UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INTEGRATED HYDROLOGIC VALIDATION OF SATELLITE PRECIPITATION OVER THE UNITED STATES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

DEVON WOODS

Norman, Oklahoma

2023

INTEGRATED HYDROLOGIC VALIDATION OF SATELLITE PRECIPITATION OVER THE UNITED STATES

A DISSERTATION APPROVED FOR THE

SCHOOL OF CIVIL ENGINEERING AND ENVIROMENTAL SCIENCE

BY THE COMMITTEE CONSISTING OF

Dr. Pierre Kirstetter, Chair

Dr. Jeffrey Basara, Co-Chair

Dr. Randall Kolar

Dr. Jennifer Koch

Dr. Jonathan Gourley

Contents

Acknowledgements

This might be simultaneously the hardest and easiest part of this entire document to write; I honestly do not know where to begin. First and foremost, I want to thank Dr. Pierre Kirstetter and Dr. Jeffrey Basara. Pierre, without your unending guidance and support and belief in my abilities I never would have accomplished as much as I have these past four years. And Jeff, I quite literally would not be here at the University of Oklahoma if not for you; thank you for your immediate enthusiasm and support from the moment I first sent you an email and taking a chance on this tattooed hydrologist from New England. To Dr. JJ Gourley, thank you so much for all you have taught and explained to me these past years on operational and research hydrology, from EF5 to ANCHOR, you've made me a better hydrologist just by proxy. To Dr. Randy Kolar and Dr. Jennifer Koch, I cannot thank you enough for your insight as I've worked my way through my dissertation, I know for a fact my research and my process is better because of it.

Next I'd like to thank the man who without a shadow of a doubt is responsible for making all of this a reality, Dr. Arthur Gold. I know we've talked about it at length already, Art, but this doctrine might be its most significant iteration yet. If I hadn't happened to take your undergrad Hydrology class, after swapping to my third major in three years, I may never have finished my degree. Without your guidance, your reality checks, and your sheer determination to see me succeed, none of this would have been possible. Your eagerness to bring me back into the fold to prove myself for my Master's after growing up and working in the field may have been the most important opportunity provided to me so far, and it is something I will never forget. Thank you, Art, for everything; where I am now would have never occurred to me in my wildest dreams when I first walked into your class ten years ago.

To my friends and my fellow comrades in research here in Oklahoma, I cannot thank you

enough for making this place feel like a home away from home. From B.E.E.R. meetings and conferences to game days and disc golf, there were always opportunities and offerings to give my brain a break for some mental health recharge. I may not be out here in the Midwest for much longer, but I promise that you have not seen the last of me.

To my friends dotted across this country and even several others, a hodgepodge of remarkable and eclectic humans all brought together solely by the internet and Discord, you all helped get me through some unbelievably difficult times both during and in the wake of the pandemic. Your friendship is an incredible gift that I never would have expected, and I am beyond privileged to have met and gotten to know you all over these past few years.

And last, but far from the least, I would never have been able to survive these world-changing ordeals or surpass every new setback and challenge if not for the constant support and encouragement from my friends and my family (and my cat) back home. I honestly don't know what else to say, any words chosen would still be an understatement. Thank you, all of you, from the bottom of my heart. I'll finally be home soon.

Abstract

Increasing precipitation extremes across the globe will naturally lead to not only increased flooding but an increase in the occurrences of flash floods, with assessment and prediction of these events becoming more critical every year. Floods are a natural facet of the water cycle and serve as an unmistakable indicator of changes in that cycle globally. As such, the characterization and monitoring of floods also needs to be undertaken at a global scale especially in the face of an ever-changing climate. A key tool in that process is understanding the inherent capabilities of satellite-based precipitation products in their ability to model flood characteristics. Hence, this work focuses on assessing the innate differences in the high-resolution Multi-Radar Multi-Sensor (MRMS) system and the Integrated Multi-satellitE Retrievals for GPM (IMERG) suite of satellite products when used as precipitation forcings to simulate hydrologic outputs through the operational Ensemble Framework for Flash Flood Forecasting (EF5) hydrologic framework. The biases of precipitation presented by satellite products has been well studied, but less has been done to assess how significant these errors persist into physical hydrologic processes especially when compared to ground-based radar estimates. By using each precipitation product as an individual forcing for EF5, the simulated hydrographs can be post-processed where discrete simulated flood events are determined and matched, assessing the ability of each precipitation product to accurately generate reliable simulated representations of flood characteristics. Relationships between these flood characteristics, such as peak flow and flood duration, and physical basin characteristics are investigated. Steps are also undertaken in reimagining the calculation for a timing characteristic of floods in order to increase the accuracy of the estimates, allowing for future integration into the methodologies presented herein. Increasing knowledge regarding the capabilities or deficiencies

of satellite precipitation products with respect to flash flood modeling will have implications on the ability to characterize flood events over areas with little or no coverage by ground-based precipitation monitoring networks and subsequently improve flood forecasting operations globally.

Chapter 1: Introduction

Floods have always been a devastating force to reckon with across the world. In the Continental United States (CONUS) alone, over 1988 to 2017 increasing precipitation totals and precipitation extremes have contributed to more than $73 billion dollars in damages (Davenport et al., 2021). The occurrence of extreme precipitation events over the CONUS are also projected to increase nonlinearly, subsequently increasing the exposure of more and more of the population to flood hazards (Swain et al., 2020). Even more concerning is that these increasing precipitation extremes will naturally lead to not only increased flooding but an increase in the number of flash floods as well. Assessment and prediction of these events will be critical through the 21st century.

A key tool for hydrologists in understanding the physical processes of floods and basins remains to be hydrologic models. The Ensemble Framework for Flash Flood Forecasting (Flamig et al., 2020), more commonly known as EF5, is a premiere example of this. EF5 is an open-source distributed hydrologic modeling framework that allows the user to utilize several different water balance models and routing schemes to generate hydrologic output data such as discharge, specific discharge (i.e., discharge at each pixel normalized by upstream basin area, also referred to as unit discharge), and return periods, all while offering flexibility in the usable format of precipitation input data. In recent years, EF5 has been combined with the Multi-Radar Multi-Sensor (MRMS) system (Zhang et al., 2016) to create the Flooded Locations And Simulated Hydrographs (FLASH) operational flash flood forecasting network over the CONUS (Gourley et al., 2017). Using 176 operational radars, MRMS provides accurate high-quality data at 1km spatial resolution and 2-minute temporal resolution.

The rest of the world, however, does not have the luxury that the suite of MRMS

precipitation products can provide. As such, those areas with lacking radar coverage instead rely on data provided through satellite precipitation products. The Global Precipitation Measurement Mission (GPM) has within its mission statement the explicit desire to not only improve our understanding of Earth's water cycle, but to also improve the forecasting capabilities for natural hazards such as floods (gpm.nasa.gov, 2020). This dream can be realized in the realm of hydrologic modeling through the use of the Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm (Huffman et al., 2014) and associated global precipitation dataset (half-hourly temporal and 0.1-degree spatial resolution, from 90N to 90S latitude). Previous studies have shown that intercomparison of precipitation products (i.e., satellite vs. ground-based) is possible through the use of ground sensors; for example, Gebregiorgis et al. (2018) used the Gauge-Validated MRMS (GV-MRMS) as reference data (Kirstetter et al., 2012, 2020). Yet there still remains much to learn about the accuracy of these products when dealing with flooding and their application in flood forecasting, especially in order to address the call for "integrated validation" of hydrologic products put forward by Hou et al. (2014). A systematic, statistical assessment is warranted.

Previous studies in this realm have investigated hydrologic modeling using the Tropical Rainfall Measurement Mission (TRMM) and its associated Multisatellite Precipitation Analysis (TMPA) algorithm for singular or clusters of basins with favorable results towards the potential of using satellite precipitation for hydrologic applications (Su et al., 2008; Xue et al., 2013; Yong et al., 2010). Importantly, these studies consistently found benefits while operating at the native TMPA resolutions of 0.25-degree (~27km) spatial resolution and 3-hour temporal resolution, and at a range of significantly different basin areas (from the 3550 km^2 Wangchu Basin in Xue et al., 2013, to the massive 3.2 x 106 km^2 La Plata Basin in Su et al., 2008). All three studies, however, cited basin calibration issues, with Xue et al. (2013) and Yong et al. (2010) adding that while better

statistics were achieved through additional calibration this came at the cost of realistic basin parameters. This sentiment of caution towards overcalibration extends into more recent single-basin studies using GPM and IMERG, although all distinctly agree that the IMERG suite of products perform better for hydrologic applications than TRMM and TMPA (Jiang et al., 2018; Yuan et al., 2018; Zhang et al., 2019). These studies, however, lack the representativeness needed to draw robust conclusions on satellite precipitation products that cover a wide variety of watersheds with their global coverage, all while both resampling IMERG to the more coarse TMPA spatial resolution and using aggregated daily precipitation. Given the variability of precipitation and hydrologic processes at all scales, it is key to perform hydrologic assessment at the native scale of precipitation products in order to draw more robust conclusions relevant to both algorithm developers and end users. One final major caveat to establishing the level of integrated hydrologic validation desired by the GPM mission creators is that traditional assessment approaches become distinctly limited when dealing with diverse catchments and large sample sizes (Clark et al., 2021; Lamontagne et al., 2020; Nanding et al., 2021; Newman et al., 2015). This research hypothesizes that significant understanding of product accuracies and tendencies can be drawn instead from the signals of discrete flood characteristics exhibited by the simulations. However, in order to truly begin to understand larger scale impacts of the use of satellite precipitation products, a reliable, region-scale benchmark representing diverse basin characteristics and hydrological conditions is required.

This proposed research targets the following objectives:

Using the regional perspective of the full CONUS, the first research aspect focuses on establishing a working dataset containing simulated streamflow timeseries generated through EF5 with precipitation forcings from MRMS and IMERG at different temporal and spatial resolutions.

With native resolution MRMS precipitation as a benchmark, we can begin an intercomparison and analysis regarding the accuracies and biases present between the products. This research will serve as one of the first major forays into integrated hydrologic validation of the Global Precipitation Measurement (GPM) mission.

After completion of the dataset, the second phase of research will focus on: (1) hydrologic assessments of simulated hydrograph characteristics between native products; (2) error budgeting of the native products and hydrologic model with respect to ground observations; (3) intercomparison of simulations between resampled products; and (4) an assessment and improvement of lag time calculation for future use as an additional dimension in flash flood studies with this methodology.

Chapter 2: Peak Discharge, Duration, and Timing

Currently published in the Journal of Hydrology

2.1 Introduction

With a greater understanding of the limitations of IMERG and how they translate through hydrologic models, the ability to study flash floods at the global scale becomes significantly easier. EF5 in tandem with IMERG has already been shown to be an effective combination over ungauged locations of Namibia (Clark et al., 2017). Extrapolating this framework and methodology globally allows for the development of improved warning capabilities in areas without high-quality radar data or gauge networks. With a changing climate causing increases in both flood frequencies and intensities globally (IPCC 2013), the importance of gaining the knowledge of hydrologic cycles in these gray areas continues to rise.

This study addresses these limitations through a novel approach. Instead of looking at the behavior of IMERG over a single basin, cluster of basins, or single region, this research looks at the entire CONUS simultaneously. Using a quality-controlled selection of over 3000 well-documented gauged basins, a robust variety of sizes, shapes, terrains, and regimes are incorporated into one dataset. Additionally, in place of only examining outputs generated through the use of one precipitation product, results from IMERG are compared directly against benchmark modeling results from the more accurate and higher resolution MRMS, an already established product used for operational flood forecasting. Comparing IMERG and MRMS through the same hydrologic model allows for the factoring out of model impacts, thus providing a more direct comparison between precipitation products. Importantly, this study is one of the few currently looking to perform integrated validation of IMERG at its native resolution as a means to avoid undesired

resampling effects on the assessment. With both products operating at their native resolutions, baseline assessments can be made about where modeling through IMERG performs well, where it performs poorly, and what conditions are associated with both.

Further novelty comes from the generated streamflow timeseries themselves. Traditionally, hydrologic modeling studies focus on one or more bulk error metrics as a way to quantify and score the accuracy of one entire timeseries to another. This extensive usage comes at the cost of substantially increased uncertainty in the calculation of these metrics themselves, leading to potential contamination of conclusions and murkier definitions of what may be considered a "good" model (Clark et al., 2021). In this study, however, instead of comparing streamflow values directly, the precipitation products are assessed based on their ability to model distinct features of floods (i.e., signals such as peak magnitude, total duration, and timing). This approach not only circumvents the need for bulk metrics, but provides a more unique, robust, and tangible assessment of the differences between the products and their ability to present meaningful flood characteristics at the event scale.

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.

2.2 Data and Methods

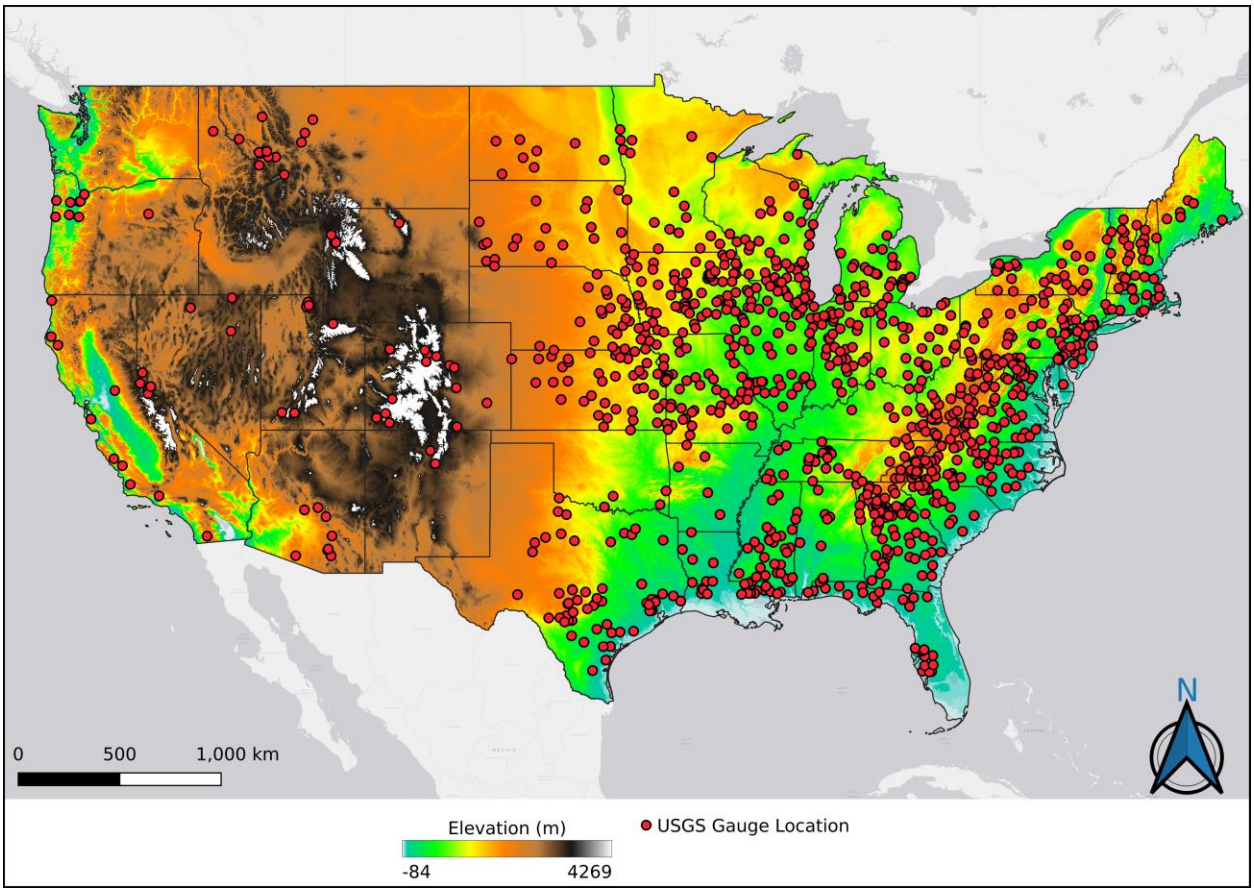

Figure 1. Map of gauge locations utilized across the Continental United States.

For the past several years, an MRMS mosaic precipitation reanalysis over the CONUS between 2004-2011 has been used with EF5 as the basis of several large-scale studies (Zhang and Gourley, 2018; Flamig et al., 2020; Gourley et al., 2017). Not only have these studies shown that EF5 can effectively capture flood hydrographs when using the high-resolution precipitation forcing from MRMS, but they have led to the utilization of EF5/MRMS operationally in real time as part of the FLASH system. Gourley et al. (2017) extensively quality-controlled a selection of over 3000 basins, removing any deemed by the USGS to have any anthropogenic influence as well as any basins where snowmelt processes are dominant (i.e., basins where snowfall contributes to >30% of annual precipitation). This same selection of basins has been used in the current study,

and their locations can be seen in **Figure 1**. Previous studies have also shown the capability of using GV-MRMS as a benchmark to assess the quality of satellite precipitation products (Derin & Kirstetter, 2022; Gebregiorgis et al., 2017; Gebregiorgis et al., 2018; Upadhyaya et al., 2020) but a more in-depth assessment has yet to be completed, especially in the context of flood modeling.

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)

This study focuses on the use of the Version 06 IMERG Early run (IMERG-E) for the satellite forcing as it has the lowest data latency (4 hours) of the suite of IMERG products, giving it the highest potential to be used operationally for flood prediction in a similar way to MRMS FLASH in the future. Each precipitation forcing runs with EF5 using the Coupled Routing and Excess STorage (CREST; Wang et al., 2011) distributed hydrologic model combined with kinematic wave routing (Vergara et al., 2016). The simulations span the same period of time and features the same gauges as those used in Gourley et al. (2017) in order to build an effective control and benchmark. Timeseries are composed of stream discharge values in cubic meters per second (cms).

Traditionally, hydrologic modeling involves calibration and validation through the use of one or more bulk metrics, such as the Nash-Sutcliffe Efficiency (NSE) or the Kling-Gupta Efficiency (KGE) (Nash and Sutcliffe, 1970; Gupta et al., 2009), in order to assess the generated timeseries as a whole across diverse hydrologic conditions and processes. Despite being used extensively in hydrologic applications throughout the years, these performance metrics have been scrutinized of late as their inherent sampling uncertainty cannot be overcome when modeling at large scales (Lamontagne et al., 2020; Newman et al., 2015). Bulk metrics also provide limited insight in hydrologic discrepancies. First, metrics such as correlation and the Kling-Gupta Efficiency are often applied without necessarily checking their applicability or relevance (Gupta et al. 2009; Liu, 2020). For example, the linear correlation generally insufficiently describes non-linear and heteroscedastic dependence between the hydrologic estimates and reference. Second, because model estimates are often assumed to display homogeneous properties over the domain of comparison, bulk metrics are computed over samples that gather a variety of hydrologic situations for which the model is likely to behave differently through its assumptions. Bulk error metrics lack specificity and depict averaged properties hence the representativeness of these hydrologic assessments is confined to the domain over which they are performed, with limited extension over other watersheds, regimes, regions, seasons, etc. These issues are, to an extent, common to those of all geophysical variables (see for instance Kirstetter et al., 2020, for a discussion on precipitation and Radice et al., 2022 for a discussion on atmospheric water vapor).

This is only exacerbated when dealing with the sheer volume of gauges available to be investigated simultaneously across an area as diverse as the CONUS, where errors associated with precipitation across vastly different terrains can propagate even further into the modeling and provide a less robust analysis of the intricacies of the results (Nanding et. al, 2021). This was seen

in a study by Jiang & Bauer-Gottwein (2019), where over 300 basins in China were evaluated simultaneously between the entire suite of IMERG products; hydrologic simulations were shown to have "satisfactory" KGE values (median of 0.60) during calibration, but these metrics plummeted to near-zero or worse during validation. Commentary on these results was made at face value, but due to the inherent nature of the bulk metric calculation little else can be gleamed from the timeseries. Moving away from the confines of bulk metric analysis in hydrologic modeling is a growing sentiment in the greater hydrologic community, with a larger push for it coming after a recent paper by Clark et al. (2021). As such, this study focuses on how well specific flood event characteristics are captured by the hydrologic simulations, instead of analyzing the timeseries as a whole, in order to extract a physically meaningful assessment on the ability of precipitation products to model distinct features of floods.

Additionally, the version of EF5/CREST employed in this study uses the same configuration utilized by the FLASH system for flash flood warning operations in the United States National Weather Service. Although the system was not calibrated against time series of streamflow observations, its parameters were estimated through a robust methodology based on geospatial datasets describing physiographic attributes such as soil types and land cover/use. A machine-learning technique was also employed to estimate hydraulic parameters for the kinematic wave flow routing model. Comprehensive evaluations of this configuration have been presented in Vergara et al. (2016), Gourley et al. (2017) and Flamig et al. (2020).

For this study, every timeseries generated was post-processed in order to isolate individual flood events. Each gauge has a designated "action-level" discharge value set by the United States Geological Survey (USGS) and local stakeholders, which was used 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. An example of how this may look graphically is provided in **Figure 2**, with a zoomed-in look at an arbitrary USGS gauge in Indiana (Gauge 03358000). With each raw event logged by start and end time, events were then procedurally matched one-to-one between the products using multiple levels of cross-referencing criteria. This new dataset of more than 63,000 matched events is the basis of this study.

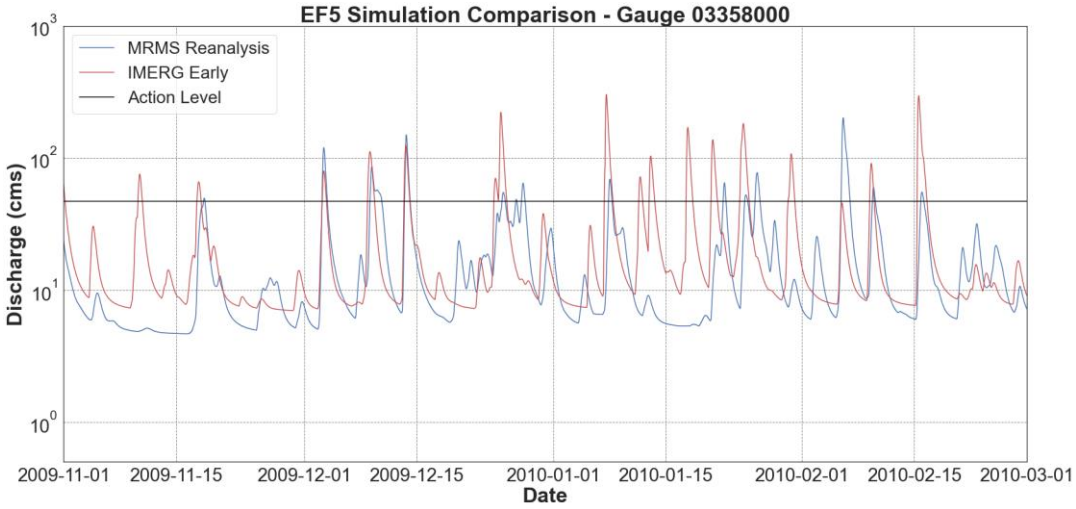

Figure 2. An example of a modeled timeseries comparison, with included USGS action level.

Since all the discharge information is retained for each event, as well as each individual time step, peak discharge values can be obtained as well as relative timings. As such, three aspects of these events will be evaluated in this study: 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), with the idea being that agreement between products on these factors provides a more robust assessment of modeling quality across the study area than traditional methods.

2.3 Results & Discussion

2.3.1 Magnitude (Peak Discharge)

Understanding how well a hydrologic simulation estimates the magnitude of any given flood is significantly important both operationally for emergency management and practically for the purposes of building engineered controls or mitigation strategies. **Figure 3a** shows the accuracy of IMERG simulated peak discharge values plotted against the corresponding MRMS simulated peak discharge values with a density scatterplot in order to assess both the structure of the scatter field as well as the underlying density of points. Note that peak discharge strongly depends on the basin size because the amount of collected water from the sky mechanically increases with size. To filter out this dependence, specific peak discharge (i.e., the peak discharge at a gauge normalized by its associated basin area) was also calculated and plotted to facilitate further depth of comparison between the products **(Figure 3b**).

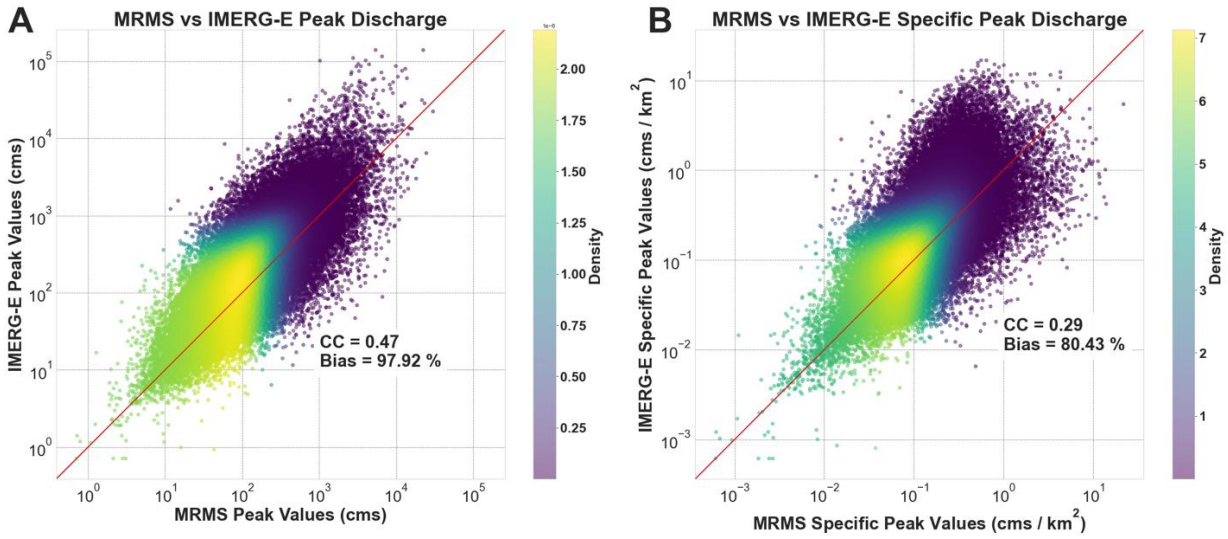

Figure 3. Scatterplots of IMERG peak discharge (A) and specific peak discharge (B) values compared against respective MRMS values. The red diagonal line indicates the 1:1 line.

The IMERG-E peak values show general agreement with MRMS simulated peak discharges, with the pairs gathering around the 1:1 line. In both plots provided in **Figure 3**, overestimation can be seen on the part of IMERG, with the bulk of the densities tending to fall higher above the 1:1 line. Using specific peak discharge, however, brings the data slightly closer to the 1:1 line overall and further centers the densities.

To further dissect the results from the density plots and make the investigation more robust, the data was taken to generate plots of conditional distributions. This particular style of plot examines an independent variable (i.e., IMERG-simulated peak discharge) through quantiles associated with bins of dependent variable values (i.e., MRMS-simulated peak discharge). Quantile plots in Figs. 3a–b display conditional quantiles (10th, 25th, 50th, 75th, 90th) highlighting the variability of the dependency of IMERG-E specific peak discharge conditioned on MRMS specific peak discharge. This exercise was once again performed on both peak discharge (**Figure 4a**) and specific peak discharge values (**Figure 4b**). The conditional median provides the first-order trend of the dependency, while the interquartile area 25th-75th estimates the uncertainty in the relationship and the 10th and 90th quantiles describe the extreme values of IMERG-E specific peak discharge for a given MRMS value.

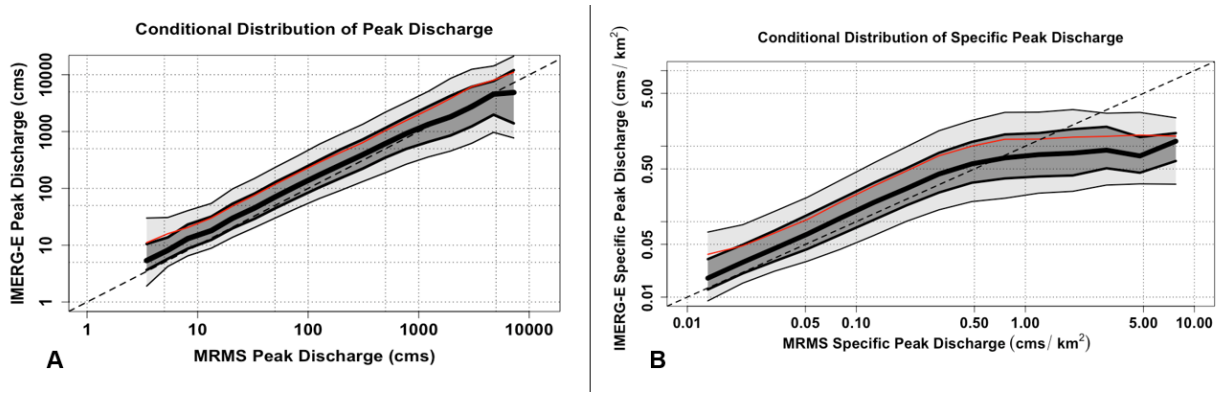

Figure 4. Conditional distribution plots of IMERG peak discharge (A) and specific peak discharge (B) values compared against respective MRMS values. The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 90th and 10th. The red line is the mean value, with the dashed line serving as the 1:1 line.

Similar to what was seen in the density scatterplots, the conditional distribution investigation shows that IMERG tends to overestimate simulated values of peak discharge with most of the distribution of simulated peak discharges, as well as the median and mean lines, falling above the 1:1 line. Note that the conditional distribution of IMERG simulated peaks grow broader as peak discharge increases, showing noticeably greater uncertainties in capturing higher peak discharges. Also worth noting is a distinct plateau effect at the highest MRMS values of specific peak discharge (above 0.5 cms.km-2). These extremely high values are associated with flash floods, often caused by extreme levels of precipitation over small periods of time and over smaller basins. Such events are unlikely to be resolved by IMERG both spatially (at 0.1° resolution) and temporally (30-min resolution). This inability to resolve extreme values of precipitation may also be caused by limitations in the algorithm itself.

2.3.2 Flood Duration

The expected duration of any given flood is another important metric for emergency managers and planners alike. Following the same methodology as the magnitude investigation, a density scatterplot of simulated IMERG-E and MRMS duration values was created first (**Figure 5**). While the points tend to gather around the 1:1 line, the scatter increases for shorter duration events, highlighting challenges for IMERG-E to capture short-lived precipitation events causing floods. The overall spread of the data does tend to reduce as durations increase, but at much lower densities. Once again, it can be seen that IMERG-E tends to overestimate its simulated flood durations, with most of the density of values falling above the 1:1 line.

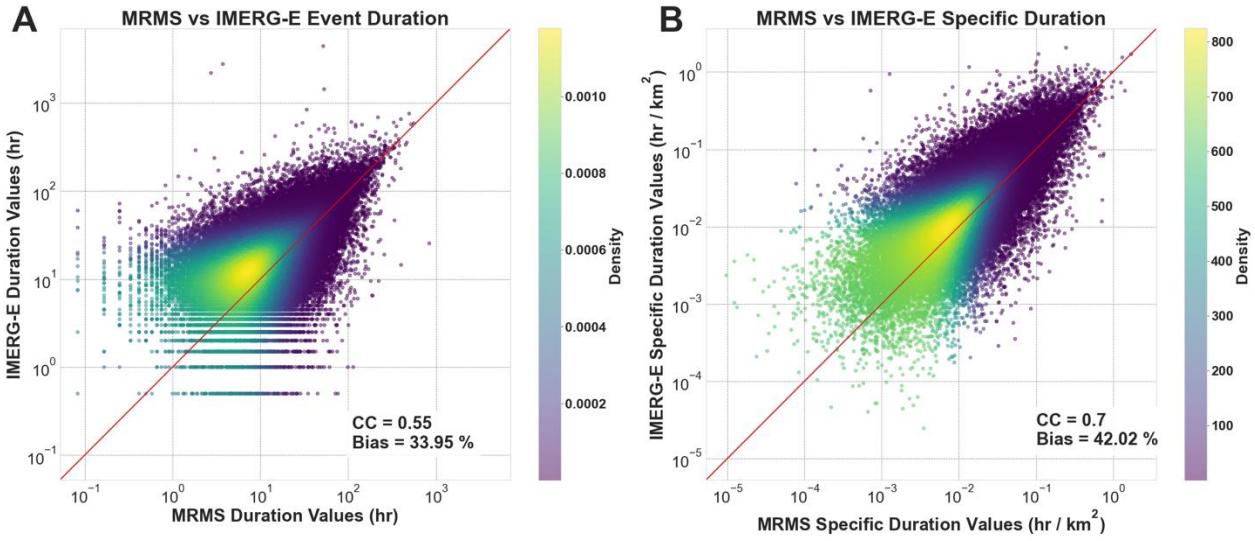

Figure 5. Density scatterplots of calculated durations of IMERG simulated floods (in hours) compared against respective MRMS flood durations (A), and normalized duration values based on associated basin area (B). The red line indicates the 1:1 line.

The conditional distribution of duration values was also assessed and can be seen in **Figure 5**. The results behave similarly to the density plot, with distinct overestimation being shown by

IMERG-E especially at shorter MRMS event durations, but this plotting method also yields additional features.

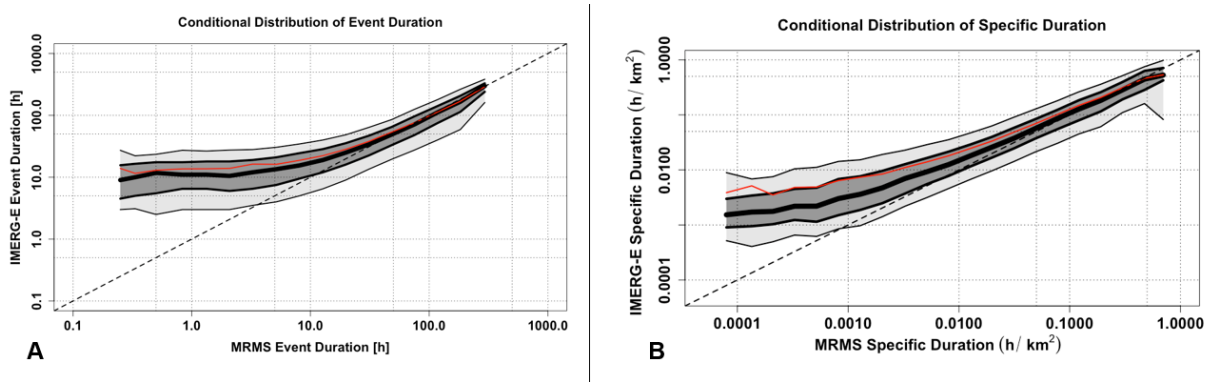

Figure 6. Conditional distribution plots of IMERG simulated event durations compared against respective MRMS durations (A), and normalized durations (B). The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 90th and 10th. The red line is the mean value, with the dashed line serving as the 1:1 line.

What can be seen more clearly in the conditional distribution plot is that IMERG-E shows distinct bias associated with overestimation of shorter flood durations (< 5h), but this bias decreases as flood duration times increase. The spread of the quantiles also decreases as duration increases, meaning increased agreement between products and decreased uncertainty between them. This again could be a result of the IMERG spatial resolution, with potentially more rain falling within the confines of the larger IMERG grids, but this discrepancy is more likely caused by the coarser 30-min temporal resolution of IMERG with respect to MRMS, which this study resolves at 5 minutes and can be resolved operationally as low as 2 minutes. The longer the duration, the less this temporal difference becomes a factor. It is worth noting that all the exact same trends are seen in **Figure 6B**, in which durations were normalized by their respective basin

area (in the same way as peak discharge) in order to remove any potential influence on the results.

2.3.3 Flood Timing

The ability to determine the difference in IMERG-E event timing with respect to MRMS has also provided interesting insights, especially some critical for the flood forecasting community. Since the dataset provides timestamps for both the start time and end time of each event (i.e., the point where the flow exceeds then subsequently falls below the gauge's action level), the time differences between the products are computed and analyzed. The absolute IMERG-E start and end times were subtracted from their associated absolute MRMS start and end times, giving either a positive or negative time difference value in hours. A positive (negative) value indicates that the IMERG-E event occurs earlier (later) than its MRMS absolute time counterpart. **Figure 7** shows that overall IMERG-E tends to simulate floods that begin earlier and end later than those simulated by MRMS. This tracks with what was seen with event duration, as one would expect to have longer duration floods if simulations are generating floods that both start earlier and end later. Like duration, this is likely also caused by differences in both spatial and temporal resolutions.

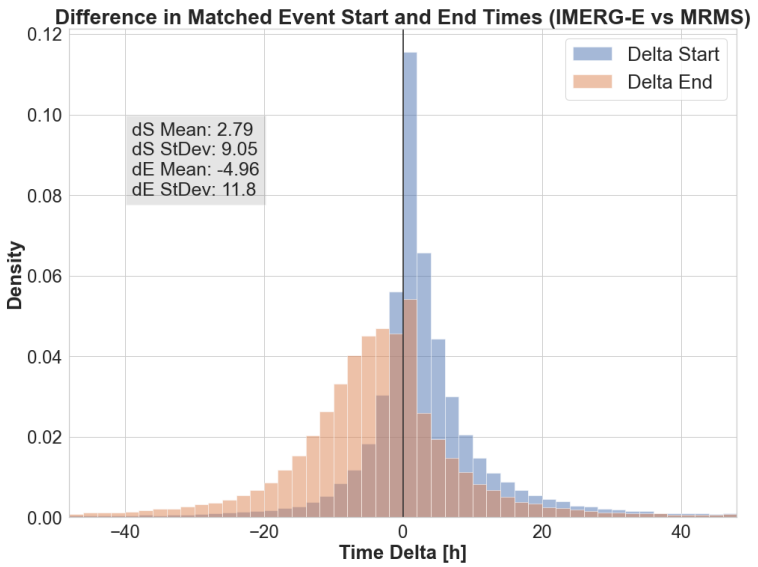

Figure 7. Histogram of the time deltas of matched flood start times (blue) and end times (orange) between IMERG-E and MRMS simulations, with associated means and standard deviations.

Once again, conditional distribution plots were generated in order to better understand underlying associations of hydrologic conditions within the data. The size of a basin naturally ties into flood timing, so area was chosen to compare against start time (**Figure 8A**) and end time (**Figure 8B**). At their core, both plots show the same results as seen in **Figure 7**, with positive values associated with earlier IMERG-E start time in **Figure 8A** and more negative values associated with later IMERG-E end times in **Figure 8B**. It can also be seen in both plots that these respective positive and negative features tend to be accentuated at the smallest and largest basin sizes, with the latter being noticeably more severe. These effects are once again likely due to the coarser resolutions of IMERG-E, but for different reasons. At smaller basin sizes (< 100 km^2), the resolution of IMERG-E is simply too large to generate the more necessary precise precipitation-flood responses. Specifically, such basins are covered by a few IMERG pixels given the nominal spatial resolution of IMERG (0.1°). For larger basins (> 100 km^2), the impact of the smoothing of

the precipitation fields by the IMERG effective resolution (Guilloteau and Foufoula-Georgiou, 2020) is likely aggregated at the basin scale, which translates into higher systematic departures and increased uncertainty in the flood timings. As basin size increases, the significance of the precipitation spatial distribution increases, so by "painting with a large brush", so to speak, more uncertainty is added into the system as seen by the increasing spread of the quantiles.

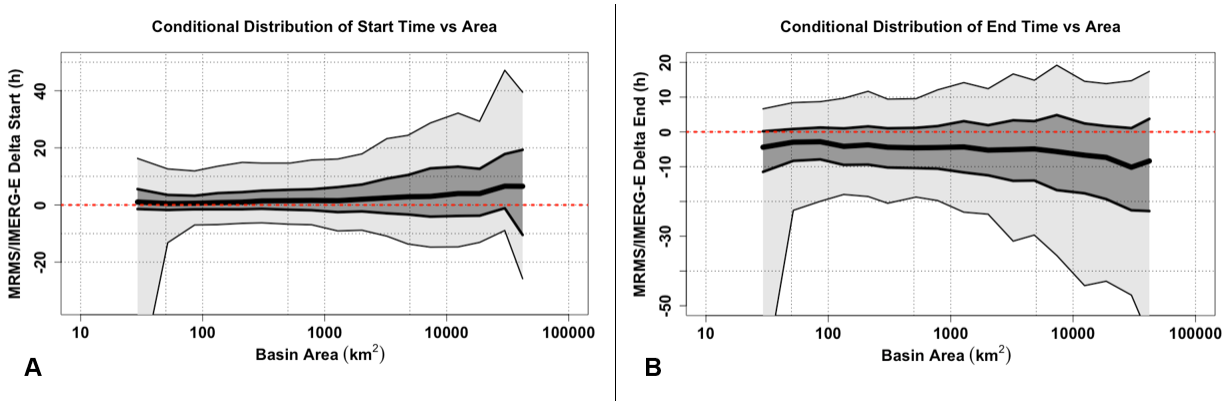

Figure 8. Conditional distribution plots of calculated event delta start (A) and delta end (B) times compared against associated basin areas. The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 90th and 10th. The dashed red line is the zero line, signifying matching timing of events. A positive (negative) value indicates that the IMERG-E event occurs earlier (later) than its MRMS absolute time counterpart.

Despite the uncertainty seen at larger basin sizes, however, the start time discrepancies in **Figure 8A** remain well centered around zero. End times fall farther away from the zero-hour center line in **Figure 8B**, with approximately 75% of the total events ending later than MRMS. Overall, start time is the more accurate of the two parameters, with the median line only slightly deviating from zero at larger basin sizes as well as maintaining tighter uncertainty bounds compared to end

times. This would be welcome news for flood forecasters, as predicting the start time of a flood event is significantly more important than when it ends. That point aside, the end times are still fairly well represented, with the median remaining within 5h at smaller basins despite wider uncertainty bounds overall compared to start times. This increased uncertainty is likely also a byproduct of the coarser and effective IMERG-E resolutions, with potentially larger volumes of precipitation falling over larger periods of time, generating longer receding limbs as the additional water works its way through the system.

2.4 Conclusions

In this study, IMERG-E satellite precipitation was used as a forcing in the EF5 hydrologic model framework in order to better understand its ability to accurately simulate and predict floods and flash floods characteristics. Through a nontraditional assessment of event-based flood characteristics (magnitude, duration, and timing) it was found that more work needs to be done to tune the IMERG algorithm for direct hydrologic applications and studies. IMERG-E was shown to generally overestimate peak discharge values up until a point when it can no longer resolve the highest and most extreme cases of precipitation likely associated with flash floods over small watersheds, leading to severe underestimation of peak values associated with flash floods. The duration of floods simulated by IMERG-E also tend to be longer than those simulated by the higher-resolution MRMS, exacerbated by the product's tendency to simulate earlier start times and later end times. All of this information is crucial and useful to both current and future end users from hydrologists to emergency managers and forecasters, but more work is necessary to further flesh out these results especially when it comes to flood timing.

With a greater understanding of the limitations of IMERG and how they translate through

hydrologic models, the ability to study flash floods at the global scale becomes significantly easier. EF5 in tandem with IMERG has already been shown to be an effective combination over ungauged locations of Namibia (Clark et al., 2017). Extrapolating this framework and methodology globally allows for the development of improved warning capabilities in areas without high-quality radar data or gauge networks. With a changing climate causing increases in both flood frequencies and intensities globally (IPCC 2013), the importance of gaining the knowledge of hydrologic cycles in these gray areas continues to rise. Upcoming research aims to use this same methodology to generate intercomparisons between the entire suite of IMERG products (Early, Late, and Final) as well as how different versions of the algorithm itself compare.

Chapter 3: Error Budget and Model Analyses

Currently accepted and in production in the Journal of Hydrology

3.1 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 (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.

3.2 Data and Methods

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 precipitation forcing data. Each precipitation forcing was run with EF5 using the Coupled Routing and Excess STorage (CREST; Wang et al., 2011) distributed hydrologic model combined with kinematic wave routing (Vergara et al., 2016). This scheme of EF5/CREST is the same configuration utilized by the FLASH system for flash flood warning operations in the United States National Weather Service and is built off extensive geospatial datasets of parameters which remove the need for timeseries-centered model calibration (Vergara et al., 2016; Gourley et al, 2017; Flamig et al., 2020).

Using a previously extensively quality-controlled selection of over 3000 gauges (Gourley et al., 2017), simulations were run across the CONUS for both precipitation forcings from 2004 to 2011. United States Geological Survey (USGS) data for each gauge was also taken for the time period simulated. Each time series was post-processed in order to isolate individual flood events based on its designated USGS "action-level" discharge value. 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. Each raw event was then procedurally matched one-to-one between the products and the USGS observations, respectively, using multiple levels of cross-referencing criteria. This new and representative dataset of more than 20,000 matched events per product serves as the basis of this study.

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).

3.3 Results & Discussion

3.3.1 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 9** 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 9a** and **9c** display peak discharge values whereas **Figures 9b** and **9d** show specific peak discharge. Note that specific peak discharge was calculated and provided as a means to filter out the natural dependence of peak discharge values with basin area; it is also a vital metric when

dealing with flash floods.

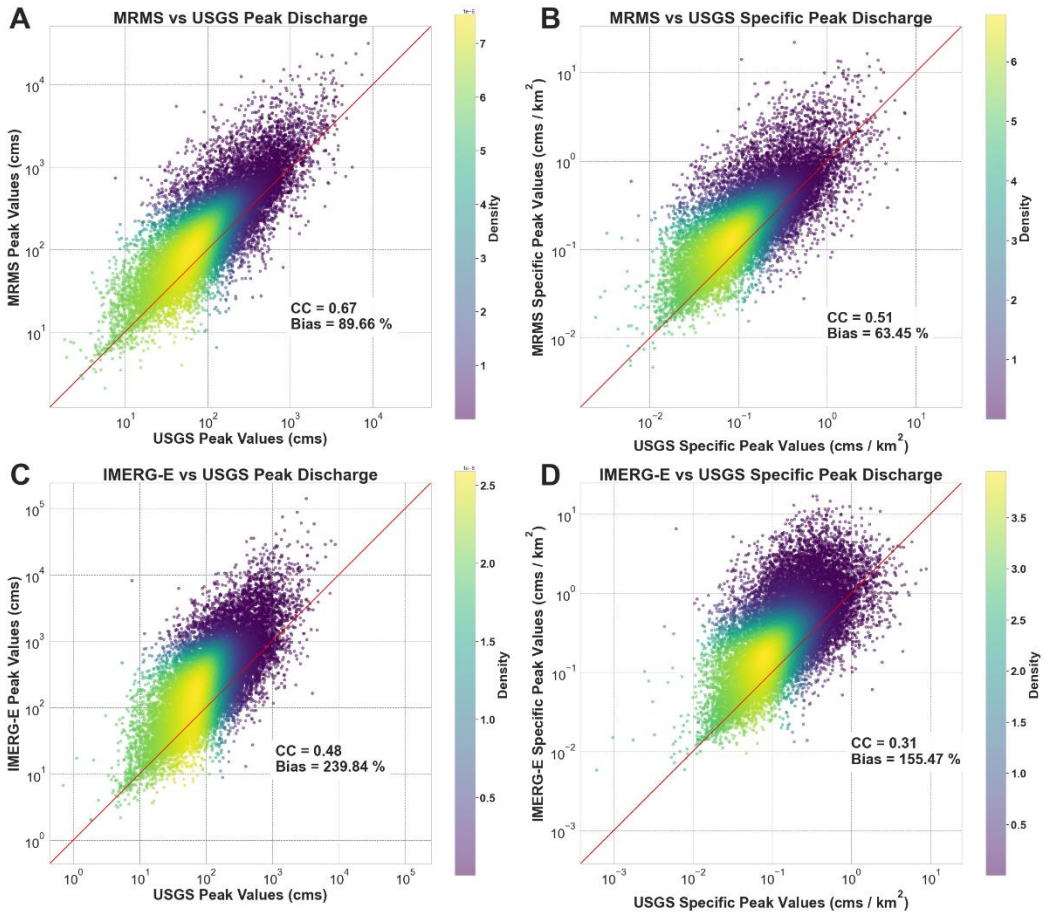

Figure 9. Scatterplots of MRMS peak discharge (A), MRMS specific peak discharge (B), IMERG-E peak discharge (C) and IMERG-E specific peak discharge values compared against USGS reference values. The red diagonal line indicates the 1:1 line.

While the points tend to gather around the one-to-one line, a distinct conditional bias can be seen across both products and discharge types, with an increasing overestimation of higher (specific) discharges. Both MRMS and IMERG-E overestimate with respect to USGS, 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 10**). 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.

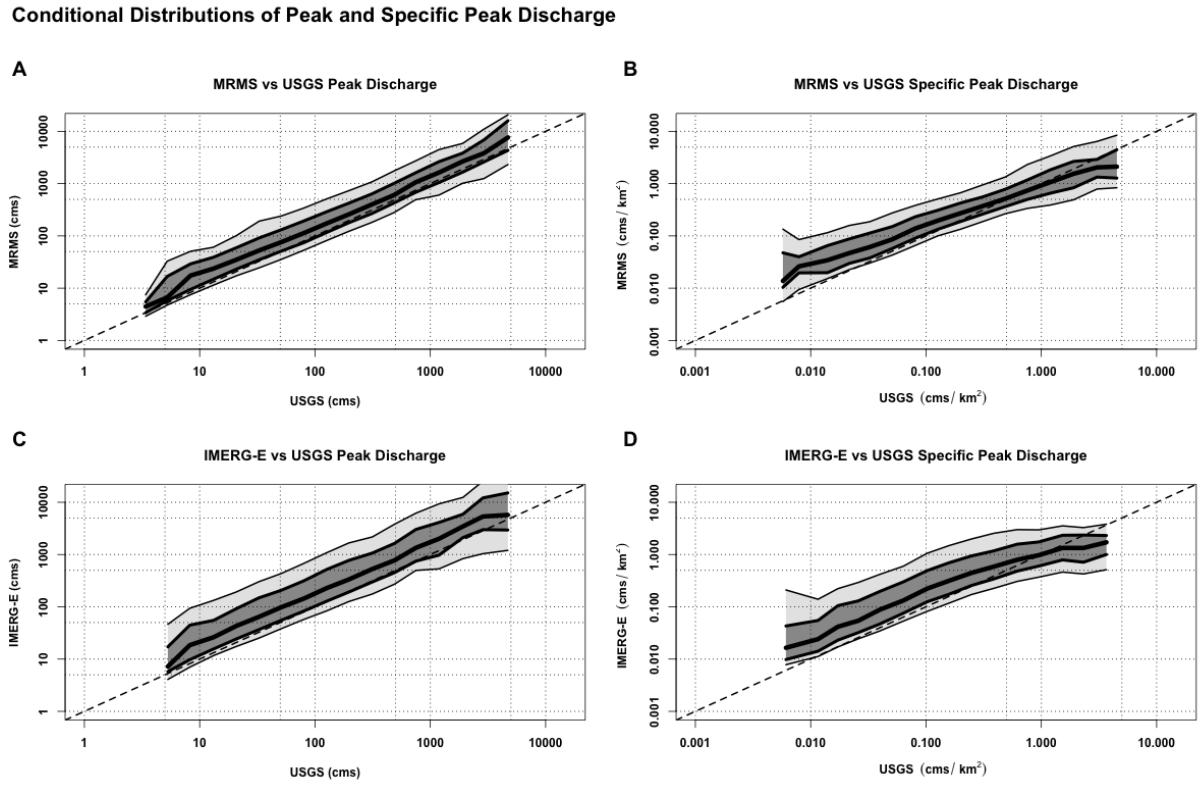

Figure 10. Conditional distribution plots of MRMS and IMERG-E peak discharges (A and C) and specific peak discharges (B and D) compared against USGS references. The thick center line

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 is the 1:1 line.

The conditional distribution investigation in **Figure 10** reiterates what was seen in the density scatterplots: distinct overestimation on the part of both MRMS and IMERG-E with respect to the USGS observations. Again, as expected, the uncertainties associated with MRMS (i.e., the overall spread of the quantiles) are smaller than those associated with IMERG-E; the effects of resolution certainly play a role here. Interesting to note, however, is how the specific peak discharge of both products (**Figure 10b** and **Figure 10d**) trend from overestimation at lower values towards the 1:1 line and eventually into slight underestimation at the highest values to the point where IMERG-E begins to plateau out. This plateau effect was similarly seen in Woods et al. (2023a) and attributed to the coarser spatial and temporal resolutions of IMERG, with these resolutions prohibiting the algorithm's ability to resolve the highest levels of instantaneous precipitation and therefore being unable to resolve the highest specific peak discharges often associated with them. Seeing the effect appear when compared to the gauged USGS reference corroborates this idea, suggesting that the shortcoming lies within the ability of IMERG to resolve the highest values and locations of extreme precipitation events (i.e., those responsible for flash floods associated with these high specific peak discharges) as opposed to errors generated within the hydrologic model itself. This plateau effect can also slightly be seen in the MRMS specific discharge plot, a phenomenon likely associated with the time period of this study. It was found that the MRMS algorithm employed in the single-polarization reanalysis, the product used in this methodology, generated underestimation associated with the highest rainfall rates – though the most recent version of this algorithm (utilizing present day dual-polarization data) has been shown

to mitigate these biases (Gourley and Vergara, 2021; Zhang et al., 2020).

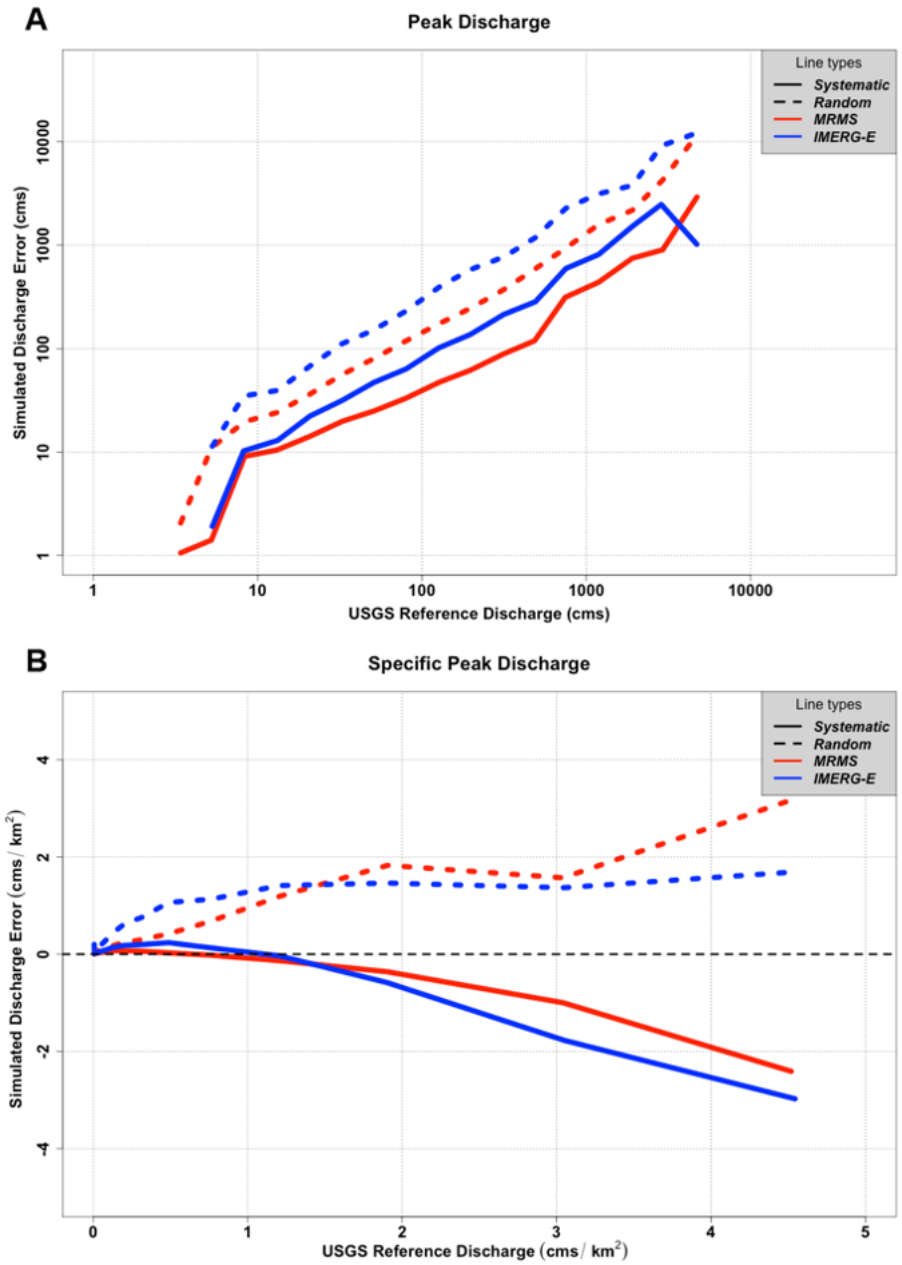

Figure 11. Error calculations for simulated flood peak discharge and specific peak discharge from MRMS (red) and IMERG-E (blue) with respect to USGS. Solid lines represent systematic error while dashed lines represent random error.

Building upon the quantile analysis, as well as to further inform on the abilities of the products, an error analysis was conducted (**Figure 11**). For both products, and for both discharge types, the systematic error (simulated median minus observed median) and random error (75^{th} quantile minus 25^{th} quantile) were calculated and plotted against the USGS reference values. In **Figure 11a,** distinct increasing trends in systematic (positive bias) and random error are seen for both MRMS and IMERG-E 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 consistently showing higher systematic and random biases compared to MRMS.

When looking at specific peak discharge (**Figure 11b**) similar stories can be seen. While both products now trend into underestimation of specific peak discharges compared to USGS, IMERG-E still shows more negative systematic bias than 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 IMERG-E random error. 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 rainfall events normally associated with these extreme values of specific discharge better than IMERG can, naturally leading to increased random error in the system. The underestimation of extreme events present in this version of the MRMS algorithm may also be contributing to these higher random error values. It is worth noting that accuracies in flash flood discharge estimation have been shown to improve significantly as precipitation products become more sophisticated (Gourley and Vergara, 2021), so future research is warranted to better dissect and diagnose the behavior of EF5 with post-2011 MRMS values for a more in-depth understanding of potential model deficiencies.

3.3.2 Flood Duration

Further critical to emergency management efforts and flood operations is an understanding of the expected duration of a flooding event, real or simulated. As such, the analyses utilized for peak discharge were also undertaken for simulated flood duration. First, density scatterplots were created and can be found in **Figure 12**. As with discharge, event durations were normalized by basin area to generate specific duration values as an additional method of assessment. What can be seen is surprising; overall, MRMS simulations of floods tend to underestimate their durations with respect to their USGS counterparts. Longer flood durations are increasingly underestimated (conditional bias). This conditional bias is related to basin size, as it is less significant with unit flood durations (see also **Figure 13b** and **13d**). This is likely explained by the routing scheme used; 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 (in **Figures 12c** and **12d**) is

also interesting, with durations being closer to the 1:1 line with respect to USGS than MRMS. 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 by IMERG-E's propensity to overestimate precipitation durations and resulting floods.

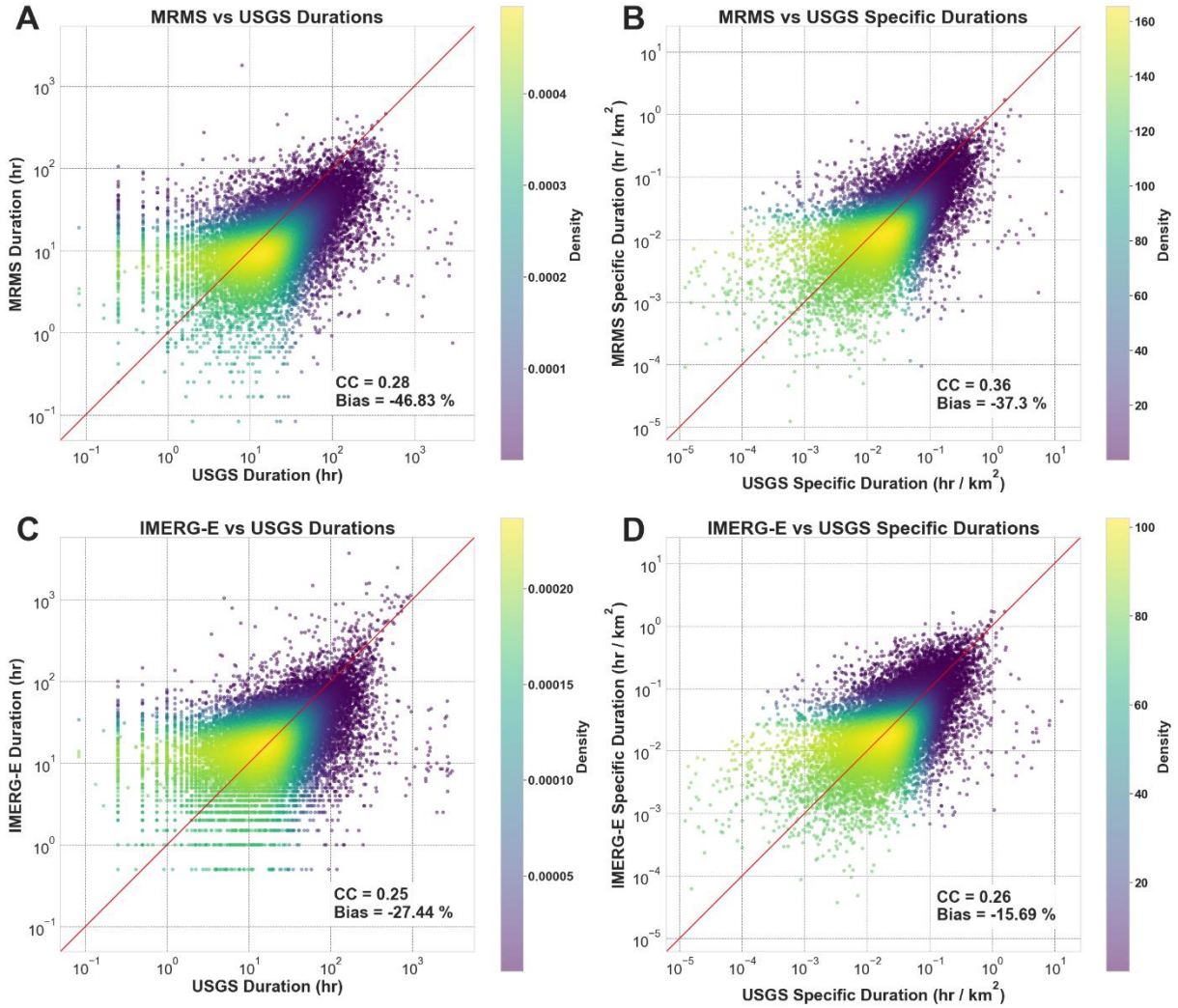

Figure 12. Density scatterplots of MRMS and IMERG-E calculated flood durations (A and C) and normalized duration values based on associated basin area (B and D), all plotted against USGS references. The red line indicates the 1:1 line.

The conditional distribution plots (**Figure 13**) tell a similar tale, with noticeable underestimations seen for both products, but several additional features can be extracted. For instance, despite the core of MRMS durations in the density plot showing underestimation, there are distinct regions of overestimation at the shortest of flood durations (<5 hr). This feature is consistent across both products as well as both duration types, as well as both products trending from overestimation to underestimation as flood durations increase. Unlike with peak discharge, however, there is no noticeable difference in error spread between MRMS durations and IMERG durations with respect to USGS. Both products also behave similarly when normalized by basin area, though with a somewhat closer spread of quantiles from MRMS simulations. This is more consistent with expectations regarding the higher resolutions associated with MRMS.

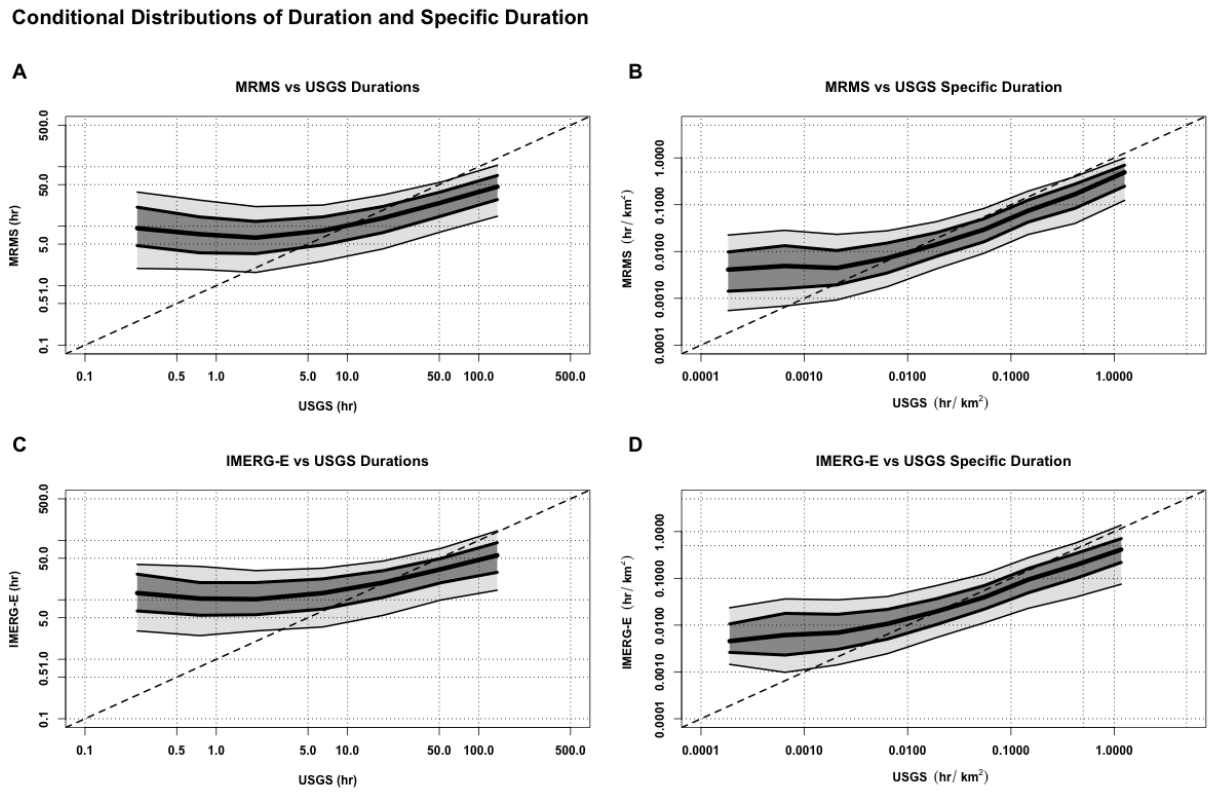

Figure 13. Conditional distribution plots of MRMS and IMERG-E simulated event durations (A and C) and normalized duration values (B and D), all plotted against USGS references. The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, 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 14**. When looking at the duration of events (**Figure 14a**), 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 than MRMS, a byproduct of the difference in product resolution. Systematic error is a different story; IMERG-E overestimates more than MRMS at shorter durations (again, a factor of resolution) but at longer durations MRMS is the product with higher underestimation. This corroborates what was seen in the density scatterplots (**Figure 12**) where IMERG-E simulated durations fall closer on the 1:1 line with respect to USGS than MRMS simulated durations.

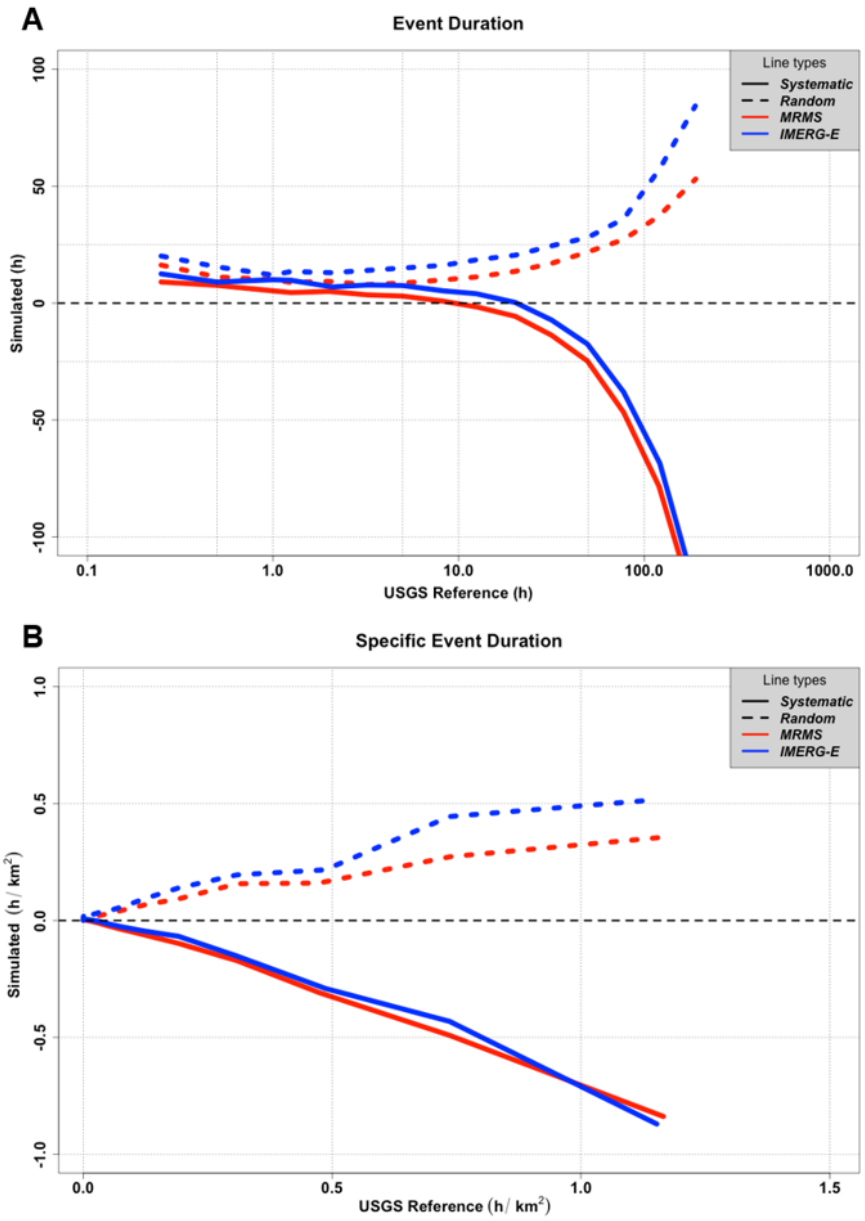

Figure 14. Error calculations for simulated flood durations and specific durations from MRMS (red) and IMERG-E (blue) with respect to USGS. Solid lines represent systematic error while dashed lines represent random error.

The errors associated with specific duration (**Figure 14b**) largely mirror what was seen with duration; the systematic error of IMERG-E remains slightly less negative than MRMS while

the random error of IMERG-E remains higher than 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.

3.3.3 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 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 15**. For both MRMS and IMERG-E 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 values also have an average start time closer to zero and with a smaller standard deviation than 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).

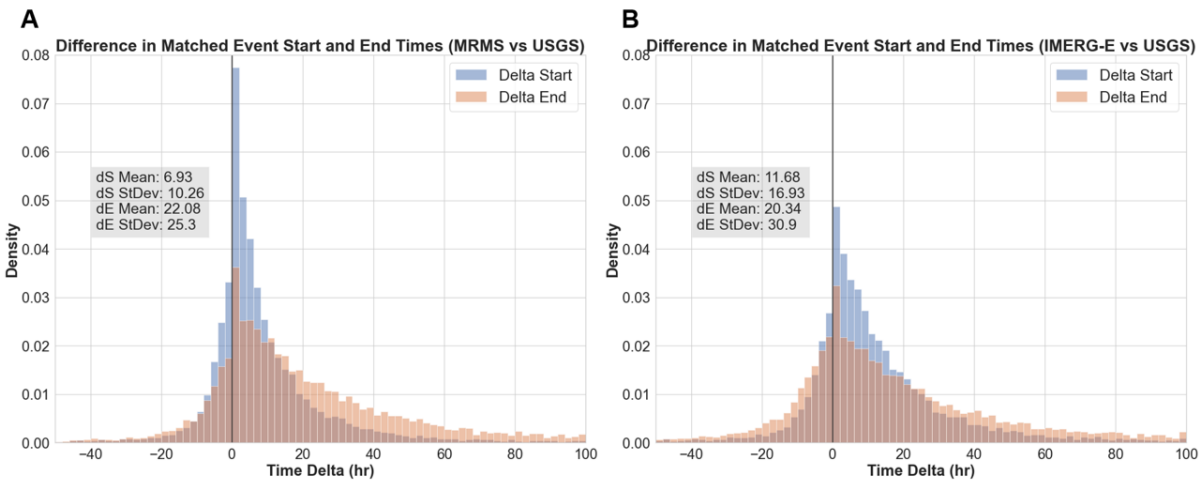

Figure 15. Histograms of the time deltas of matched flood start times (blue) and end times (orange) for MRMS, IMERG-E, and USGS, with associated means and standard deviations. A positive (negative) value indicates that the simulated event (MRMS or IMERG-E) event occurs earlier (later) than its USGS absolute time counterpart.

In investigating the conditional distribution plots, found in **Figure 16**, 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 15**, 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. IMERG-E is shown to have significantly higher extreme error quantiles associated with smaller basin sizes than 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 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 at large basins suggests routing from the model itself is likely also a contributor in this case.

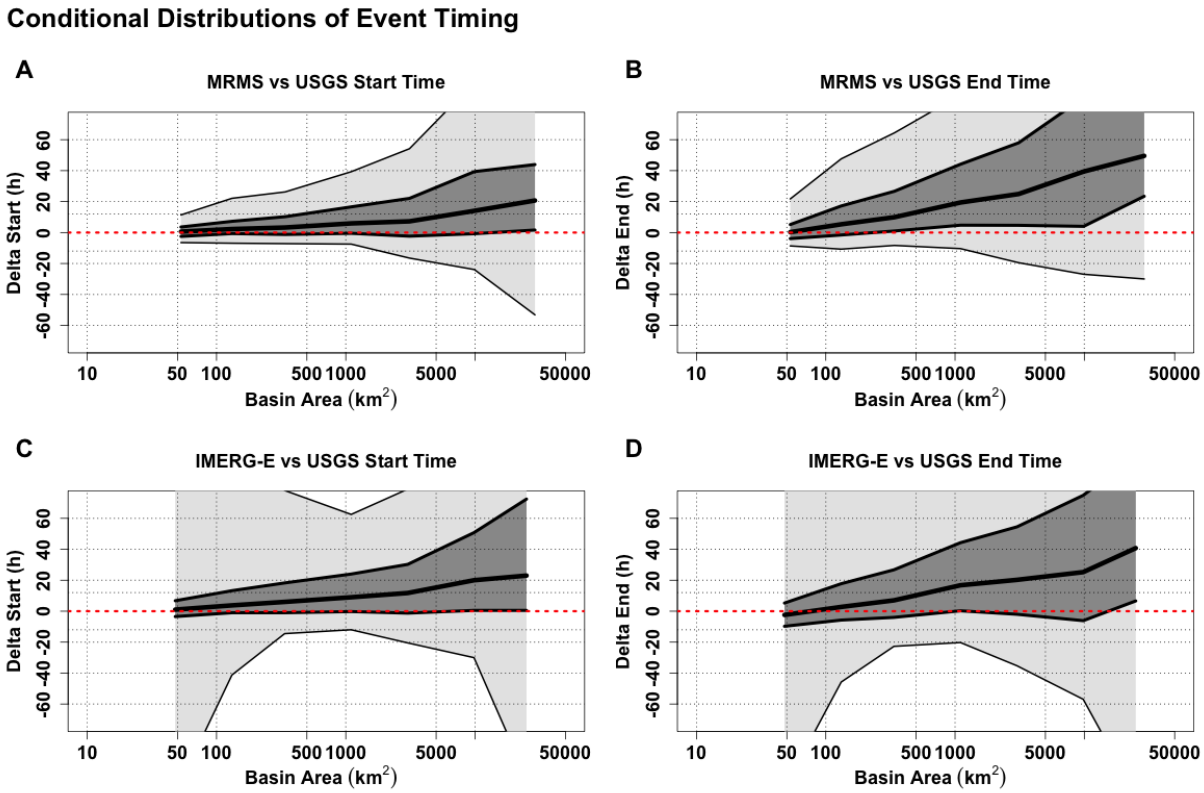

Figure 16. Conditional distribution plots of calculated event delta start (A and C) and delta end (B and D) times compared against associated basin areas. The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 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.

Despite increasing uncertainty with basin size (as well as significant extreme quantiles associated with IMERG-E), median values and $25^{th}/75^{th}$ error quantiles remain noticeably less expansive for both products at areas <1000 km^2. End time values lose effectiveness sooner, before reaching 500 km^2 in both cases, and the overall spread ends noticeably wider than final spreads for start time. Regardless, event start time is inherently a more important statistic to predict accurately more often, especially in the case of flood forecasting and emergency response.

The error budgets of the products with regard to event timing (**Figure 17**) are in agreement with overall trends seen throughout this analysis but are able to provide important insight into accuracies at different scales. Before discussion, however, it is important to establish an understanding of what timing error means in this context. Throughout this section, the positive and negative deltas have been associated with absolute times. With regards to error, this instead translates to positive values signifying an overall trend towards earlier times (both start and end) while negative values signify an overall trend towards later times. As can be seen across both time delta plots, the overwhelming majority of errors for both products tend to push start and end times earlier than USGS. This effect is likely caused by routing within the EF5 model, with water more likely to flow faster through the system (especially at larger basin areas) than more slowly. For end times there also exists a small window at basins < 100 km^2 where IMERG-E has negative systematic error values, meaning that at smaller basins IMERG-E tends to try to pull end times later. Overall, this suggests that there is an inherent competition between routing and resolution being exhibited; this trend to counteract end times and extend the total duration of events ties into what was seen in the previous section (Section 3.2) and **Figure 12**, where IMERG-E produces

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

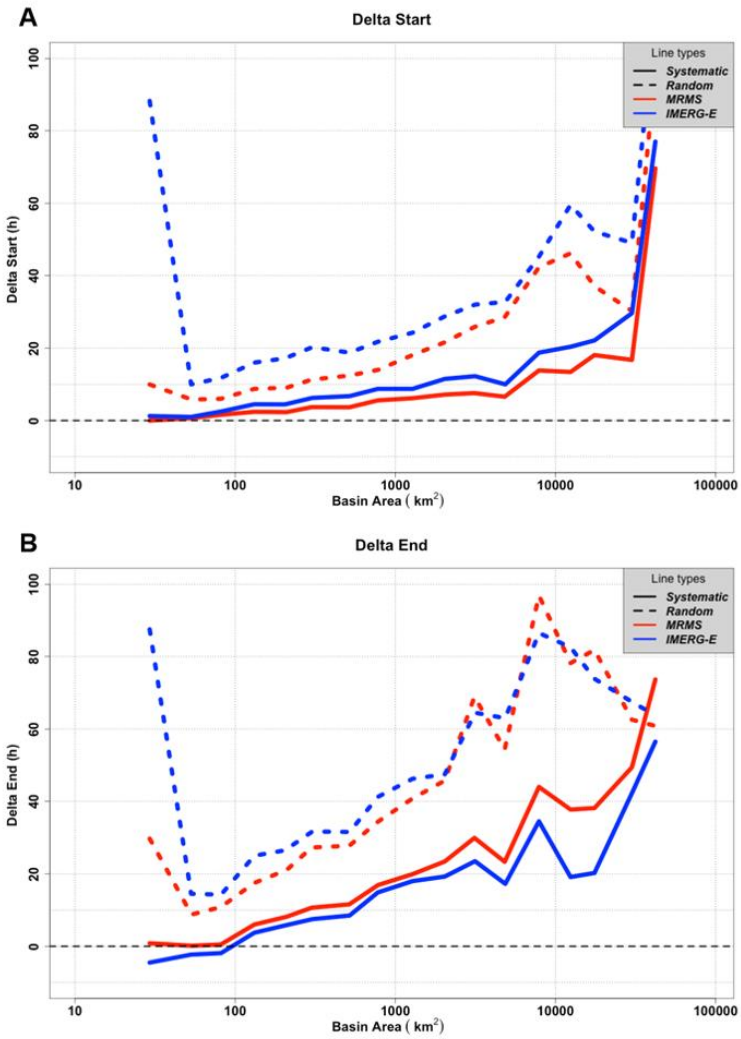

Figure 17. Error calculations for start time and end time deltas from MRMS (red) and IMERG-E (blue) events with respect to USGS, plotted against associated basin area. Solid lines represent systematic error while dashed lines represent random error.

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

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, with MRMS showing higher systematic errors at all basin sizes. MRMS still maintains 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 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).

3.3.4 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 model and focus solely on the influence of the products themselves. The ability to include USGS data as the true reference in this study allows for a more robust analysis and diagnosis of both hydrologic outputs and model tendencies, benefitting extensively from what was found in the simulation-only research.

In order to characterize the joint peak and duration errors that can be influenced by the

model itself, observed values for peak discharge and flood duration were subtracted from their respective simulated values and the resulting discharge and duration deltas were plotted (**Figure 18**), with each error quadrant signifying a different hydrologic tendency within the model. Points in the top left quadrant (positive peak deltas and negative duration deltas) indicate simulated floods with higher peaks and shorter durations than USGS, a signal of influence from kinematic wave where the water is being pushed through the system too quickly. In the top right quadrant (positive peak deltas and positive duration deltas) points are found where both the peak and the duration are higher than USGS, indicating positive water balance errors (i.e. there is too much water in the system, with greater areas under the theoretical hydrograph). The bottom left quadrant (both negative deltas) is again dominated by water balance, but instead with too little water simulated. The bottom right quadrant shows simulations with smaller peaks but longer durations than the reference, signifying flood attenuation by the model.

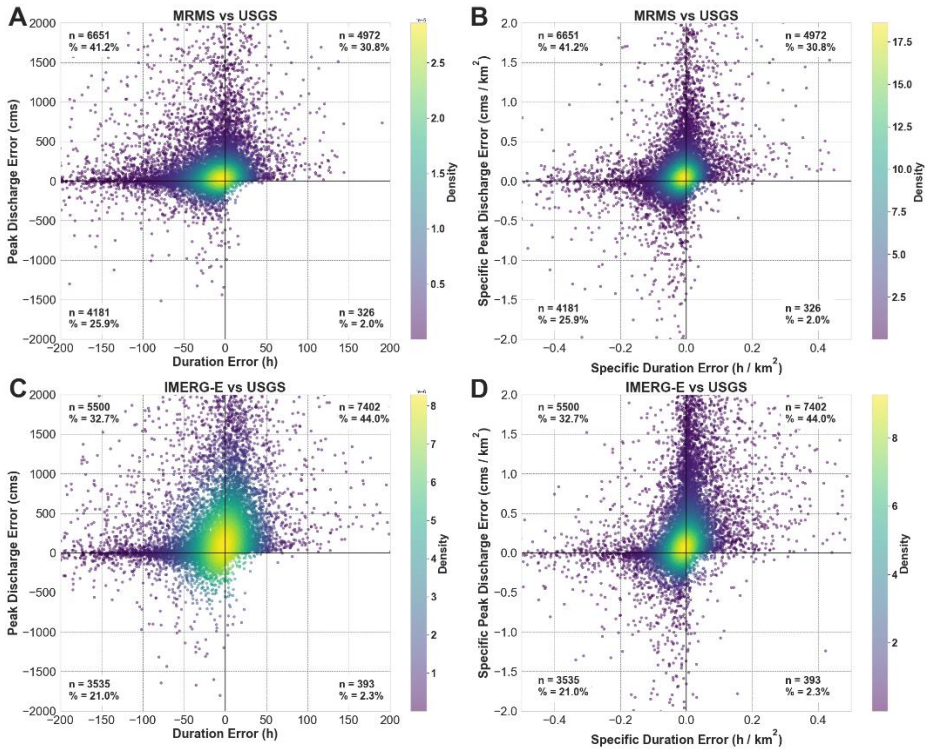

Figure 18. Density scatterplots of discharge and duration errors for MRMS and IMERG-E simulations with respect to USGS observations. Total numbers of points in each quadrant are provided, as well as each quadrant's percentage of the total points.

In the MRMS plots, the highest percentage of points fall into the top left quadrant (41.2%), highlighting increased influence on simulations by the kinematic wave scheme. This corroborates what has been seen throughout this study, where MRMS simulations are routinely more likely to underestimate flood durations than IMERG-E. There is influence from the water balance dominated quadrants as well (56.7 %), meaning there are discrepancies with how or where water is entering the system. In the case of IMERG-E, these quadrants are where the majority of points are found (65%), with most falling into positive water balance error (44%). For IMERG-E this is to be expected because coarser spatial and temporal resolutions naturally tend to add excess water to the system through a combination of both smoothing over larger pixel sizes and more limited accuracy in precipitation values themselves, leading to hydrographs that are taller and longer than those of USGS. Kinematic wave is still a factor, but the increased tendency towards water balance overestimation counteracts its effects and explains why IMERG-E maintains lower systematic errors in duration and flood timing than MRMS. Additionally, neither product had a significant number of points in the bottom right quadrant, reiterating that the physics of the model performs well, and that flood attenuation is not a factor here.

These results show that FLASH/EF5's model design choice on kinematic wave was correct because for the overwhelming majority of the territory the assumptions of this model apply. The fact that the highest densities are near the (0,0) point speaks well of the modeling system. Such small numbers of points are seen on the bottom right quadrant due to several factors. First,

kinematic wave does not have as much capability to attenuate the flood wave at higher resolutions; it can, however, if the pixel resolution is coarser, which is a result of numerical diffusion/attenuation (i.e., an artifact of the numerical approximation). Second, because for most of the terrain over the CONUS, kinematic wave applies. And third, because most of the basins and subsequent events being considered in this study do not have the geomorphology and hydraulics necessary to lead to significant flood attenuation.

In order to determine if there were any additional unforeseen tendencies within the model, the same approach was taken by contrasting the MRMS and IMERG-E simulations themselves. MRMS values were subtracted from IMERG-E, and the same discharge-duration plots are provided in **Figure 19**. As expected, almost all of the points fall within the water balance quadrants, with the distinct majority in overestimation (60.8%). When the influence of the model itself is removed, the effects of resolution difference between precipitation products is expected to be dominant; IMERG again naturally puts more water into the system than its higher-resolution counterpart. There is still influence from underestimation, however, likely caused by a combination of spatial variability and variability in the accuracy of precipitation estimates, which in turn is exacerbated by the algorithm's smoothing of rainfall itself (i.e, the correct volume of rainfall is not always falling over the right area or basin).

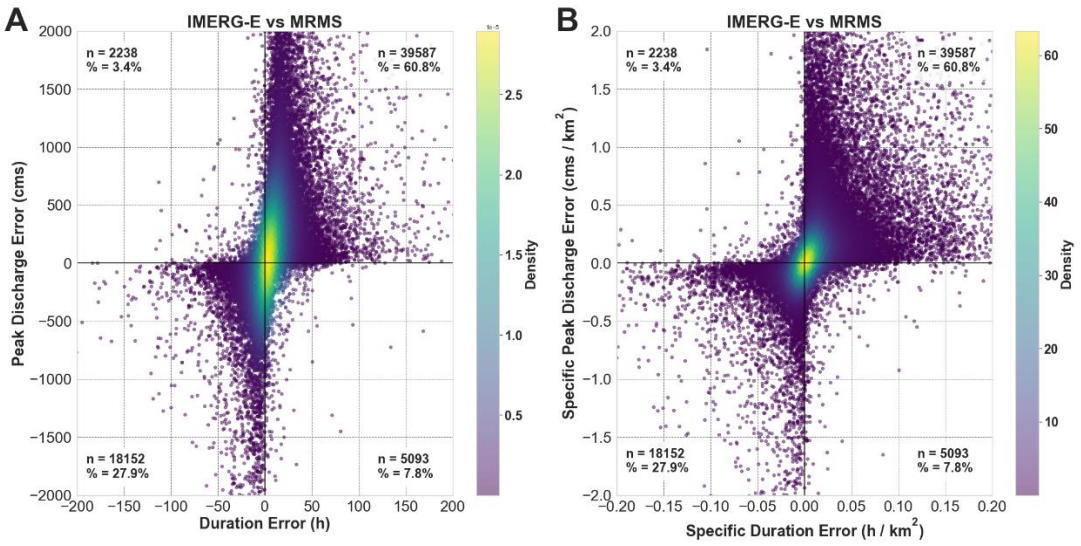

Figure 19. Density scatterplots of discharge and duration errors for IMERG-E with respect to MRMS. Total numbers of points in each quadrant are provided, as well as each quadrant's percentage of the total points.

Between duration (**Figure 19a**) and specific duration (**Figure 19b**) themselves, the plots behave similarly, though there is a more asymptotic spread across the duration scatter than specific duration. Both plots maintain higher densities closer to the (0,0) point, with that spread becoming even tighter when normalized by basin area.

3.4 Conclusions

In this study, precipitation forcings from IMERG-E and MRMS were run through the EF5 hydrologic modeling framework, broken down into discrete flood characteristics (magnitude, 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 were shown to overestimate values with respect to the USGS reference, with IMERG-E peak values being attributed to greater uncertainties. IMERG-E was also shown to have more difficulty resolving higher-end 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 also had consistently lower systematic and random errors than IMERG-E, with the exception of specific peak discharge where MRMS was higher.

When looking at the simulated flood durations, interesting interactions surfaced: MRMS consistently underestimated durations with respect to USGS, with underestimation further increasing with basin size, while IMERG-E was 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 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 having a higher systematic error than MRMS at smaller basin sizes but transferring to a less negative error than MRMS as basin size increases. Overall, however, IMERG-E retained higher random errors than MRMS 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 is shown to have significantly higher extreme error quantiles associated with smaller basin sizes than MRMS, 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 showcases both higher systematic and higher random error values than MRMS. At larger basin sizes, these errors shown by IMERG-E can be attributed to systematic biases and uncertainty caused by basin-scale aggregations, but similar trends from MRMS 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 resolutions naturally tend to add excess water to the system, leading to hydrographs that are taller and longer than those of USGS. The increased tendency towards water balance overestimation counteracts the tendency of kinematic wave to push water through the system too quickly and explains why IMERG-E maintains lower systematic errors in duration and flood timing than MRMS. Additionally, it was shown across both products that the physics of the model performs well, and that flood attenuation is not a factor in the results.

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

Chapter 4: Impact of Scaling on Flood Characteristics

4.1 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. EF5 has been well established in research and operations as a viable modeling tool and has been combined 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.

Without reliable radar coverage globally, however, researchers and forecasters in these darker areas 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 (Huffman et al., 2014). To date, a great deal of research and focus has been undertaken with the intent to assess and intercompare the returns of satellite precipitation products relative to those generated 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 only in recent years has the need for “integrated hydrologic validation” of GPM (Hou et al., 2014) been addressed at large within the community.

A preliminary dive into the realm of integrated validation was made in Woods et al. (2023a) where MRMS and IMERG at their native resolutions were used as precipitation forcings through EF5, and their extracted flood characteristics were directly compared using the MRMS-forced simulations as a benchmark. In a follow-up study, both product simulations were compared instead to observational data gathered from the United States Geological Survey (USGS) in order to develop an understanding of the error tendencies of the simulations. Both approaches were made with the broader calls of the hydrologic community in mind, with the use of model and output assessment methods less reliant on "bulk metrics". This idea was championed by Clark et al. (2021), as "traditional" approaches to hydrologic model assessment 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).

There are still areas to shed light on along this research avenue, however, chief of which being how the differences in spatiotemporal scaling between these satellite and ground-based precipitation products impacts the subsequent accuracy of hydrologic model outputs. It is no surprise that the resolution of the precipitation field is inherently tied to the accuracy of hydrologic simulations (Seo et al., 2023; Vergara et al., 2014). Previous studies have shown that downscaling of the coarser resolution product is the more simple but effective approach to benefit hydrologic model accuracy (Kay et al., 2023; Ma et al., 2018), but it has also been shown that when working across larger spatial and temporal scales upscaling will actually serve to reduce errors in rain rate fields from satellite precipitation products (Tan et al., 2017). Little information exists, however, within the hydrologic literature on the impacts of upscaling the higher-resolution reference precipitation data instead of downscaling the coarser resolution precipitation data, and it is here

where the novelty of this current study flourishes.

Building upon the results from and assessments made in both previous studies (Woods et al., 2023a and 2023b), this research aims to further develop those findings within the context of spatiotemporal scaling. Looking laterally and independently at how spatial, temporal, and spatiotemporal upscaling of MRMS precipitation data to native IMERG resolutions as a means of hydrologic model forcings provides a comprehensive look at which facets of the product's resolutions impact streamflow simulations the most, an invaluable asset for the broader puzzle of integrated hydrologic validation. This endeavor is aided by the heterogeneity of having a CONUS-wide expanse of reference gauges, allowing for a truly robust assessment of the behaviors of these simulations and their associated discrete flood characteristics (i.e., peak magnitude, flood duration, and event timing) at different resolutions.

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.

4.2 Data and Methods

Numerous large-scale studies utilizing a CONUS-wide MRMS precipitation reanalysis dataset (Zhang and Gourley, 2018; Flamig et al., 2020; Gourley et al., 2017) have been used as a basis for integrated hydrologic validation studies (Woods et al., 2023a; Woods et al., 2023b) featuring extensive use of the EF5 hydrologic modeling framework. 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). More importantly, EF5 also allows the user flexibility in the format of its input precipitation forcing data. Woods et al. (2023a) 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. Each precipitation forcing was run with EF5 using the Coupled Routing and Excess STorage (CREST; Wang et al., 2011) distributed hydrologic model combined with kinematic wave routing (Vergara et al., 2016). This scheme of EF5/CREST is the same configuration utilized by the FLASH system for flash flood warning operations in the United States National Weather Service and is built off extensive geospatial datasets of parameters which remove the need for timeseries-centered model calibration (Vergara et al., 2016; Gourley et al, 2017; Flamig et al., 2020). Woods et al. (2023b) then dug deeper and compared the simulations from both products against observational data retrieved from the United States Geological Survey (USGS) in order to not only provide and develop error budgets for the products but to also use this as a means to investigate tendencies and deficiencies within the EF5 model itself.

The selection of over 3000 gauges used for this study have been extensively quality-controlled (Gourley et al., 2017), where any gauges deemed by the USGS to have any anthropogenic influence as well as any basins where snowmelt processes are dominant (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), as well as three resampled versions of MRMS precipitation

(1-km spatial and 30-min temporal, 10-km spatial and 5-min temporal, and 10-km spatial and 30-min temporal), from the time period of 2004 to 2011.

For this exercise, MRMS rainfall rate fields at 1-km and 5-min resolution were processed to emulate IMERG's 10-km, 30-min accumulation resolution. The Geospatial Data Abstraction Library (GDAL; GDAL 2022) was used for the spatial resampling. The resampling method used was the "average" method, which computes the weighted average of all valid data points (i.e., all non-NODATA contributing pixels) within the 10-km IMERG pixel. To accumulate MRMS rainfall data, the fields were first converted from rates to totals for each 5-min interval. Accumulations were then computed through summation of the 5-min fields within the 30-min interval covered by each IMERG data file. In cases where missing MRMS 5-min data files existed, estimates from data files immediately prior or after the missing interval were used. In cases with no immediate data file availability, an inverse weighted average interpolation was used to fill gaps (see **Figure 20** below). For gaps larger than one hour (of 5-min MRMS files), no interpolation was executed, and so those 30-min IMERG-like estimates are deemed as No Data.

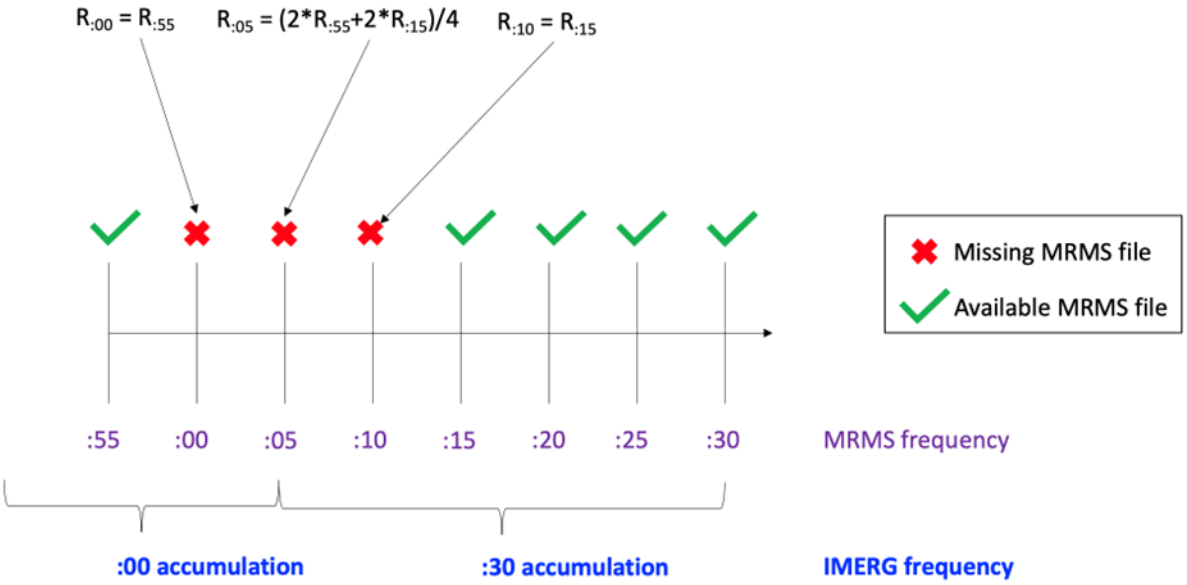

Figure 20. Schematic of gap-filling methodology for MRMS estimates accumulations to 30-min intervals.

After simulations for each product were complete, each time series was post-processed in order 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 attempted to be procedurally matched one-to-one between the products and the USGS observations, respectively, using multiple levels of cross-referencing criteria such as the amount of total overlap of a simulated event with respect to an observed event and the proximity of start and end times of events that do not overlap. These

criteria also serve 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. When all is said and done, each individual simulated event that could be successfully matched to an individual observed event generates a fixed pair of overall peak discharge values, respective event durations, and overall event start and end times while the remaining unmatched events are archived.

Maintaining a focus on new methods of model assessment (Clark et al., 2021), this study continues the path put forward by Woods et al. (2023a,b) in investigating discrete flood characteristics provided by the simulations themselves as opposed to more "traditional" methods of hydrologic evaluation such as the Nash-Sutcliffe Efficiency (NSE) and Kling-Gupta Efficiency (KGE) bulk metrics (Nash and Sutcliffe, 1970; Gupta et al., 2009). This base set of flood characteristics includes 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).

4.3 Results and Discussion

4.3.1 Magnitude (Peak Discharge)

As has been done in Chapters 2 and 3, the first focus of the results will be on the differences in flood peak discharges between the native resolutions and the resamples. **Figure 21** shows how each resampled MRMS product simulation fares when compared to the native resolution IMERG-E simulation, while **Figure 22** shows the resampled product simulations compared against

simulations from native MRMS resolutions. What is immediately striking in **Figure 21** is how the structure of all four plots is functionally the same. There is some fluctuation within the density structure itself as evidenced by the bias and correlation statistics, however. Correlation only slightly improves when moving to coarser spatial resolutions (0.47 to 0.49), while bias significantly grows at coarser temporal resolutions (97% to 140%).

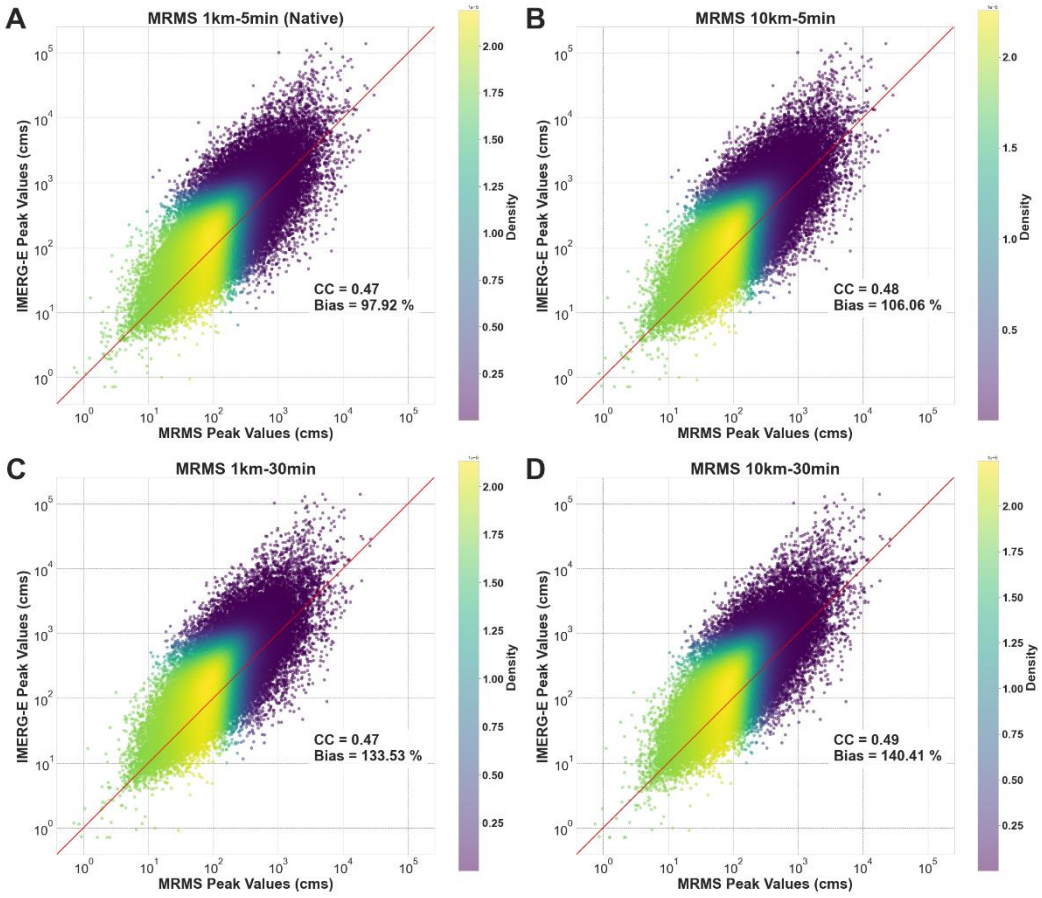

Figure 21. Scatterplots of IMERG-E peak discharge compared to: (A) MRMS Native peak discharge, (B) MRMS 10km-5min resampled peak discharge, (C) MRMS 1km-30min resampled peak discharge, and (D) MRMS 10km-30min resampled peak discharge. Pearson correlation coefficient and percent bias statistics are provided. The red diagonal line indicates the 1:1 line.

When looking at the MRMS native vs MRMS resampled plots, however (**Figure 22**), some underlying features begin to appear. As one would expect, comparing between the same sensing product results in extremely correlated distributions (compared to the IMERG-E vs MRMS native simulation plot for reference in **22A**) but allows for a better view of the effects of the resampled resolutions on the simulations. **Figure 22B** has the lowest bias of the resampled simulations, suggesting minimal impact on peak discharge values when upscaling to coarser spatial resolutions within this methodology, although the data spread is slightly wider than in **22C** which maintained the 1km spatial resolution of native MRMS. In **Figures 22C** and **22D**, where upscaling of temporal resolutions were added, a noticeable shift in bias appears. By resampling MRMS to the coarser 30-min resolution of IMERG-E, the simulations generated by the resampled products begin to underestimate peak discharge values with respect to the simulations forced by native MRMS. The increase in data spread from the 10-km spatial resolution can also be seen in **22D**. A likely explanation for this underestimation is that the temporal resampling procedure results in negative effects to the overall water balance of the forcing/model system, but this will be further investigated in Section 4.3.4.

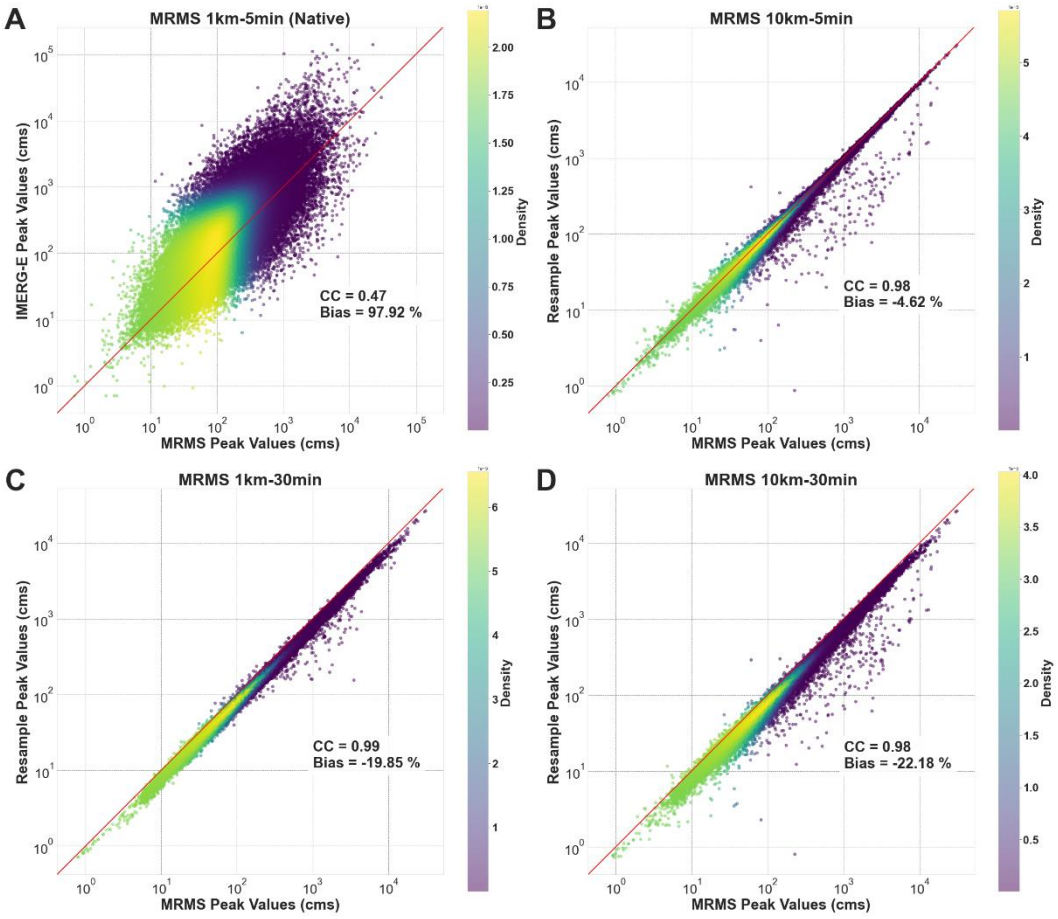

Figure 22. Scatterplots of MRMS peak discharge compared to: (A) IMERG-E Native peak discharge, (B) MRMS 10km-5min resampled peak discharge, (C) MRMS 1km-30min resampled peak discharge, and (D) MRMS 10km-30min resampled peak discharge. Pearson correlation coefficient and percent bias statistics are provided. The red diagonal line indicates the 1:1 line.

Moving on to conditional distributions, it is clear from **Figure 23** just how similar the structures of the density plots are between the IMERG-E, MRMS native, and the MRMS resampled simulations. It is increasingly likely that the inherent conditional biases and errors within IMERG are completely masking the errors and biases coming from the resampled MRMS product simulations. It is interesting to note, however, that the median line reaches the 1:1 line

sooner in both **23A** and **23B**, showing a slight tendency in **23C-D** of increased IMERG-E simulation overestimation at peak discharges above 1000 cms.

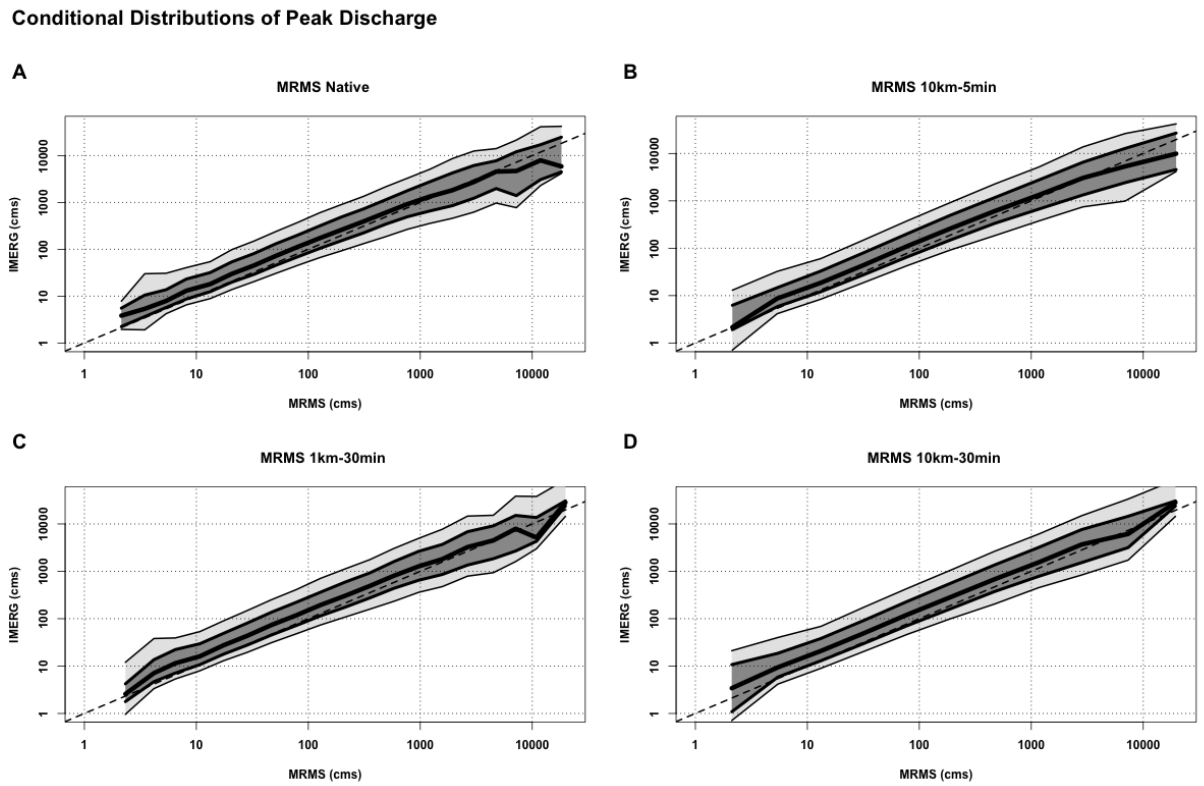

Figure 23. Conditional distribution plots of IMERG-E peak discharge compared to: (A) MRMS Native peak discharge, (B) MRMS 10km-5min resampled peak discharge, (C) MRMS 1km-30min resampled peak discharge, and (D) MRMS 10km-30min resampled peak discharge. The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 90th and 10th. The dashed line is the 1:1 line.

This behavior likely pairs with the increasing underestimation at higher peak discharges showcased by the simulations from the MRMS resampled products in **Figures 24C-D**, but the

signal is much less clear from the IMERG-E plots. The effect of the increased data spread from the spatially resampled simulations can also be seen in **Figures 24B-D**, but it is unclear if or how much that translates into the IMERG-E comparisons.

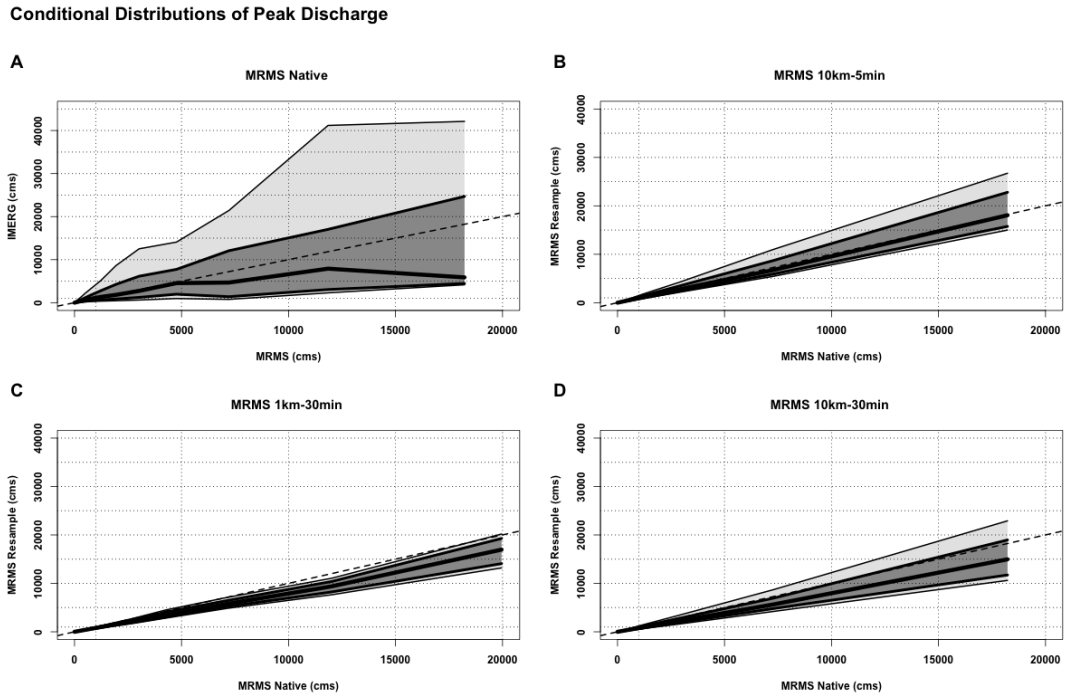

Figure 24. Conditional distribution plots of MRMS peak discharge compared to: (A) IMERG-E Native peak discharge, (B) MRMS 10km-5min resampled peak discharge, (C) MRMS 1km-30min resampled peak discharge, and (D) MRMS 10km-30min resampled peak discharge. The thick center line shows the 50th quantile (median), with the dark grey section extending to the 75th and 25th quantiles, then light gray to the 90th and 10th. The dashed line is the 1:1 line.

The bias masking exhibited by the IMERG-E simulations does not appear to be influenced by basin area, as seen by the specific discharge plots in **Figure 25**. The same general behaviors

are seen as in the peak discharge plots, with a slight improvement in simulation correlation when resampled spatially and a larger increase in simulation bias when resampled temporally.

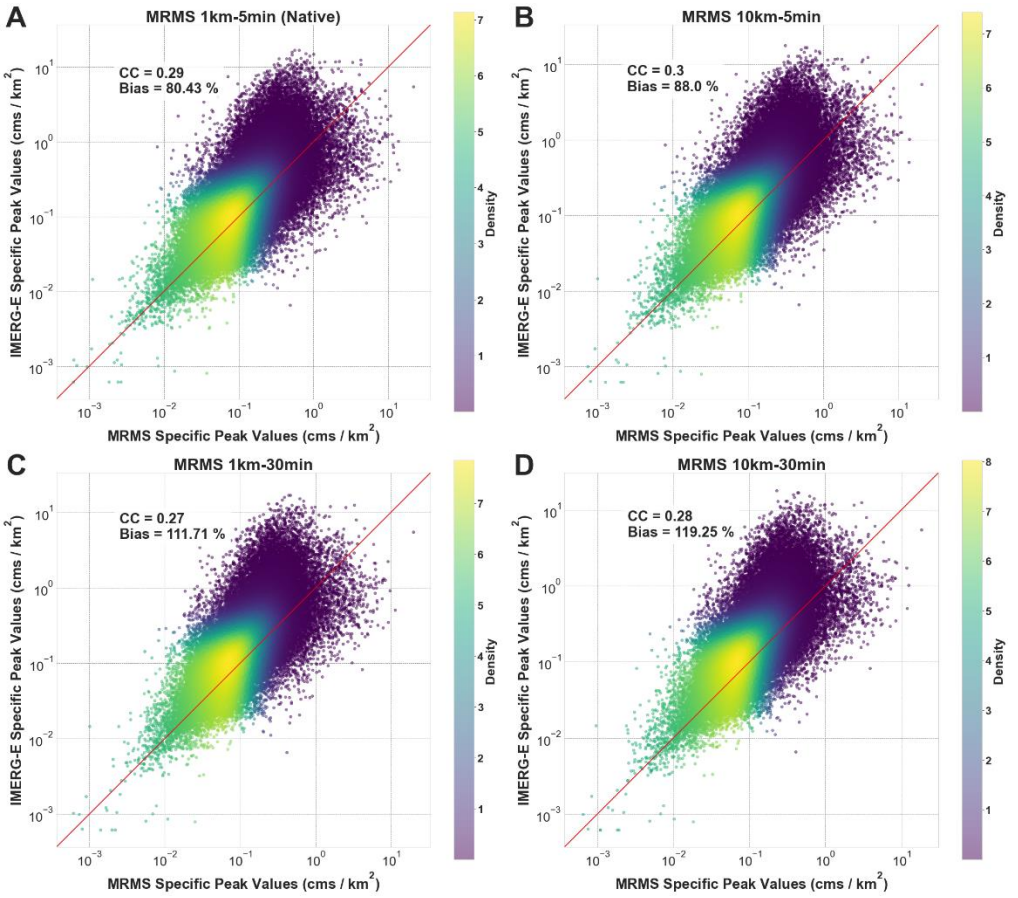

Figure 25. Scatterplots of IMERG-E specific peak discharge compared to: (A) MRMS Native specific peak discharge, (B) MRMS 10km-5min resampled specific peak discharge, (C) MRMS 1km-30min resampled specific peak discharge, and (D) MRMS 10km-30min resampled specific peak discharge. Pearson correlation coefficient and percent bias statistics are provided. The red diagonal line indicates the 1:1 line.

This remains true for the MRMS native vs MRMS resampled simulations, with the bias being added into the system for the simulations that were resampled temporally. In **Figure 26**, however, to showcase the different structures of the data a natural scale was elected to be used instead of a logarithmic scale. So while the shift in bias was still seen in **Figure 22**, a clearer picture of the skew generated by the temporal resampling can be seen in the natural scale.

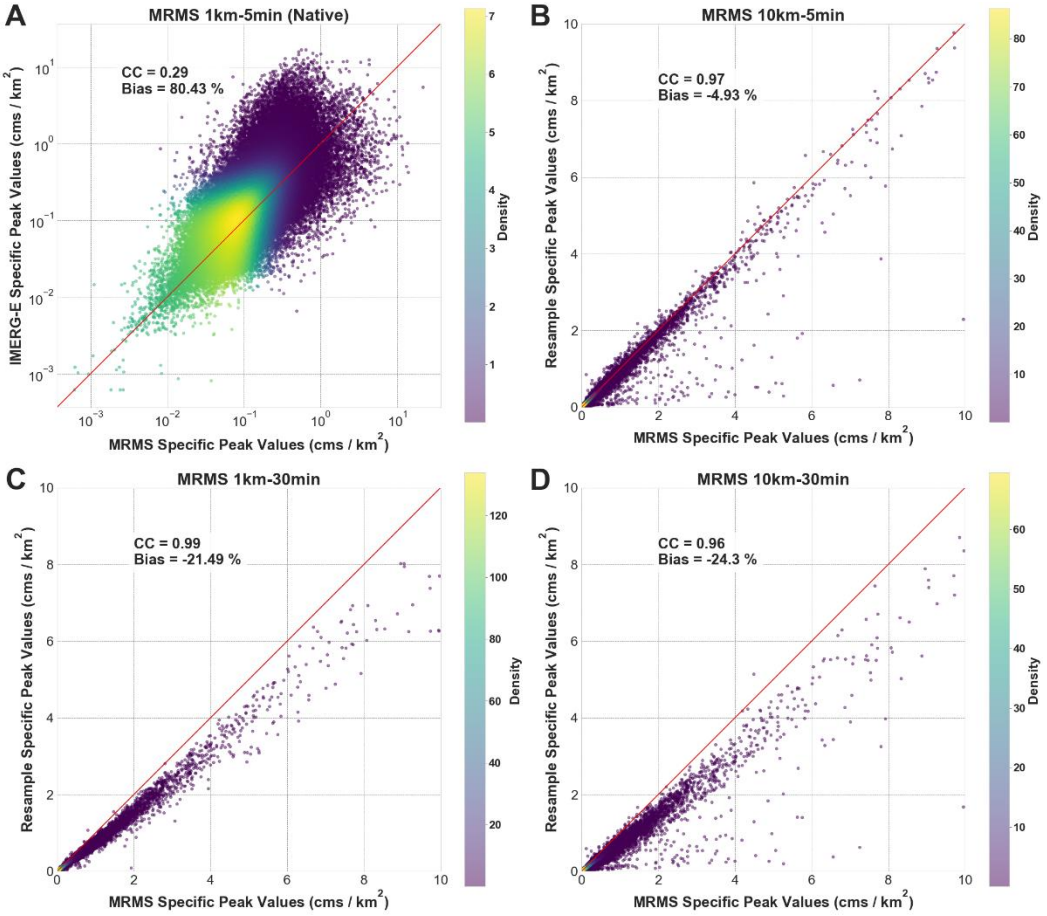

Figure 26. Scatterplots of MRMS specific peak discharge compared to: (A) IMERG-E Native specific peak discharge, (B) MRMS 10km-5min resampled specific peak discharge, (C) MRMS 1km-30min resampled specific peak discharge, and (D) MRMS 10km-30min resampled specific peak discharge. Pearson correlation coefficient and percent bias statistics are provided. The red diagonal line indicates the 1:1 line.

What is seen in the conditional distribution plots in **Figure 27** is the same behavior that was found in Chapter 2, with a slight overestimation of IMERG-E simulations at lower specific peak discharges followed by a plateau of underestimation at high values of specific peak discharge. Little variation can be seen across the four plots once again, with the exception of a minor reduction in quantile spreads at values greater than 1 cms/km^2 for plots **27B-C-D** compared to native in **27A**.

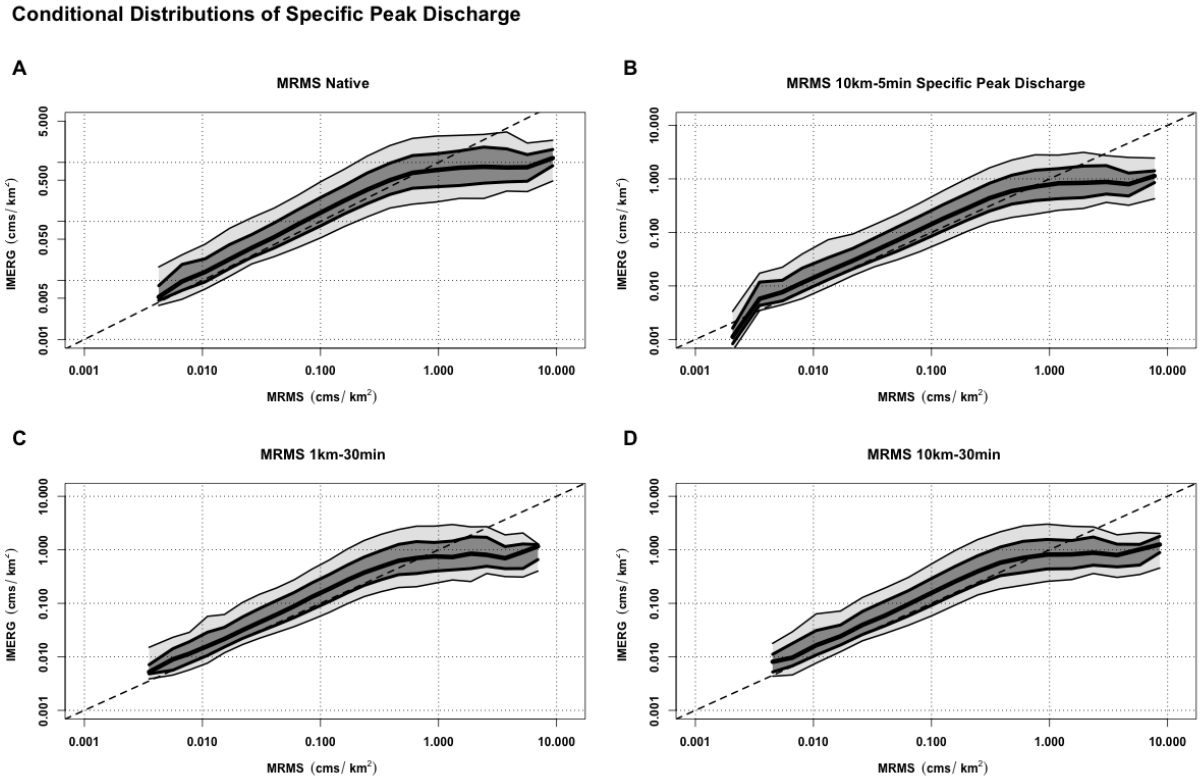

Figure 27. Conditional distribution plots of IMERG-E specific peak discharge compared to: (A) MRMS Native specific peak discharge, (B) MRMS 10km-5min resampled specific peak discharge, (C) MRMS 1km-30min resampled specific peak discharge, and (D) MRMS 10km-30min resampled specific peak discharge. The 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 is the 1:1 line.

For the quantile plots in **Figure 28**, those showing the MRMS native simulations vs the MRMS resampled simulations, it was again decided to use natural scale instead of logarithmic in order to showcase the features of the plots. The characteristics of **Figure 24** are much more visible once normalized by basin area, with less bias but significantly more quantile spread at high specific peak discharges in **28B**, an increase in bias but a far tighter quantile spread in **27C**, and both increased bias and spread in **28D**. This shows consistency in that, at least from the perspective of native vs resampled MRMS simulations, upscaling of spatial resolution tends to increase uncertainty when dealing with higher peak and specific peak discharges (as has been seen when using coarser resolutions throughout this study) while upscaling of temporal resolution leads to overall underestimations of peak and specific peak discharges, likely caused by the loss of some of the more intense rain rates through the upscaling process – those that normally contribute to the higher discharges currently being underestimated.

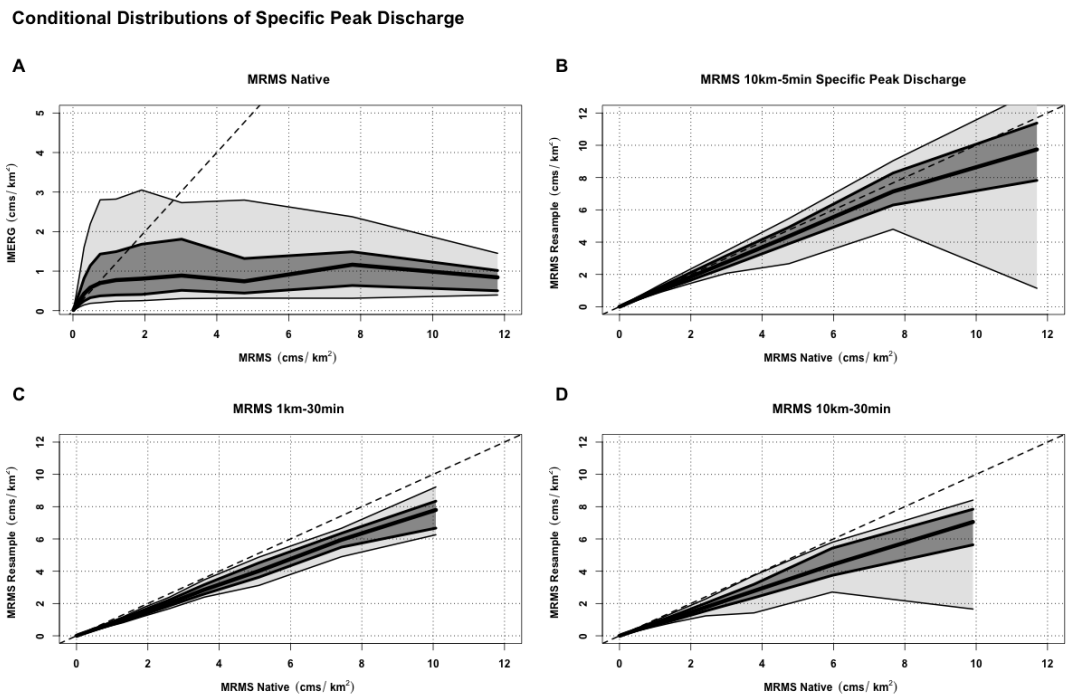

Figure 28. Conditional distribution plots of MRMS specific peak discharge compared to: (A) IMERG-E Native specific peak discharge, (B) MRMS 10km-5min resampled specific peak discharge, (C) MRMS 1km-30min resampled specific peak discharge, and (D) MRMS 10km-30min resampled specific peak discharge. The 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 is the 1:1 line.

4.3.2 Flood Duration

Through continuing the exercise with simulated flood durations, continued error masking can be seen from the IMERG-E results with respect to those from the MRMS resampled products (**Figure 29**). Small insights can still be gleamed from these results, however. Contrary to what was seen with the peak discharge simulations, where correlation increased with spatial upscaling and bias increased with temporal upscaling, the duration results show an increase in correlation with the resampled MRMS products at IMERG temporal resolution and an increase in bias when resampled spatially. This result is not surprising, as simulated flood durations will naturally be more impacted by temporal resolutions of the products at hand than spatial resolutions.

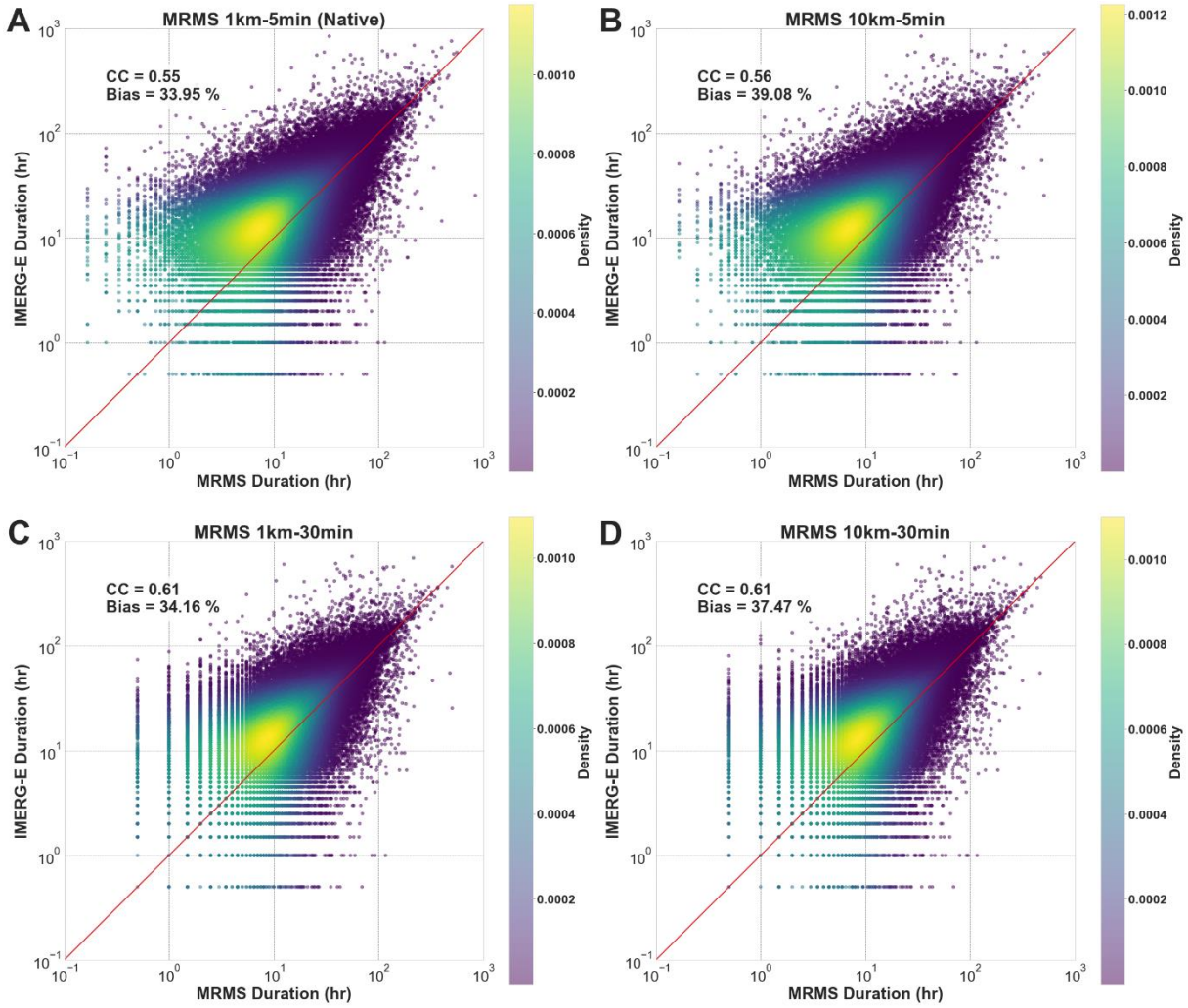

Figure 29. Density scatterplots of IMERG-E simulated flood durations compared to: (A) MRMS Native simulated durations, (B) MRMS 10km-5min resampled simulated durations, (C) MRMS 1km-30min resampled simulated durations, and (D) MRMS 10km-30min resampled simulated durations. Pearson correlation coefficient and percent bias statistics are provided. The red line indicates the 1:1 line.

When looking at the MRMS and MRMS-resampled results of the simulated flood durations (**Figure 30**), continued underestimation can be seen on the part of the resampled product simulations. While correlations between the native and resampled product data remain locked, all three resampled versions show increasing underestimation as they approach the temporal and spatial resolutions of IMERG. With both peak discharge and now flood duration showing underestimation, it is increasingly likely that the resample is resulting in negative water balance errors (i.e., there is less water entering the system after resampling than there is with the native MRMS product simulation). An investigation into this will be present with the quadrant plots in Section 4.3.4.

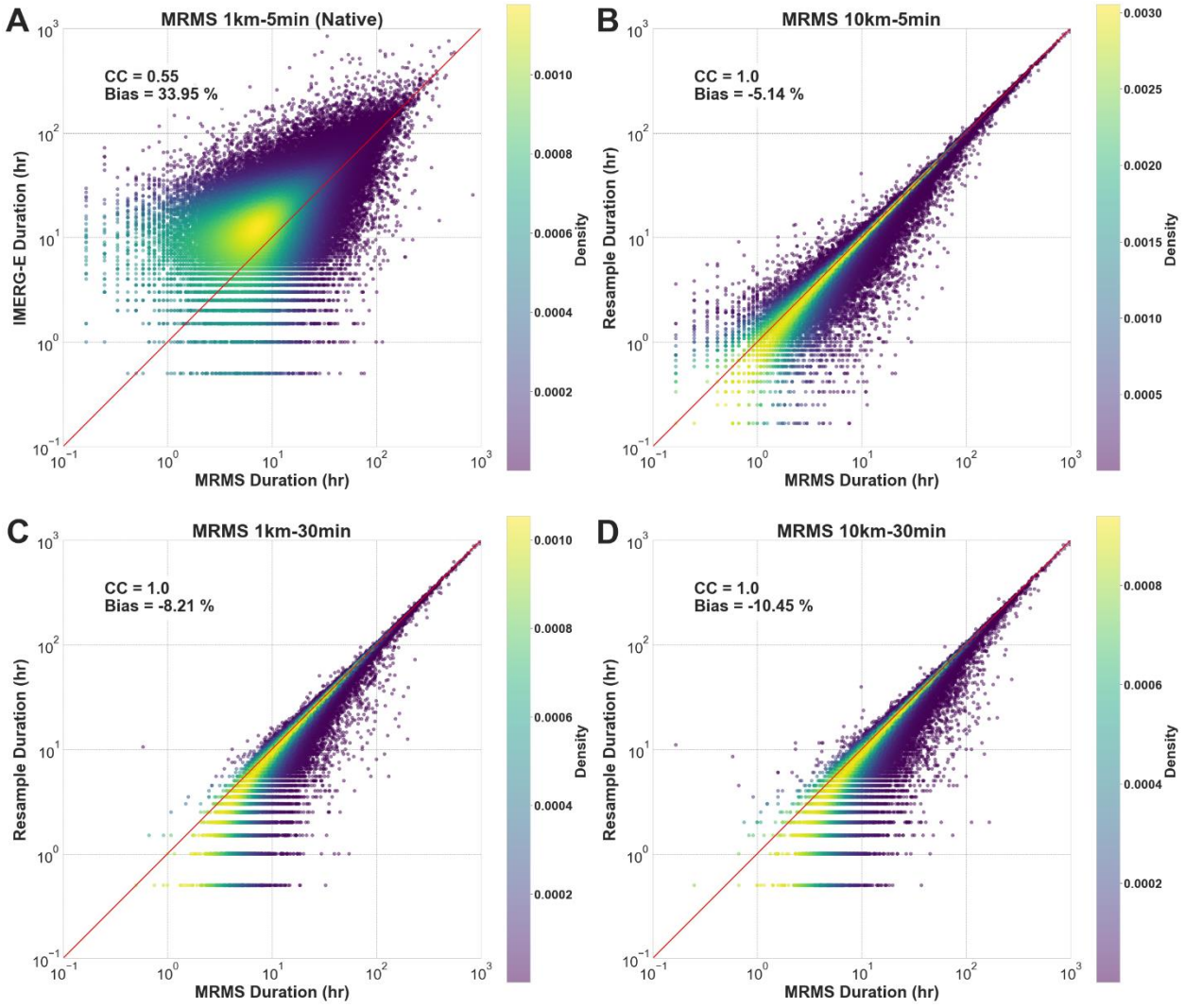

Figure 30. Density scatterplots of MRMS simulated flood durations compared to: (A) IMERG-E Native simulated durations, (B) MRMS 10km-5min resampled simulated durations, (C) MRMS 1km-30min resampled simulated durations, and (D) MRMS 10km-30min resampled simulated durations. Pearson correlation coefficient and percent bias statistics are provided. The red line indicates the 1:1 line.

The conditional distribution plots of IMERG-E vs MRMS and the resamples (**Figure 31**) again show little variation between themselves, save the longer tails of overestimation for shorter duration floods in **31A-B**. The tails are present in these two plots because both are operating at

shorter temporal resolutions, allowing the simulations to catch finer-scale flood durations at 5min temporal than those operating at 30min temporal. The symmetrical nature of all four plots, however, would suggest that the tails in **31C-D** would follow the same overestimation trend if it were possible to catch those shorter simulated floods. Otherwise, all four comparisons behave the same, with overestimation at shorter durations and increasing agreement as simulated durations grow longer and longer.

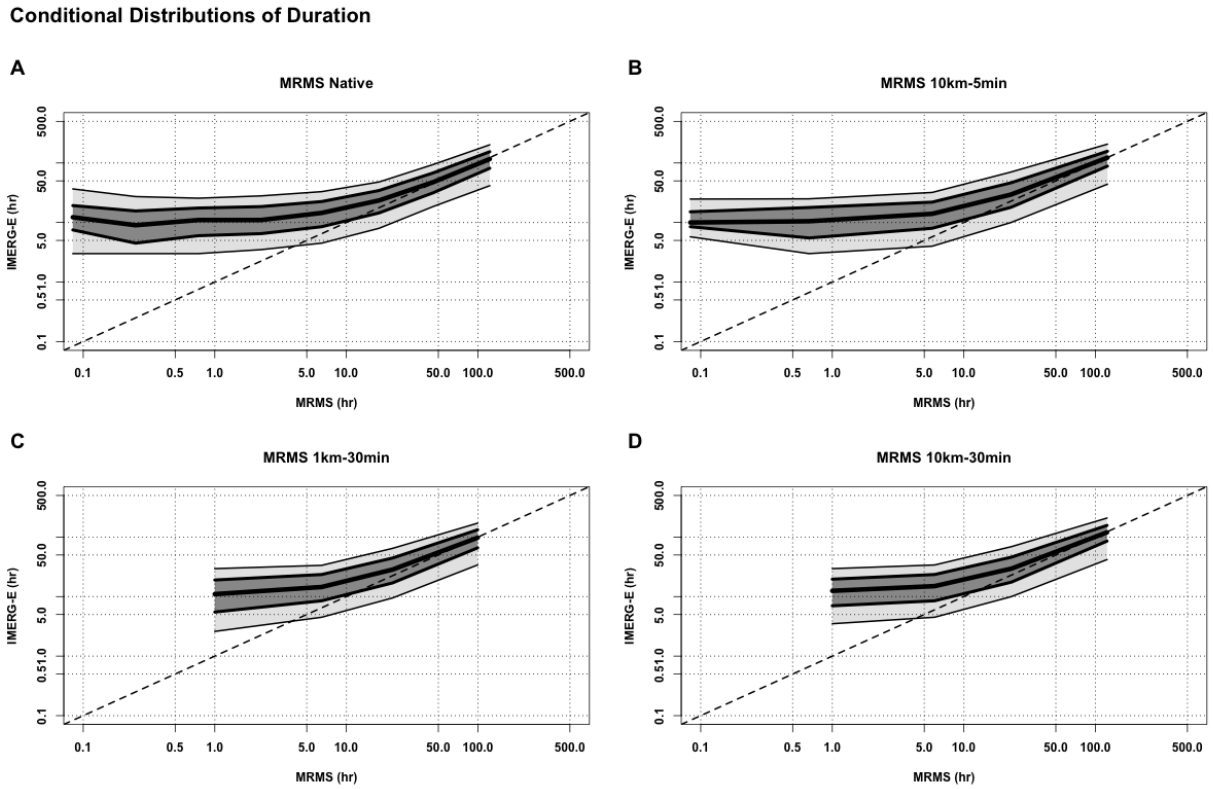

Figure 31. Conditional distribution plots of IMERG-E simulated durations compared to: (A) MRMS Native simulated durations, (B) MRMS 10km-5min resampled simulated durations, (C) MRMS 1km-30min resampled simulated durations, and (D) MRMS 10km-30min resampled simulated durations. The thick center line shows 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.

Looking at the conditional distributions of the MRMS native simulations and the MRMS resampled simulations in **Figure 32**, a noticeable difference can be seen. When MRMS is only resampled spatially (10km-5min, **32B**) the quantile distribution actually behaves similarly with respect to the native MRMS simulations as IMERG-E does, with a tail of overestimation at shorter durations of simulated floods (albeit with significantly less overestimation overall). When resampled temporally (**32C-D**) the quantiles of the simulations begin to exhibit the underestimation trends seen in the density scatterplots, reaffirming the assumption that this is a direct result of the upscaling of temporal resolution on simulated volumes of water in the system.

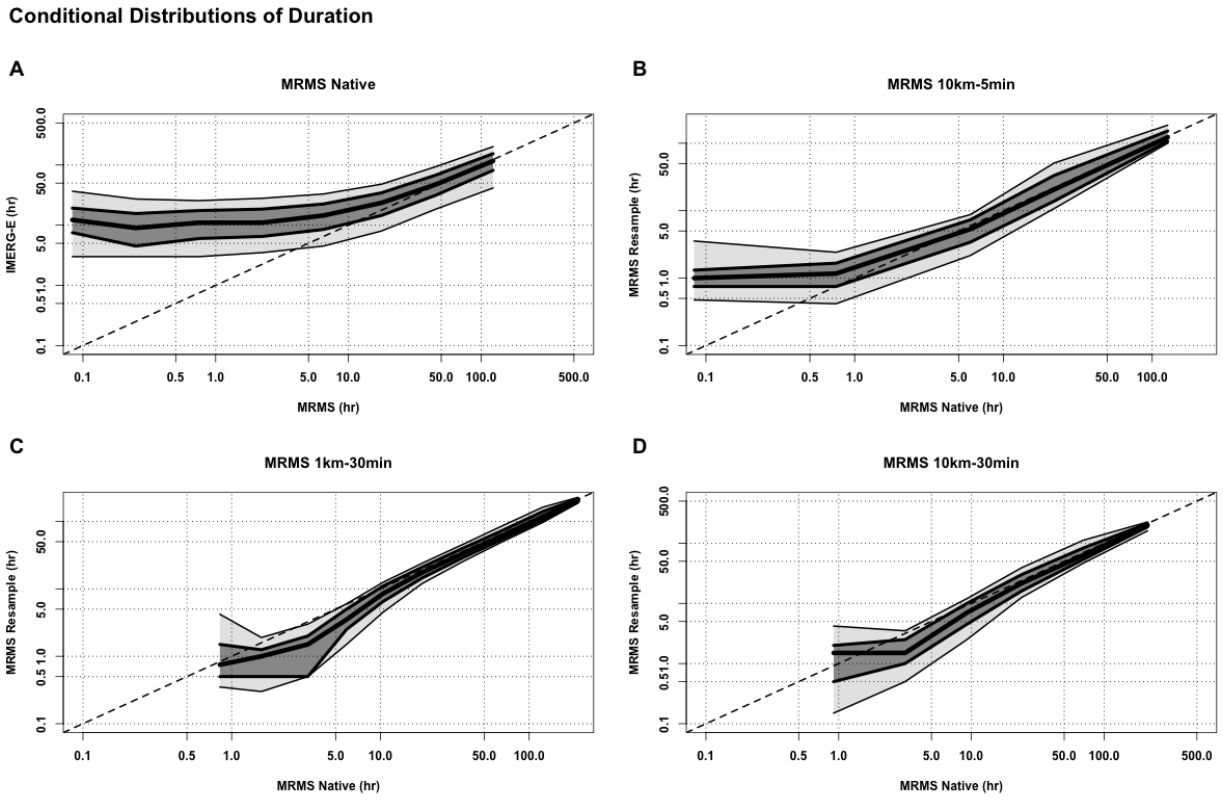

Figure 32. Conditional distribution plots of MRMS simulated durations compared to: (A) IMERG-E Native simulated durations (B) MRMS 10km-5min resampled simulated durations, (C) MRMS

1km-30min resampled simulated durations, and (D) MRMS 10km-30min resampled simulated durations. 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.

While still unconventional, normalizing duration values by each basin's respective area serves to provide a more robust look at the effects of resolution on simulated flood durations. As can be seen in **Figure 33**, this error masking phenomenon present in the IMERG-E simulation comparisons is completely independent of basin area with once again all four plots exhibiting the same structure. The best statistics are again present in the simulations generated by temporally resampled MRMS products. The resampled simulations compared against MRMS native (**Figure 34**) also tell the same story when normalized by basin area as the normal flood durations, with the best overall simulation agreement coming from the 10km-5min resample and increasing underestimation coming from the temporally resampled simulations.

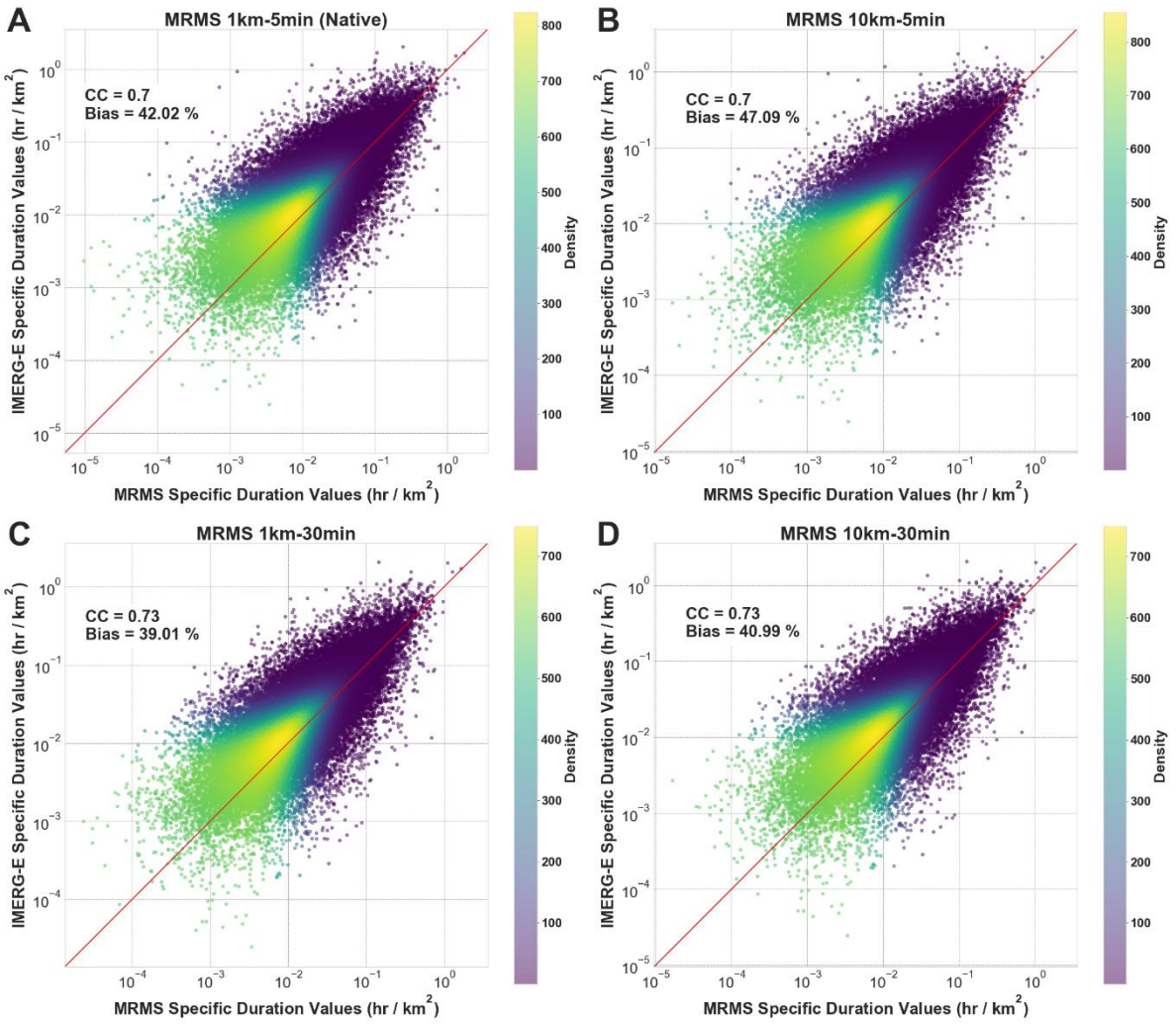

Figure 33. Density scatterplots of IMERG-E simulated specific flood durations compared to: (A) MRMS Native simulated specific durations, (B) MRMS 10km-5min resampled simulated specific durations, (C) MRMS 1km-30min resampled simulated specific durations, and (D) MRMS 10km-30min resampled simulated specific durations. Pearson correlation coefficient and percent bias statistics are provided. The red line indicates the 1:1 line.

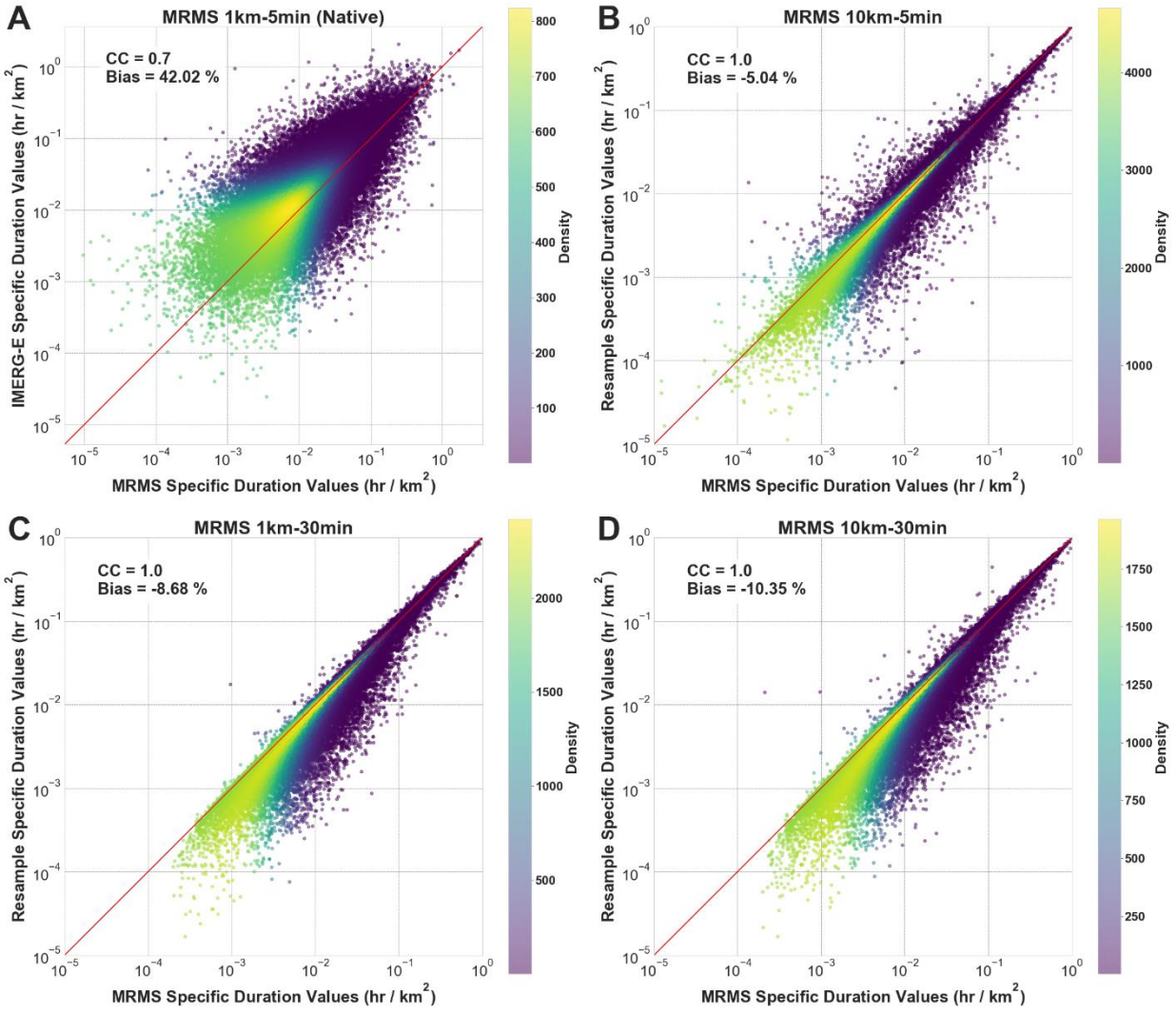

Figure 34. Density scatterplots of MRMS simulated specific flood durations compared to: (A) IMERG-E Native simulated specific durations, (B) MRMS 10km-5min resampled simulated specific durations, (C) MRMS 1km-30min resampled simulated specific durations, and (D) MRMS 10km-30min resampled simulated specific durations. Pearson correlation coefficient and percent bias statistics are provided. The red line indicates the 1:1 line.

Perhaps the most interesting result from this chapter's look into the normalized flood durations is how remarkably similar all of the quantile plots are (**Figure 35** and **Figure 36**). Across the board, both figures exhibit the same trends, with median values largely being centered on the

1:1 line and increasing uncertainty and quantile spread as values of specific duration increase. The only major difference between the IMERG-E simulation comparisons to the resamples and the MRMS native simulation comparisons is the expected decrease in overall quantile spread from the MRMS-only comparisons. This highlights that even when upscaling the only effect that basin area has on the results of simulated flood durations is overall uncertainty, it does not appear to generate any significant amount of conditional bias.

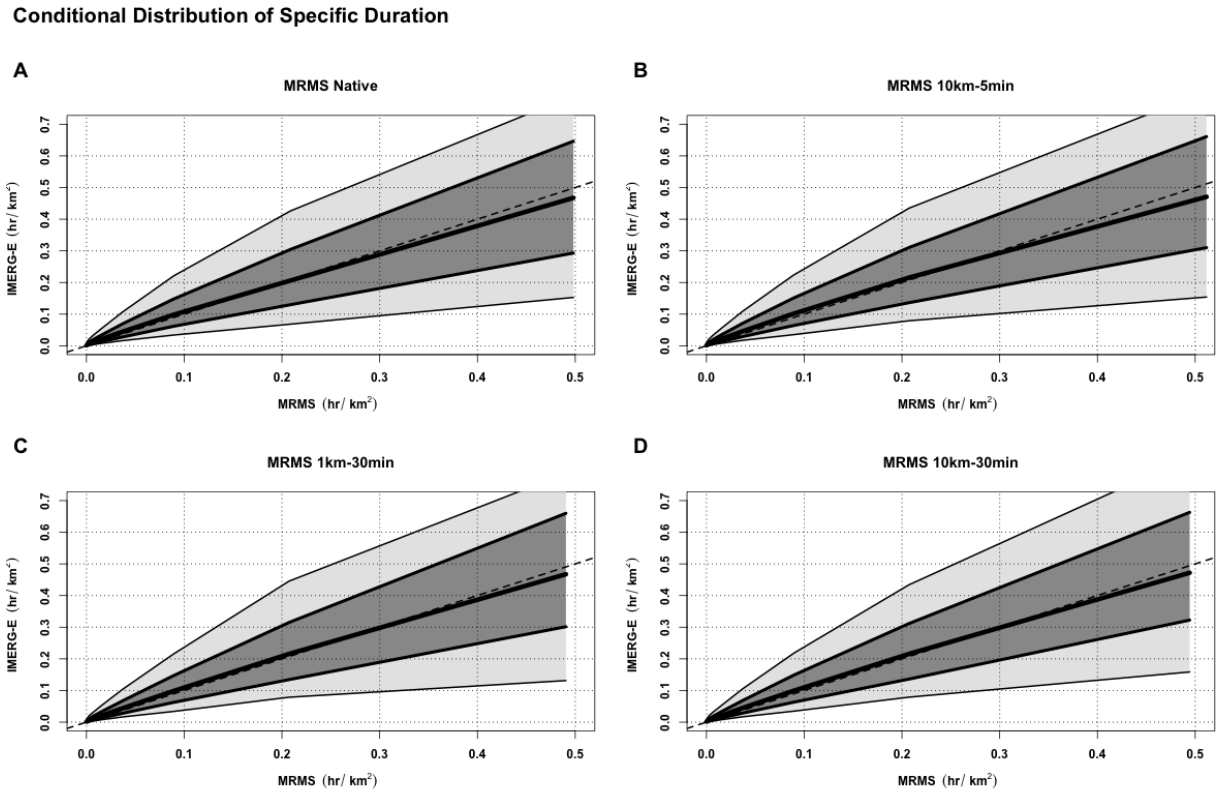

Figure 35. Conditional distribution plots of IMERG-E simulated specific durations compared to: (A) MRMS Native simulated specific durations, (B) MRMS 10km-5min resampled simulated specific durations, (C) MRMS 1km-30min resampled simulated specific durations, and (D) MRMS 10km-30min resampled simulated specific durations. 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.

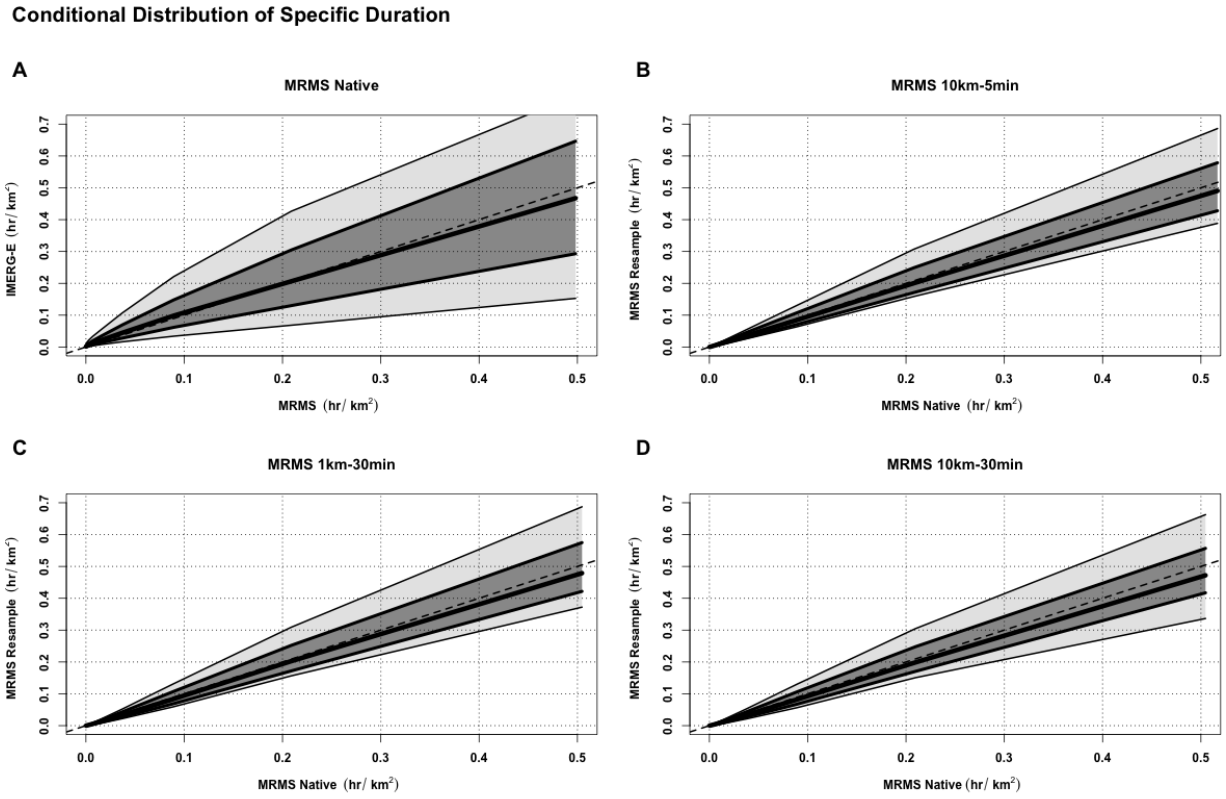

Figure 36. Conditional distribution plots of MRMS simulated specific durations compared to: (A) IMERG-E Native simulated specific durations (B) MRMS 10km-5min resampled simulated specific durations, (C) MRMS 1km-30min resampled simulated specific durations, and (D) MRMS 10km-30min resampled simulated specific durations. 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.

4.3.3 Flood Timing

The overall methodology for the process of investigating flood timings remains the same

as in previous chapters, utilizing the timestamps generated for both the start of an event and end of an event. For **Figure 37**, the absolute start and end times for the IMERG-E simulations were subtracted from their associated MRMS-simulated (native or resampled) 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. This process was also repeated for the MRMS vs resampled simulations (**Figure 38**), where each respective resampled value was subtracted from its MRMS native-simulated counterpart.

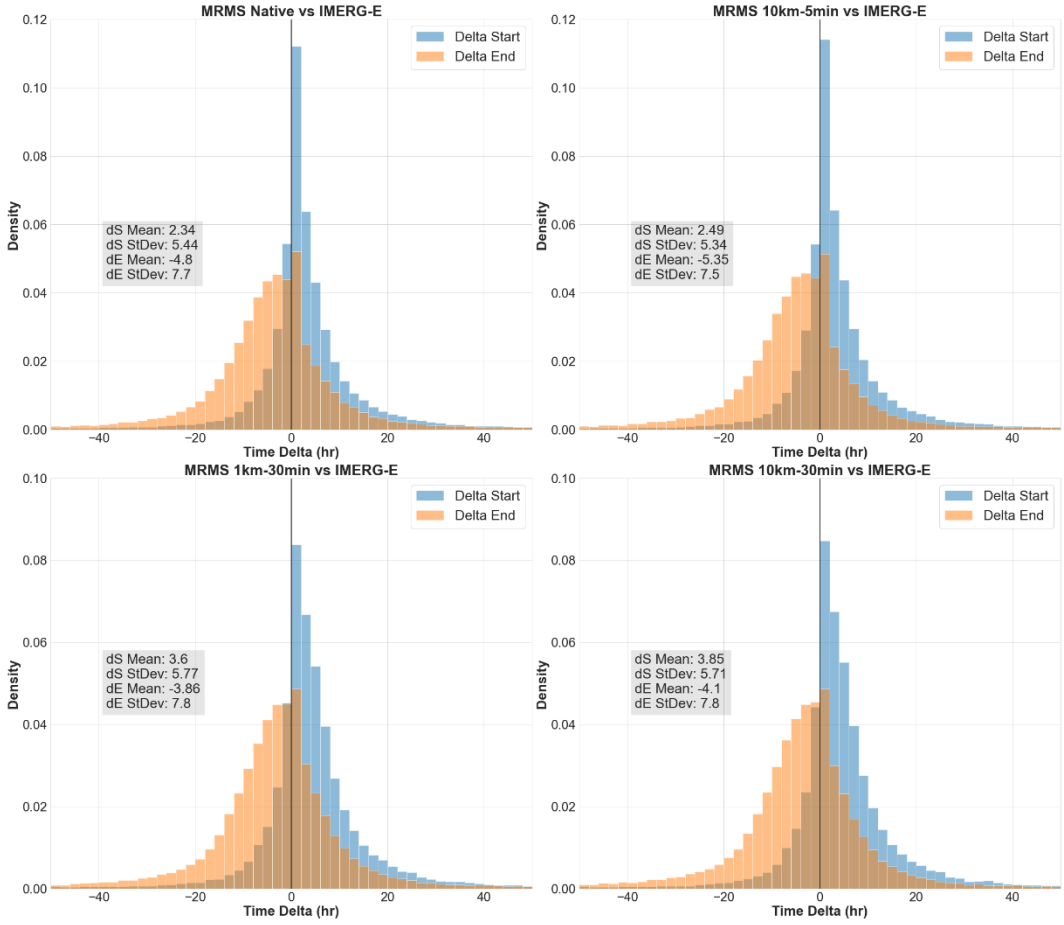

Figure 37. Histograms of the time deltas of matched flood start times (blue) and end times (orange) for IMERG-E native simulations compared to: (A) MRMS native, (B) MRMS 10km-5min resampled, (C) MRMS 1km-30min resampled, and (D) MRMS 10km-30min resampled, all with associated means and standard deviations. A positive (negative) value indicates that the simulated event (IMERG-E) occurs earlier (later) than its MRMS native (or MRMS resampled) absolute time counterpart. Note that in order to better see the histogram features not all y-axis values are the same.

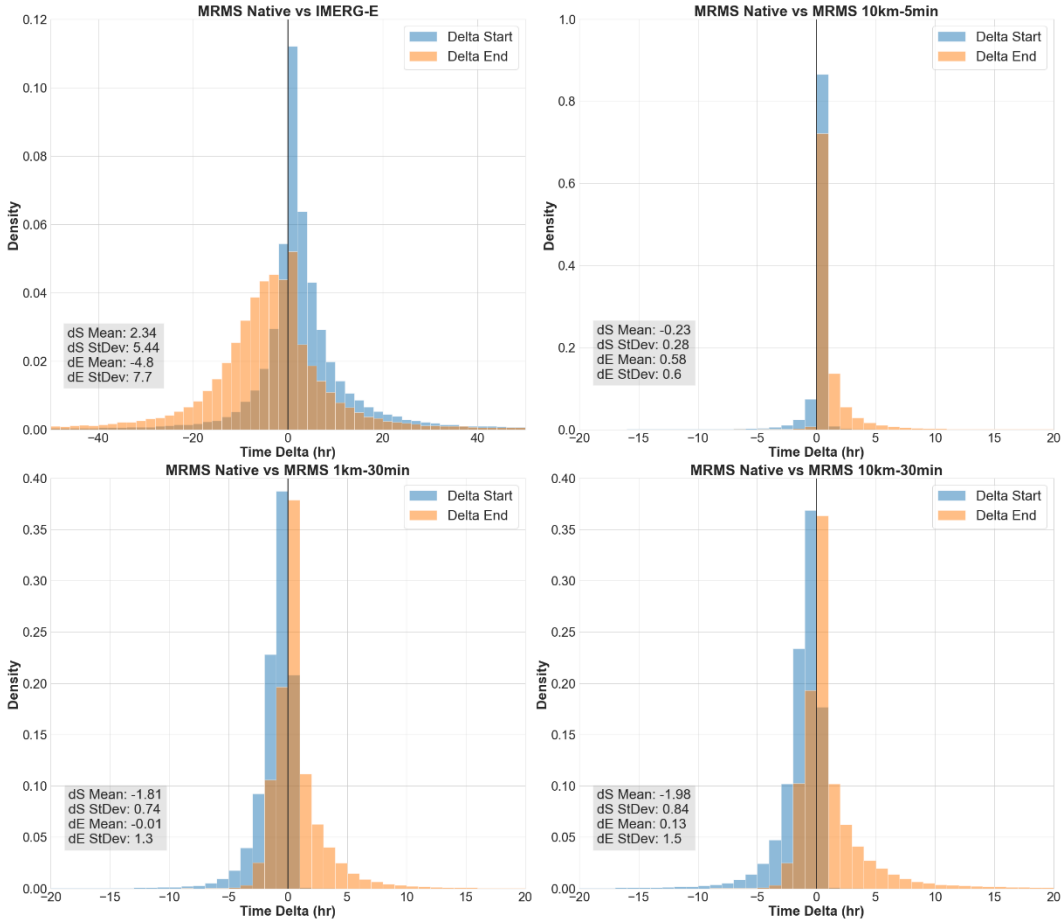

Figure 38. Histograms of the time deltas of matched flood start times (blue) and end times (orange) for MRMS native simulations compared to: (A) IMERG-E native, (B) MRMS 10km-5min resampled, (C) MRMS 1km-30min resampled, and (D) MRMS 10km-30min resampled, all with associated means and standard deviations. A positive (negative) value indicates that the simulated event (MRMS resampled IMERG-E) occurs earlier (later) than its MRMS native absolute time counterpart. Note that in order to better see the histogram features not all y-axis values are the same.

The trend of similarities continues for the IMERG-related histograms, with the same overall structure consistent across all four comparisons. The results remain consistent with what was seen in the previous two chapters, with IMERG-E tending to generate flood simulations that both start earlier and end later than its MRMS native or MRMS resampled counterparts. Statistically, the best performing product of the four is the 10km-5min resampled MRMS (**Figure 37B**) whose simulations resulted in the lowest set of both delta start and delta end means and standard deviations. Both temporally-resampled MRMS products (**Figure 37C-D**) performed statistically worse than the others, though adding in the spatial resample (**37D**) did not meaningfully impact the scores between the two. This suggests that, much like the duration plots in the previous subsection, less concern is necessary for the accuracy of flood timing results when upscaling spatially than is necessary when upscaling temporally.

The histograms from the MRMS-simulated only analysis, however, tell a different story. Despite significantly higher agreement between the MRMS-simulated and MRMS resample-simulated events (as evidenced by the much shorter tails and much higher densities than those seen with the IMERG-E simulations) the timings instead trend the opposite direction, with an increased

likelihood of the resampled products simulating events that start later and end earlier than the MRMS native simulated reference. This trend is only slight when resampled spatially (**Figure 38B**) but pushes progressively further when resampled temporally (**38C**) and then spatially and temporally (**38D**). A noticeable increase in the extent of the histograms' tails appears in both temporally-resampled product simulations, again signifying that caution must be taken in this regard when utilizing temporally-downscaled forcings.

Continuing the procedure established in the previous two chapters, conditional distribution plots were made for IMERG-E simulated delta start times (**Figure 39**) and delta end times (**Figure 40**), both compared against associated basin area, for each MRMS-forced product (native and resampled). Consistent with previous chapters, simulated start times and end times across all product forcings show increasing uncertainty and quantile spread both at small basin scales and large basin scales, with the most accurate ranges coming between 100 km^2 and 1000 km^2. Otherwise, both figures show results consistent with those seen in the histograms; start times overall trend more positive (earlier than the reference) while end times trend more negative (later than the reference).

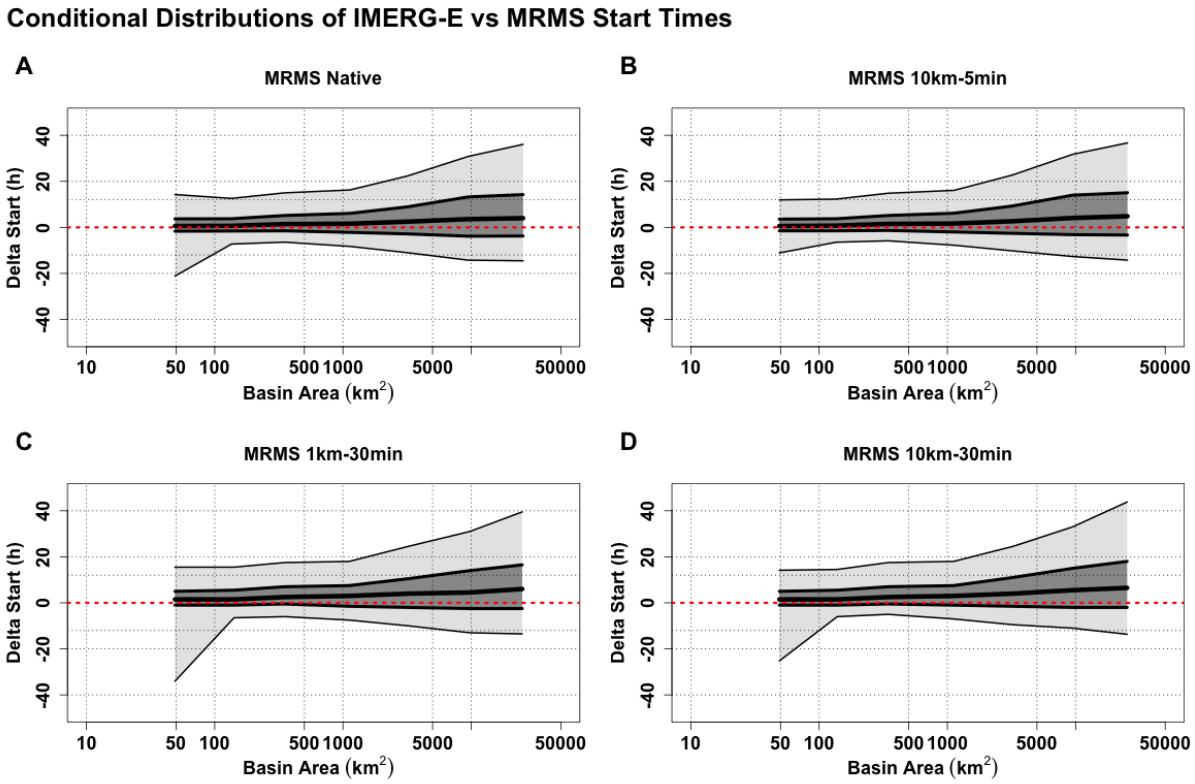

Figure 39. Conditional distribution plots of IMERG-E simulated event start timings compared to: (A) MRMS native simulated event start timings (B) MRMS 10km-5min resampled simulated event start timings, (C) MRMS 1km-30min resampled simulated event start timings, and (D) MRMS 10km-30min resampled simulated event start timings 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 (IMERG-E) occurs earlier (later) than its MRMS native/resampled absolute time counterpart.

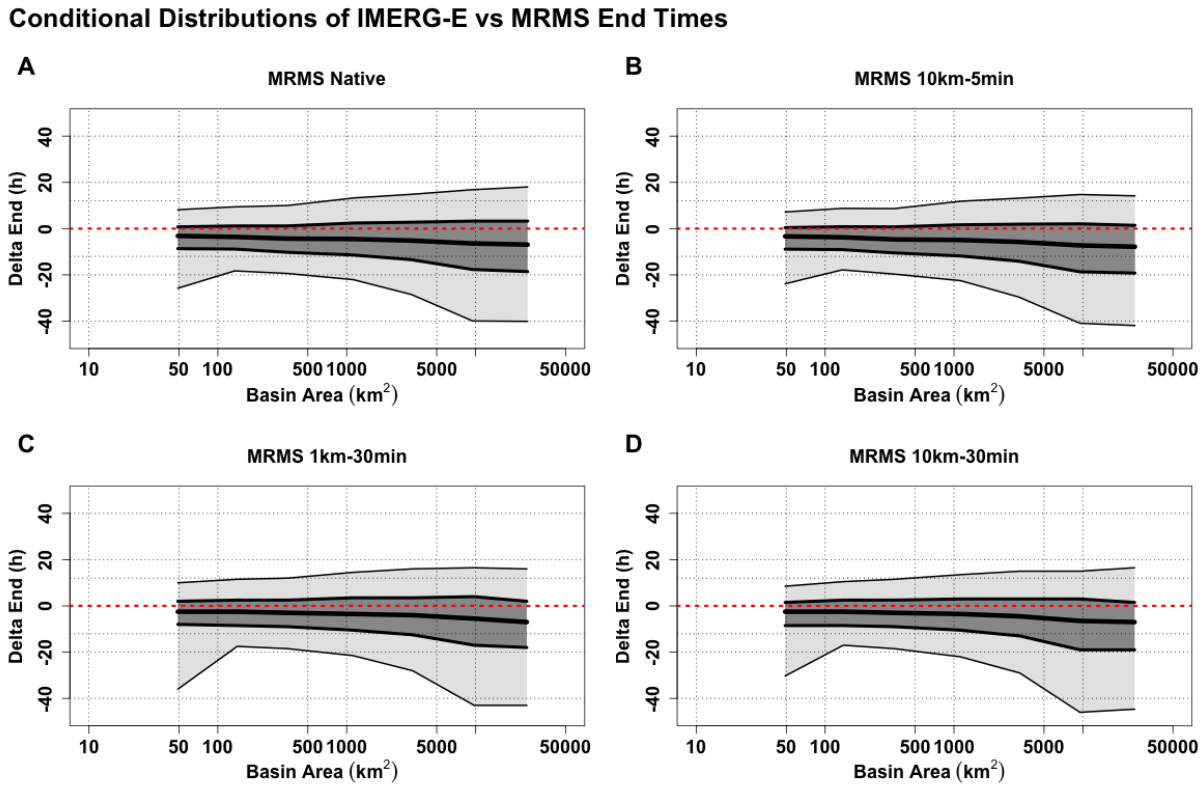

Figure 40. Conditional distribution plots of IMERG-E simulated event end timings compared to: (A) MRMS native simulated event end timings (B) MRMS 10km-5min resampled simulated event end timings, (C) MRMS 1km-30min resampled simulated event end timings, and (D) MRMS 10km-30min resampled simulated event end timings 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 (IMERG-E) occurs earlier (later) than its MRMS native/resampled absolute time counterpart.

The next focus is on the MRMS-related conditional bias plots (**Figure 41** for simulated start times and **Figure 42** for simulated end times). What is immediately striking is that compared to the MRMS native simulations all three resampled simulations exclusively exhibit negative delta

start times, meaning that no matter how you upscale the product it will result in simulated floods across all basin sizes that begin later than those simulated by native MRMS. This issue becomes noticeably exacerbated at increasing basin size, with quantiles increasing in spread across all three sets of resampled simulations. The longer tails present in the histograms in **Figure 40C-D** translate into overall larger quantile spreads in **Figure 41C-D** and **Figure 42C-D**, showing that most of the uncertainty in this aspect of the simulations comes from the upscaling of temporal resolution. Contrary to the start times, however, this uncertainty actually serves to work in favor of the simulated end times. In **Figure 42C-D** the spread of quantiles falls far more evenly across the zero-line of deltas than in **Figure 42B** where it exists exclusively in positive delta values. Though this does appear to come at a cost at large basin sizes, where the quantile spread is significantly larger for the simulations from temporally resampled products.

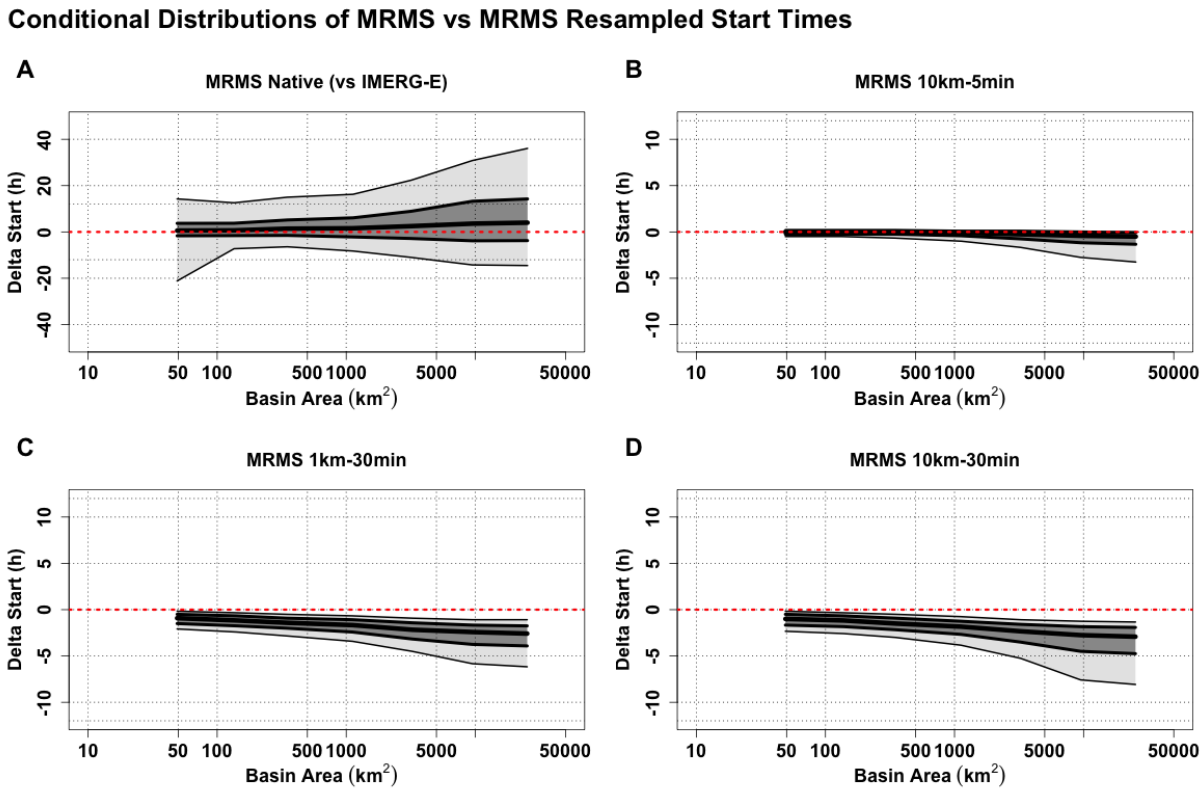

Figure 41. Conditional distribution plots of MRMS native simulated event start timings compared to: (A) IMERG-E simulated event start timings (B) MRMS 10km-5min resampled simulated event start timings, (C) MRMS 1km-30min resampled simulated event start timings, and (D) MRMS 10km-30min resampled simulated event start timings 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 resampled or IMERG-E) occurs earlier (later) than its MRMS native absolute time counterpart.

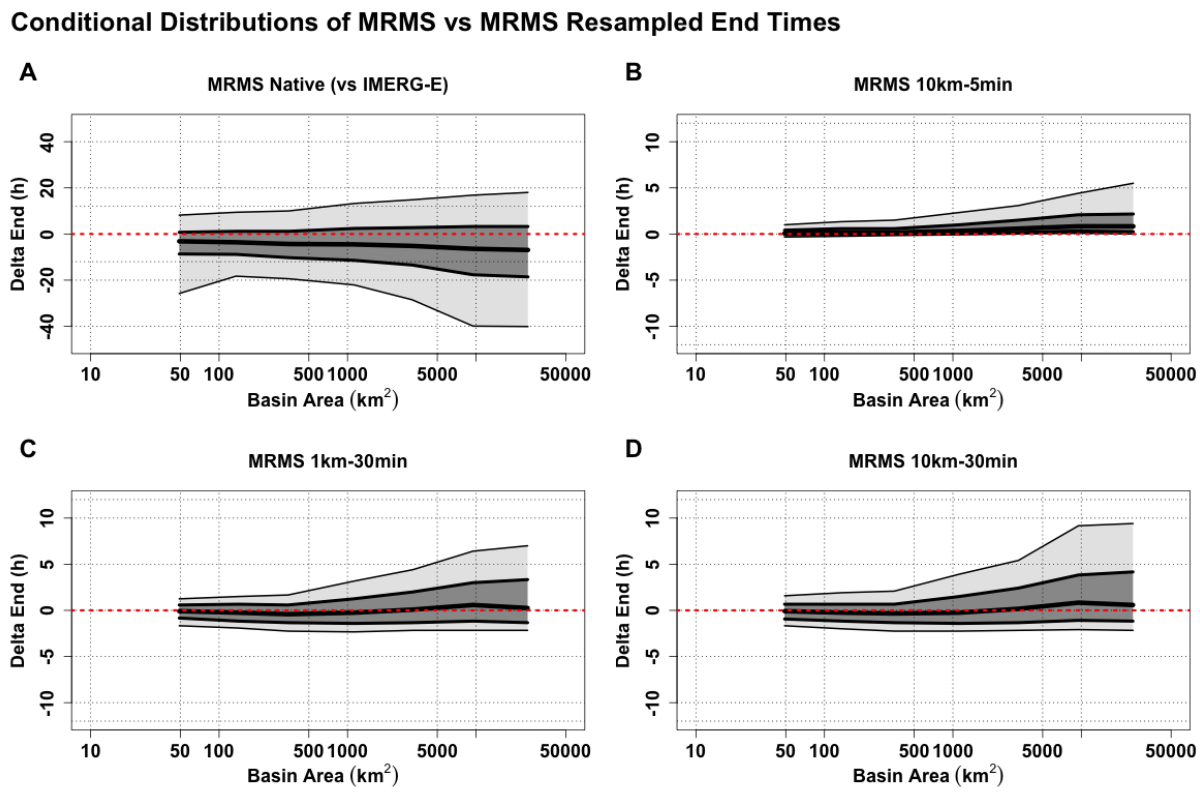

Figure 42. Conditional distribution plots of MRMS native simulated event end timings compared to: (A) IMERG-E native simulated event end timings (B) MRMS 10km-5min resampled simulated event end timings, (C) MRMS 1km-30min resampled simulated event end timings, and (D) MRMS 10km-30min resampled simulated event end timings 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 resampled) occurs earlier (later) than its MRMS native absolute time counterpart.

4.3.4 Error Contributions

With the apparent masking of IMERG-E simulation uncertainties when compared against the MRMS resampled simulations, an additional analysis was undertaken in order to assess the relative contribution of uncertainties from the upscaled resolutions to the overall uncertainties of the IMERG-E comparisons. The error separation variance concept (e.g., Kirstetter et al. 2010 for precipitation) makes it possible to quantify this contribution by introducing the reference peak discharge value, P, in the expression of the variance of the residuals between the IMERG-E peak discharge values (P*) and the MRMS 10km-30min peak discharge value, P_r:

$$Var\ (\varepsilon) = Var(P^* - P_r) = Var((P^* - P) - (P_r - P)) \quad \text{(Eq. 1)}$$

$$Var\ (\varepsilon) = Var(P^* - P) + Var(P_r - P) - 2Cov((P^* - P), (P_r - P)) \quad \text{(Eq. 2)}$$

Assuming the errors (P* - P) and (P_r - P) to be uncorrelated (given the differences in the nature of the radar and satellite precipitation estimates used as forcing to estimate the peak discharge values, the covariance term vanishes and the standard deviation of the residual (P* - P) can be expressed as:

$$\sigma(P^* - P) = \sqrt{Var(\varepsilon) - Var(P_r - P)} \quad \text{(Eq. 3)}$$

We are interested in the ratio (Eq. 4) quantifying the relative contribution of scaling uncertainties to the overall uncertainties of the IMERG-E comparisons.

$$100 * \sqrt{\frac{Var(P_r - P)}{Var(P^* - P)}} \quad \text{(Eq. 4)}$$

By dividing the uncertainty of MRMS 10km-30min vs MRMS native simulations by the uncertainty of IMERG-E vs MRMS native simulations, taking the square root of the result and multiplying by 100%, the percent contribution of uncertainty can be logged and then plotted with respect to basin area as seen in **Figure 43**.

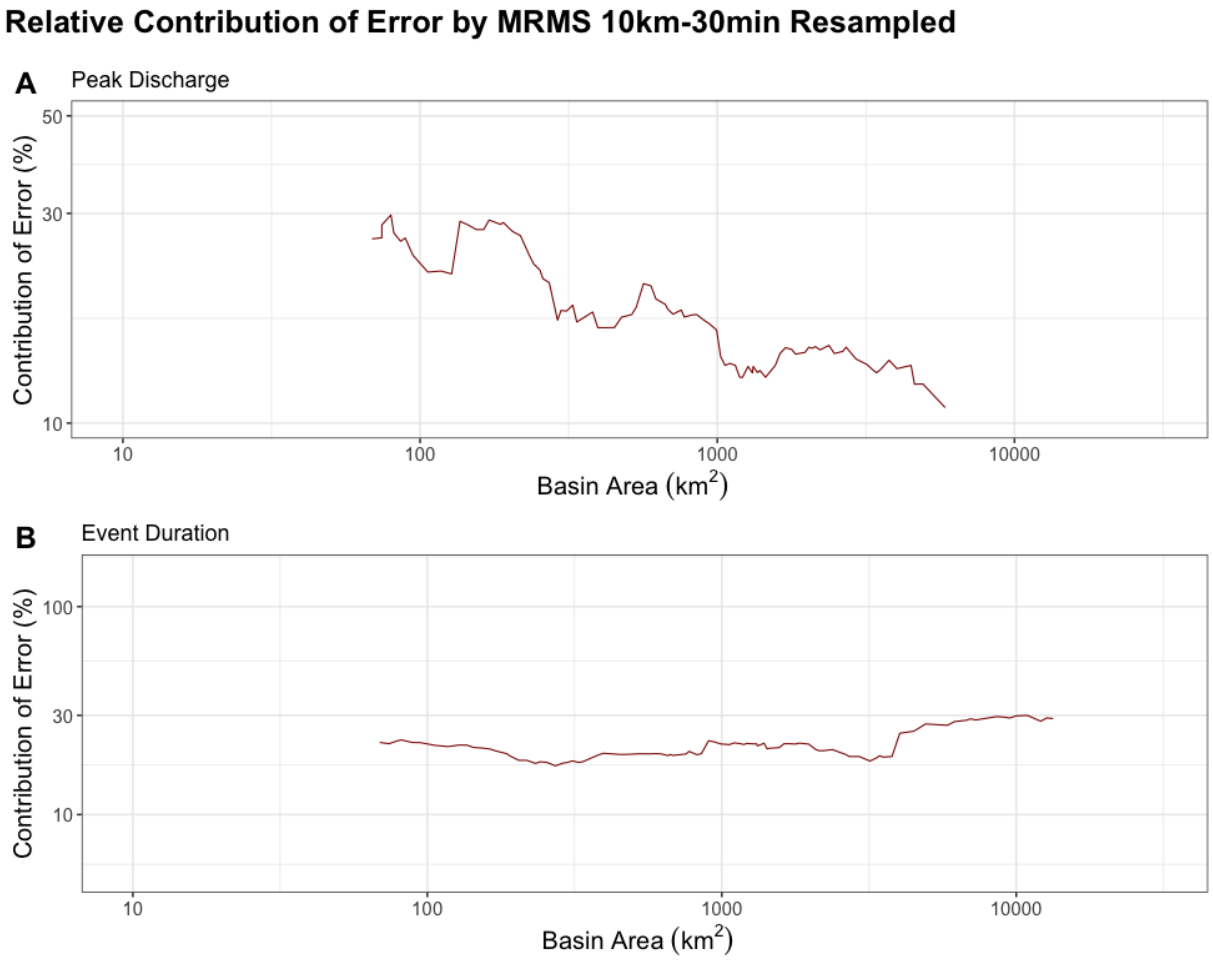

Figure 43. Relative percent contribution of error to peak discharge values (A) and event duration values (B) by MRMS 10km-30min resampled simulations with respect to IMERG-E simulations, plotted against basin area.

Both investigations of the error contributions yield different but important results. For peak discharge simulations (**Figure 43A**) the greatest scaling error contributions come from basins less than 1000 km^2, where contributions range from 20% to as much as 30%. These contributions are distinctly scaling-based, however, falling to almost 10% contribution as basin size increases. When looking at simulated event durations (**Figure 43B**) this scaling effect disappears, with a relatively flat curve of error contribution between 20% and 30% across the entire range of basin areas. Based on the quantile plots from the duration subsection (Section 4.3.2) this result is not surprising, as across both duration plots and specific duration plots the quantile spreads with respect to basin area remained relatively consistent.

4.3.5 Hydrologic Model Performance Analysis (Quadrant Plots)

To confirm that the behaviors being seen throughout the data analysis were consistent with previous chapters and were not being caused by any discrepancies between the model runs, quadrant plots were generated in the same manner as in Chapter 3. To reiterate this process and significance, observed values for peak discharge and flood duration were subtracted from their respective simulated values and the resulting discharge and duration errors were plotted. Each error quadrant on the plot then signifies a different hydrologic tendency within the model. Points in the top left quadrant (positive peak deltas and negative duration deltas) indicate simulated floods with higher peaks and shorter durations than USGS, a signal of influence from kinematic wave where the water is being pushed through the system too quickly. In the top right quadrant (positive peak deltas and positive duration deltas) points are found where both the peak and the duration are

higher than USGS, indicating positive water balance errors (i.e., there is too much water in the system, with greater areas under the theoretical hydrograph). The bottom left quadrant (both negative deltas) is again dominated by water balance, but instead with too little water simulated. The bottom right quadrant shows simulations with smaller peaks but longer durations than the reference, signifying flood attenuation by the model.

As can be seen in the plots where IMERG-E simulations were compared to MRMS native and MRMS resampled simulations (**Figure 44** and **Figure 45**), regardless of the upscaling each plot looks similar to its respective IMERG-E/MRMS native reference plot. Strong tendencies exist in these simulation errors in both the top-right (positive water balance) and bottom-left (negative water balance) quadrants, with the majority of points falling in that positive water balance error quadrant. As was seen in Chapter 3, this tracks with the consistent overestimation of IMERG-E simulations with respect to MRMS (native and resampled) simulations due to the increased tendency to generate too much water within the system.

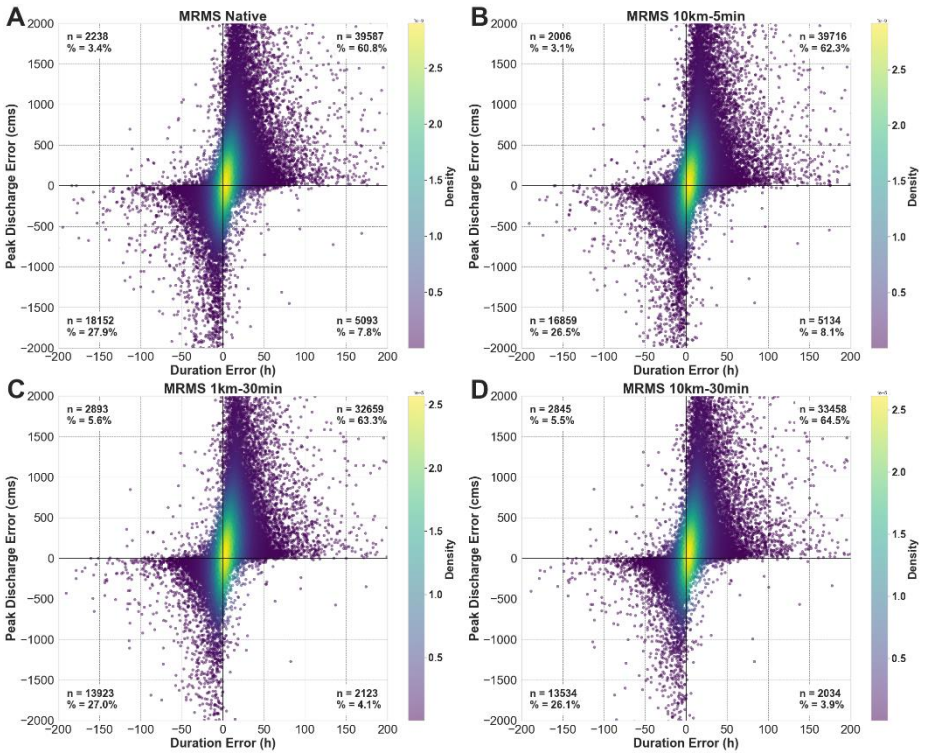

Figure 44. Density scatterplots of discharge and duration errors for IMERG-E simulations with respect to: (A) MRMS Native errors, (B) MRMS 10km-5min resampled errors, (C) MRMS 1km-30min resampled errors, and (D) MRMS 10km-30min resampled errors. Total numbers of points in each quadrant are provided, as well as each quadrant's percentage of the total points.

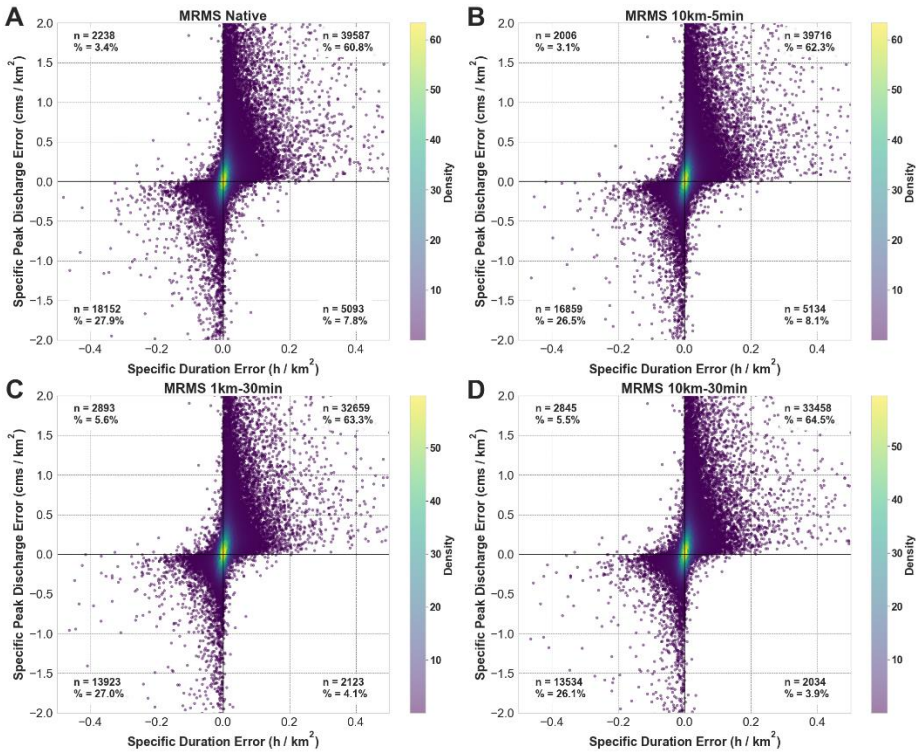

Figure 45. Density scatterplots of specific discharge and duration errors for IMERG-E simulations with respect to: (A) MRMS Native errors, (B) MRMS 10km-5min resampled errors, (C) MRMS 1km-30min resampled errors, and (D) MRMS 10km-30min resampled errors. Total numbers of points in each quadrant are provided, as well as each quadrant's percentage of the total points.

When investigating the MRMS native simulations vs the MRMS resampled simulations (**Figure 46** and **Figure 47**), however, two observations become clear. The most apparent is that the majority of points in all instances fall into the lower-left quadrant (negative water balance). Given the consistency of underestimation and negative bias seen previously in this chapter by the resampled product simulations with respect to the MRMS native simulations, this is not surprising to see. What is surprising, however, is the emergence of a larger percentage of points falling into the bottom-right quadrant (flood attenuation within the model) when the products are resampled

temporally (**Figure 46C-D**). Even when normalized by basin area this problem persists (**Figure 47C-D**), though this provides an enhanced look at the structure of the points in this quadrant. As of this writing, there is no previous research to look to in order to better explain this result. Currently, it is believed that this increased flood attenuation when the products are upscaled temporally is being caused by an artifact in the numerical solution of the kinematic wave function within EF5. Due to the nature of the points falling densely within a small range of minor error values and not being widely spread across the error quadrant, it is safe to assume that this is not being caused by a major flaw in the framework. More focused and targeted experiments into this phenomenon will be needed in the future to further diagnose the root causes.

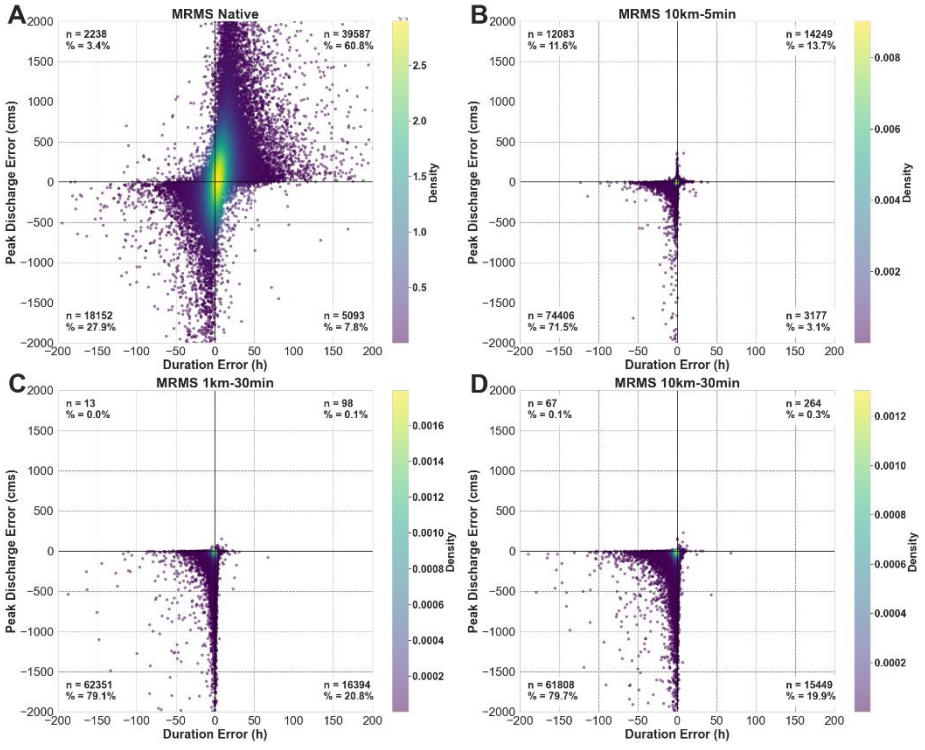

Figure 46. Density scatterplots of discharge and duration errors for MRMS simulations with respect to: (A) IMERG-E Native errors, (B) MRMS 10km-5min resampled errors, (C) MRMS 1km-30min resampled errors, and (D) MRMS 10km-30min resampled errors. Total numbers of

points in each quadrant are provided, as well as each quadrant's percentage of the total points.

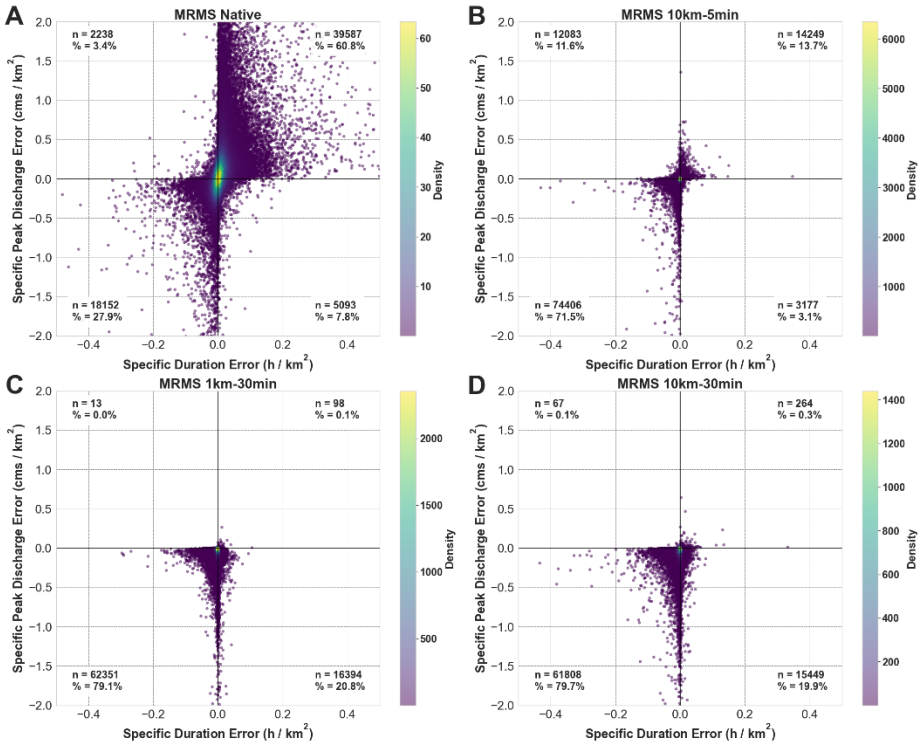

Figure 47. Density scatterplots of specific discharge and duration errors for MRMS simulations with respect to: (A) IMERG-E Native errors, (B) MRMS 10km-5min resampled errors, (C) MRMS 1km-30min resampled errors, and (D) MRMS 10km-30min resampled errors. Total numbers of points in each quadrant are provided, as well as each quadrant's percentage of the total points.

4.4 Conclusions

In this chapter, precipitation forcings from MRMS at native resolution were upscaled spatially, temporally, and spatiotemporally to the native resolutions of IMERG-E and run through a hydrologic modeling framework with the intent to assess how the different resolutions impact the simulated outputs of flood magnitude, event duration, and timing. Looking at how upscaling can affect the modeled output desired allows for improved understanding of which resolutions lead to the greatest changes in simulation accuracy, increasing the potential use and utility of current or future satellite products in regions where high-resolution ground radars are sparse.

Perhaps the most surprising result of this study is the potency of the errors put into the hydrologic system by the IMERG-E precipitation forcings. These tendencies have been well catalogued in the previous two chapters, but in every instance when the IMERG-E simulations were compared against the resampled MRMS simulations (magnitude, duration, timing, and quadrant) the plots looked highly similar to the reference plots of IMERG-E vs MRMS at native resolution. Some changes in statistics can be seen, such as slight increases in correlation or overall bias in peak discharge when MRMS is upscaled to 10-km or in duration when upscaled to 30-min, hinting at underlying changes in behavior but overall offering little of robust significance.

When comparing simulations from the upscaled MRMS products against simulations from MRMS at native resolution, however, more noteworthy observations can be made. Despite exceptionally high correlations from comparing between the same base product, spatial upscaling tended to increase overall uncertainty and quantile spread for both peak discharge and event duration while temporal upscaling led to underestimation of both flood characteristics. Additionally, floods simulated with the upscaled MRMS products tended to both start later and end earlier than those simulated by MRMS native, again with increased uncertainty coming from the temporally resampled products especially as basin size increases.

The discrepancy between the signals in flood characteristics between the two sets of product comparisons led to the need for an error contribution investigation between IMERG-E and the upscaled MRMS products. After extracting the variances associated with all three product simulation pairs, it was found that the structures of relative contribution of errors by MRMS at 10-km 30-min differ between peak discharge and event duration. When looking at peak discharge, a distinct scaling effect can be seen with greater contributions of error at basin sizes less than 1000 km^2 and decreasing further as sizes increase past 10000 km^2. Duration, on the other hand, maintained a relatively steady contribution of error across the entire range of basin sizes.

As was done in Chapter 3, the relative errors of discharge and duration were used to generate quadrant plots of model tendencies between IMERG-E and the MRMS native and resampled products. Once again all of the comparisons involving IMERG-E simulations looked similar, with most of the points falling in the quadrant associated with positive water balance error, consistent with the overestimation seen previously. When the MRMS native vs MRMS resampled errors were plotted, however, the majority of points instead fell in the negative water balance quadrant. Between the underestimations of peak discharge and event duration associated with the upscaled simulations, as well as the overall shortening of event timing, this result was not surprising. The upscaling of resolutions distinctly tended to remove too much water from the system, resulting in these negative water balance errors. These plots also unearthed an unexpected artifact within the model itself, with the upscaled temporal resolution causing an issue within the numerical solution to the kinematic wave equation and resulting in an increased rate of flood attenuation in the system. It is believed that this is not a critical issue within the framework, but does warrant more targeted investigation in the future.

Chapter 5: Lag Time

5.1 Introduction

For decades it has been established that time parameters are an important and fundamental factor when dealing with hydrologic analyses of rainfall and runoff. Changes in timing alone potentially determine whether a basin will flood or not given a particular volume of water (McCuen, 1998). Time governs (flash) flood forecasting operations and subsequent warnings. Lag time, defined as the time difference between the center of mass of precipitation and the resultant peak discharge, is one such time parameter. Yet it remains difficult to model and predict due to complexities arising from variability in basin geomorphology, precipitation characteristics, and climatology (Beven, 2020). McCuen (2009) points to a lack of uniformity in the literature when it comes to terminology regarding time parameters (e.g., time of concentration, another parameter often interchanged with lag time) as well as surprisingly high variance in estimates produced across different methods (Grimaldi et al., 2012). This is echoed in a study by Sauer et al. (1983) who acknowledged that while relationships involving basin characteristics can be used to derive lag time estimates, it is not recommended when analyzing the estimates of flood frequencies because "the error introduced by estimated lag time negates any advantage gained from using the equations". There is also uncertainty regarding the extrapolation of lag time equations from gauged to ungauged basins. For example, it has been shown that Desbordes' formula (Desbordes, 1974), an equation built from 21 urban watersheds in France, Europe, and the United States, can be rendered inadequate when applied to watersheds in Quito, Ecuador, and needed to be manually corrected to improve the simulations (Lhomme et al., 2004). In another study, it was found the 1971 SCS lag time equation (Kent, 1972) should be carefully applied when assessing hydrographs

and lag time, as estimated lag times were both highly variable across events as well as overestimated despite having specifically adjusted Curve Numbers (CN, a composite value representing precipitation excess as a function of precipitation, soil cover, land use, and antecedent moisture; Kent, 1972) values for each event modeled (Beskow et al., 2018). These findings bolster the idea that synthetic unit hydrographs may go out of favor compared to newer unit hydrograph methods that take into account geomorphological and climatological factors (Khaleghi et al., 2011) or the rainfall spatial distribution (Andrieu et al. 2021). Increasing the collective understanding of lag time will go a long way in improving overall knowledge of floods and flood processes.

Moreover, several studies have been undertaken across watersheds in South Africa investigating both the applicability of and relationships between estimates provided through different catchment response time parameters (lag time, time of concentration, and time to peak; Gericke & Smithers, 2014) and the variability and inconsistencies between estimates of even one time parameter (Gericke & Smithers, 2016). The former provides credence to the concerns put forth by McCuen (2009) that the inherent confusion in the literature regarding the definitions and interchangeability of different time parameters not only leads to "conceptual and computational misinterpretations" but also "results in significantly different estimates in most cases" (Gericke & Smithers, 2014). The authors go on to recommend that considerable effort is needed to develop more appropriate approaches to ensure both representative and consistent estimations of these time parameters. The latter study, more focused on the estimation of time of concentration, also showcased the range of variability of estimates that can be obtained through the use of different equations even applied over the same subset of basins. This variability was interpreted to be caused by the equations "being applied outside of the bounds of their original developmental regions without the use of local correction factors," though the authors also determined that because no

one equation performed significantly better or worse than the others that the issue is not attributed to the use of "inappropriate catchment response variables" (e.g., basin size and shape, flow distance, and slope) but likely caused by differences in catchment geomorphology (Gericke & Smithers, 2016).

The goal of this study is to assess the significance of factors described in the literature as most important to lag time (i.e., basin area, percent imperviousness, basin length, and slope) (Rao & Delleur, 1974; Sauer et al., 1983). Another goal is to assess the potential further use of geomorphological and climatological parameters to increase the accuracy of lag time calculation, with a byproduct of this involving an evaluation and expansion of the capabilities of the SCS equation. Across the literature countless variations of lag time equations have been developed by researchers and fine-tuned to their regions and local basins, leading to a disconnect in cross-applicability and viability in other regions (as seen by Gericke and Smithers, 2016). Through use of a robust dataset spanning 10 years of lag time values estimated over a large variety of basin characteristics (1,113 basins) across the continental United States (CONUS), this study aims to overcome the limitations of previous studies that were based on datasets lacking representativeness. We propose that an objective assessment of lag time using a data-driven approach on such dataset will generate a clearer picture of the relationships between lag time and basin characteristics, both geophysical and climatological, and provide a definitive step forward towards the creation of a general lag time equation viable across regions.

The paper is organized as follows. Section 2 describes the flood and physiographic data and the methods. Results and discussions are provided concurrently for each case in Section 3, with Section 4 establishing final conclusions.

5.2 Dataset and Methods

The dataset for this study consists of 21,143 observations for 133 variables from 17,491 rainfall and flood events across 1,113 stations (902 unique basins) over the CONUS, compiled from the FLASH Flood Observation Database (Gourley et al., 2013) and extensively quality-controlled from previous studies (Potdar et al., 2021; Saharia et al., 2021). These variables cover a wide range of hydrological aspects such as geomorphological, meteorological, bioclimatic and climatological data, which makes the dataset representative over the U.S. Selected stream gauge locations are used (1) with stages corresponding to action, minor, moderate, and major flooding (2) devoid of regulation or diversion. A flood event is defined at a gauge as the period when streamflow is above the corresponding action stage. A 24-h period with discharge values below action stage is used to separate events. While intense precipitation or snowmelt generate natural floods, the transport of water through the channel network and the lag time are regulated by the physiography of the basin. The catchment behavior is described through geomorphologic and climatological characteristics that represent the integrated contribution of hydrologic processes at the basin outlet. The database comprises attributes representing various properties such as geomorphology, topography, and climatology (see **Table 3** in Section 5.3.2). The main focus for this study will be on geomorphology as these values are known to play a role in the prediction of lag time and will allow for intercomparison between basins, but basin climatological variables will be investigated in tandem to assess their suitability as additional predictors of flood responses.

The distribution of precipitation over the basin is computed from the Multi-Radar Multi-Sensor (MRMS) precipitation reanalysis (Zhang et al. 2016; Zhang and Gourley, 2018) that derives precipitation data at fine spatial and temporal scale (0.01° and 5 min) for a period from 2001 to 2011 over the CONUS from the NEXRAD data archive available from Amazon Web Services

(https://aws.amazon.com/public-datasets/nexrad/). Only high-quality MRMS rainfall data are included by retaining only events that fall in basins with mean radar beam height lower than 2 km above the ground level, helping to reduce potential effects of input uncertainties on modeling results (Potdar et al., 2021). For the purposes of this study, lag time is defined as the duration between the time of the centroid of effective rainfall over the basin and the peak time of the hydrograph (Saharia et al., 2021). The characteristic lag-time variable for a given basin is summarized by the median value computed from all flooding events observed at that station, and it is associated with the corresponding basin physiographic characteristics (e.g., area, slope, river length).

5.3 Results and Discussions

This extension of the estimation of lag time utilizes a wider diversity, and especially wider range, of basins sizes than typically considered in the literature. It also uses a larger pool of geomorphological and climatological variables. The analysis is performed in the following four steps. First, the widely used SCS lag time formula (Kent 1972; **Eq. 1**) is analyzed over its established range of viable basin sizes (i.e., sizes lower than 55 km^2) to evaluate its bias with respect to observed lag time values. A revised equation using the same parameters with updated coefficients is also proposed. As a second step, lag time equations are derived for larger basin size categories to extend the application of lag time estimation. Additional geomorphological and climatological variables are considered. This step provides a more robust breakdown of the effects that basin size has on the coefficients themselves as well as the resulting performance statistics of these lag time equations (correlation coefficient, bias, and percent bias). After the variables undergo a selection process and are locked for each basin size class (**Table 2**), the third step aims

at introducing more specificity to lag time estimation by considering smaller basin size ranges through a moving window to update the variable coefficients through regression (as used in step 1). The prediction performance is contrasted between equations using the coefficients identified in step 2 per basin size category and the specific moving window coefficients. All of these steps serve to not only create a "general equation" utilizing the selected-for pool of variables, but to also provide enhanced flexibility of the general equation by understanding which of the coefficients can be set to zero to increase predictive accuracy based on the user's chosen basin sizes.

Table 2. Final pool of geomorphological (noted Geo as a variable class) and climatological (noted Climo as a variable class) variables generated through stepwise selection and removal.

Variable	**Variable Class**	**Definition**
el	Geo	Elongation ratio; a measure of basin shape
rl	Geo	River length
rr	Geo	Relief ratio; average drop in elevation per unit length of river
imperv	Geo	Basin total surface imperviousness
kfact	Geo	K-Factor; relative index of susceptibility of bare, cultivated soil to particle detachment and transport by rainfall
G1	Geo	First-order moment of flow distance (catchment averaged flow distance)
Bio_2	Climo	Mean diurnal range
Bio_5	Climo	Max temperature of the warmest month
Bio_7	Climo	Temperature annual range

Bio_13	Climo	Precipitation of the wettest month
Bio_15	Climo	Precipitation seasonality (coefficient of variation)
Bio_18	Climo	Precipitation of the warmest quarter

To round out the investigations of lag time equations, the addition of flashiness ("f") is considered in addition to the already identified geomorphological and climatological variables as a compound variable built for improved flood characterization by Saharia et al. (2017). Created to serve as measure of flood severity, flashiness is defined as "the difference between the peak discharge and action stage discharge normalized by the flooding rise time and basin area", providing a metric that exhibits both the magnitude and timing aspects of floods (Saharia et al., 2017).

The results of these investigations into the calculation of lag time within this section are divided into subsections by analysis. As such, discussions pertaining to each analysis will be represented in each corresponding section in order to preserve clarity and cohesion. The subsections focus on the small-scale basin investigation, the overall ability of the equations to predict lag time, the response of the equations to a moving window, and the addition of the flashiness variable, respectively.

5.3.1 Updated SCS Equation on Basins Smaller Than 55 km^2

This analysis focuses first on the basin sizes that fall within the range of viability of the SCS lag time equation (**Eq. 5**; from 1 mi^2 up to approximately 22 mi^2 or 55 km^2; Folmar et al.,

2007; Kent, 1972). The dataset contains 22 basins with less than 55 km^2 of area, with the smallest of those being ~22 km^2. In this equation the lag time, T_{lag}, is a function of the river length of the watershed (L), the curve number (CN), and the slope index (i.e., the slope between two points along the main channel upstream from the outlet of the basin at distances equal to 10 and 85 percent of the total main-channel length; SI). For the sake of this study, however, and for ease of calculation the entire curve number term in **Eq. 6** (i.e., 1000 / CN – 9) will be replaced by 'S' as it often is in the literature, which denotes the overall maximum retention in the watershed.

$$\boldsymbol{T_{lag} = \frac{L^{0.8}[\left(\frac{1000}{CN}\right)-9]^{0.7}}{1900*SI^{0.5}}} \qquad \text{(Eq. 5)}$$

$$\boldsymbol{\ln\left(T_{lag}\right) = 0.8\ln(L) + 0.7\ln\left[\left(\frac{1000}{CN}\right) - 9\right] - \ln(1900) - 0.5\ln\left(SI\right)} \qquad \text{(Eq. 6)}$$

A multivariate linear regression approach with leave-one-out cross validation (LOOCV) is used after taking the natural log of both sides of the equation (**Eq. 6**). LOOCV cross validation is applied to identify the best model and to improve the model representativeness by using all observations for both training and validation. The goal is to generate a revised equation using the same parameters but with updated coefficients that can be applied in the same manner as the SCS equation.

For comparison with subsequent analyses using larger datasets in the following sections, another fit was performed with a set of geomorphological and climatological variables identified in Section 5.3.2 for various classes of basin sizes and listed in **Table 2**. The selected parameters and their associated coefficients values can be seen in **Table 3**.

Table 3. Coefficient values used for each equation for lag time prediction in basins <50 km^2.

Asterisks (*) represent significance at ***p*-values < 0.05.**

Variable	SCS	Revised SCS	General
SI	-0.23	-0.183	--
S	1.318	0.280	--
rl	0.554	0.108	3.075*****
el	--	--	2.712*****
rr	--	--	-0.430
imperv	--	--	-0.219*****
kfact	--	--	0.038
G1	--	--	--
Bio_2	--	--	3.396
Bio_5	--	--	-5.865
Bio_7	--	--	--
Bio_13	--	--	0.487
Bio_15	--	--	-0.490
Bio_18	--	--	0.595
Intercept	-7.55	-0.615	-22.17

All models are evaluated and intercompared with respect to the observed lag times using

LOOCV as shown in **Figure 48**, with provided statistics of correlation coefficient (CC), overall bias, and percent bias (p-bias).

The SCS equation performed as expected (**Figure 48A**), with distinct underestimation of lag time values compared to observations. Improvement is seen in the revised equation (**Figure 48B**), removing a significant amount of bias but providing negligible improvements to overall correlation. Note that the parameter coefficients have consistent signs across the SCS equation and the revised version. The greatest improvements in calculation are seen in the application of the general equation (**Figure 48C**), with this form providing an even improved bias correction from the revised SCS equation as well as significant improvements to overall correlation (+40% explained variance). Obviously, these improvements stem from the larger set of parameters compared to the SCS equation, but it also illustrates the applicability of the general model to small basins.

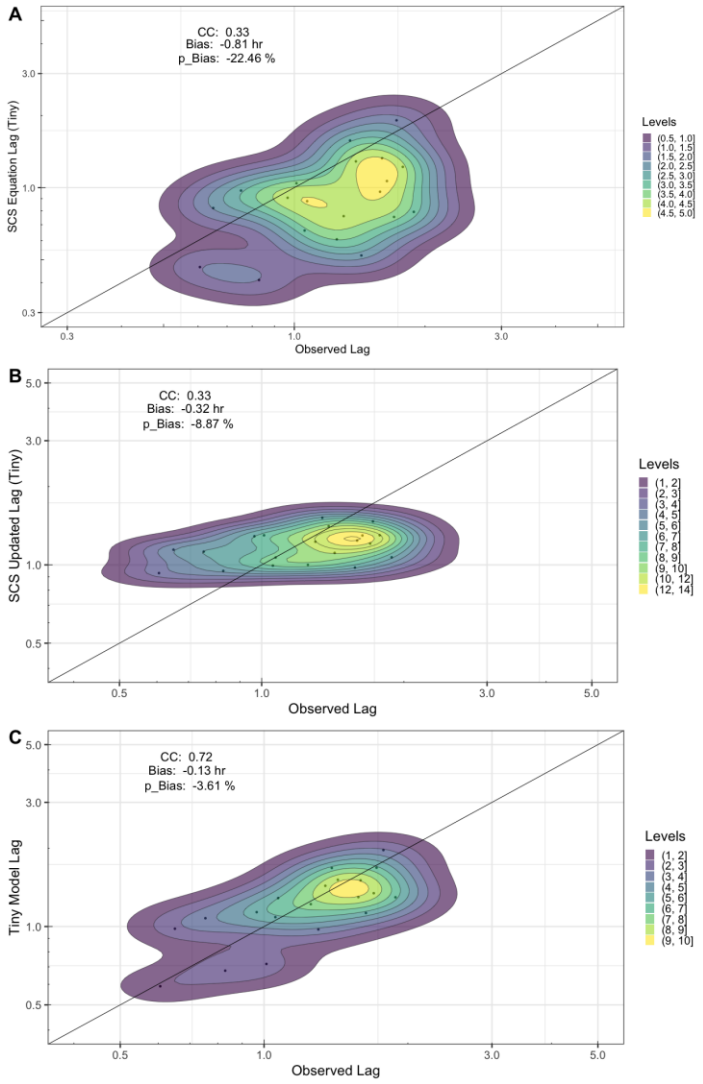

Figure 48. Density scatterplots and statistics of calculated versus observed lag times for the <50 km^2 basins using the original SCS equation (A), the revised SCS equation (B), and the stepwise-selected equation (C). The black line represents the 1:1 line.

These results show that for those working in forecasting or planning at this basin scale, even small improvements to the SCS equation itself can result in more accurate flood predictions. While it is apparent that the proposed general equation has even higher returns for statistics, if all of the required basin information for successful calculation is not known then forecasters can still

settle for the revised SCS equation (as it is assumed that they would already have access to the same three variables). More importantly, however, these results do serve as a good baseline for the rest of this study as larger and larger basin sizes are examined.

5.3.2 Lag Time Investigations for Larger Basin Categories

In this second step, lag time equations are derived for larger basin size categories to extend the application of lag time estimation. A stepwise investigation of geomorphological and climatological variables is performed to allow for a much more robust understanding of how the calculation of lag time can be optimized or better built over a range of basins sizes. The distribution of basin areas with respect to their associated index value is provided in **Figure 49**. A preliminary covariance analysis was performed on the 43 available geomorphological and climatological variables in order to make sure the starting pool of variables maintained as little covariance as possible. The resulting final pool contained 26 variables. The bidirectional stepwise regression was run across the dataset using a moving window approach. With the basins arranged as an index from smallest area to largest, a 100-basin window was applied at each index point (50 basins before and after each indexed basin), generating a series of 1113 logged stepwise variable selections. Once a full run was completed, the number of times each variable was selected was tallied and the least-selected variable in the pool was removed. This iteration of stepwise selection and variable removal continued until any further removal created a detriment to the equation's performance. What was left at the end of this process was a more parsimonious pool of six (6) geomorphological and six (6) climatological variables. These final variables and their definitions can be seen in **Table 3**.

The stepwise analysis was further broken down to understand how variables were being selected relative to basin size. By looking at the stepwise selection for each iteration of the moving window, each instance where an individual variable was selected can be logged. From there, a curve can be generated using each variable's logged coefficient value at each basin index in order to showcase where and when the variable starts or stops being selected for. This allows for a natural breakdown of the indexes into potential size classes, representing more parsimonious subsets of application of the general equation. The results from the selection test can be seen in **Figure 50**. From the plot, three distinct sets of selected variables can be associated with three ranges of basin sizes: indexes 1 – 400 (smaller scale basins, less than 518 km^2; see **Fig. 49**), indexes 401 – 900 (medium scale basins, between 518 km^2 and 2712 km^2), and the remaining indexes 901 – 1113 (the largest basins, with areas greater than 2712 km^2). With this understanding of variable behavior as a function of basin area, even more parsimonious equations can be generated for each range.

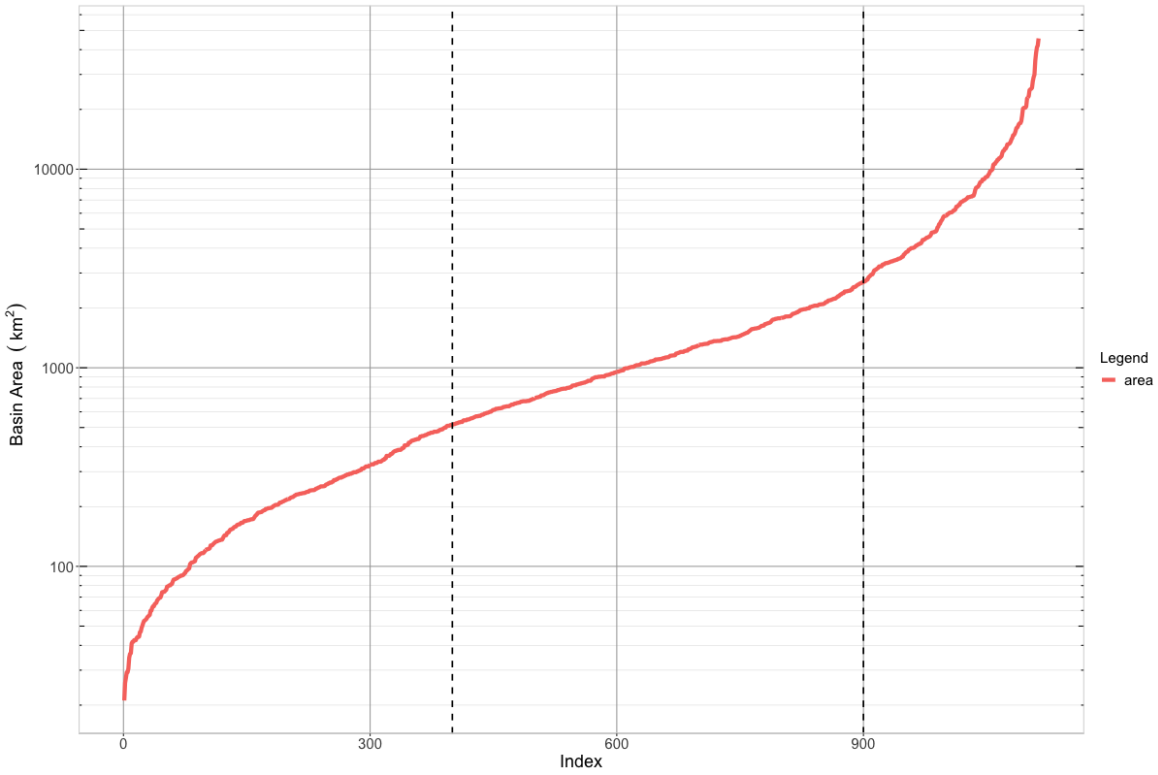

Figure 49. Distribution of basin areas by index value. Dashed lines represent cutoff locations for the Small (left), Medium (middle), and Large (right) basin size classes.

With the variables and respective coefficients for each size range locked in (see **Table 3**), the same linear model regression approach from the SCS equation analysis can be applied.

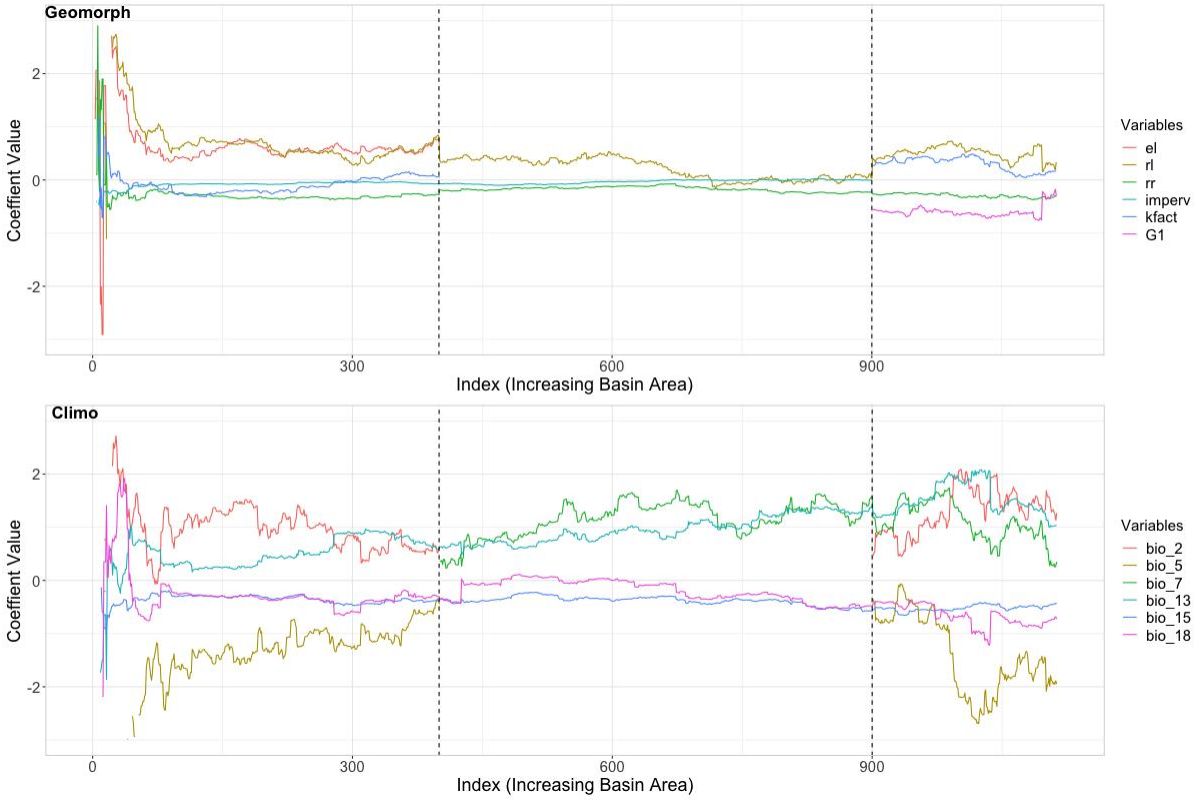

Figure 50. Curves of variable coefficient values as a function of basin index (i.e., basin area). Dashed lines represent the index boundaries for each basin size class.

Table 4. Coefficient values for each variable used in each final equation, as well as the X-intercept generated by the regression. One asterisk (*) represents significance at p-values < 0.05, two asterisks (**) represent significance at p-values < 0.01, and three asterisks (***) represent significance at *p*-values < 0.001.

Variable	Small	Medium	Large
el	0.534***	--	--
rl	0.544***	0.079	0.285
rr	-0.276***	-0.173***	-0.338***

imperv	-0.073***	-0.019	--
kfact	-0.080	--	0.159
G1	--	--	-0.253
Bio_2	0.576	--	1.315
Bio_5	-0.488	--	-1.833*
Bio_7	--	1.525***	0.377
Bio_13	0.563***	1.235***	1.034**
Bio_15	-0.345***	-0.468***	-0.458**
Bio_18	-0.356***	-0.255**	-0.765**
Intercept	-4.699***	-8.060***	2.317

With results successfully extracted from the smallest class of basins (those basins with areas < 55 km^2, henceforth referred to as the "Tiny" class; see section 5.3.1), the more robust aspects of this study of the relationships between lag time, basin size, and basin characteristics were investigated. This second phase follows the same procedure as with the Tiny basins (predicting with each equation) but with the dataset being split up by the pre-established basin size classes (Small, Medium, and Large). No changes were made to the calculation using the original SCS equation, but both the revised and general equations had their respective regressions performed only with the subsets of gauges pertaining to each size class. Also note that for calculations involving the Small class of basins, this subset also contains the 22 basins from the Tiny class. The coefficients for these equations can be found in **Table 4**. This allows for not only an assessment of how the equations change within each size class, but also provides a more accurate equation based on the size class that one may be utilizing. Once again, the predicted

results from each equation using each gauge established characteristics were plotted against the basin's median observed lag time value, with the results shown in **Figure 51**.

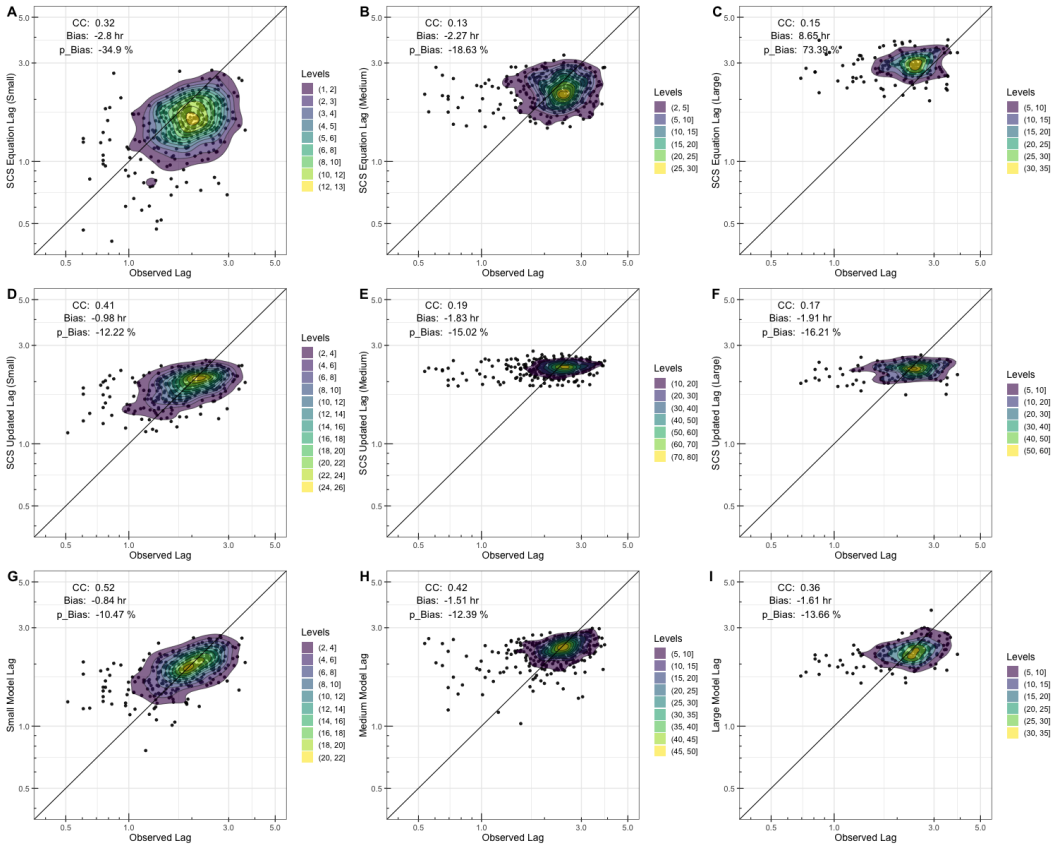

Figure 51 Density scatterplots and statistics of calculated versus observed lag times for each basin class (Small, Medium, Large) using the original SCS equation (A,B,C), the revised SCS equation (D,E,F), and the stepwise-selected equation (G,H,I). The black line represents the 1:1 line.

Overall, a similar story is seen here to what was seen with the Tiny basin benchmark. The SCS equation underestimates lag time values, with the exception of the Large basin class where it instead switches to significant overestimation. The revised SCS equation again successfully lowers the overall bias of predictions across each basin class, also noticeably returning the predictions of

the Large class to a position of underestimation. Significant improvements across all statistics are seen with the predictions from the proposed general equations, though diminishing returns can be seen as basin size increases.

5.3.3 Moving Window with Basin Size

With a baseline understanding of the behavior of the error statistics from discrete values calculated for each class of basin size, the next step to delve deeper was to investigate how these statistics change across finer basin scale ranges using the same 100-basin moving window technique. Each equation was applied across their respective size class ranges, with the exception of the original and revised SCS equations which were run across the full range of basins. In addition, this approach was split into two parts; one run of moving windows that utilized fixed equation coefficients (established in Section 5.3.2) that simply calculated statistics in the same manner as those in **Figure 51** but for each index point, and a second run of the moving window where a new regression of coefficients was performed within the window at each index before subsequently calculating the statistics. The rationale behind this approach was to look at a finer range of basin areas at a time, instead of the entire range of the size class. This increases the granularity of the results, providing a more specific representation of the underlying variable behaviors. **Figure 52** shows the results of this investigation, with the top half (plots A, B, and C) corresponding to the run with fixed equation coefficients and the bottom half (plots D, E, and F) representing the run with “unlocked” equation coefficients.

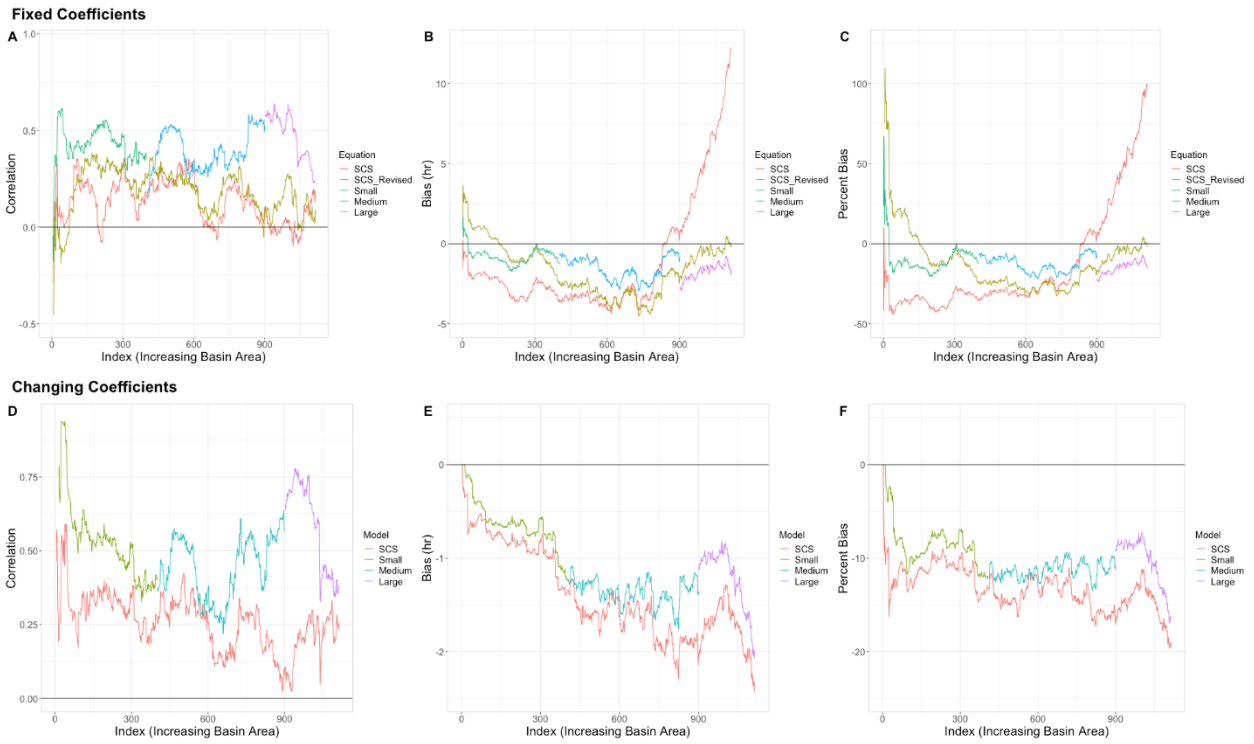

Figure 52. Line plots of correlation, bias, and percent bias as a function of basin area for each lag time equation. Upper plots utilized a 100-basin moving window with fixed equation coefficients, while lower plots utilized the same moving window while calculating new regression coefficients at each step.

Starting with correlation (**Figure 52A** and **52D**), the same trends are seen in the moving window plots as were seen with the density scatters: all three classes of the general equation obtain higher correlation values than the SCS equations (greater than 10% increase in explained variability in most places), with only a few points where they perform similarly (index ~300 and index ~600). Expectedly, the revised SCS equation also tends to have higher correlation overall than the original equation but because both versions still only have access to the same three variables, they behave similarly across the range of basin sizes. What is interesting, however, is that there is noticeably significant correlation for both SCS equations far beyond their established

range of basin area viability (from less than 50 km^2 all the way to ~2000 km^2, or approximately index 600). This gives further credence to the underlying assumption of this study, that a general equation for lag time is very much obtainable; the viability of the wealth of available lag time equations is less about the size range for which they were originally calculated and more about the variables and basin characteristics represented within them. Additionally, correlation values for all the equations increase overall when using the "unlocked" coefficient approach, meaning even more optimization of coefficients is possible should a form of these general equations be developed as a function of basin area itself.

Trends within the bias (**Figure 52B** and **52E**) and percent bias plots (**Figure 52C** and **52F**) behave similarly as well, with the original SCS equation having the most conditional bias, the revised SCS equation improving on bias, and both the Small and Medium class general equations performing with the least bias. Interestingly, when looking at the fixed coefficient plots (**Figure 52B** and **52C**) the Large class equation ends up with more bias than the revised SCS equation. This does not hold true for when the coefficients are unlocked in the moving window (**Figure 52E** and **52F**), however, suggesting that the increased bias in the fixed coefficient equation is tied to the higher variability across these largest basins and that finer tuning is required at this range.

5.3.4 Lag Time and Flashiness

The final step of these investigations involves the addition of flashiness ("f"), the compound variable built for improved flash flood forecasting and lag time calculation (Saharia et al., 2017). Knowing that flashiness is designed to be well correlated with lag time, it was decided that it should be left out of the original stepwise selection procedure to guarantee a more

representative assessment of the most important discrete basin characteristics. The dataset has several iterations of flashiness available for each event, but the one chosen for this study is the "basin median flashiness" value as the lag times being used as observations are median values as well. Median flashiness was subsequently added into the general equation pool, outside of stepwise selection, bringing the total number of variables used to thirteen (13). As such, the updated regression-calculated coefficients for each of the four general equations (Tiny, Small, Medium, and Large) are provided in **Table 5**.

Table 5. Coefficient values for each variable used in each flashiness-included equation, as well as the X-intercept generated by the regression. One asterisk (*) represents significance at p-values < 0.05, two asterisks (**) represent significance at p-values < 0.01, and three asterisks (***) represent significance at *p*-values < 0.001.

Variable	Tiny	Small	Medium	Large
el	0.364	0.307**	--	--
rl	0.716	0.267***	-0.024	0.256
rr	0.306	-0.063*	-0.094**	-0.229*
imperv	0.056	-0.055***	-0.031	--
kfact	0.235	0.054	--	0.188
G1	--	--	--	-0.249
Bio_2	0.464	0.011	--	1.091
Bio_5	6.645	0.851**	--	-1.362
Bio_7	--	--	1.605***	0.588
Bio_13	-2.096*	0.612***	1.363***	1.181***

Bio_15	0.177	-0.297***	-0.474***	-0.467**
Bio_18	3.990**	-0.219**	-0.273**	-0.744**
f	-0.399***	-0.207***	-0.115***	-0.076*
Intercept	-41.42***	-5.686***	-7.938***	0.192

Following the methodology of the previous two subsections, density scatterplots were made first and are represented in **Figure 53**. For reference, and to show the full story of equation results trends, the first three rows contain the data previously established in Sections 5.3.1 and 5.3.2. The fourth row (**Figure 53M** and **53P**) represents the flashiness-included general equation results. What can be seen is a striking improvement in correlation from the addition of flashiness, though this appears to come with diminishing returns as basin size increases (i.e., correlation jumps from 0.72 in the Tiny class to 0.93, whereas the Large class only increases from 0.36 to 0.37). This makes sense, as flashiness plays a more important role in smaller basins (that are naturally more susceptible to flash floods in general). With the largest of basin sizes, however, there are more factors that play into the lag time of a flood wave than simply the basin's flashiness, like what was seen in the stepwise analysis where more climatic, larger-scale variables were selected. The flashiness equations also maintain some conditional bias, albeit slightly improved over their predecessors. For potential forecasters, however, this still equates to almost half an hour of bias correction in smaller basins, which is an invaluable amount of warning time and accuracy.

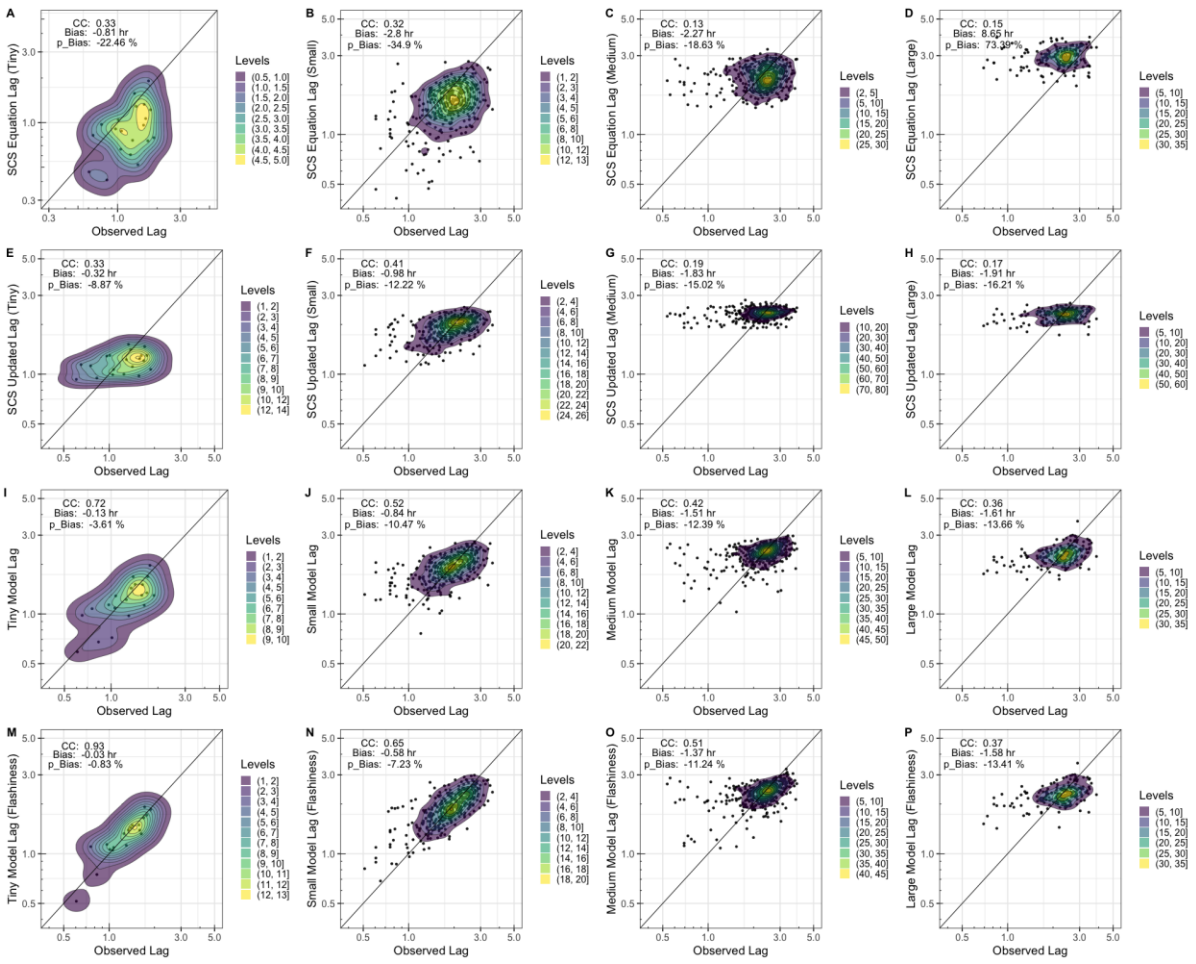

Figure 53. Density scatterplots and statistics of calculated versus observed lag times for each basin class (Tiny, Small, Medium, Large) using the original SCS equation (A,B,C,D), the revised SCS equation (E,F,G,H), the stepwise-selected equation (I,J,K,L), and the stepwise equation with included flashiness (M,N,O,P). The black line represents the 1:1 line.

The moving window approaches were also utilized with the new flashiness-included equations, but for simplicity **Figure 54** only plots the two sets of general equations in order to more clearly represent the differences and improvements between the two. Much like the density scatterplots, **Figure 54A** and **54D** show both the significant improvements in correlation and subsequent diminishing returns provided through the addition of flashiness. Interestingly, not even

the moving window with “unlocked” coefficients was able to provide significant improvements to correlations in the Large class, confirming the idea that at this scale flashiness is simply just not as important to lag time calculation as other variables. The small improvements to overall bias and p-bias can also be seen in their respective plots, but more research will be necessary in order to ascertain why these improvements are more muted.

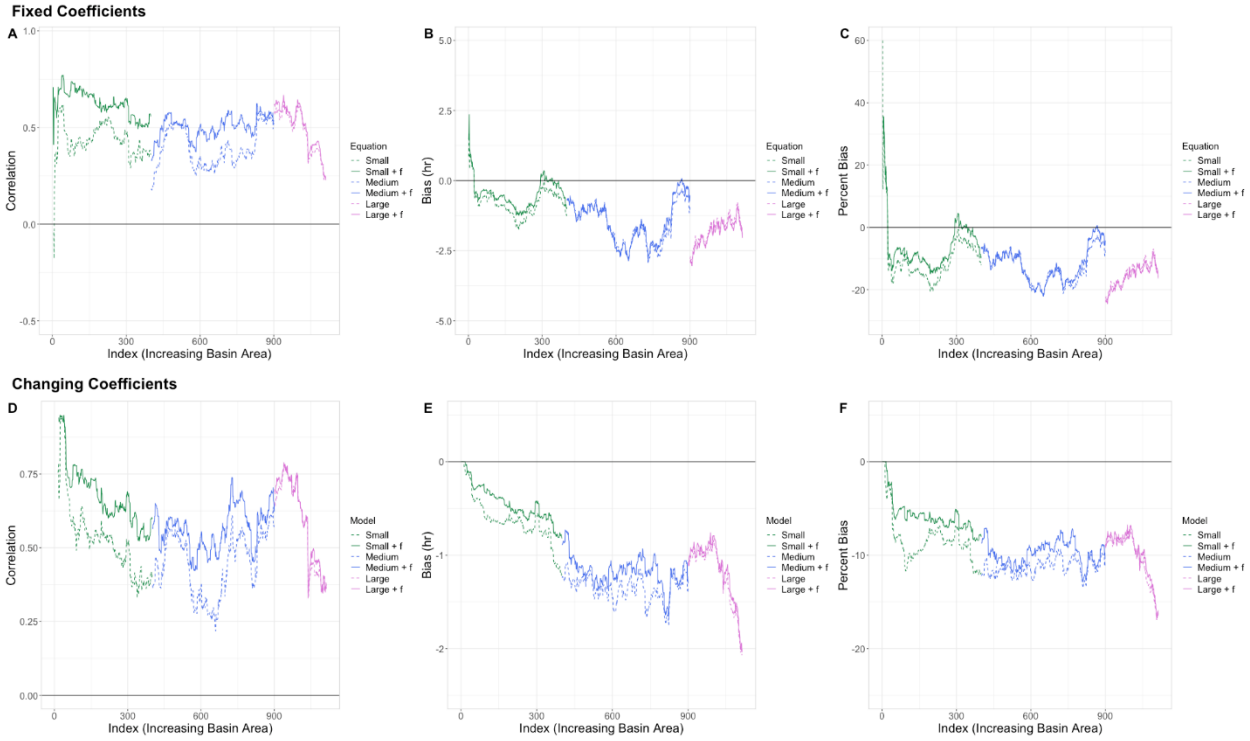

Figure 54. Line plots of correlation, bias, and percent bias as a function of basin area for the general lag time equations (with and without flashiness). Lines are color-coded by size class (green for Small, blue for Medium, and magenta for Large), with the solid lines representing the equation with flashiness and dashed lines representing the equation without. Upper plots utilized a 100-basin moving window with fixed equation coefficients, while lower plots utilized the same moving window while calculating new regression coefficients at each step.

5.4 Conclusions

This study utilized an expansive dataset of flash flood events and their respective geomorphological and climatological characteristics to not only improve a long-standing equation for the calculation of lag time but to also provide a robust new equation of discrete characteristics that can be fine-tuned based on the user's basin size class needs. The analysis presented in this study was performed across four different steps. First, the widely used SCS lag time formula was analyzed over its established range of viable basin sizes (i.e., sizes lower than 55 km^2) to evaluate its bias with respect to observed lag time values. A revised SCS equation using the same parameters with updated coefficients was also calculated and proposed, with the results showing a distinct removal of the bias normally seen with the original SCS equation.

As a second step, lag time equations were derived for larger basin size categories as a means of extending the application of lag time estimation through the consideration of additional geomorphological and climatological variables. This step provided a more robust breakdown of the effects that basin size has on the coefficients themselves as well as the resulting performance statistics of these lag time equations (correlation coefficient, bias, and percent bias). It was found that smaller basin scales favored more basin-specific variables, such as elongation ratio and k-factor, while larger basins instead favored the climatological variables as well as the catchment-averaged flow distance (G1). Across all size classes, this new general equation noticeably improved all three performance statistics. It was also found that the revised SCS equation can still perform acceptably at ranges far beyond what it was considered viable for.

The third step was aimed at introducing more specificity to lag time estimation by considering even narrower basin size ranges through a moving window approach to update the

variable coefficients through regression (as was done in the first step). The prediction performance was contrasted between equations using the coefficients identified in the second step both per basin size category and per the specific moving window coefficients. Once again, the general equation was shown to almost exclusively outperform across all basin sizes with the exception of the Large class containing more bias than the revised SCS equation.

To round out the investigations of lag time equations, flashiness was added to the pool of already identified geomorphological and climatological variables and the previous two steps were re-run. The addition of flashiness to the general equation significantly improved metrics across the board, albeit with diminishing returns as basin size increased. This was not concerning, as flashiness is known to be a more prevalent variable at smaller basin sizes.

All of these steps served to not only create a general equation utilizing the selected-for pool of variables, but to also provide enhanced flexibility of the general equation by understanding which of the coefficients can be set to zero to increase predictive accuracy based on the user's chosen basin sizes. Developed through a gamut of selection processes, regressions, and validity investigations, this prototype general equation is designed to serve as a stepping-stone for the unification of lag time calculation across the broader hydrologic community. As such, the authors understand that the data utilized in this study was only collected over the Continental United States and implore researchers in other regions to test this general equation in tandem with their own observations and datasets in the hope that a "true" global general equation can continue to be refined and developed.

Chapter 6: Conclusions

6.1 Summary

In recent years, a great amount of research has been done towards evaluating precipitation data generated by satellites, but less has focused on how these estimates and their uncertainties manifest further into the water cycle. This dissertation serves as a means to investigate these tendencies within the lens of simulated discrete flood characteristics while simultaneously addressing needs presented by the developers of the Global Precipitation Measurement mission for integrated hydrologic validation of their precipitation products.

In Chapter 2, the first phase of this study, ten years of satellite-based and ground-based radar data are used as forcings for a distributed hydrologic model across the Continental United States. They are compared using a methodology designed to assess the flood signals and characteristics generated by the model. By focusing on how well the model reproduces flood characteristics rather than fits traditional bulk statistics, this research provides robust insights into satellite precipitation deficiencies. It is found that satellite data has greater success at resolving lower magnitude flood events while tending to generate floods of longer durations. Additionally, flood managers should note that satellites tend to generate floods that characteristically both begin earlier and end later than the ground radar reference. Subsequent research is recommended into other satellite data products in order to better understand these discrepancies and mitigate or plan for them in the future.

Chapter 3 further investigated the hydrologic utility of satellite precipitation estimates from the Global Precipitation Measurement mission by comparing flood signals produced across the Continental United States to an observational ground reference. The flood characteristics

generated with radar and satellite precipitation through a distributed hydrologic model were 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 was 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.

Chapter 4 focused on the effects of spatially, temporally, and spatiotemporally upscaling precipitation forcings from MRMS at native resolution to the native resolutions of IMERG-E. These upscaled products were then run through the hydrologic modeling framework, using the same methodology from Chapters 2 and 3, to assess how the different resolutions impact the simulated outputs of flood magnitude, event duration, and timing. It was found that, despite being compared to products at similar resolutions, the uncertainties generated by the IMERG-E simulations overpowered those associated with the upscaled products across all flood characteristics. When compared to native MRMS simulations, however, the upscaled products tended to underestimate peak discharges and event durations, associated with distinct negative water balance errors. It was also found that the upscaled product simulations exhibited a scaling effect with regards to error contribution of peak discharge, with higher contributions at smaller basin sizes and decreasing contributions as basin sizes increased. Looking at how upscaling can

affect the modeled output desired allows for improved understanding of which resolutions lead to the greatest changes in simulation accuracy, increasing the potential use and utility of current or future satellite products in regions where high-resolution ground radars are sparse.

Chapter 5 constitutes a deeper dive into the inherent variables and calculations involved in flood and flash flood predictions and forecasting. Lag time is a significant metric in this regard, and this chapter served to further develop understanding of how it is calculated and the important basin characteristics to target when refining it. First, the original Soil Conservation Service (SCS) lag time equation was examined within its known range of validity and updated, removing its inherent bias. Further modeling was performed using a stepwise-selected pool of variables to create a new lag time equation. This process was further broken down by assessing the size classes of basins where certain variables are selected for, increasing the specificity and utility of the new general equation for more fine-tuned applications. Furthermore, the entire process was repeated using the compound variable “flashiness”, significantly improving the accuracy of predictions using the general equation. Overall, it was found that this equation performs well in terms of correlation and bias across a wide range of basin sizes. The redevelopment of this important part of flash flood forecasting allows for more accurate future integration into the methodologies of Chapters 2-4, adding an additional layer of integrated validation potential across the Continental United States.

6.2 Final Remarks and Future Work

This research has served to not only be novel in its overall approach, through the use of discrete flood characteristics instead of traditional bulk metric comparisons, but also through the investigations of some of its finer points such as the generation of the model-investigative quadrant

plots or the error contributions from the upscaled-to-IMERG-resolution MRMS product. Integrated hydrologic validation of the GPM mission is a vast undertaking, one certainly not achieved through this study alone, but it is the hope of this research that a noteworthy chunk has been taken out of the process and work can be done to better understand and improve how errors and uncertainty within the satellite products and algorithms are translated through the hydrologic cycle into estimates of flood characteristics. These revelations provide not only timely information to the research-focused areas of the world, but also hope towards better understanding and prediction of floods in ungauged locations globally.

There is always more to be done, however. With the recent release of IMERG Version 7.0, the methodology of this research can be reapplied to provide a direct comparison between the current results (those using Version 6) and simulations generated by forcings from the new algorithm. Another research avenue involves the assessment of solely the infrared contribution of IMERG precipitation as a forcing into EF5, further helping to diagnose where the most important deficiencies may lie when considering translation through the hydrologic cycle. Once a future reanalysis of MRMS through the dual-pol era is finalized (2016 – Present) an entire new timeframe of assessment will be unlocked, one that involves the advancements of not only ground-based radar references but also the true GPM-collected era of satellite measurements, provided a far more modern view of the state of the art when it comes to satellite-forced hydrologic model simulations. And finally, an entire new equation now exists regarding the estimation of the lag time of a flood event, one that once integrated into this research's methodology will allow for the generation of an entirely new set of density and quantile plots and an unprecedented look at the variable itself.

References

Andrieu, H., Moussa, R. & Kirstetter, P.E. (2021). The Event-specific Geomorphological Instantaneous Unit Hydrograph (E-GIUH): The basin hydrological response characteristic of a flood event. Journal of Hydrology, 603, p.127158. https://doi.org/10.1016/j.jhydrol.2021.127158

Beskow, S., Nunes, G.S., MELLO, C.R., Caldeira, T.L., Norton, L.D., Steinmetz, A.A., Vargas, M.M. & Ávila, L.F. (2018). Geomorphology-based unit hydrograph models for flood risk management: case study in Brazilian watersheds with contrasting physiographic characteristics. Anais da Academia Brasileira de Ciências, 90(2), pp.1873-1890, https://doi.org/10.1590/0001-3765201820170430.

Beven, K.J., 2020. A history of the concept of time of concentration. Hydrology and Earth System Sciences, 24(5), pp.2655-2670. https://doi.org/10.5194/hess-24-2655-2020

Clark, M.P., Vogel, R.M., Lamontagne, J.R., Mizukami, N., Knoben, W.J., Tang, G., Gharari, S., Freer, J.E., Whitfield, P.H., Shook, K.R. and Papalexiou, S.M., 2021. The abuse of popular performance metrics in hydrologic modeling. Water Resources Research, 57(9), p.e2020WR029001. https://doi.org/10.1029/2020WR029001

Clark, R.A., Flamig, Z.L., Vergara, H., Hong, Y., Gourley, J.J., Mandl, D.J., Frye, S., Handy, M., Patterson, M., 2017. Hydrological Modeling and Capacity Building in the Republic of Namibia, Bulletin of the American Meteorological Society, 98, 1697-1715. https://doi.org/10.1175/BAMS-D-15-00130.1

Davenport, F.V., Burke, M. and Diffenbaugh, N.S., 2021. Contribution of historical precipitation change to US flood damages. Proceedings of the National Academy of Sciences, 118(4). https://doi.org/10.1073/pnas.2017524118

Derin, Y., Kirstetter, P.E. and Gourley, J.J., 2021. Evaluation of IMERG satellite precipitation over the land–coast–ocean continuum. Part I: Detection. Journal of Hydrometeorology, 22(11), pp.2843-2859. https://doi.org/10.1175/JHM-D-21-0058.1

Derin, Y. and Kirstetter, P.E., 2022. Evaluation of IMERG over CONUS Complex Terrain Using Environmental Variables. Geophysical Research Letters, p.e2022GL100186. https://doi.org/10.1029/2022GL100186.

Desbordes, M., 1974. Re´flexions sur les me´thodes de calcul des re´seaux urbains d'assainissement pluvial. PhD thesis, University of Montpellier 2, France, 224 p.

Duarte García, J.A., 2019. Probabilistic Characterization of Floods from Catchment-Scale Precipitation Moments, (Master's Thesis). Retrieved from ShareOK. (https://hdl.handle.net/11244/321052). Norman, OK: University of Oklahoma.

England Jr, J.F., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr, W.O., Veilleux, A.G., Kiang, J.E. and Mason Jr, R.R., 2019. Guidelines for determining flood flow frequency—Bulletin 17C (No. 4-B5). US Geological Survey.

Flamig, Z.L., Vergara, H. and Gourley, J.J., 2020. The Ensemble Framework For Flash Flood Forecasting (EF5) v1. 2: description and case study. Geoscientific Model Development, 13(10), pp.4943-4958. https://doi.org/10.5194/gmd-13-4943-2020

Folmar, N.D., Miller, A.C. & Woodward, D.E. (2007). History and Development of the NRCS Lag Time Equation 1. JAWRA Journal of the American Water Resources Association, 43(3), pp.829-838. https://doi.org/10.1111/j.1752-1688.2007.00066.x

Gebregiorgis, A.S., Kirstetter, P.E., Hong, Y.E., Carr, N.J., Gourley, J.J., Petersen, W. and Zheng, Y., 2017. Understanding overland multisensor satellite precipitation error in TMPA-RT products. Journal of Hydrometeorology, 18(2), pp.285-306. https://doi.org/10.1175/JHM-D-15-0207.1

Gebregiorgis, A.S., Kirstetter, P.E., Hong, Y.E., Gourley, J.J., Huffman, G.J., Petersen, W.A., Xue, X. and Schwaller, M.R., 2018. To what extent is the day 1 GPM IMERG satellite precipitation estimate improved as compared to TRMM TMPA-RT?. Journal of Geophysical Research: Atmospheres, 123(3), pp.1694-1707. https://doi.org/10.1002/2017JD027606

Gericke, O.J. and Smithers, J.C., 2014. Review of methods used to estimate catchment response time for the purpose of peak discharge estimation. Hydrological sciences journal, 59(11), pp.1935-1971. https://doi.org/10.1080/02626667.2013.866712

Gericke, O.J. and Smithers, J.C., 2016. Are estimates of catchment response time inconsistent as

used in current flood hydrology practice in South Africa?. Journal of the South African Institution of Civil Engineering, 58(1), pp.02-15. http://dx.doi.org/10.17159/2309-8775/2016/v58n1a1

GDAL/OGR contributors (2022). GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation. URL https://gdal.org. DOI: 10.5281/zenodo.5884351

Gourley, J. J., Flamig, Z.L., Vergara, H., Kirstetter, P., Clark, R.A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J.M., Yang, H., 2017. The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. Bulletin of the American Meteorological Society, 98, pp.361–372. https://doi.org/10.1175/BAMS-D-15-00247.1

Gourley, J.J., Hong, Y., Flamig, Z.L., Arthur, A., Clark, R., Calianno, M., Ruin, I., Ortel, T., Wieczorek, M.E., Kirstetter, P.E. & Clark, E. (2013). A unified flash flood database across the United States. Bulletin of the American Meteorological Society, 94(6), pp.799-805. https://doi.org/10.1175/BAMS-D-12-00198.1

G. Huffman, D. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, P. Xie, 2014: Integrated Multi-satellitE Retrievals for GPM (IMERG), version 4.4. NASA's Precipitation Processing Center

Grimaldi, S., Petroselli, A., Tauro, F. and Porfiri, M., 2012. Time of concentration: a paradox in modern hydrology. Hydrological Sciences Journal, 57(2), pp.217-228. https://doi.org/10.1080/02626667.2011.644244

Guilloteau, C., Foufoula-Georgiou, E., & Kummerow, C. D. (2017). Global Multiscale Evaluation of Satellite Passive Microwave Retrieval of Precipitation during the TRMM and GPM Eras: Effective Resolution and Regional Diagnostics for Future Algorithm Development, Journal of Hydrometeorology, 18(11), 3051-3070. https://doi.org/10.1175/JHM-D-17-0087.1

Guilloteau, C., Foufoula-Georgiou, E. (2020). Multiscale Evaluation of Satellite Precipitation Products: Effective Resolution of IMERG. In: Levizzani, V., Kidd, C., Kirschbaum, D., Kummerow, C., Nakamura, K., Turk, F. (eds) Satellite Precipitation Measurement. Advances in Global Change Research, vol 69. Springer, Cham. https://doi.org/10.1007/978-3-030-35798-6_5

Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), pp.80-91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hartke, S.H., Wright, D.B., Quintero, F. and Falck, A.S., 2023. Incorporating IMERG Satellite Precipitation Uncertainty into Seasonal and Peak Streamflow Predictions using the Hillslope Link Hydrological Model. Journal of Hydrology X, p.100148. https://doi.org/10.1016/j.hydroa.2023.100148

Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K. and Iguchi, T., 2014. The global precipitation measurement mission. Bulletin of the American Meteorological Society, 95(5), pp.701-722. https://doi.org/10.1175/BAMS-D-13-00164.1

Kay, A.L., Rudd, A.C. and Coulson, J., 2023. Spatial downscaling of precipitation for hydrological modelling: assessing a simple method and its application under climate change in Britain. Hydrological Processes, 37(2), p.e14823. https://doi.org/10.1002/hyp.14823

Kent, K. M. (1972). National engineering handbook. Hydrology Section, 4. Soil Conservation Service, U.S.D.A., Washington, D.C.

Khaleghi, M.R., Gholami, V., Ghodusi, J. & Hosseini, H. (2011). Efficiency of the geomorphologic instantaneous unit hydrograph method in flood hydrograph simulation. Catena, 87(2), pp.163-171, https://doi.org/10.1016/j.catena.2011.04.005.

Kirstetter, P.E., Delrieu, G., Boudevillain, B. and Obled, C., 2010. Toward an error model for radar quantitative precipitation estimation in the Cévennes–Vivarais region, France. Journal of Hydrology, 394(1-2), pp.28-41. https://doi.org/10.1016/j.jhydrol.2010.01.009

Kirstetter, P.E., Hong, Y., Gourley, J.J., Chen, S., Flamig, Z., Zhang, J., Schwaller, M., Petersen, W. and Amitai, E., 2012. Toward a framework for systematic error modeling of spaceborne precipitation radar with NOAA/NSSL ground radar–based National Mosaic QPE. Journal of Hydrometeorology, 13(4), pp.1285-1300. https://doi.org/10.1175/JHM-D-11-0139.1

Kirstetter, P.E., Y. Hong, J.J. Gourley, M. Schwaller, W. Petersen and J. Zhang, 2013: Comparison of TRMM 2A25 Products Version 6 and Version 7 with NOAA/NSSL Ground Radar-based

National Mosaic QPE. Journal of Hydrometeorology, 14(2), 661-669. doi:10.1175/JHM-D-12-030.1

Kirstetter, P.E., Y. Hong, J.J. Gourley, Q. Cao, M. Schwaller, and W. Petersen, 2014: A research framework to bridge from the Global Precipitation Measurement mission core satellite to the constellation sensors using ground radar-based National Mosaic QPE.
In L. Venkataraman, in Remote Sensing of the Terrestrial Water Cycle (eds V. Lakshmi, D. Alsdorf, M. Anderson, S. Biancamaria, M. Cosh, J. Entin, G. Huffman, W. Kustas, P. van Oevelen, T. Painter, J. Parajka, M. Rodell and C. Rüdiger). AGU books Geophysical Monograph Series, Chapman monograph on remote sensing. John Wiley & Sons, Inc, Hoboken, NJ. doi: 10.1002/9781118872086.ch4

Kirstetter, P.E., Petersen, W.A., Kummerow, C.D. and Wolff, D.B., 2020. Integrated multi-satellite evaluation for the global precipitation measurement: Impact of precipitation types on spaceborne precipitation estimation. Satellite Precipitation Measurement: Volume 2, pp.583-608. https://doi.org/10.1007/978-3-030-35798-6_7

Lamontagne, J.R., Barber, C.A. and Vogel, R.M., 2020. Improved estimators of model performance efficiency for skewed hydrologic data. Water Resources Research, 56(9), p.e2020WR027101. https://doi.org/10.1029/2020WR027101

Lhomme, J., Bouvier, C. & Perrin, J.L. (2004). Applying a GIS-based geomorphological routing model in urban catchments. Journal of Hydrology, 299(3-4), pp.203-216,

https://doi.org/10.1016/j.jhydrol.2004.08.006.

Liu, D., 2020. A rational performance criterion for hydrological model. Journal of Hydrology, 590, p.125488. https://doi.org/10.1016/j.jhydrol.2020.125488

Ma, Z., Tan, X., Yang, Y., Chen, X., Kan, G., Ji, X., Lu, H., Long, J., Cui, Y. and Hong, Y., 2018. The first comparisons of IMERG and the downscaled results based on IMERG in hydrological utility over the Ganjiang River basin. Water, 10(10), p.1392. https://doi.org/10.3390/w10101392

McCuen, R. H. (1998). Hydrologic Analysis and Design. Prentice Hall, Englewood Cliffs, NJ, USA

McCuen, R.H. (2009). Uncertainty analyses of watershed time parameters. Journal of Hydrologic Engineering, 14(5), pp.490-498. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000011

Nanding, N., Wu, H., Tao, J., Maggioni, V., Beck, H.E., Zhou, N., Huang, M. and Huang, Z., 2021. Assessment of Precipitation Error Propagation in Discharge Simulations over the Contiguous United States. Journal of Hydrometeorology, 22(8), pp.1987-2008. https://doi.org/10.1175/JHM-D-20-0213.1

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology, 10(3), pp.282-290. https://doi.org/10.1016/0022-1694(70)90255-6

Newman, A.J., Clark, M.P., Sampson, K., Wood, A., Hay, L.E., Bock, A., Viger, R.J., Blodgett, D., Brekke, L., Arnold, J.R. and Hopson, T., 2015. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. Hydrology and Earth System Sciences, 19(1), pp.209-223. https://doi.org/10.5194/hess-19-209-2015

NRCS (2010). "Time of Concentration," USDA Natural Resources Conservation Service, National Engineering Manual, Chapter 15, 210-VI-NEH.

Potdar, A.S., Kirstetter, P.E., Woods, D. and Saharia, M., 2021. Toward Predicting Flood Event Peak Discharge in Ungauged Basins by Learning Universal Hydrological Behaviors with Machine Learning. Journal of Hydrometeorology, 22(11), pp.2971-2982. https://doi.org/10.1175/JHM-D-20-0302.1

Rao, R.A. and Delleur, J.W. (1974). Instantaneous unit hydrographs, peak discharges and time lags in urban basins. Hydrological Sciences Journal, 19(2), pp.185-198, https://doi.org/10.1080/02626667409493898.

Reed, S., Schaake, J. and Zhang, Z., 2007. A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. Journal of Hydrology, 337(3-4), pp.402-420. https://doi.org/10.1016/j.jhydrol.2007.02.015

Saharia, M., Kirstetter, P. E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping flash flood severity in the United States. Journal of Hydrometeorology, 18(2), 397-411, https://doi.org/10.1175/JHM-D-16-0082.1.

Saharia, M., Kirstetter, P.E., Vergara, H., Gourley, J.J., Emmanuel, I. & Andrieu, H. (2021). On the impact of rainfall spatial variability, geomorphology, and climatology on flash floods. Water Resources Research, 57(9), https://doi.org/10.1029/2020WR029124.

Sauer, V.B., Thomas Jr, W.O., Stricker, V.A. & Wilson, K.V. (1983). Flood characteristics of urban watersheds in the United States (No. 2207). USGPO, https://doi.org/10.3133/wsp2207

Seo, B.C., Quintero, F. and Krajewski, W.F., 2023. Hydrologic Assessment of IMERG Products Across Spatial Scales Over Iowa. Journal of hydrometeorology, 24(6), pp.997-1015. https://doi.org/10.1175/JHM-D-22-0129.1

Sivakumar, B. (2000). Chaos theory in hydrology: important issues and interpretations. Journal of Hydrology, 227(1-4), pp.1-20. https://doi.org/10.1016/S0022-1694(99)00186-9

Smith, A., Sampson, C. and Bates, P., 2015. Regional flood frequency analysis at the global scale. Water Resources Research, 51(1), pp.539-553. https://doi.org/10.1002/2014WR015814

Su, F., Hong, Y. and Lettenmaier, D.P., 2008. Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin. Journal of

hydrometeorology, 9(4), pp.622-640. https://doi.org/10.1175/2007JHM944.1

Swain, D.L., Wing, O.E.J., Bates, P.D., Done, J.M., Johnson, K.A. and Cameron, D.R., 2020. Increased flood exposure due to climate change and population growth in the United States. Earth's Future, 8(11), p.e2020EF001778. https://doi.org/10.1029/2020EF001778

Tan, J., Petersen, W.A., Kirstetter, P.E. and Tian, Y., 2017. Performance of IMERG as a function of spatiotemporal scale. Journal of Hydrometeorology, 18(2), pp.307-319. https://doi.org/10.1175/JHM-D-16-0174.1

Upadhyaya, S.A., Kirstetter, P.E., Gourley, J.J. and Kuligowski, R.J., 2020. On the propagation of satellite precipitation estimation errors: from passive microwave to infrared estimates. Journal of hydrometeorology, 21(6), pp.1367-1381. https://doi.org/10.1175/JHM-D-19-0293.1

Vergara, H., Hong, Y., Gourley, J.J., Anagnostou, E.N., Maggioni, V., Stampoulis, D. and Kirstetter, P.E., 2014. Effects of resolution of satellite-based rainfall estimates on hydrologic modeling skill at different scales. Journal of Hydrometeorology, 15(2), pp.593-613. https://doi.org/10.1175/JHM-D-12-0113.1

Vergara, H., Kirstetter, P.E., Gourley, J.J., Flamig, Z.L., Hong, Y., Arthur, A. and Kolar, R., 2016. Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the Conterminous United States. Journal of Hydrology, 541, pp.421-433. https://doi.org/10.1016/j.jhydrol.2016.06.011

Wang, J., Hong, Y., Li, L., Gourley, J.J., Khan, S.I., Yilmaz, K.K., Adler, R.F., Policelli, F.S., Habib, S., Irwn, D. and Limaye, A.S. (2011). The coupled routing and excess storage (CREST) distributed hydrological model. Hydrological sciences journal, 56(1), pp.84-98. https://doi.org/10.1080/02626667.2010.543087

Woods, D., Kirstetter, P.E., Vergara, H., Duarte, J.A. and Basara, J., 2023. Hydrologic evaluation of the Global Precipitation Measurement Mission over the US: flood peak discharge and duration. Journal of Hydrology, p.129124. https://doi.org/10.1016/j.jhydrol.2023.129124

Xue, X., Hong, Y., Limaye, A.S., Gourley, J.J., Huffman, G.J., Khan, S.I., Dorji, C. and Chen, S., 2013. Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins?. Journal of Hydrology, 499, pp.91-99. https://doi.org/10.1016/j.jhydrol.2013.06.042

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S. and Arthur, A., 2016. Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. Bulletin of the American Meteorological Society, 97(4), pp.621-638. https://doi.org/10.1175/BAMS-D-14-00174.1

Zhang, J., and J. Gourley, 2018: Multi-Radar Multi-Sensor Precipitation Reanalysis (Version 1.0). Open Commons Consortium Environmental Data Commons, accessed 15 August 2021,

https://doi.org/10.25638/EDC.PRECIP.0001.

Zhang, J., Tang, L., Cocks, S., Zhang, P., Ryzhkov, A., Howard, K., Langston, C. and Kaney, B., 2020. A dual-polarization radar synthetic QPE for operations. Journal of Hydrometeorology, 21(11), pp.2507-2521. https://doi.org/10.1175/JHM-D-19-0194.1.

Zoccatelli, D., Borga, M., Viglione, A., Chirico, G. B., & Blöschl, G. (2011). Spatial moments of catchment rainfall: rainfall spatial organisation, basin morphology, and flood response. Hydrology and Earth System Sciences, 15(12), 3767-3783, https://doi.org/10.5194/hess-15-3767-2011.

Zoccatelli, D., Borga, M., Zanon, F., Antonescu, B., & Stancalie, G. (2010). Which rainfall spatial information for flash flood response modelling? A numerical investigation based on data from the Carpathian range, Romania. Journal of Hydrology, 394(1-2), 148-161, https://doi.org/10.1016/j.jhydrol.2010.07.019.