manuscript submitted to *Journal of Hydrology*

Hydrologic Evaluation of the Global Precipitation Measurement Mission over the U.S.: Error Budget Analysis

Devon Woods[1,2,4], Pierre-Emmanuel Kirstetter[1,3,4], Humberto Vergara[2,4], Jorge A. Duarte[2,4], and Jeffrey Basara[1,3]

[1]School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, USA

[2]Cooperative Institute for Severe and High-Impact Weather Research and Operations, Norman, OK, USA

[3]School of Meteorology, University of Oklahoma, Norman, OK, USA

[4]NOAA/National Severe Storms Laboratory, Norman, OK, USA

Corresponding author: Pierre Kirstetter (pierre.kirstetter@noaa.gov) and Devon Woods (woods.devon.j@ou.edu)

Abstract

This study investigates the hydrologic utility of satellite precipitation estimates from the Global Precipitation Measurement mission by comparing flood signals produced across the Continental United States by a ten-year span of in-situ, ground-based radar and satellite-based precipitation data. The flood characteristics generated with radar and satellite precipitation through a distributed hydrologic model are contrasted against reference stream gauge data as a method of integrated validation to assess and quantify error budgets between precipitation products by highlighting precipitation products' accuracy, hydrologic scaling effects, and the impact of the hydrologic model. It is found that systematic and random errors associated with flood characteristics behave similarly to trends previously seen in precipitation rate errors between precipitation products, establishing a clear link through propagation of errors into the water cycle. Additionally, behaviors associated with both water balance and routing schemes within the hydrologic model were shown to affect outputs. Errors generated by water balance tend to cause overestimation of peak discharge values, while errors associated with routing tend to cause underestimation of flood durations and push flood timings earlier than the stream gauge reference.

Plain Language Summary

This study investigates how effectively rainfall estimates from the Global Precipitation Measurement mission can generate models of floods observed by stream gauges across the Continental United States. By comparing these modeled floods to actual gauge data, assessments can be made regarding the overall trends in error associated with the rainfall products themselves, the hydrologic model used, and the scales at which these errors are detected the most. It is found that, overall, the trends in hydrologic error between the products behave similarly to previously

established errors in rainfall between products, showing a clear link as these errors move through the water cycle. The analysis also found that different components of the hydrologic model itself can affect the characteristics of the floods modeled, with one tending to cause overestimation of flood peaks and the other leading to underestimation of flood durations.

1 Introduction

In research and operations alike, hydrologic models are the keystone for flood assessment, understanding, and forecasting. This remains especially true in the realm of flash floods, with one well-known model being the Ensemble Framework for Flash Flood Forecasting (Flamig et al., 2020) or EF5, an open-source distributed hydrologic modeling framework. To date, EF5 has been established in tandem with the Multi-Radar Multi-Sensor (MRMS) system (Zhang et al., 2016) to build an operational flash flood forecasting network over the CONUS: the Flooded Locations And Simulated Hydrographs (FLASH) system (Gourley et al., 2017). The MRMS network of 176 ground-based radars provides high-quality precipitation data at a spatial resolution of 1-km and temporal resolutions as low as 2 minutes, with FLASH subsequently operating at 1-km spatial and 10-minute temporal.

The same boast cannot be said across most of the world, however. Without reliable radar coverage, researchers and forecasters instead turn to satellite precipitation products, such as those provided through the Global Precipitation Measurement mission (GPM). This program generates a global dataset of precipitation at half-hourly temporal and 0.1-degree spatial resolution, from 90N to 90S latitude, through use of the Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm Version 6 (Huffman et al., 2014). Great lengths of research have been undertaken to assess and intercompare satellite precipitation product returns to those provided by ground-based

products (Gebregiorgis et al., 2018; Kirstetter et al., 2012; Kirstetter et al., 2020; Derin et al., 2021; Derin and Kirstetter, 2022), but until recently less has been done to forward the need for "integrated hydrologic validation" of GPM (Hou et al., 2014). A foray into this was made in Woods et al. (2023) where MRMS and IMERG were used as precipitation forcings through EF5, and their extracted flood characteristics were directly compared. This approach also took heed to answer calls put forward in the greater hydrologic community, premier of which by Clark et al. (2021), to assess hydrologic models and hydrograph outputs through new methods less reliant on "bulk metrics", as these traditional approaches become increasingly limited when expressed simultaneously over large sample sizes and more diverse ranges of catchment and flood characteristics (Clark et al., 2021; Lamontagne et al., 2020; Nanding et al., 2021; Newman et al., 2015).

The research put forth here continues this premise, but with the addition of observational flood data provided by the United States Geological Survey (USGS) as a benchmark. As such, focus can now be shifted from initial relative assessment of the products to a more objective and in-depth analysis of error trends and model behaviors. Error budgets and analyses have been done previously between precipitation products (satellite and ground-based), but again have focused less on how this propagates further into the water cycle. This information in the literature, however, can still provide valuable insights towards what to expect from a more hydrology-focused error budget. For example, studies have consistently highlighted increasing underestimation and random error in estimates of satellite precipitation products at higher reference rain rates (Kirstetter et al., 2013; Kirstetter et al., 2014; Uphadyaya et al., 2020). Links have also been shown between errors generated by IMERG precipitation and errors in the performance of streamflow simulations when compared to observations at basin scales (e.g. Hartke et al., 2023, investigating six years of data

over Iowa), so by association there are already grounds for significant propagation of errors into the hydrologic system and subsequent flood characteristics, especially at the continental modeling scale.

This study seeks to build upon the results and assessments made in Woods et al. (2023) and bring them fully into the context of on-ground observations. The quality-controlled selection of gauged USGS basins provides an unprecedented look at model behaviors across the entire CONUS at once, as opposed to basin or region-scale studies. Additionally, the results of this research not only aim to better understand the appearance and root causes of water cycle-related simulation errors but also better inform algorithm developers and end-users alike about potential ways to mitigate for and model these errors. This is especially important to undertake with both precipitation products operating at their native resolutions, helping to establish clear benchmarks in behavior without having to account for resampling. The approach put forth here and in Woods et al. (2023) is novel in its ability to assess these precipitation products on their capability to model distinct signals of features associated with floods (i.e. peak magnitude, flood duration, and event timing) as opposed to directly comparing streamflow time series. Results from this process serve to provide more robust and tangible information regarding the behavior of these products when held up against observed reference data.

The rest of the paper is organized as follows: Section 2 describes the dataset generation and methodology, Section 3 provides the results for and immediate discussion of each of the three flood characteristics investigated, and Section 4 constitutes the final conclusions.

2 Data and Methods

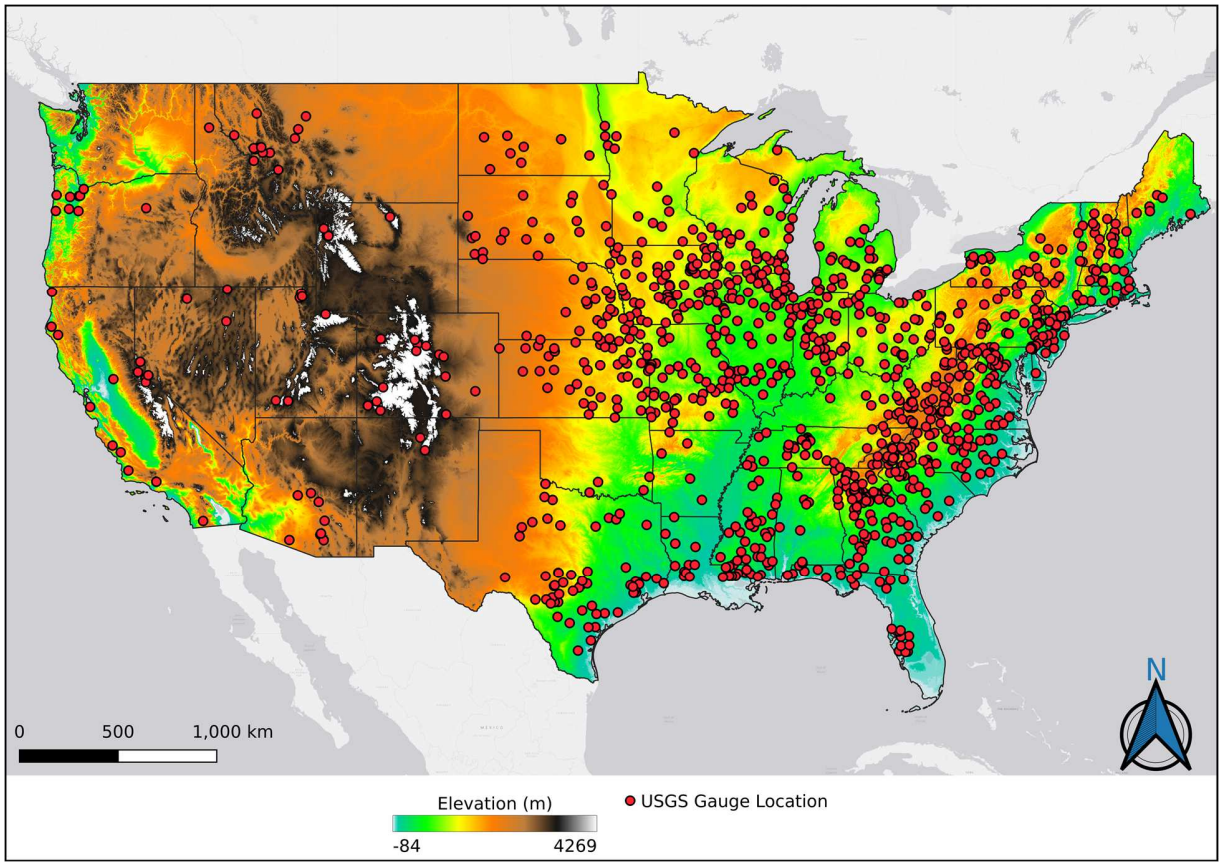


Figure 1. Map of gauge locations utilized across the Continental United States.

This study continues to build upon the body of work featuring numerous large-scale studies utilizing a CONUS-wide MRMS precipitation reanalysis dataset (Zhang and Gourley, 2018; Flamig et al., 2020; Gourley et al., 2017). Woods et al. (2023) focused on the use of the Version 06 IMERG Early run (IMERG-E) for a satellite forcing compared against the MRMS mosaic as a ground-based benchmark to highlight the impact of satellite precipitation resolution and accuracy. EF5 allows its user to arbitrarily select from and utilize several different options of both water balance models and routing schemes to generate hydrologic outputs such as return period indexes, streamflow discharge, and specific/unit discharge (i.e. the discharge at a pixel normalized by its upstream basin area). Importantly, EF5 also allows the user flexibility in the format of its input

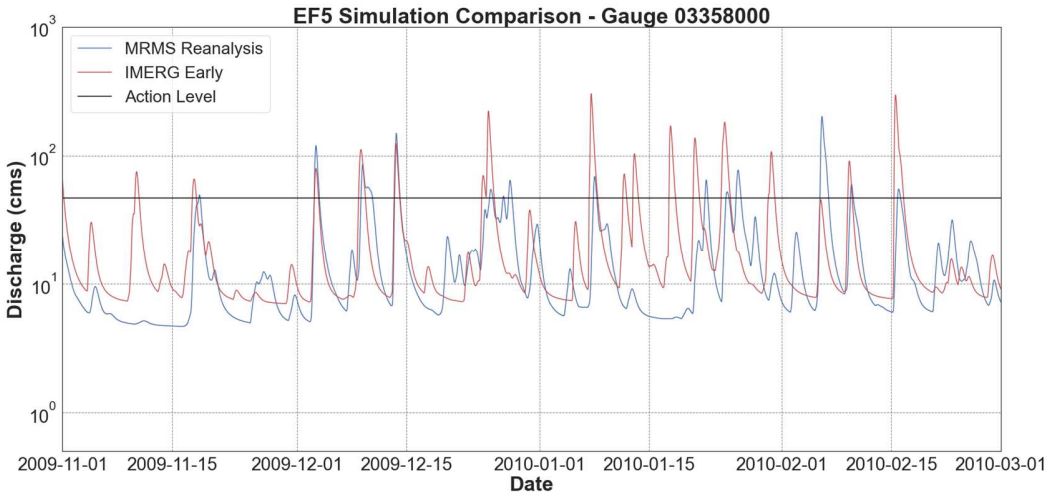
124 precipitation forcing data. For this study, each precipitation forcing was run with EF5 using the
125 Coupled Routing and Excess STorage (CREST; Wang et al., 2011) distributed hydrologic model
126 combined with kinematic wave routing (Vergara et al., 2016). This scheme of EF5/CREST is the
127 same configuration utilized by the FLASH system for flash flood warning operations in the United
128 States National Weather Service and is built off extensive geospatial datasets of parameters which
129 remove the need for timeseries-centered model calibration (Vergara et al., 2016; Gourley et al,
130 2017; Flamig et al., 2020).
131
132 **Table 1.** Associated general basin characteristics of gauges selected for analysis.

Basin Characteristic	**Value Range**
Area	21.11 – 45557.9 (km^2)
Slope Index	0.00013 – 0.08999
Relief Ratio	0.00043 – 0.16836
Basin Average Imperviousness	0.0 – 1.074 (%)
Basin Average Curve Number	48.2 – 89.4
Annual Precipitation	261.1 – 2841.2 (mm)

133
134
135 This study utilizes a previously extensively quality-controlled selection of over 3000
136 gauges (Gourley et al., 2017), where any gauges deemed by the USGS to have any anthropogenic
137 influence, where at least 80% of the basin falls within an area where the MRMS radar beam height
138 is 1 km above ground level or less, as well as any basins where snowmelt processes are dominant
139 (i.e. basins where snowfall contributes to >30% of annual precipitation) were removed. The

locations of these gauges can be seen in **Figure 1**, while the associated basin characteristics of these gauges can be found in **Table 1**. Simulations were run across the CONUS for both precipitation forcings at their native resolutions (i.e., MRMS-forced at 1-km spatial and 5-min temporal, and IMERG-forced at 10-km spatial and 30-min temporal) from 2004 to 2011. United States Geological Survey (USGS) data for each gauge was also taken as reference data for the time period simulated. Each time series was post-processed to isolate individual flood events based on its designated USGS "action-level" discharge value, which is the lowest threshold value provided by the USGS at each specific basin denoting the water level at which a given event is considered a flood. This also serves to denote the start time (i.e., the time point where discharge exceeded the threshold) and end time (i.e., the point where discharge fell back below the threshold) of each event. For an example of how this may look graphically, see **Figure 2** which provides a zoomed-in look at an arbitrary USGS gauge in Indiana (Gauge 03358000). Each raw event was then matched one-to-one between the simulated streamflow time series and the USGS observations, respectively, using an algorithm of cross-referencing criteria. The algorithm first looks for and matches events that overlap, i.e. where an observed event shares timesteps with a simulated event. Where there is an unmatched observed event with no overlap, the algorithm then uses the start and end times of the unmatched observed event to attempt to locate an unmatched simulated event in proximity (i.e., within a window of 100 hours) that has both the closest start time and closest end time to the observed event. These criteria also served to remove outliers where multiple simulated events appear to be logged over the time period of one observed event, caused by the wobbling of the timeseries above and below the flood threshold. Each individual simulated event that was successfully matched to an individual observed event generates a fixed pair of overall peak discharge values (observed and simulated), respective event durations, and overall event start and

163 end times while the remaining unmatched events are archived. Differences in the simulated and
164 observed characteristics are used to compute errors with respect to the USGS reference and analyze
165 errors in the simulated flood characteristics. Specifically for each event, (1) the difference in peak
166 discharge indicates whether the simulation overestimates (positive error) or underestimates
167 (negative error) the observed flood peak; (2) the difference in flood duration indicates whether the
168 simulated flood is shorter (negative error) or longer (negative error) than the observed flood; (3) a
169 simulated flood that starts (ends) earlier (later) than the observed flood will be associated with a
170 positive (negative) start (end) time error. This new and representative dataset of more than 20,000
171 matched events per product serves as the basis of this study. Given the diversity of basins and
172 climatologies gathered in this study, errors in peak discharge, duration, and timing are expected to
173 characterize representative behaviors associated with the precipitation forcing (MRMS and
174 IMERG-E) as well as from the hydrologic model. Specifically, error samples will be used to
175 quantify separate systematic errors and random errors.
176



177

178 **Figure 2.** An example of a modeled timeseries comparison, with included USGS action level.

179

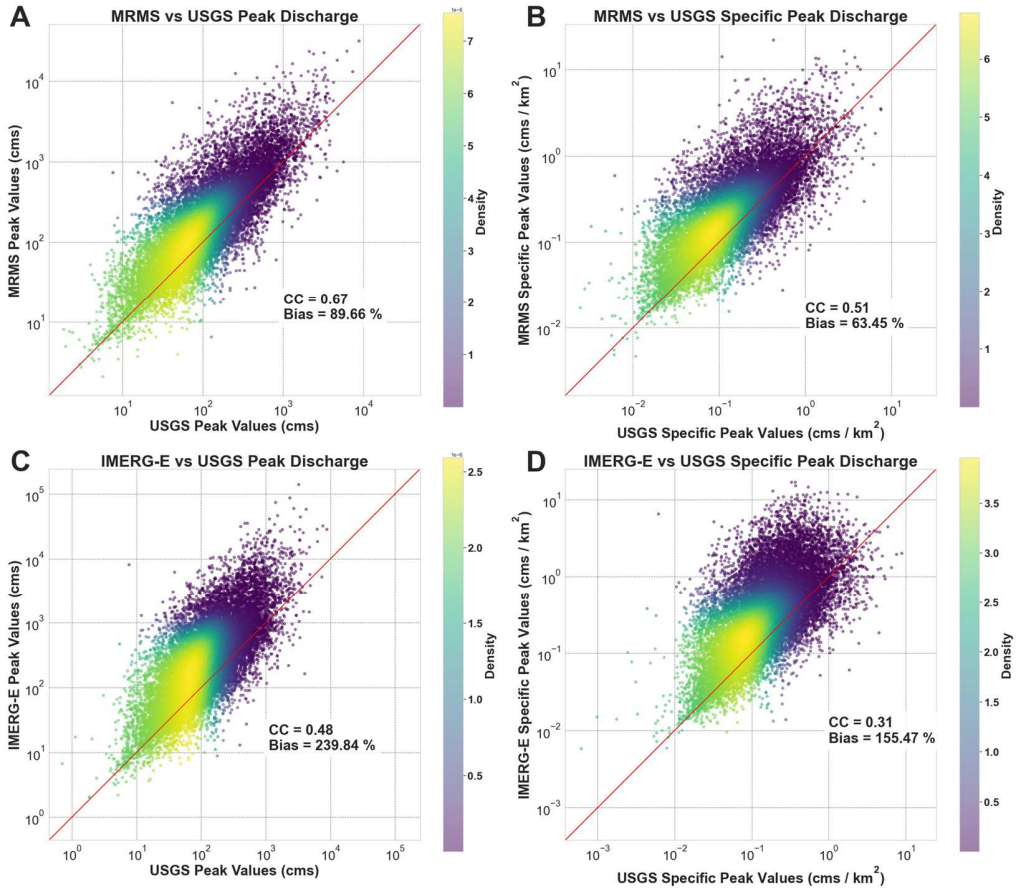
All three flood characteristics evaluated in Woods et al. (2023) will again be evaluated in this study in the context of USGS observations: the flood magnitude (peak discharge), the flood duration (total time elapsed from start to end), and the flood timing (the relative difference in start and end times between products). This continues to delve into the growing sentiment in the greater hydrologic community to move away from traditional methods of hydrologic evaluation, bulk metrics such as the Nash-Sutcliffe Efficiency (NSE) or the Kling-Gupta Efficiency (KGE) (Nash and Sutcliffe, 1970; Gupta et al., 2009), and focus on new methods of model assessment (Clark et al., 2021). The idea here is that agreement between the products and observations on these flood characteristics from discrete events can provide a far more robust assessment of modeling quality across the study area than traditional methods. For a more in-depth explanation of this reasoning, please refer to Woods et al. (2023).

3 Results and Discussion

3.1 Magnitude (Peak Discharge)

Critical to the development of flood mitigation strategies and engineered controls, as well as for emergency managers and real-time flood forecasters, is the understanding of how well the magnitude of a simulated flood behaves with respect to what is observed in the underlying basin. **Figure 3** provides a comprehensive representation of the accuracy of MRMS-forced and IMERG-forced flood peak discharge simulations, respectively. Of the density scatter plots provided, **Figures 3a** and **3c** display peak discharge values whereas **Figures 3b** and **3d** show specific peak discharge. Note that specific peak discharge was calculated and provided as a means to filter out

202 the natural dependence of peak discharge values with basin area; it is also a vital metric when
203 dealing with flash floods.
204

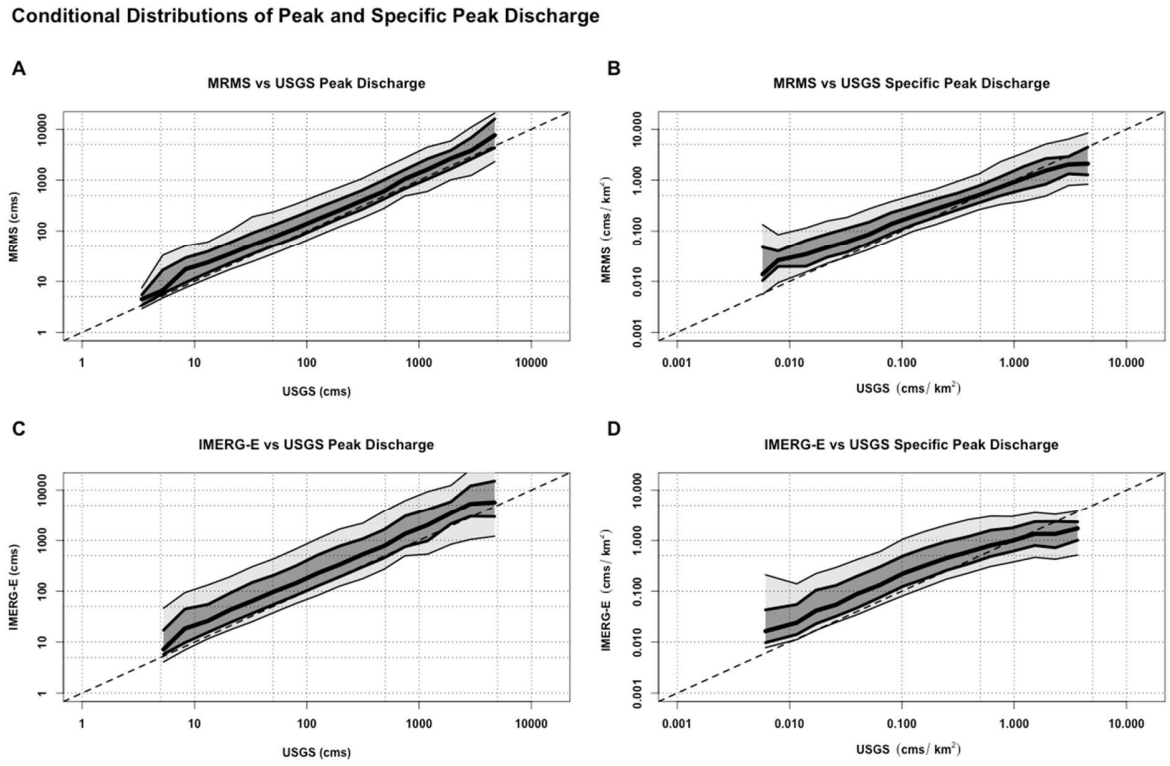


205

206 **Figure 3.** Scatterplots of MRMS-forced simulated peak discharge (A), MRMS-forced simulated specific peak
207 discharge (B), IMERG-E-forced simulated peak discharge (C) and IMERG-E-forced simulated specific peak
208 discharge values compared against USGS reference values. The red diagonal line indicates the 1:1 line.
209

210 While the points tend to gather around the one-to-one line, a distinct conditional bias can
211 be seen across both products and discharge types, with an increasing overestimation of higher
212 (specific) discharges. Both MRMS and IMERG-E simulations overestimate with respect to USGS,
213 though a tighter spread can be seen in the MRMS simulations. This is to be expected, with MRMS

operating at higher spatial and temporal resolutions than IMERG-E. Additional conditional bias can also be seen in the peak discharges, with point densities tending to fall more vertical on the plots as opposed to following the 1:1 line. To further dissect these results, the data was converted into plots of conditional distributions (provided in **Figure 4**). This style of plot was highlighted in Woods et al. (2023) as a more direct way of assessing conditional biases and random error. The process examines an independent variable through binned quantiles (10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}) of values from a chosen dependent variable. For the figure shown here (as well as in subsequent sections) the conditional median (50^{th} quantile) provides the first-order trend of the dependency, the interquartile area (25^{th} to 75^{th}) estimates the uncertainty in the relationship between the variables, and the 10^{th} and 90^{th} quantiles describe the range of extreme values between the variables.



227 **Figure 4.** Conditional distribution plots of MRMS-forced and IMERG-E-forced peak discharges (A and C) and
228 specific peak discharges (B and D) compared against USGS references. The thick center line shows the 50th quantile
229 (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 90th and 10th. The
230 dashed line is the 1:1 line.
231

232 The conditional distribution investigation in **Figure 4** reiterates what was seen in the
233 density scatterplots: distinct overestimation on the part of both MRMS and IMERG-E simulations
234 with respect to the USGS observations. Again, as expected, the uncertainties associated with
235 MRMS simulations (i.e., the overall spread of the quantiles) are smaller than those associated with
236 IMERG-E; the effects of resolution certainly play a role here. Interesting to note, however, is how
237 the specific peak discharge of both products (**Figure 4b** and **Figure 4d**) trend from overestimation
238 at lower values towards the 1:1 line and eventually into slight underestimation at the highest values
239 to the point where IMERG-E simulations begin to plateau out. This plateau effect was similarly
240 seen in Woods et al. (2023) and attributed to the coarser spatial and temporal resolutions of
241 IMERG, with these resolutions prohibiting the algorithm's ability to resolve the highest levels of
242 instantaneous precipitation and therefore being unable to resolve the highest specific peak
243 discharges often associated with them. Seeing the effect appear when compared to the gauged
244 USGS reference corroborates this idea, suggesting that the shortcoming lies within the ability of
245 IMERG to resolve the highest values and locations of extreme precipitation events (i.e., those
246 responsible for flash floods associated with these high specific peak discharges) as opposed to
247 errors generated within the hydrologic model itself.
248

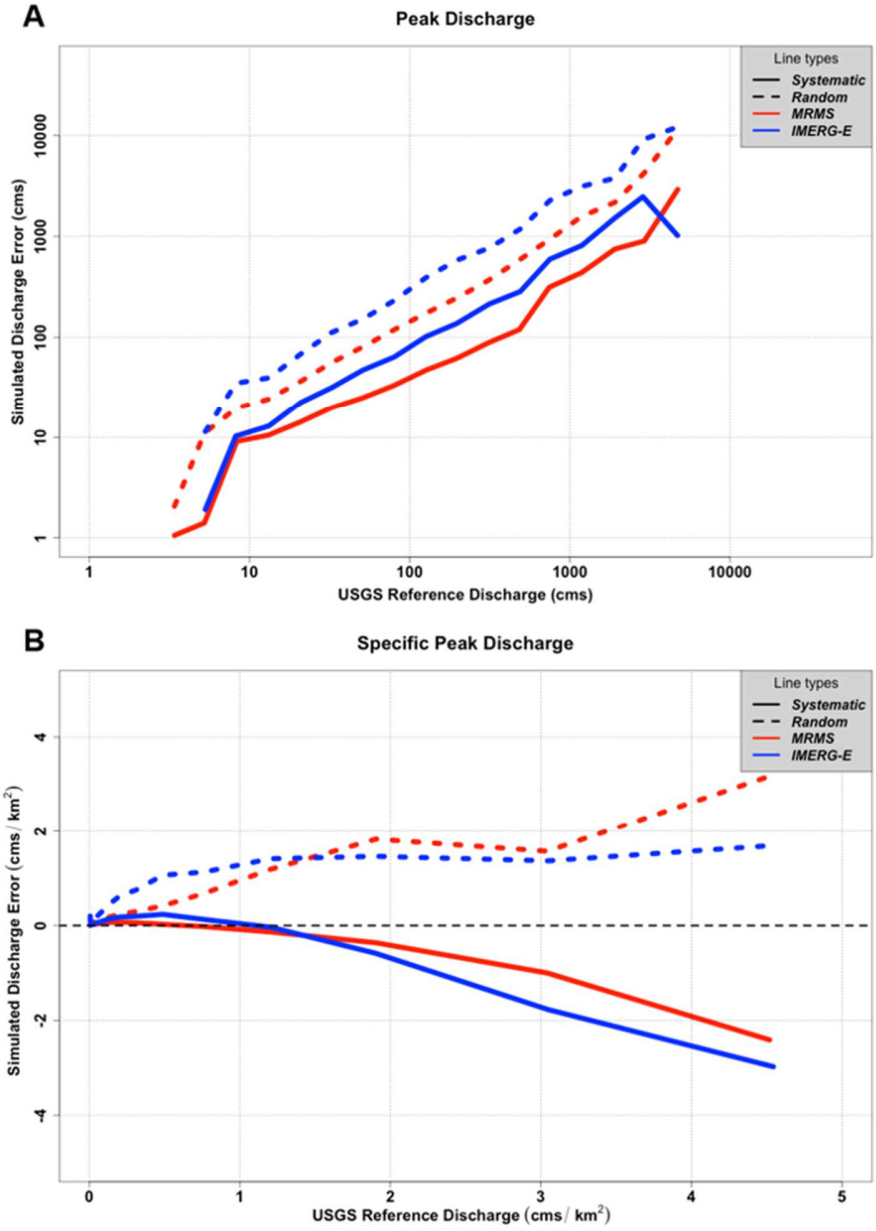


Figure 5. Error calculations for simulated flood peak discharge and specific peak discharge from MRMS (red) and IMERG-E (blue) with respect to USGS. Solid lines represent systematic error while dashed lines represent random error.

Building upon the quantile analysis, as well as to further inform on the abilities of the products, an error analysis was conducted (**Figure 5**). For both products, and for both discharge

types, the systematic error (simulated median minus observed median) and random error (75th quantile minus 25th quantile) were calculated and plotted against the USGS reference values. In **Figure 5a,** distinct increasing trends in systematic (positive bias) and random error are seen for both MRMS-forced and IMERG-E-forced simulations with respect to increasing associated USGS peak discharge values. This is likely associated with the behavior of EF5 itself with the generation of larger floods at larger basin sizes; there could potentially be issues with the water balance model and the sheer volume of water, but it is also known that kinematic wave routing becomes less effective than more dynamic routing schemes when modeling larger rivers (Vergara et al., 2016). The effects of satellite product resolution and accuracy can be seen between the simulations themselves, with IMERG-E simulations consistently showing higher systematic and random biases compared to MRMS simulations.

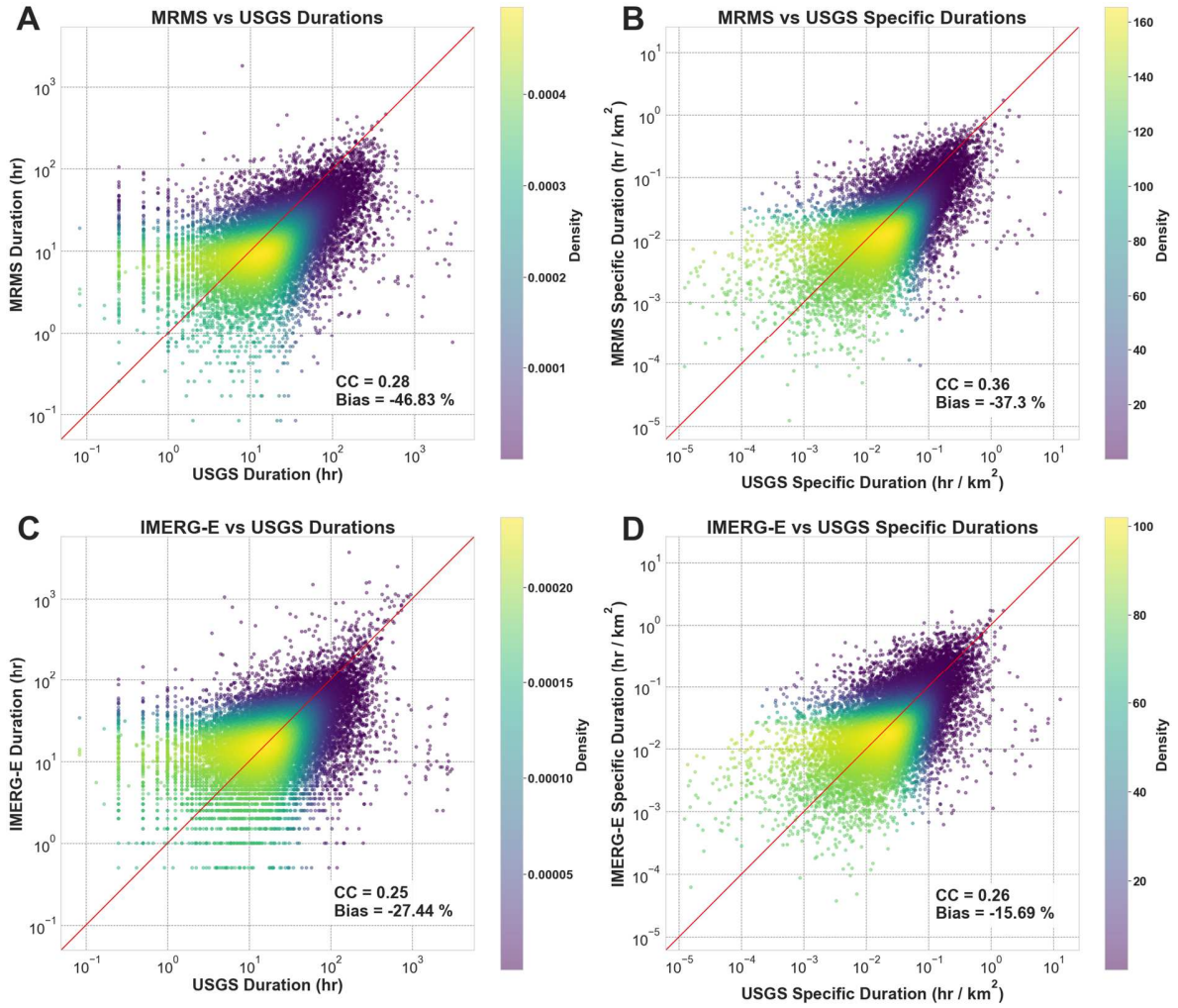
When looking at specific peak discharge (**Figure 5b**) similar stories can be seen. While both products now trend into underestimation of specific peak discharges compared to USGS, simulations generated by IMERG-E still show more negative systematic bias than those generated by MRMS. From a model perspective, this overall underestimation at the highest specific discharges is likely associated with the water balance component, CREST, as opposed to routing. To generate flash floods of these magnitudes there needs to be considerably high rainfall rates; if precipitation products are already underestimating these rates, errors are likely going to propagate even further when combined with basin characteristics and model physics. Random error provides a new interesting look, however; at increasing values of specific discharge (> 1.5 cms/km^2) the random error associated with MRMS simulations overtakes the random error of those associated with IMERG-E. This is likely due to smoothing effects of IMERG resolution as well as algorithm limitations; MRMS, with its higher resolutions, has a better chance of capturing the high-intensity

279 rainfall events normally associated with these extreme values of specific discharge better than
280 IMERG can, naturally leading to increased random error in the system. It is worth noting that
281 accuracies in flash flood discharge estimation have been shown to improve significantly as
282 precipitation products become more sophisticated (Gourley and Vergara, 2021), so future research
283 is warranted to better dissect and diagnose the behavior of EF5 with the improvements that have
284 been made to both MRMS and IMERG precipitation products in the years after the time period of
285 this study. Namely, MRMS forcings generated with weather radar data that have been upgraded
286 and processed using dual-polarization technology (i.e., after 2013) and IMERG forcing data that
287 has been retrieved using the spaceborne sensors launched with the GPM constellation itself (i.e.,
288 after 2014). These updated products will only serve to enhance the results of this study and provide
289 for a more in-depth understanding of potential hydrologic model deficiencies.
290

291 3.2 Flood Duration

292 Further critical to emergency management efforts and flood operations is an understanding
293 of the expected duration of a flooding event, real or simulated. As such, the analyses utilized for
294 peak discharge were also undertaken for simulated flood duration. First, density scatterplots were
295 created and can be found in **Figure 6**. As with discharge, event durations were normalized by basin
296 area to generate specific duration values as an additional method of assessment. What can be seen
297 is surprising; overall, MRMS simulations of floods tend to underestimate their durations with
298 respect to their USGS counterparts. Longer flood durations are increasingly underestimated
299 (conditional bias). This conditional bias is related to basin size, as it is less significant with unit
300 flood durations (see also **Figure 7b** and **7d**). This is likely explained by the routing scheme used;
301 the accuracy of the kinematic wave routing employed by this version of EF5 is known to degrade

as basin size and river size increases, where more dynamic routing schemes typically perform better (Vergara et al., 2016). What is seen from IMERG-E simulations (in **Figures 6c** and **6d**) is also interesting, with durations being closer to the 1:1 line with respect to USGS than MRMS simulations. This behavior is likely due to the inherent overestimation of IMERG-E durations with respect to MRMS, as was seen in Woods et al., 2023, meaning the underestimation exhibited by EF5 is instead counteracted in the simulations by IMERG-E's propensity to overestimate precipitation durations and resulting floods.



311 **Figure 6.** Density scatterplots of MRMS and IMERG-E simulated flood durations (A and C) and normalized
312 duration values based on associated basin area (B and D), all plotted against USGS references. The red line indicates
313 the 1:1 line.
314

315 The conditional distribution plots (**Figure 7**) tell a similar tale, with noticeable
316 underestimations seen for both products, but several additional features can be extracted. For
317 instance, despite the core of MRMS-simulated durations in the density plot showing
318 underestimation, there are distinct regions of overestimation at the shortest of flood durations (<5
319 hr). This feature is consistent across both products as well as both duration types, as well as both
320 products trending from overestimation to underestimation as flood durations increase. Unlike with
321 peak discharge, however, there is no noticeable difference in error spread between MRMS-
322 simulated durations and IMERG-simulated durations with respect to USGS. Both products also
323 behave similarly when normalized by basin area, though with a somewhat closer spread of
324 quantiles from MRMS simulations. This is more consistent with expectations regarding the higher
325 resolutions associated with MRMS.
326

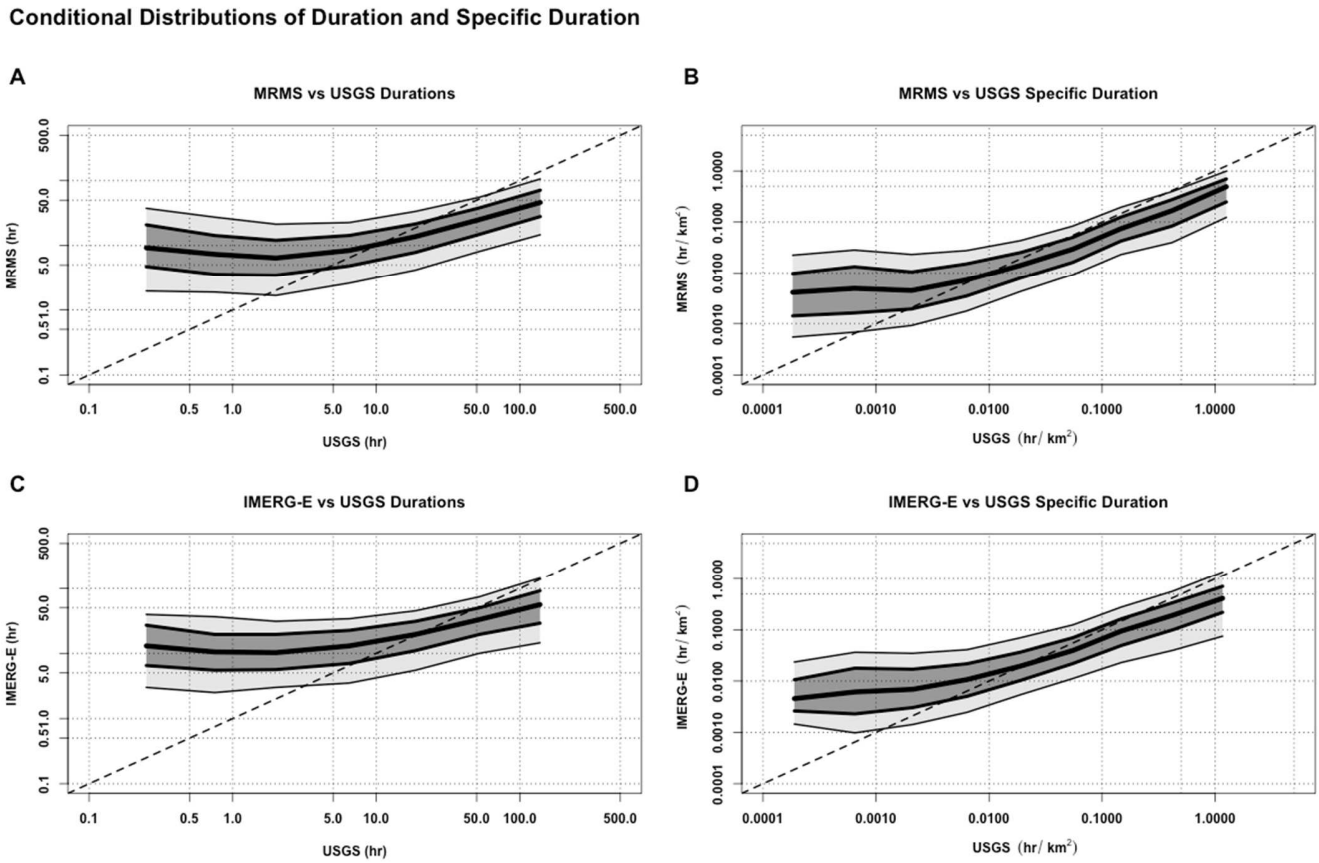


Figure 7. Conditional distribution plots of MRMS and IMERG-E simulated event durations (A and C) and normalized duration values (B and D), all plotted against USGS references. The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 90th and 10th. The dashed line indicates the 1:1 line.

Like with discharge, representations of error for duration and specific duration are shown in **Figure 8**. When looking at the duration of events (**Figure 8a**), the errors remain fairly regular (overestimation) for shorter events (< 10 hr) before a steep drop-off into large underestimation as durations increase. The overestimation at lower durations is likely associated with EF5's tendency to start flood events earlier and with potentially longer trailing limbs and ends (seen in Section 3.3). The intense underestimation of longer durations is again likely an artifact generated by the

breakdown of efficiency of kinematic wave routing at larger basins and rivers, the usual culprits responsible for floods of these long lengths.

For intercomparison between the products themselves, some interesting features arise. Random error is as expected, with consistently higher random error associated with IMERG-E-forced simulations than MRMS-forced simulations, a byproduct of the difference in product resolution. Systematic error is a different story; IMERG-E simulations overestimate more than MRMS at shorter durations (again, a factor of resolution) but at longer durations MRMS is the product with higher underestimation in simulations. This corroborates what was seen in the density scatterplots (**Figure 6**) where IMERG-E simulated durations fall closer on the 1:1 line with respect to USGS than MRMS simulated durations.

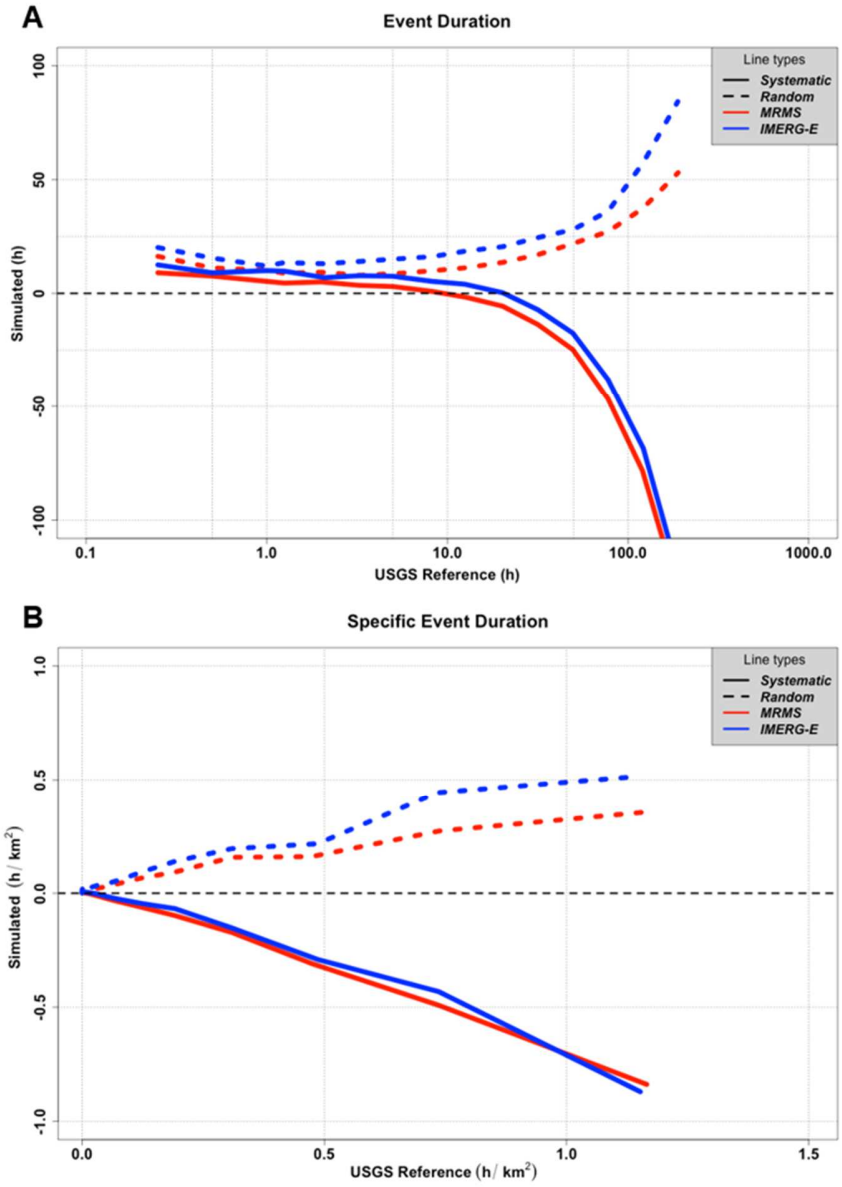


Figure 8. Error calculations for simulated flood durations and specific durations from MRMS (red) and IMERG-E (blue) with respect to USGS. Solid lines represent systematic error while dashed lines represent random error.

The errors associated with specific duration (**Figure 8b**) largely mirror what was seen with duration; the systematic error of IMERG-E simulations remain slightly less negative than those generated by MRMS while the random error of IMERG-E simulations remain higher than those

of MRMS. Due to the quasi-linear nature of the systematic biases we see from the products for specific duration, it will be simple to make an error model in the future.

3.3 Flood Timing

Perhaps the most critical information for flood and flash flood forecasting generated by this study are the computations of event timings. When events are logged and matched as part of the overall methodology, they are naturally associated with timestamps for both the start of the event and end of the event. As such, the difference between the observed and simulated start (and end) times can also be calculated and logged. For this process, the absolute start and end times for MRMS and IMERG-E simulations were subtracted from their associated USGS event absolute start and end times, giving either a positive or negative time difference value in hours. A positive (negative) value in this regard indicates that the simulated event occurs earlier (later) than its reference counterpart.

Histograms of both products with respect to USGS can be found in **Figure 9**. For both MRMS and IMERG-E simulations most events are associated with both positive start and positive end times, meaning that the simulated events for both products tend to start early and end early with respect to their matched USGS event. This is likely associated with the routing component of EF5, with water overall moving through the system faster than what is observed at the gauge. MRMS-forced simulations values also have an average start time closer to zero and with a smaller standard deviation than those forced by IMERG-E, which remains consistent with the higher temporal resolution available to the product. The end times for both products behave similarly statistically, however, which is interesting to note. Larger time deltas are likely associated with longer duration floods, which in turn are associated with larger basins and flow lengths – an area

where the kinematic wave routing scheme utilized in this study's EF5 scheme becomes less effective (Vergara et al., 2016).

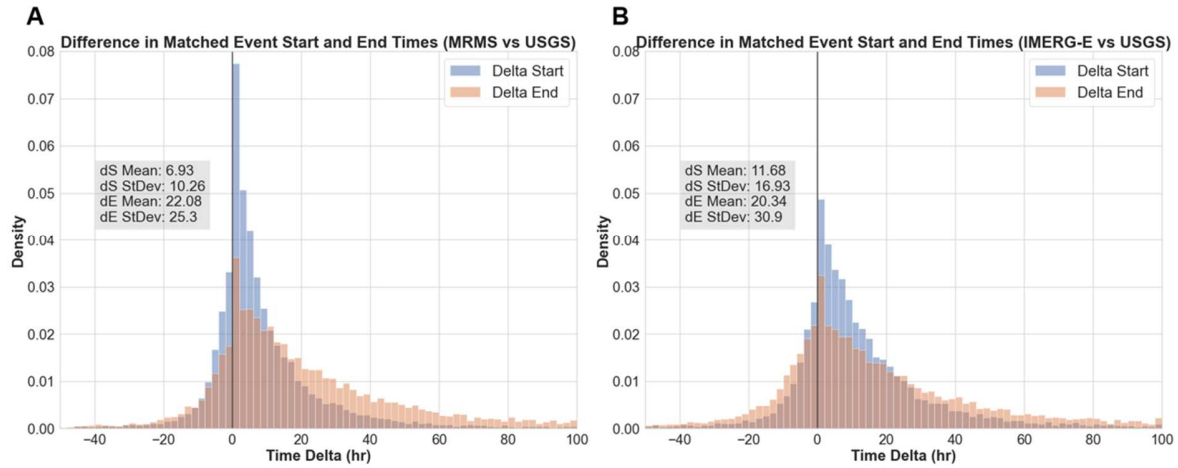


Figure 9. Histograms of the time deltas of matched flood start times (blue) and end times (orange) for MRMS simulations, IMERG-E simulations, and USGS observations, with associated means and standard deviations. A positive (negative) value indicates that the simulated event (MRMS or IMERG-E) event occurs earlier (later) than its USGS absolute time counterpart.

In investigating the conditional distribution plots, found in **Figure 10**, these same trends can be seen. Since the size of a basin is naturally associated with flood timing, area was chosen to be the dependent variable to draw for the quantiles of start time and end time. All four sets of quantiles track well with the overlying conclusions from **Figure 9**, that both products tend to simulate floods that start and end earlier than the reference. This also corroborates the idea that the higher means and standard deviations seen with end time are more often associated with the largest basins, scales where kinematic wave routing begins to struggle. Simulations forced by IMERG-E are shown to have significantly higher extreme error quantiles associated with smaller basin sizes than those forced by MRMS, an effect similarly seen in Woods et al., 2023, understood to likely

be associated with the coarse resolution of IMERG-E being unable to generate more precise precipitation-flood responses. At larger basin sizes, these errors shown by IMERG-E simulations can be attributed to systematic biases and uncertainty caused by basin-scale aggregations, with an increasing importance falling on precipitation spatial distributions (Woods et al., 2023), but similar trends from MRMS simulations at large basins suggests routing from the model itself is likely also a contributor in this case.

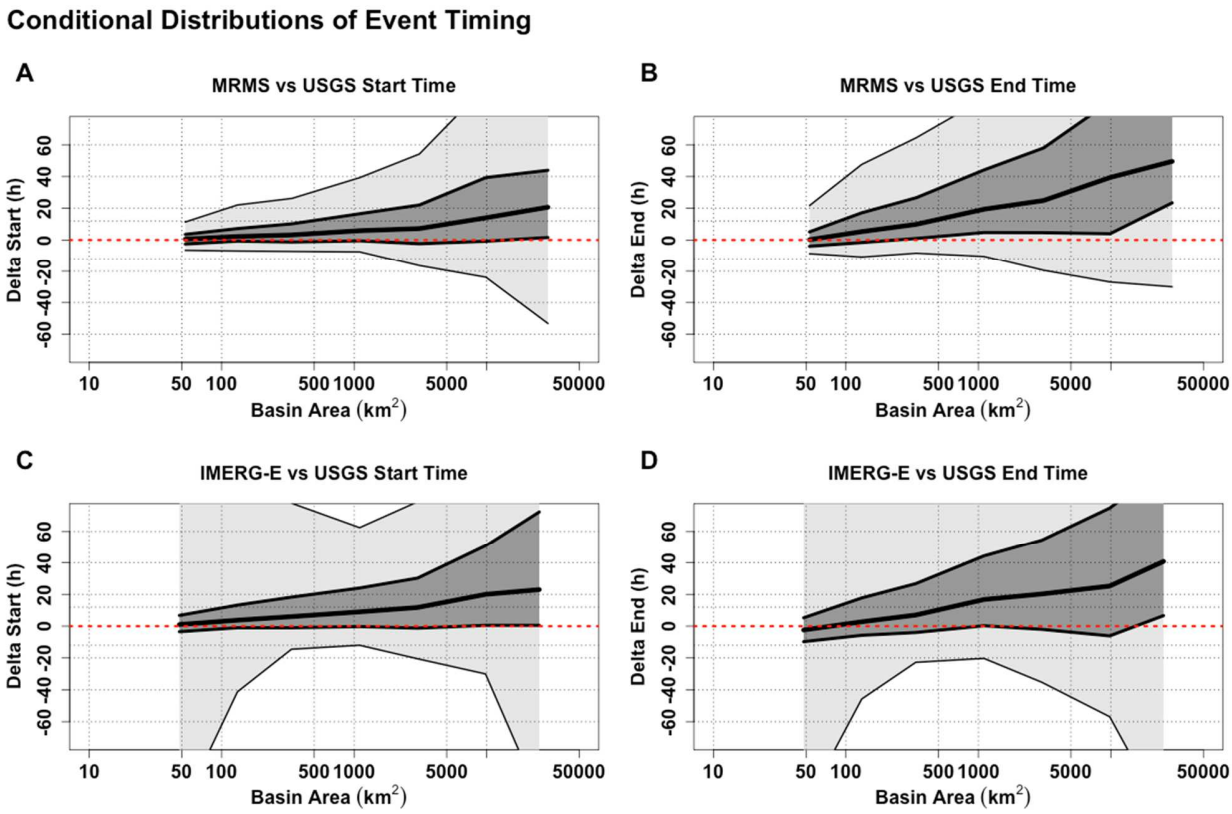


Figure 10. Conditional distribution plots of calculated event delta start (A and C) and delta end (B and D) times compared against associated basin areas. The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 90th and 10th. The dashed red line is the zero line, signifying matching timing of events. A positive (negative) value indicates that the simulated event (MRMS or IMERG-E) event occurs earlier (later) than its USGS absolute time counterpart.

410

411 Despite increasing uncertainty with basin size (as well as significant extreme quantiles
412 associated with IMERG-E simulations), median values and 25th/75th error quantiles remain
413 remarkably tame for both products at areas <1000 km^2. End time values lose effectiveness sooner,
414 before reaching 500 km^2 in both cases, and the overall spread ends noticeably wider than final
415 spreads for start time. Regardless, event start time is inherently a more important statistic to predict
416 accurately more often, especially in the case of flood forecasting and emergency response.
417 The error budgets of the products with regard to event timing (**Figure 11**) are in agreement
418 with overall trends seen throughout this analysis but are able to provide important insight into
419 accuracies at different scales. Before discussion, however, it is important to establish an
420 understanding of what timing error means in this context. Throughout this section, the positive and
421 negative deltas have been associated with absolute times. With regards to error, this instead
422 translates to positive values signifying an overall trend towards earlier times (both start and end)
423 while negative values signify an overall trend towards later times. As can be seen across both time
424 delta plots, the overwhelming majority of errors for both products tend to push start and end times
425 earlier than USGS. This effect is likely caused by routing within the EF5 model, with water more
426 likely to flow faster through the system (especially at larger basin areas) than more slowly. For
427 end times there also exists a small window at basins < 100 km^2 where IMERG-E simulations have
428 negative systematic error values, meaning that at smaller basins IMERG-E-forcings tend to try to
429 pull end times later. Overall, this suggests that there is an inherent competition between routing
430 and resolution being exhibited; this trend to counteract end times and extend the total duration of
431 events ties into what was seen in the previous section (Section 3.2) and **Figure 6**, where IMERG-

E produces more consistent simulated event durations with respect to USGS than the underestimation of durations simulated by MRMS.

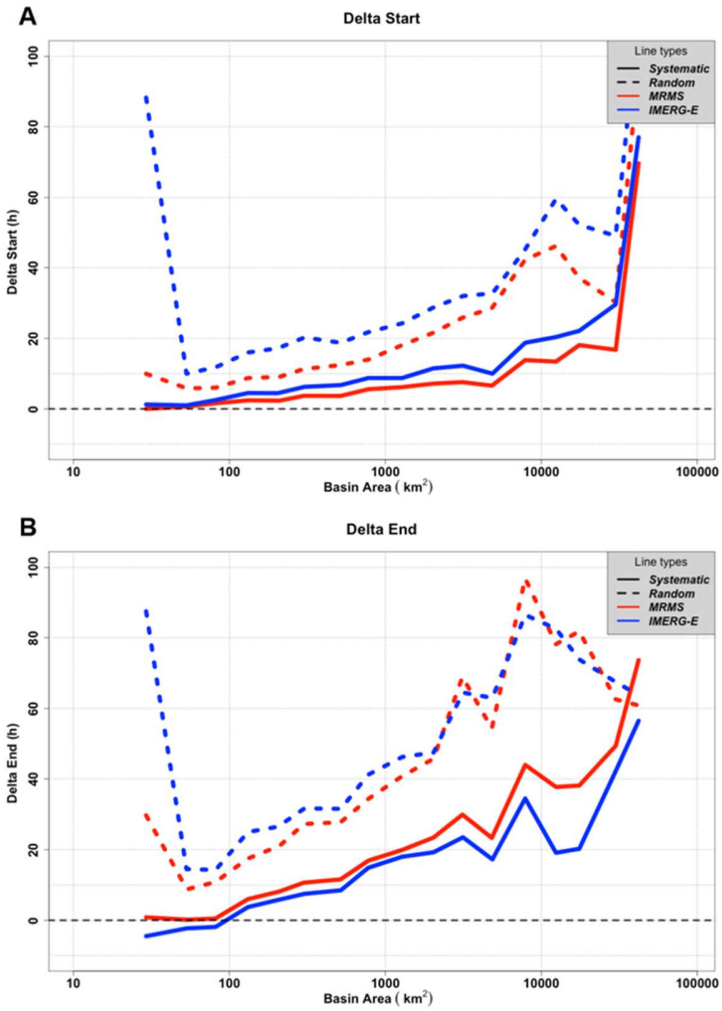


Figure 11. Error calculations for start time and end time deltas from MRMS-simulated (red) and IMERG-E-simulated (blue) events with respect to USGS, plotted against associated basin area. Solid lines represent systematic error while dashed lines represent random error.

For delta start errors, what can be seen is consistent with the other characteristics previously discussed; IMERG-E simulations showcase both higher systematic and higher random error values

than those simulated by MRMS. Both products, however, perform well at smaller basins with minimal systematic error; welcome news for the potential to utilize IMERG-E for operational flood prediction purposes. With small basins naturally more susceptible to flash flooding, having a reliable benchmark for predicting the timing of when these events will begin significantly improves the ability of forecasters and emergency managers to protect life and property.

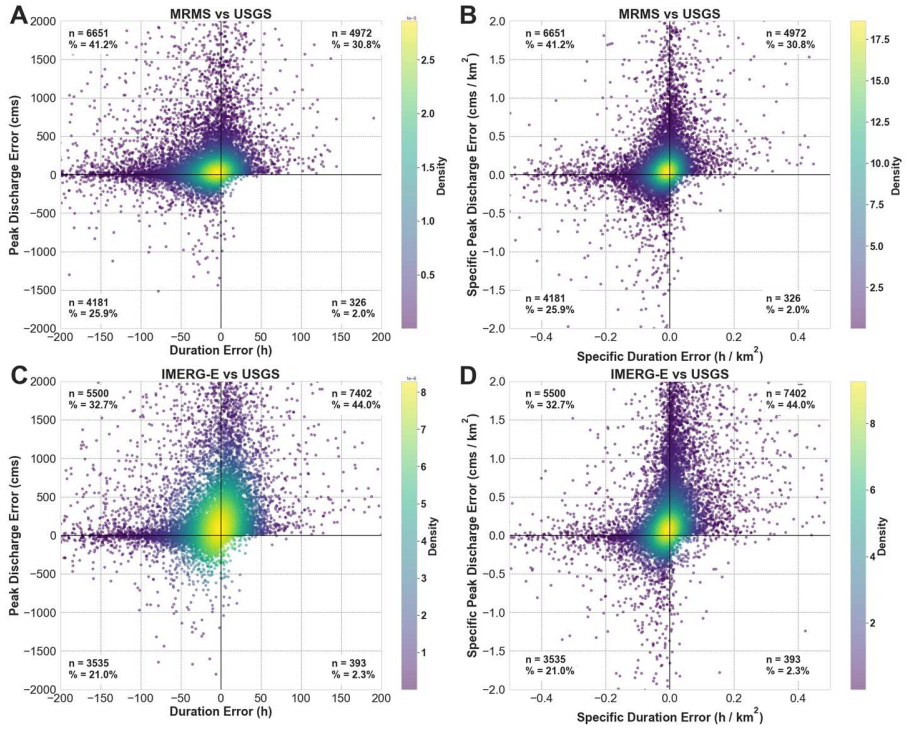
Contrary to delta start, errors seen with delta end are more favorable to IMERG-E simulations, with MRMS simulations showing higher systematic errors at all basin sizes. MRMS simulations still maintain a lower random error, up until the larger basins where the random error of the two products becomes noisier and essentially evens out. Another interesting feature is the sharp decrease in random error from IMERG-E simulations from ~50 km^2 to ~75 km^2; this likely points to the location of the effective resolution of IMERG-E for flood simulation utility (Guilloteau et al., 2017; Guilloteau et al., 2020).

3.4 Hydrologic Model Performance Analysis (Quadrant Plots)

Given the increased influence of simulated flood tendencies attributed to the hydrologic model itself with respect to USGS observations that have been highlighted so far in this study, further error characterization into EF5 was undertaken. Model influence on outputs was expected, to a degree, which was a core reasoning behind why Woods et al. (2023) elected to directly compare only simulated events against each other, with MRMS simulations serving as the reference, in order to specifically remove any effects from the hydrologic model and focus solely on the influence of the precipitation products themselves. The ability to include USGS data as the reference in this study allows for a more robust analysis and diagnosis of both hydrologic outputs

464 and model tendencies, benefitting extensively from what was found in the simulation-only
465 research.

466 More insight can be gained by characterizing the joint peak and duration errors that can be
467 influenced by the precipitation forcings and the hydrologic model components (i.e., water balance
468 and routing) . A quadrant plot displays the duration (x-axis) and peak discharge (y-axis) errors
469 (**Figure 12**), with each error quadrant signifying a different tendency within the hydrologic model
470 outputs. Points in the top left quadrant (positive peak errors and negative duration errors) indicate
471 simulated floods with higher peaks and shorter durations than USGS, a signal of influence from
472 kinematic wave where the water is being pushed through the system too quickly. In the top right
473 quadrant (positive peak errors and positive duration errors) points are found where both the peak
474 and the duration are higher than USGS, indicating positive water balance errors (i.e. there is too
475 much water in the system, with greater areas under the theoretical hydrograph). The bottom left
476 quadrant (both negative errors) is again dominated by water balance, but instead with too little
477 water simulated. The bottom right quadrant shows simulations with smaller peaks but longer
478 durations than the reference, signifying flood attenuation by the model.
479



480

481 **Figure 12**. Density scatterplots of discharge and duration errors for MRMS and IMERG-E simulations with respect
482 to USGS observations. Total numbers of points in each quadrant are provided, as well as each quadrant's percentage
483 of the total points.

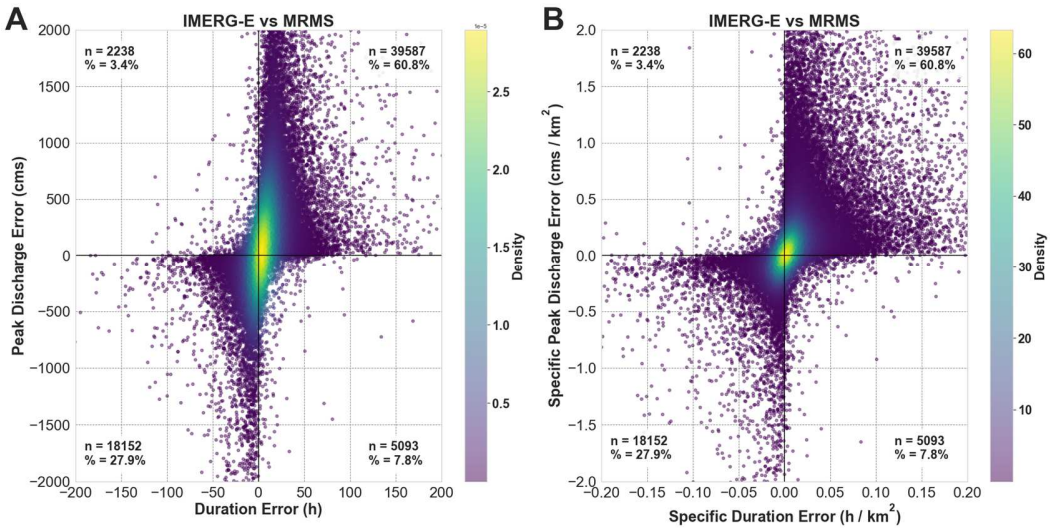
484 In the MRMS plots, the highest percentage of points fall into the top left quadrant (41.2%),
485 highlighting increased influence on simulations by the kinematic wave scheme. This corroborates
486 what has been seen throughout this study, where MRMS simulations are routinely more likely to
487 underestimate flood durations than IMERG-E. There is influence from the water balance
488 dominated quadrants as well (56.7 %), meaning there are discrepancies with how or where water
489 is entering the system. In the case of IMERG-E simulations, these quadrants are where the majority
490 of points are found (65%), with most falling into positive water balance error (44%). For IMERG-
491 E this is to be expected because coarser spatial and temporal resolutions naturally tend to add
492 excess water to the system through a combination of both smoothing over larger pixel sizes and
493 more limited accuracy in precipitation values themselves, leading to hydrographs that are taller

494 and longer than those of USGS. Kinematic wave is still a factor, but the increased tendency towards
495 water balance overestimation counteracts its effects and explains why IMERG-E maintains lower
496 systematic errors in simulated duration and flood timing than MRMS. Additionally, neither
497 product had a significant number of points in the bottom right quadrant, reiterating that the physics
498 of the model performs well, and that flood attenuation is not a factor here.

499 These results show that FLASH/EF5's model design choice on kinematic wave was correct
500 because for the overwhelming majority of the territory the assumptions of this model apply. The
501 fact that the highest densities are near the (0,0) point speaks well of the modeling system. Such
502 small numbers of points are seen on the bottom right quadrant due to several factors. First,
503 kinematic wave does not have as much capability to attenuate the flood wave at higher resolutions;
504 it can, however, if the pixel resolution is coarser, which is a result of numerical
505 diffusion/attenuation (i.e., an artifact of the numerical approximation). Second, because for most
506 of the terrain over the CONUS, kinematic wave applies. And third, because most of the basins and
507 subsequent events being considered in this study do not have the geomorphology and hydraulics
508 necessary to lead to significant flood attenuation.

509 In order to determine if there were any additional unforeseen tendencies within the model,
510 the same approach was taken by contrasting the MRMS and IMERG-E simulations themselves.
511 MRMS-simulated values were subtracted from IMERG-E-simulated values, and the same
512 discharge-duration plots are provided in **Figure 13**. As expected, almost all of the points fall within
513 the water balance quadrants, with the distinct majority in overestimation (60.8%). When the
514 influence of the model itself is removed, the effects of resolution difference between precipitation
515 products is expected to be dominant; IMERG again naturally puts more water into the system than
516 its higher-resolution counterpart. There is still influence from underestimation, however, likely

517 caused by a combination of spatial variability and variability in the accuracy of precipitation
518 estimates, which in turn is exacerbated by the algorithm's smoothing of rainfall itself (i.e, the
519 correct volume of rainfall is not always falling over the right area or basin).
520



521
522 **Figure 12**. Density scatterplots of discharge and duration errors for IMERG-E with respect to MRMS. Total
523 numbers of points in each quadrant are provided, as well as each quadrant's percentage of the total points.
524

525 Between duration (**Figure 13a**) and specific duration (**Figure 13b**) themselves, the plots
526 behave similarly, though there is a more asymptotic spread across the duration scatter than
527 specific duration. Both plots maintain higher densities closer to the (0,0) point, with that spread
528 becoming even tighter when normalized by basin area.

529 **4 Conclusions**

530 In this study, precipitation forcings from IMERG-E and MRMS were run through the EF5
531 hydrologic modeling framework, broken down into discrete flood characteristics (magnitude,
532 duration, and timing) and compared against reference observation data from USGS stream gauges

in order to develop an understanding of error trends and overall error budgets between the products. While consistent overall with previously established results (Woods et al., 2023), this study provides a more robust outlook into the hydrologic behaviors and accuracies of the products themselves and how they translate into the greater push towards integrated hydrologic validation of the GPM mission itself.

For flood peak discharge and specific peak discharge, both IMERG-E and MRMS simulations were shown to overestimate values with respect to the USGS reference, with IMERG-E simulated peak values being attributed to greater uncertainties. IMERG-E was also shown to have more difficulty resolving higher-end simulated specific peak discharge values than MRMS, which is attributed to the coarser spatial and temporal resolutions of the product as well as the lower accuracy ceiling associated with these resolutions. From a model perspective, this overall underestimation at the highest specific discharges is also likely associated with the water balance component. Both products showed similar error trends, with increasing systematic and random errors as basin size increases. MRMS simulations also had consistently lower systematic and random errors than IMERG-E simulations, with the exception of specific peak discharge where MRMS was higher.

When looking at the simulated flood durations, interesting interactions surfaced: MRMS consistently underestimated simulated durations with respect to USGS, with underestimation further increasing with basin size, while IMERG-E simulations were found to more closely fit the 1:1 line. In this scenario, the overall underestimation created by the products with respect to USGS is being counteracted by the inherent overestimation of simulated flood durations by IMERG-E with respect to MRMS (Woods et al., 2023). The consistent underestimation is associated with the accuracy of the kinematic wave routing scheme, which is known to degrade as basin size and river

size increases, where more dynamic routing schemes typically perform better. The error budgets of the products reflect this interaction, with IMERG-E simulations having a higher systematic error than MRMS simulations at smaller basin sizes but transferring to a less negative error than MRMS as basin size increases. Overall, however, IMERG-E simulations retained higher random errors than MRMS simulations across the board.

In the case of flood timing, simulated events for both products tend to both start early and end early with respect to their matched USGS event, a net earlier shift in timing for both products. Additionally, IMERG-E simulations are shown to have significantly higher extreme error quantiles associated with smaller basin sizes than MRMS simulations, an effect associated with the coarser resolution of IMERG-E being unable to generate more precise precipitation-flood responses. In regard to the systematic and random errors, both products have a tendency to push start and end times earlier than USGS, though IMERG-E simulations showcase both higher systematic and higher random error values than MRMS simulations. At larger basin sizes, these errors shown by IMERG-E simulations can be attributed to systematic biases and uncertainty caused by basin-scale aggregations, but similar trends from MRMS simulations at large basin sizes suggests routing from the hydrologic model itself is likely also a contributor in this case. Both products, however, perform well at smaller basins with minimal systematic error, a result that directly affects the potential to utilize IMERG-E for operational flood prediction purposes.

With instances of model behavior being shown to have an effect on simulation outputs at all three phases of this investigation, an additional analysis into the model's tendencies was also undertaken, where it was found that MRMS simulations were more likely to be impacted by the kinematic wave routing component while IMERG-E simulations were more likely to be impacted by water balance. For IMERG-E this is to be expected because coarser spatial and temporal

579 resolutions naturally tend to add excess water to the system, leading to hydrographs that are taller
580 and longer than those of USGS. The increased tendency towards water balance overestimation
581 counteracts the tendency of kinematic wave to push water through the system too quickly and
582 explains why simulations forced by IMERG-E maintain lower systematic errors in duration and
583 flood timing than those forced by MRMS. Additionally, it was shown across both products that
584 the physics of the model performs well, and that flood attenuation is not a factor in the results.

585 Based on these findings, it is recommended that further, more concentrated studies be
586 undertaken into the tendencies of EF5 in order to more accurately diagnose and quantify its
587 tendencies. Additional research is also being planned to assess how more recent product and
588 algorithm improvements translate into flood simulations, allowing for a trend to be established
589 regarding the state of improving hydrologic validation in advance of the Atmosphere Observing
590 System (AOS) mission.

591

592 **Acknowledgments**

593 We are very much indebted to the teams responsible for the MRMS and IMERG precipitation
594 products. Funding for this research was provided by the Joint Technology Transfer Initiative
595 program, which provided support to the Cooperative Institute for Severe and High-Impact Weather
596 Research and Operations at the University of Oklahoma under Grant NA20OAR4590354. P.
597 Kirstetter acknowledges support from the National Aeronautics and Space Administration Global
598 Precipitation Measurement Ground Validation program under Grant 80NSSC21K2045 and
599 Precipitation Measurement Missions program under Grant 80NSSC19K0681.

600

601 **Data Availability**

602 This reanalysis was performed on the raw, publicly available NEXRAD data archive available
603 from Amazon Web Services (https://aws.amazon.com/public-datasets/nexrad/).
604
605 **References**
606
607 Clark, M.P., Vogel, R.M., Lamontagne, J.R., Mizukami, N., Knoben, W.J., Tang, G., Gharari, S.,
608 Freer, J.E., Whitfield, P.H., Shook, K.R. and Papalexiou, S.M., 2021. The abuse of popular
609 performance metrics in hydrologic modeling. *Water Resources Research, 57(9),*
610 p.e2020WR029001. https://doi.org/10.1029/2020WR029001
611
612 Derin, Y., Kirstetter, P.E. and Gourley, J.J., 2021. Evaluation of IMERG satellite precipitation
613 over the land–coast–ocean continuum. Part I: Detection. Journal of Hydrometeorology, 22(11),
614 pp.2843-2859. https://doi.org/10.1175/JHM-D-21-0058.1
615
616 Derin, Y. and Kirstetter, P.E., 2022. Evaluation of IMERG over CONUS Complex Terrain Using
617 Environmental Variables. Geophysical Research Letters, p.e2022GL100186.
618 https://doi.org/10.1029/2022GL100186.
619
620 Flamig, Z.L., Vergara, H. and Gourley, J.J., 2020. The Ensemble Framework For Flash Flood
621 Forecasting (EF5) v1. 2: description and case study. *Geoscientific Model Development, 13(10),*
622 pp.4943-4958. https://doi.org/10.5194/gmd-13-4943-2020
623

624 Gebregiorgis, A.S., Kirstetter, P.E., Hong, Y.E., Gourley, J.J., Huffman, G.J., Petersen, W.A.,
625 Xue, X. and Schwaller, M.R., 2018. To what extent is the day 1 GPM IMERG satellite
626 precipitation estimate improved as compared to TRMM TMPA-RT?. *Journal of Geophysical*
627 *Research: Atmospheres*, 123(3), pp.1694-1707. https://doi.org/10.1002/2017JD027606
628
629 Gourley, J. J., Flamig, Z.L., Vergara, H., Kirstetter, P., Clark, R.A., Argyle, E., Arthur, A.,
630 Martinaitis, S., Terti, G., Erlingis, J.M., Yang, H., 2017. The FLASH Project: Improving the
631 Tools for Flash Flood Monitoring and Prediction across the United States. *Bulletin of the*
632 *American Meteorological Society, 98*, pp.361–372. https://doi.org/10.1175/BAMS-D-15-00247.1
633
634 Guilloteau, C., Foufoula-Georgiou, E., & Kummerow, C. D. (2017). Global Multiscale
635 Evaluation of Satellite Passive Microwave Retrieval of Precipitation during the TRMM and
636 GPM Eras: Effective Resolution and Regional Diagnostics for Future Algorithm Development,
637 Journal of Hydrometeorology, 18(11), 3051-3070.
638
639 Guilloteau, C., Foufoula-Georgiou, E. (2020). Multiscale Evaluation of Satellite Precipitation
640 Products: Effective Resolution of IMERG. In: Levizzani, V., Kidd, C., Kirschbaum, D.,
641 Kummerow, C., Nakamura, K., Turk, F. (eds) Satellite Precipitation Measurement. Advances in
642 Global Change Research, vol 69. Springer, Cham. https://doi.org/10.1007/978-3-030-35798-6_5
643
644 Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F., 2009. Decomposition of the mean
645 squared error and NSE performance criteria: Implications for improving hydrological

646 modelling. *Journal of hydrology*, *377*(1-2), pp.80-91.
647 https://doi.org/10.1016/j.jhydrol.2009.08.003
648
649 Hartke, S.H., Wright, D.B., Quintero, F. and Falck, A.S., 2023. Incorporating IMERG Satellite
650 Precipitation Uncertainty into Seasonal and Peak Streamflow Predictions using the Hillslope
651 Link Hydrological Model. Journal of Hydrology X, p.100148.
652 https://doi.org/10.1016/j.hydroa.2023.100148
653
654 Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R.,
655 Nakamura, K. and Iguchi, T., 2014. The global precipitation measurement mission. *Bulletin of*
656 *the American Meteorological Society, 95(5),* pp.701-722. https://doi.org/10.1175/BAMS-D-13-
657 00164.1
658
659 G. Huffman, D. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, P. Xie, 2019: Integrated Multi-
660 satellitE Retrievals for GPM (IMERG), version 06. NASA's Precipitation Processing Center,
661 ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/
662
663 Kirstetter, P.E., Hong, Y., Gourley, J.J., Chen, S., Flamig, Z., Zhang, J., Schwaller, M., Petersen,
664 W. and Amitai, E., 2012. Toward a framework for systematic error modeling of spaceborne
665 precipitation radar with NOAA/NSSL ground radar–based National Mosaic QPE. Journal of
666 Hydrometeorology, 13(4), pp.1285-1300. https://doi.org/10.1175/JHM-D-11-0139.1
667

668 Kirstetter, P.E., Y. Hong, J.J. Gourley, M. Schwaller, W. Petersen and J. Zhang, 2013:
669 Comparison of TRMM 2A25 Products Version 6 and Version 7 with NOAA/NSSL Ground
670 Radar-based National Mosaic QPE. Journal of Hydrometeorology, 14(2), 661-669.
671 doi:10.1175/JHM-D-12-030.1
672
673 Kirstetter, P.E., Y. Hong, J.J. Gourley, Q. Cao, M. Schwaller, and W. Petersen, 2014: A research
674 framework to bridge from the Global Precipitation Measurement mission core satellite to the
675 constellation sensors using ground radar-based National Mosaic QPE.
676 In L. Venkataraman, in Remote Sensing of the Terrestrial Water Cycle (eds V. Lakshmi, D.
677 Alsdorf, M. Anderson, S. Biancamaria, M. Cosh, J. Entin, G. Huffman, W. Kustas, P. van
678 Oevelen, T. Painter, J. Parajka, M. Rodell and C. Rüdiger). AGU books Geophysical Monograph
679 Series, Chapman monograph on remote sensing. John Wiley & Sons, Inc, Hoboken, NJ. doi:
680 10.1002/9781118872086.ch4
681
682 Kirstetter, P.E., Petersen, W.A., Kummerow, C.D. and Wolff, D.B., 2020. Integrated multi-
683 satellite evaluation for the global precipitation measurement: Impact of precipitation types on
684 spaceborne precipitation estimation. Satellite Precipitation Measurement: Volume 2, pp.583-608.
685 https://doi.org/10.1175/JHM-D-19-0293.1
686
687 Lamontagne, J.R., Barber, C.A. and Vogel, R.M., 2020. Improved estimators of model
688 performance efficiency for skewed hydrologic data. Water Resources Research, 56(9),
689 p.e2020WR027101. https://doi.org/10.1029/2020WR027101
690

691 Liu, D., 2020. A rational performance criterion for hydrological model. Journal of Hydrology,
692 590, p.125488. https://doi.org/10.1016/j.jhydrol.2020.125488
693
694 Nanding, N., Wu, H., Tao, J., Maggioni, V., Beck, H.E., Zhou, N., Huang, M. and Huang, Z.,
695 2021. Assessment of Precipitation Error Propagation in Discharge Simulations over the
696 Contiguous United States. *Journal of Hydrometeorology, 22(8),* pp.1987-2008.
697 https://doi.org/10.1175/JHM-D-20-0213.1
698
699 Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A
700 discussion of principles. *Journal of hydrology*, *10*(3), pp.282-290. https://doi.org/10.1016/0022-
701 1694(70)90255-6
702
703 Newman, A.J., Clark, M.P., Sampson, K., Wood, A., Hay, L.E., Bock, A., Viger, R.J., Blodgett,
704 D., Brekke, L., Arnold, J.R. and Hopson, T., 2015. Development of a large-sample watershed-
705 scale hydrometeorological data set for the contiguous USA: data set characteristics and
706 assessment of regional variability in hydrologic model performance. Hydrology and Earth
707 System Sciences, 19(1), pp.209-223. https://doi.org/10.5194/hess-19-209-2015
708
709 Saharia, M., Kirstetter, P.E., Vergara, H., Gourley, J.J., Hong, Y. and Giroud, M., 2017.
710 Mapping flash flood severity in the United States. *Journal of Hydrometeorology, 18(2),* pp.397-
711 411. https://doi.org/10.1175/JHM-D-16-0082.1
712

713 Upadhyaya, S.A., Kirstetter, P.E., Gourley, J.J. and Kuligowski, R.J., 2020. On the propagation
714 of satellite precipitation estimation errors: from passive microwave to infrared estimates. Journal
715 of hydrometeorology, 21(6), pp.1367-1381. https://doi.org/10.1175/JHM-D-19-0293.1
716
717 Vergara, H., Kirstetter, P.E., Gourley, J.J., Flamig, Z.L., Hong, Y., Arthur, A. and Kolar, R.,
718 2016. Estimating a-priori kinematic wave model parameters based on regionalization for flash
719 flood forecasting in the Conterminous United States. *Journal of Hydrology*, 541, pp.421-433.
720 https://doi.org/10.1016/j.jhydrol.2016.06.011
721
722 Wang, J., Hong, Y., Li, L., Gourley, J.J., Khan, S.I., Yilmaz, K.K., Adler, R.F., Policelli, F.S.,
723 Habib, S., Irwn, D. and Limaye, A.S., 2011. The coupled routing and excess storage (CREST)
724 distributed hydrological model. *Hydrological Sciences Journal*, 56(1), pp.84-98.
725 https://doi.org/10.1080/02626667.2010.543087
726
727 Woods, D., Kirstetter, P.E., Vergara, H., Duarte, J.A. and Basara, J., 2023. Hydrologic
728 evaluation of the Global Precipitation Measurement Mission over the US: flood peak discharge
729 and duration. Journal of Hydrology, p.129124. https://doi.org/10.1016/j.jhydrol.2023.129124
730
731 Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks,
732 S., Martinaitis, S. and Arthur, A., 2016. Multi-Radar Multi-Sensor (MRMS) quantitative
733 precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological*
734 *Society, 97(4)*, pp.621-638. https://doi.org/10.1175/BAMS-D-14-00174.1
735

736 Zhang, J. and Gourley, J., 2018. (2018). Multi-Radar Multi-Sensor Precipitation Reanalysis
737 (Version 1.0). Open Commons Consortium Environmental Data Commons.
738 https://doi.org/10.25638/EDC.PRECIP.0001