framework. The streamflow observation is widely accepted to have the least uncertainty, that is, <5% (Oberg and Mueller 2007), among the measurements of all hydrological water budget components. The (satellite) ET measurement usually has more uncertainty than the streamflow while being the dominant part in partitioning the precipitation, meaning that ET is likely the major uncertainty source in the reference precipitation we constructed. Aside from the small variations in the model-simulated ETs with different QPE inputs, Table 2 also shows that the ET is less sensitive to QPE input when the annual precipitation is above 1000 mm. With the highest QPEs...
from the two satellite-only products, the corresponding ET estimation appears to be saturated at 657 mm, which likely approximates the upper bound for the mean annual ET of the river basin with a 4.8% bias against the reference ET. The lowest ET estimated from the DRIVE model is 636 mm using the CMORPH-adj QPE of 764 mm, although we believe CMORPH-adj underestimates the precipitation. Furthermore, considering the fact that more summer soil water stress was experienced in the central United States during the last decade (thus more ET) and both NLDAS-2 and the MODIS product MOD16 (Mu et al. 2011) estimated the mean annual ET of ~600 mm, the reference ET of 627 mm is most likely within ±5% of the truth. The residual (runoff + ET − precipitation) in the modeled water balance as indicated by Table 2 can be attributed to the slight overestimation of ET by the DRIVE model and the small changes in annual water storage. According to the best simulation using NLDAS-2, the residual in the water balance is 6% (of annual precipitation); considering the likely overestimation of ET (~22 mm or ~2.5% of annual precipitation), the uncertainty in the reference precipitation caused by water storage change should be negligible (~3%) in this study, which agrees with the above assumption. Therefore, we have good confidence in the validity of the reference precipitation for this study, with the reference precipitation having an uncertainty of ~5%.

The model simulations using the various observed precipitation products result in the calculated mean annual ET and mean annual outlet streamflow values in Table 2. According to the reference precipitation and observed streamflow, the bias in the simulated mean annual runoff by different products responded very well

FIG. 3. Accumulated precipitation estimations (mm) for IFloodS period from multiple data sources. Numbers are the basin-averaged accumulated precipitation of the ICRB.
to the precipitation bias (Table 2). All ground-based precipitation estimations (i.e., NLDAS-2, CPC-U, and stage IV) with small bias (compared to the reference) tend to produce very good simulations of both runoff and ET compared to the reference. The two satellite-only products with larger positive annual bias lead to the lowest model performance scores with larger bias than all gauge-adjusted products. The model-simulated ET has small variations among simulations (i.e., bias <5%), which indicates that the ET estimation is controlled not only by the wetness but also by the energy availability. This is consistent with Voisin et al. (2008) that evapotranspiration is much less sensitive to precipitation differences than is runoff. The annual water budget analysis provides an overall evaluation of the system and indicates that the model has reasonable estimation of precipitation partitioning at annual scale.

b. Evaluation of the long-term simulations according to in situ discharge observations

The method to construct a “reference” precipitation for the mean annual scale cannot be directly applied to seasonal and finer time scales because of the larger uncertainty from soil moisture and surface water storage. The routing time estimation, seasonal agriculture practice, and other human regulation of the water systems make it more complicated to calculate the precipitation from streamflow and ET observations. Since the NLDAS-2 has the lowest bias (~4%) and the best model performance (Table 2) in the long-term simulation, it is selected to disaggregate the mean annual precipitation “reference value” for finer-scale analysis. The NLDAS-2 is adjusted by simply adding 4% more to its original estimation at every 3-h time step, producing a zero annual bias against the mean annual reference. The negative 4% bias in the NLDAS-2 may be related to the winter undercatch as indicated in the seasonal analysis (section 5e). Hereafter, the adjusted 12-yr NLDAS-2 dataset is referred to as NLDAS-2-Ref, which will be further evaluated throughout the intercomparisons discussed below.

The model performance is then evaluated by comparing the simulated discharge (m³ s⁻¹) to the observations from in situ river gauges. A total of 10 USGS gauges located within the ICRB (Fig. 4) with continuous daily observations during 2002–13 are selected for the evaluation. All these gauges are well geolocated in the dominant-river-tracing (DRT) river network used by the model, according to the same criteria adopted by Alfieri et al. (2013) and W2014, with the difference in drainage area between the DRT and the USGS report being less than 10%.

Large differences exist in the model performance in reproducing the observed discharge, corresponding to various precipitation products (see Fig. 5). The long-term simulations (2002–13) driven by the “reference” dataset NLDAS-2-Ref have the overall best model performance with the highest daily NSE (0.53–0.83) and monthly NSE (0.70–0.90) scores for almost all gauge locations. The same long-term simulations driven by the gauge-only (CPC-U) or gauge-adjusted (NLDAS-2, stage IV, TMPA-RP, and CMORPH-adj) products show better performance than the ones driven by the two satellite-only products, gauged by both NSE and MARE scores (Table 3). As shown in Fig. 4, for the 10 gauges located in the ICRB, the NLDAS-2 yields a mean daily NSE of 0.68 (ranging from 0.51 to 0.81), monthly NSE of 0.83 (0.67–0.88), and MARE of ~0.6% (from ~18% to 23%). At the outlet of the entire ICRB (USGS 05465500), these numbers become 0.81, 0.88, and ~2.1%, respectively.

The two satellite-only products have all negative NSCs and larger MARE (not shown in Fig. 5 but shown in Table 3). However, the two gauge-adjusted satellite products (TMPA-RP and CMORPH-adj) have significantly better statistics for both monthly and daily scale hydrological performance, as previously reported by Pan et al. (2010) and W2014, among others. The model performance scores using the two satellite-only products

---

Table 2. Mean annual water budget (mm), NSC, and MARE from 2002 to 2013 for the entire ICRB (USGS 05465500, close to basin outlet with drainage area of 32 381 km²).

<table>
<thead>
<tr>
<th>Component</th>
<th>Reference</th>
<th>NLDAS-2</th>
<th>CPC-U</th>
<th>Stage IV</th>
<th>TMPA-RP</th>
<th>CMORPH-adj</th>
<th>TMPA-RT</th>
<th>CMORPH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation (mm)</td>
<td>925</td>
<td>888 (-4%)</td>
<td>870 (-5.9%)</td>
<td>868 (-6.2%)</td>
<td>1004 (8.5%)</td>
<td>764 (-17.4%)</td>
<td>1207 (30.5%)</td>
<td>1045 (13%)</td>
</tr>
<tr>
<td>ET (mm)</td>
<td>627 (RS ET)</td>
<td>649 (3.5%)</td>
<td>654 (4.3%)</td>
<td>646 (3.0%)</td>
<td>652 (4.0%)</td>
<td>636 (1.4%)</td>
<td>657 (4.8%)</td>
<td>657 (4.8%)</td>
</tr>
<tr>
<td>Runoff (streamflow at outlet (mm))</td>
<td>298 (USGS)</td>
<td>292 (-2.0%)</td>
<td>259 (-11.3%)</td>
<td>268 (3.4%)</td>
<td>417 (55.5%)</td>
<td>188 (-54.8%)</td>
<td>670 (256%)</td>
<td>447 (-33.3%)</td>
</tr>
<tr>
<td>Residual (mm)</td>
<td>0</td>
<td>55 (6.0%)</td>
<td>43 (4.9%)</td>
<td>46 (5.3%)</td>
<td>65 (6.5%)</td>
<td>60 (7.8%)</td>
<td>120 (9.9%)</td>
<td>59 (5.6%)</td>
</tr>
<tr>
<td>Monthly NSC</td>
<td>—</td>
<td>0.88</td>
<td>0.82</td>
<td>0.81</td>
<td>0.78</td>
<td>0.38</td>
<td>-1.07</td>
<td>-4.22</td>
</tr>
<tr>
<td>MARE</td>
<td>—</td>
<td>-2.1%</td>
<td>-13%</td>
<td>-10%</td>
<td>40%</td>
<td>-37%</td>
<td>125%</td>
<td>50%</td>
</tr>
</tbody>
</table>
are significantly poorer for 2007 and 2010 in that both products dramatically overestimate precipitation. For the years from 2008 to 2013 but excluding 2010, the model with TMPA-RT yields positive daily and monthly NSCs for all gauges, with values ranging from 0.03 to 0.45 and 0.20 to 0.63, respectively.

The streamflow simulations also generally show decreasing accuracy in relation to the QPE types at all seasons according to the left-to-right order illustrated in Table 4. For example, the model performance in terms of daily NSC and bias generally decays from left to right (from ground-based observations to satellite-only products) at the Cedar River outlet (USGS 05465000). Gauge-adjusted products manifest more consistent seasonal NSC performance than the satellite products. All products (except TMPA) significantly underestimate streamflow in winter responding to the underestimations of precipitation, leading to considerable negative effects on spring runoff results. Flood peaks are generally underestimated by NLDAS-2 (Fig. 6), although there is a slight overestimation (1.8%) of the year-round streamflow. The NLDAS-2-Ref corresponds to more overestimation (14%) or larger bias in the overall streamflow; however, it can maintain or slightly increase the daily NSC (from 0.77 to 0.79) as NSC is sensitive to peak simulation. The seasonal performance indicates that CPC-U has the best performance in summer and fall, while it tends to underestimate streamflow during winter and spring, likely due to the underestimation of snow in the gauge-based dataset.

All products tend to have better model performance in terms of NSC scores for downstream gauges with larger drainage areas (Fig. 5). Based on the criteria described in Boone et al. (2004), the model has produced reasonable year-round performance over the whole long-term period with TMPA-RP (0.5 ≤ daily NSC ≤ 0.7) and shown very good performance (daily NSC ≥ 0.7) with NLDAS-2, CPC-U, and stage IV at the outlets of both Iowa River and Cedar River subbasins (Fig. 5 and “Annual” row in Table 4). With the ground-based QPE products, the good statistical performance has been achieved not only at the river basin outlet but also for internal points or subbasins (Figs. 4, 5). The results indicate that the satellite-only precipitation products have substantial room for improvement in hydrological applications and that this type of comparison and validation is valuable. However, TMPA-RP shows much lower skills for summer and fall, suggesting that the monthly adjustment using gauge information does not work well for these two seasons.

c. Temporal evaluation with the long-term hydrographs

Intercomparison of hydrographs can furnish a temporal perspective of the evaluation of precipitation products and their impacts on the accuracy of simulated streamflow. Simulated long-term monthly and daily streamflow time series (or hydrographs) using all the precipitation products are further compared to the observations at each of the 10 USGS gauges. The long-term simulations from all the products (except CMORPH) capture the interannual, seasonal variations in streamflow responding to the precipitation variations. All gauge-adjusted products also show good capability in reproducing daily variations. For instance, at the Cedar River outlet (USGS 05465000), the simulated monthly hydrographs resulting from gauge-only or gauge-adjusted (radar and

---

**Fig. 4.** The daily and monthly NSC scores at the 10 USGS river gauge locations within the ICRB derived by the DRIVE model with NLDAS-2 precipitation inputs. The 1/8° resolution river network in blue was derived by DRT from the HydroSHEDS 1-km river network (black lines). The black dots are the dams in this area.
satellite) products have magnitude and variations very close to the observations (Fig. 6). The monthly hydrograph from TMPA-RT generally covaries with the observations, though overestimation appears in most of the years. The notable 2008 flood that occurred in Iowa is well captured by the DRIVE model at all 10 locations when using all these products (Fig. 6). All gauge or gauge-adjusted products reproduce this extreme event well, especially in terms of the relative magnitude compared to other years. Both satellite-only products capture the 2008 event, with the precipitation overestimated 16% by TMPA-RT (991 mm) and 32% by CMORPH (1122 mm) compared to the reference precipitation of 851 mm over the main rain season (April–September) of the year. However, the satellite-only products have less temporal consistency in quality and produce two false events in 2007 and 2010, respectively (Fig. 6), which are confirmed through the examination of the precipitation products and the simulations. The reasons for the overestimation of rain by both satellite-only products for two such extreme cases is worthy of further careful investigation, including a detailed examination of QPE algorithms, and such an investigation may lead to significant improvements for real-time, satellite-based precipitation products.

d. DRIVE model performance versus precipitation inputs

According to the evaluation results and analyses presented in sections 5a–c, the model performance in terms of streamflow is closely related to the accuracy of precipitation products. The in situ gauge products are generally better than the radar-based products, while the radar products are generally better than the satellite products. Such a relative order of model performance is essentially driven by the areal-averaged precipitation bias at river basin and subbasin scales, which is further investigated in this section.

A total of 80 simulated monthly streamflow time series (or hydrographs) are selected for bias analyses for the 10 gauge locations in the ICRB, with the DRIVE
<table>
<thead>
<tr>
<th>Gauge No.</th>
<th>NLDAS-2-Ref</th>
<th>NLDAS-2</th>
<th>CPC-U</th>
<th>Stage IV</th>
<th>TMPA-RP</th>
<th>CMORPH-adj</th>
<th>TMPA-RT</th>
<th>TMPA-RT&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CMORPH</th>
<th>CMORPH&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Daily NSC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5457700</td>
<td>0.53</td>
<td>0.51</td>
<td>0.47</td>
<td>0.45</td>
<td>0.38</td>
<td>0.21</td>
<td>-0.41</td>
<td>0.22</td>
<td>-0.85</td>
<td>0.17</td>
</tr>
<tr>
<td>5460400</td>
<td>0.56</td>
<td>0.59</td>
<td>0.64</td>
<td>0.45</td>
<td>0.33</td>
<td>0.28</td>
<td>-2.01</td>
<td>0.24</td>
<td>-7.42</td>
<td>-0.15</td>
</tr>
<tr>
<td>5451500</td>
<td>0.73</td>
<td>0.68</td>
<td>0.67</td>
<td>0.61</td>
<td>0.62</td>
<td>0.24</td>
<td>-0.35</td>
<td>0.30</td>
<td>-3.49</td>
<td>-0.03</td>
</tr>
<tr>
<td>5453100</td>
<td>0.70</td>
<td>0.64</td>
<td>0.63</td>
<td>0.59</td>
<td>0.64</td>
<td>0.27</td>
<td>-0.28</td>
<td>0.45</td>
<td>-3.20</td>
<td>-0.01</td>
</tr>
<tr>
<td>5454500</td>
<td>0.64</td>
<td>0.63</td>
<td>0.64</td>
<td>0.62</td>
<td>0.39</td>
<td>0.35</td>
<td>-1.49</td>
<td>0.11</td>
<td>-6.20</td>
<td>-0.13</td>
</tr>
<tr>
<td>5455700</td>
<td>0.78</td>
<td>0.77</td>
<td>0.75</td>
<td>0.75</td>
<td>0.53</td>
<td>0.45</td>
<td>-1.38</td>
<td>0.03</td>
<td>-6.03</td>
<td>-0.00</td>
</tr>
<tr>
<td>5464000</td>
<td>0.64</td>
<td>0.65</td>
<td>0.67</td>
<td>0.57</td>
<td>0.47</td>
<td>0.27</td>
<td>-0.74</td>
<td>0.20</td>
<td>-2.63</td>
<td>0.09</td>
</tr>
<tr>
<td>5464500</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.66</td>
<td>0.51</td>
<td>0.36</td>
<td>-0.76</td>
<td>0.19</td>
<td>-2.69</td>
<td>-0.56</td>
</tr>
<tr>
<td>5465000</td>
<td>0.79</td>
<td>0.77</td>
<td>0.76</td>
<td>0.72</td>
<td>0.62</td>
<td>0.35</td>
<td>-0.70</td>
<td>0.22</td>
<td>-2.78</td>
<td>-0.57</td>
</tr>
<tr>
<td>5465500</td>
<td>0.83</td>
<td>0.81</td>
<td>0.79</td>
<td>0.78</td>
<td>0.62</td>
<td>0.40</td>
<td>-1.09</td>
<td>0.24</td>
<td>-4.38</td>
<td>-0.38</td>
</tr>
<tr>
<td><strong>Monthly NSC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5457700</td>
<td>0.71</td>
<td>0.67</td>
<td>0.53</td>
<td>0.57</td>
<td>0.62</td>
<td>0.15</td>
<td>-0.60</td>
<td>0.25</td>
<td>-1.31</td>
<td>0.15</td>
</tr>
<tr>
<td>5460400</td>
<td>0.72</td>
<td>0.75</td>
<td>0.79</td>
<td>0.49</td>
<td>0.63</td>
<td>0.38</td>
<td>-1.84</td>
<td>0.46</td>
<td>-6.38</td>
<td>0.08</td>
</tr>
<tr>
<td>5451500</td>
<td>0.82</td>
<td>0.76</td>
<td>0.76</td>
<td>0.67</td>
<td>0.75</td>
<td>0.21</td>
<td>-0.58</td>
<td>0.45</td>
<td>-3.99</td>
<td>0.10</td>
</tr>
<tr>
<td>5453100</td>
<td>0.79</td>
<td>0.72</td>
<td>0.70</td>
<td>0.67</td>
<td>0.81</td>
<td>0.31</td>
<td>-0.24</td>
<td>0.63</td>
<td>-2.81</td>
<td>0.12</td>
</tr>
<tr>
<td>5454500</td>
<td>0.84</td>
<td>0.81</td>
<td>0.80</td>
<td>0.79</td>
<td>0.72</td>
<td>0.47</td>
<td>-1.09</td>
<td>0.20</td>
<td>-4.40</td>
<td>0.02</td>
</tr>
<tr>
<td>5455700</td>
<td>0.89</td>
<td>0.85</td>
<td>0.83</td>
<td>0.83</td>
<td>0.76</td>
<td>0.52</td>
<td>-1.14</td>
<td>0.21</td>
<td>-4.62</td>
<td>0.11</td>
</tr>
<tr>
<td>5464000</td>
<td>0.85</td>
<td>0.83</td>
<td>0.80</td>
<td>0.71</td>
<td>0.71</td>
<td>0.29</td>
<td>-0.78</td>
<td>0.40</td>
<td>-2.75</td>
<td>0.30</td>
</tr>
<tr>
<td>5464500</td>
<td>0.86</td>
<td>0.86</td>
<td>0.84</td>
<td>0.79</td>
<td>0.67</td>
<td>0.42</td>
<td>-0.94</td>
<td>0.32</td>
<td>-2.81</td>
<td>0.04</td>
</tr>
<tr>
<td>5465000</td>
<td>0.88</td>
<td>0.87</td>
<td>0.84</td>
<td>0.81</td>
<td>0.72</td>
<td>0.37</td>
<td>-0.85</td>
<td>0.34</td>
<td>-2.78</td>
<td>-0.16</td>
</tr>
<tr>
<td>5465500</td>
<td>0.90</td>
<td>0.88</td>
<td>0.84</td>
<td>0.84</td>
<td>0.73</td>
<td>0.42</td>
<td>-1.20</td>
<td>0.25</td>
<td>-3.99</td>
<td>-0.24</td>
</tr>
<tr>
<td><strong>MARE (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5457700</td>
<td>10</td>
<td>-2</td>
<td>-29</td>
<td>-7</td>
<td>31</td>
<td>-44</td>
<td>101</td>
<td>110</td>
<td>31</td>
<td>34</td>
</tr>
<tr>
<td>5460400</td>
<td>37</td>
<td>23</td>
<td>14</td>
<td>-19</td>
<td>20</td>
<td>-27</td>
<td>127</td>
<td>91</td>
<td>114</td>
<td>41</td>
</tr>
<tr>
<td>5451500</td>
<td>4</td>
<td>-9</td>
<td>-12</td>
<td>-23</td>
<td>32</td>
<td>-47</td>
<td>117</td>
<td>115</td>
<td>47</td>
<td>23</td>
</tr>
<tr>
<td>5453100</td>
<td>-7</td>
<td>-18</td>
<td>-21</td>
<td>-25</td>
<td>27</td>
<td>-44</td>
<td>118</td>
<td>97</td>
<td>48</td>
<td>30</td>
</tr>
<tr>
<td>5454500</td>
<td>9</td>
<td>-4</td>
<td>-10</td>
<td>-8</td>
<td>49</td>
<td>-32</td>
<td>158</td>
<td>131</td>
<td>77</td>
<td>29</td>
</tr>
<tr>
<td>5455700</td>
<td>9</td>
<td>-3</td>
<td>-10</td>
<td>-6</td>
<td>50</td>
<td>-31</td>
<td>164</td>
<td>122</td>
<td>81</td>
<td>50</td>
</tr>
<tr>
<td>5464000</td>
<td>15</td>
<td>2</td>
<td>-13</td>
<td>-13</td>
<td>34</td>
<td>-41</td>
<td>109</td>
<td>89</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>5464500</td>
<td>19</td>
<td>6</td>
<td>-8</td>
<td>-6</td>
<td>46</td>
<td>-36</td>
<td>127</td>
<td>92</td>
<td>48</td>
<td>62</td>
</tr>
<tr>
<td>5465000</td>
<td>14</td>
<td>1</td>
<td>-12</td>
<td>-9</td>
<td>41</td>
<td>-39</td>
<td>118</td>
<td>86</td>
<td>42</td>
<td>54</td>
</tr>
<tr>
<td>5465500</td>
<td>10</td>
<td>-2</td>
<td>-13</td>
<td>-10</td>
<td>40</td>
<td>-37</td>
<td>125</td>
<td>99</td>
<td>50</td>
<td>39</td>
</tr>
</tbody>
</table>

<sup>a</sup> Statistics are from 2008 to 2013, excluding 2010.
model driven by the seven precipitation products (Figs. 1, 2) and the NLDAS-2-Ref. From Fig. 7a, NLDAS-2-Ref, NLDAS-2, and CPC-U generally define the upper boundary of the monthly NSC scores at all places. The resulting NSC scores and bias at the 10 gauge locations can be consistently grouped according to precipitation products, indicating a spatially consistent relative precipitation quality among these products.

The model performance in streamflow simulations changes as a linear function of the bias in the mean annual precipitation (Fig. 7a), indicating that the model performance would be mainly controlled by precipitation bias. A correlation coefficient of ~0.93 (excluding CMORPH) is found between the monthly NSC and the absolute basin-averaged mean annual precipitation bias. The precipitation products with smaller annual bias tend to have smaller streamflow bias and higher NSC scores (Fig. 7). The sign of the annual precipitation bias is reversed (positive and negative) for the SRB and CRM. The resulting NSC scores and bias at the Cedar River outlet responding to the mean annual precipitation bias and mean annual streamflow bias exist for other subbasins. As an example, Fig. 8 depicts the mean annual streamflow bias at the Cedar River outlet responding to the mean annual precipitation bias and mean annual streamflow bias.

e. Precipitation bias versus streamflow bias at annual scale

Streamflow bias shows a strong linear relation to the bias in precipitation at the annual scale, with a correlation coefficient of 0.97. Also, the precipitation bias is amplified about threefold (percentagewise) in the streamflow output (Fig. 7b). This is consistent with the literature (Kobold and Su 2005; Decharme and Douville 2006; Komma et al. 2007; Decharme et al. 2012; Biemans et al. 2009). Similar correlations and slopes between the mean annual precipitation bias and mean annual streamflow bias exist for other subbasins. As an example, Fig. 8a depicts the mean annual streamflow bias at the Cedar River outlet responding to the mean annual precipitation bias.
annual precipitation bias with a correlation coefficient of 0.97.

The satellite-only precipitation estimations tend to introduce larger uncertainties in the hydrologic model results than using gauge-merged products as shown above, which is also consistent with W2014. Figures 8b and 8c illustrate the simulated annual streamflow bias versus the bias in the two satellite-only products. The annual streamflow biases are calculated against the observed data and the simulated streamflow using the NLDAS-2-Ref, respectively; both result in similar correlation relations and slopes relative to the respective precipitation biases in two satellite products.

f. Precipitation bias versus streamflow bias at seasonal and monthly scales

The mean monthly precipitation for the Cedar River basin is calculated from each product and NLDAS-2-Ref for each month of the year (Fig. 9a), along with the biases of precipitation and streamflow (both observed and simulated; Figs. 9b–d). Good correlations between the bias in precipitation and streamflow also exist at seasonal and monthly scales. For the Cedar River basin, a correlation coefficient of 0.81 was found for the mean monthly biases in precipitation and streamflow for all months of the year from all the products (see Fig. 10a). However, the slope of the bias relation between the precipitation and streamflow varies across seasons, and they start to depend on the time lags for runoff generation and the routing process. Figures 10b and 10c show the good correlation between the mean monthly precipitation and streamflow bias in February (with a 1-month lag of precipitation) and August (with no lag). For the two satellite-only products, the generally good linear correlations between the monthly streamflow bias (against observation) and precipitation bias are found for spring and summer, respectively (Fig. 11).

All ground-based observations have small and stable biases against the reference in both precipitation and streamflow across the seasons. TMPA-RT overestimates the precipitation in almost all months (except in April and May) and thus produces consistently higher...
streamflow year-round (Figs. 9c,d). A simple adjustment of the TMPA-RT by a uniform reduction of 25% is used in an additional simulation, which results in noticeable improvement (dashed blue line in Figs. 9c,d), with daily NSC of 0.46 and MARE of ~10% (comparable to stage IV). However, a seasonal correction, particularly regarding the precipitation values in April and May, would result in an even better streamflow result for this area.

6. Daily scale intercomparison (with emphasis on IFloodS)

a. Model performance at daily scale

The long-term (2002–13) evaluation at daily scales suggests that precipitation inputs would cause large variation in model daily skill measures across seasons (Table 4) at all gauge locations within the river basin. Similar to the mean annual and seasonal evaluations, the daily NSC performance is also related to precipitation bias. Figures 12a and 12b clearly show that precipitation products with smaller precipitation biases or root-mean-square errors tend to have higher daily NSC scores, and a good correlation (−0.76) between the precipitation root-mean-square error and the daily NSCs is found at the Cedar River outlet (Fig. 12b).

The MMF shows that model behavior over the 2013 IFloodS period is consistent with the long-term simulations. NLDAS-2, CPC-U, stage IV, and TMPA-RP have better overall daily NSC metrics than the other satellite-based products (Fig. 12c). For individual river gauge locations across the river basin, the best daily NSC
performance for the overall IFloodS period is achieved either by Q2 or NLDAS-2-Ref. NLDAS-2-Ref provides the best performance, with a daily NSC of 0.73 and a bias of $-11\%$ for the outlet of the entire ICRB. The best performance for the downstream locations of the Cedar River and the two upstream locations (before the reservoir) of the Iowa River are achieved by Q2 (Fig. 12e). The reservoir has significant impacts on the streamflow at the daily scale, confirmed by dramatically reduced daily NSC scores and increased bias (for all products) at the immediate downstream location (Figs. 12e,f).

Similar to the long-term simulations, most products tend to have gradually better performance scores for downstream locations, and such a trend is clearly shown along the Cedar River (i.e., river gauges outside of the dashed box in Figs. 12e,f). The daily NSCs based on Q2 consistently increase from 0.02 for the most upstream subbasin ($2816$ km$^2$) to 0.74 for the entire Cedar River basin ($20,057$ km$^2$). Partly as the result of the decrease of precipitation bias for larger downstream drainage areas, TMPA-RT yields less streamflow bias in downstream locations (Fig. 12f); however, the corresponding NSC score does not increase, implying that other factors determining streamflow timing should be further considered (to be discussed in section 6b). The TMPA-RT has the best model performance (except that computed with the reference) at the two upstream gauge locations (Figs. 12e,f) on the Cedar River in terms of both the streamflow bias and daily NSC, with the benefit from the higher values of previous winter precipitation that offset consistent underestimations in April and May (and happened to be within the IFloodS period).

At the Cedar River outlet, Q2-based results track the observed daily hydrograph best, while NLDAS-2, TMPA-RP, CPC-U, and stage IV also perform well (Fig. 13). TMPA-RT can generally capture the first flood event well, while both satellite-only products have less accuracy in magnitude and timing. Most of the products yield lower total streamflow estimates than the

---

**FIG. 9.** For the Cedar River basin according to USGS 05465000, (a) the mean monthly basin-averaged precipitation, (b) the mean monthly basin-averaged precipitation bias (according to NLDAS-2-Ref), and (c),(d) the mean monthly streamflow bias according to observed streamflow and NLDAS-2-Ref-based simulation. The dashed blue line in (c) and (d) is the streamflow simulated using the adjusted TMPA-RT (with 25% less) as input of the model.
observed data at all gauge locations (Figs. 12d,f and Fig. 13), indicating their precipitation underestimation over the period or during the preceding winter. The total underestimation of streamflow over the period is further primarily attributed to the underestimation of major flood events. Most products have significant negative bias (−20% or less) in peak flow for the largest flood events (in June; Fig. 13). However, for the largest flood event, Q2 yields the closest peak day simulation to the observation with a −13% bias (simulated 1477 m$^3$ s$^{-1}$ vs observed 1690 m$^3$ s$^{-1}$), likely because it has 34.6% more precipitation (than the reference) in the previous seven days of the peak date. The IFC radar dataset probably also underestimates the precipitation because of the reported intercalibration issue (Seo et al. 2013).

**b. Precipitation impacts on peak timing**

The hydrographs in Fig. 13 show the flood peak timing difference caused by different precipitation products, which is suggested in the daily NSC performance metrics in the sense that the “better” precipitation products tend to have better performance in peak timing as well. We select the largest flood event during the IFloodS period as an example to evaluate the precipitation impacts on the streamflow and peak timing, as this is also important for the quantification of the relation between streamflow–precipitation biases at daily and event scales (discussed in the next section). The largest flood event during IFloodS reached its peak on 3 June 2013 at the Cedar River outlet, with an observed daily discharge 1690 m$^3$ s$^{-1}$. It was mainly caused by continual rainfall from 17 May to 15 June, with a total basin-averaged amount of 283 mm (by Q2). As seen from Fig. 13, multiple rainfall events contributed to the flood event. During the period, there were 11 days with basinwide rainfall over 10 mm, with the maximum daily rainfall of 41 mm on 20 May and a 6-day (25–30 May) total rainfall maximum of 130 mm.
Lag-correlation analyses between the simulated and observed hydrographs indicate that the largest peak simulated by TMPA-RT and CMORPH is 3 days earlier than the observed peak date for this case, while NLDAS-2, TMPA-RP, and stage IV have 1-day faster peak timing than the observation, and CPC-U and Q2 can produce the same peak day as observed with only subdaily differences (Fig. 14a). Lag-correlation analyses between precipitation and observed streamflow for this event (Fig. 14b) further show that all precipitation products consistently have a 7- or 8-day lag time for this largest spring flood event (only NLDAS-2 and Q2 shown in Fig. 14b). However, in contrast to the analyses based on the observations, the calculated lags between the precipitation products and the resulting streamflow through the DRIVE model exhibit significant differences. Q2 and corresponding streamflow computed with the model have a 7-day lag (closest to the observation), while for NLDAS-2-Ref (and stage IV and CPC-U, not shown) the lag is 5 days. The model results using TMPA-RT and CMORPH have a 4-day time lag (Fig. 14b). This tends to suggest that, in addition to precipitation bias, other factors, including spatial variation and resolution (e.g., Q2 is at 0.01° while TMPA is at 0.25°, Table 1), might also strongly impact the model performance specifically in simulating peak timing and routing (e.g., Smith et al. 2004).

c. The precipitation–streamflow bias relation at daily and event scales

Precipitation bias has a good correlation with the simulated streamflow bias during the short-term IFloodS period as well. According to the total of 90 pairs of precipitation and streamflow time series from the simulations using the nine products for the 10 river gauge locations, the precipitation bias over the entire IFloodS 3-month period has a correlation coefficient of 0.85 with the bias in corresponding simulated streamflow (not shown). However, the daily streamflow bias tends to have very weak correlation with the bias in precipitation accumulated from a lead-up period shorter than the lag time, that is, the time between maximum precipitation and maximum streamflow. To quantify how the bias in daily streamflow or peak flow of a flood event may relate to the precipitation bias, a reasonable time period prior to the event has to be considered. Based on the lag time analysis, Figs. 15a and 15b suggest a strong positive relation between the peak flow bias and the bias in the antecedent 7-day precipitation for both the first and the largest flood events during the IFloodS period. The 7-day lag time determined here is consistent with Krajewski and Mantilla (2010). However, such a relation and the slope can be significantly impacted by the antecedent water storage condition of the river basin. For the first flood event, TMPA-RT showed a slight overestimation (6%) of the peak that was caused by the overestimation of precipitation in the winter season in spite of a 39% underestimation in the 7-day accumulated precipitation prior to the peak date (Fig. 15a).

To identify the streamflow–precipitation bias relation for any phase of a flood event (especially useful for a real-time flood system), a longer routing time period than the lag time between the maximum precipitation and flood peak has to be considered. Given that the source of much of the streamflow error is a fast runoff response (Nijssen and Lettenmaier 2004), and in order
to accurately relate the flood streamflow bias to the bias in recent precipitation events, which are usually the dominant water sources for flood, the number of antecedent days of precipitation is determined by checking the continuous correlation coefficient between the daily fast flow (only from surface runoff) bias and the bias in precipitation over antecedent days (Fig. 14c).

Figure 14c shows the correlation coefficient between the bias in antecedent accumulated precipitation by TMPA-RT (reference to NLDAS-2-Ref) and the bias in fast flow and slow flow (reference to the DRIVE simulation using NLDAS-2-Ref), respectively, for different seasons. The fast-flow and slow-flow components are derived by separately routing surface runoff and
subsurface runoff from VIC. As shown in Fig. 14c, the fast flow and slow flow (bias) respond to the antecedent precipitation (bias) at different temporal scales and vary across seasons. Based on the blue dashed line in Fig. 14c (for fast flow in spring), we determine a 13-day period (about twice the lag time of the basin) for quantifying the relation between the antecedent precipitation bias and the daily streamflow bias during the largest flood event (from 26 May to 15 June). Figure 14c also shows large flexibility for the selection of the number of antecedent days to achieve good correlation between precipitation and streamflow bias. Essentially, the choice of the number of antecedent days just needs to be close to the river basin routing time. For the Cedar River basin, it usually takes about 13 days for most of the basinwide rainfall runoff to reach its outlet.

The bias in simulated daily streamflow using the two satellite-only products is well correlated to the bias in the corresponding 13-day antecedent precipitation for the largest flood event, with correlation coefficients of 0.84 and 0.87 by TMPA-RT and CMORPH, respectively (Figs. 15c,d). This indicates that the bias in daily streamflow can be well predicted by the bias in antecedent precipitation and vice versa, although the bias relation slope seems to be different for the two products. Such linear relations of biases between streamflow and precipitation at subcatchment scale might be a useful component in the grid-based adjustment of satellite-only precipitation products for streamflow-related hydrological applications. At least, such a relation can provide GFMS users with information about possible underestimation (or overestimation) caused by satellite precipitation, when only historic river gauge data or reference simulations are available. However, it is noted that the precipitation–streamflow bias relation could get more complicated in winter as snow or snow–rainfall mixed events would prolong the runoff-routing process. Low correlation coefficients ($<0.4$) between fast-flow bias and accumulated precipitation bias are found for any length of antecedent period (Fig. 14c) during winter.

7. Conclusions

A GFMS has been routinely providing publically accessible flood information based on a hydrologic model driven by real-time satellite precipitation input (http://flood.umd.edu/). This study follows our previous work (W2014), which describes the GFMS, its associated
DRIVE model, and an evaluation of the system on the global basis. In this study, an MMF is further built. By using multiple QPEs, including observed ground-based and satellite precipitation products, as inputs to drive the DRIVE model, detailed evaluations of the MMF results are then made over a well-instrumented area (Iowa–Cedar River basin) and before and during the IFloodS field experiment. As part of the effort, the MMF utilizes a relatively long climatological period (2002–13) and a water budget approach using observed streamflow and a satellite-based ET estimate to construct a precipitation reference. A 3-hourly gridded reference precipitation product (NLDAS-2-Ref) is also formed by just adjusting the NLDAS-2 product by 4% (zeroing the bias compared to the mean annual reference), so that it could be applied at any time scale. The model was run at $\frac{1}{8}^\circ$ spatial resolution and 3-hourly time step (the same scheme as is used globally in real-time) for a long-term period (2002–13) and the IFloodS period (from 1 April to 30 June 2013), respectively.

The bias in simulated streamflow for the long-term (2002–13) period using NLDAS-2 is very small (~2%) based on the comparison with the observed streamflow at the ICRB outlet. With the reference precipitation dataset (NLDAS-2 + 4%) based on the independent water budget calculation, the long-term streamflow bias is still small (+10%) and the monthly and daily NSC values reach their peaks. Therefore, the DRIVE model, with no further adjustments or calibrations, is deemed capable of being used for the hydrologic evaluation of the QPEs. The use of the DRIVE model in this way is important, since it can provide a direct connection to its global application in the GFMS. Based on the results of the evaluation over the Iowa–Cedar River basin, the following conclusions can thus be made:

![Diagram](image-url)
1) Surface-based precipitation products generate the best results as a group, followed by the satellite products adjusted by surface information and then the satellite-only estimations. The QPE products (i.e., NLDAS-2, CPC-U, and stage IV), which have less bias, consistently lead to better streamflow simulations at annual, seasonal, daily, and event scales across the river basin, not just at the basin outlet, with good performance (daily NSC ≥ 0.5) for upstream subbasins and very good performance (daily NSC > 0.7) for locations close to the river mouth. The NLDAS-2 is overall the best QPE for the ICRB. Using NLDAS-2 with the lowest precipitation bias (−2.4%), the DRIVE model displays the best daily NSC, monthly NSC, and MARE scores of 0.81, 0.88, and −2.1%, respectively, for the outlet of the ICRB for the long-term simulations (2002–13). However, given that the NLDAS-2-Ref can only improve the model performance in spring and winter compared to the original NLDAS-2, a further adjustment of NLDAS-2 should consider the spatial–temporal distribution of bias. The other ground-based QPEs (CPC-U and stage IV) also perform very well.

2) The satellite-only products have lower performance than other products with negative NSCs for the overall long-term period and with temporal inconsistency. After excluding the two false extreme events, the model with TMPA-RT inputs has positive daily and monthly NSCs at all gauge locations within the river basin, suggesting the value of such precipitation inputs. Satellite products with some sort of gauge adjustment display an intermediate quality of results. The satellite product (TMPA-RP) adjusted by monthly gauge information can generate results approaching those from the ground-based products. This is true even for the daily NSC value, suggesting that the daily satellite precipitation values even only adjusted on the monthly scale can provide quality information on precipitation variations on the daily time scale. The CMORPH-adj product, which
uses a more complex adjustment process, can greatly improve the precipitation statistics of the satellite-only product. These results indicate the likely potential for the improvement of real-time satellite QPE products.

3) Streamflow bias is positively correlated to precipitation bias from annual to daily time scales. At the annual scale, precipitation bias is generally amplified in streamflow bias with a factor of 2–4. However, at finer time scales, including the monthly scale, the best correlation coefficient and the slope of the streamflow–precipitation bias relation depend on season, the individual product, river basin concentration (routing) time, and antecedent river basin water storage condition.

4) For the IFloodS period, the ground-based QPE products produce the best results in terms of calculated versus observed streamflow, with the Q2 product (only evaluated for this short period) edging out the NLDAS-2 product for best scores. The TMPA-RP with monthly gauge adjustment again can nearly match the NSC and bias values derived from the best ground-based precipitation information. The satellite-only QPEs, however, do relatively poorly for this short period, significantly underestimating the biggest flood event in early June.

5) Precipitation products also have large variations in peak timing calculations. Compared to the ground-based products, the TMPA-RT and CMORPH have a 4-day time lag, leading to an estimated flood wave for the largest event 2–3 days earlier than observed during IFloodS. Spatial–temporal resolution and other characteristics of precipitation products can impact the simulation of peak timing.

The evaluation framework developed in this study can be applied to different river basins at continental and global scales, depending on the availability of reliable observed data. It is possible that the QPE product that can be regarded as (or adjusted to be) the best estimation for a river basin could vary across regions and hence needs to be identified objectively. The quantification of the response of streamflow to the precipitation uncertainty can give more insights to other river basins at similar latitude and interior continental locations, with similar hydroclimate conditions of humid continental zone with generally hot summers, cold and dry winters, and wet springs. The bias relationship can provide useful information for GFMS users about possible underestimation or overestimation for ongoing flood events when real-time validation information is not available.

This paper is mainly focused on the bias of precipitation that may propagate into simulated streamflow. The impact of other precipitation characteristics on streamflow, such as precipitation intensity, location, intermittency, phase, and scale will be further investigated, as they are partly responsible for unexplained biases and timing errors in streamflow. Additional information on soil moisture, for example, from the Soil Moisture Active Passive (SMAP) mission, and water storage, for example, from the Gravity Recovery and Climate Experiment (GRACE) mission, will be valuable for such studies because they provide more information on the water budget.

Acknowledgments. This research is partially supported by NASA’s Precipitation Measurement Mission (PMM) program (Award NNX16AF19G). We thank Drs. Pingping Xie (NOAA/NCEP/CPC), Felipe Quintero Duque (University of Iowa), and Dr. Jun Dong (University of Maryland) for assisting in accessing and processing of CPC-U gridded gauge data, IFC, and Q2 radar data, respectively.

APPENDIX A

The DRIVE Model

While the DRIVE model is described in detail in W2014, a brief description is given in this section. Runoff generation and runoff routing are the two fundamental modules required in a hydrologic model for simulating the natural processes from precipitation to streamflow and floods. The DRIVE model thus couples VIC (for runoff generation) and a DRT-based routing (DRTR) model (for runoff routing). VIC solves for full water and energy balances with good skill for water budget estimation, based on its proven methodology, addressing both rainfall- and snowmelt-dominated runoff generation processes and soil frost dynamics (e.g., Storck et al. 2002; Cherkauer and Lettenmaier 1999; Christensen et al. 2004; Hamlet and Lettenmaier 2007; Hamlet et al. 2010; Elsner et al. 2010; Peters-Lidard et al. 2011; Wu et al. 2012b). VIC delineates the subgrid heterogeneity of infiltration capacity based on statistical variable infiltration curves (Zhao and Liu 1995) for fractional subgrid areas with different land-cover types and elevation bands. Therefore, VIC serves as a good water budget (or runoff generation) simulator in the DRIVE model. The DRTR model is a gridded and physically based routing model, which applies the kinematic wave equations to both dominant rivers at grid level and tributaries at subgrid level. The hierarchical DRT method (Wu et al. 2011, 2012c) not only provides a routing model with a full package of scale-consistent hydrographic parameters, for example, flow direction,
river network, drainage area, flow distance, slope, etc., but also supplies strong subgrid parameterization, which is critical for hydrological modeling at relatively coarser resolutions.

To better approximate the natural hierarchical runoff propagation, the DRTR model integrates the grid-level drainage network with numerical schemes based on the Strahler ordering system (Strahler 1957), which helps to improve computing efficiency. VIC has been modified from its original individual gridcell-based mode to match the DRTR model structure with all gridcell calculations completed at each time step in the Strahler order-based sequence. To be coupled with VIC, the routing model takes the VIC-estimated runoff to feed the routing scheme at each time step. In the version of the DRIVE model used for this study, the interactions between VIC and DRTR are set at the grid scale. Before the runoff enters the lateral routing process, the various subgrid runoff components resulting from the vertical processes in VIC are aggregated to a grid-scale output (for the routing model) at the end of each of VIC’s time steps.

APPENDIX B

Model Setup

This study uses the same soil and vegetation parameters at 1/8° resolution as in W2014. The hydrographic data, for example, flow direction, drainage area, flow length, channel width, channel slope, overland slope, flow fraction, and river order, are derived by the DRT algorithms (Wu et al. 2011, 2012b) based on the HydroSHEDS (Lehner et al. 2008) global 1-km baseline hydrographic data. Therefore, a priori datasets of the soil, vegetation, and hydrographic parameters were developed for this study. Other forcing data (i.e., air temperature and wind speed) are taken from the NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al. 2011).

It was reported in W2014 that the fixed value of 0.03 for the channel roughness tends to lead to faster flood waves for some cases. As discharge and stage increase, more roughness materials and obstacles (debris, tree roots, brush, weeds, etc.) are typically incorporated into the submerged area and the roughness coefficient values increase (Powell 1978; Pitlo 1982; Soong et al. 2012). The complex relationship between roughness and vegetation can be seen as the roughness varies substantially with depth (or discharge; Powell 1978). In this study, instead of using a fixed value at all times, we let the roughness coefficient vary with the magnitude of the flood streamflow with a simple linear relation, that is, when the simulated streamflow at a time step is greater than the flood threshold (representing the bank-full discharge defined by W2014), the roughness is calculated as

\[ n_i = n_{i-1} + n_{i-1}(Q_i - Q_{i-1})/Q_{i-1} \]

and

\[ n_i \leq 0.1, \]

where \( n_i \) and \( n_{i-1} \) are the roughness coefficients and \( Q_i \) and \( Q_{i-1} \) are the simulated discharge for the current and last time steps, respectively. We do not perform a local model calibration in this study based on the results of the long-term simulations (described in result sections) and the belief that the uncertainty in the model input data (particularly the precipitation) and human activity impacts have to be well addressed prior to the further diagnosis of the model itself (model structure and parameterization; Decharme and Douville 2006; Voisin et al. 2008; Biemans et al. 2009; Seo et al. 2013; Nikolopoulos et al. 2013; W2014).

REFERENCES


