Central Region Technical Attachment
Number 23-01
April 2023

MRMS and FLASH Thresholds for Assessing Flash Flood Potential in Realtime for the Western Great Lakes

W. Scott Lincoln, NWS Chicago
Sarah Marquardt, NWS Milwaukee

1.0 Introduction

At the beginning of 2020, the National Weather Service started issuing reformatted flash flood warnings which allow the ability to indicate the expected damage threat from the event. With three severity levels to these warnings (base, considerable, and catastrophic), and the ability to issue advisories for more nuisance-level events, the most extreme flash flooding situations can be separated from the lower-end situations that impact a far smaller number of people. Guidance remains somewhat limited, however, on how to best predict the severity level of a flash flood threat.

In late 2019, the National Weather Service (NWS) Chicago office began using experimental products from the Multi-Radar Multi-Sensor (MRMS; Zhang et al. 2016) system and the Flood Locations and Simulated Hydrographs (FLASH; Vergara et al. 2019) project to assist with estimating flash flood severity. Four products, 1-hr radar-only QPE, max QPE average recurrence interval (ARI), max QPE-to-GFFG ratio, and unit streamflow, were combined together into a single procedure to view simultaneously (hereafter referred to as the “4-Panel Technique"). More on this approach is discussed in section 2.3. To improve the usage of MRMS and FLASH in operations, NWS Chicago collaborated with NWS Milwaukee to collect information about notable flash flood events in the western Great Lakes region. Peak values for MRMS and FLASH products were determined, and then cases were subsequently assigned an estimate of severity. A script was written to determine the effectiveness of the 4-Panel Technique and then determine the best thresholds for each product. In this study we provide an overview of how cases were collected, how relative severity levels were assigned, how MRMS/FLASH threshold values were calibrated to improve flash flood warnings, and how forecasters can use the 4-Panel Technique to assess/forecast the potential severity of flash flooding in the flash flood warning decision process.

2.0 Methodology

2.1 Collecting Impacts for Flash Flood Cases

A list of flash flood events occurring in the NWS Chicago and NWS Milwaukee Hydrologic Service Areas (HSAs) was created covering summer 2016 through fall 2022. Impacts from each event were reviewed to find instances of roadway flooding, structure flooding, or water rescues, similar to the method presented by Lincoln and Thomason (2018) and summarized by Table 1. A subjective assessment of flash flood severity (using the impact-based warning (IBW) levels base, considerable, catastrophic) was then assigned to each flash flood case. In instances where a single heavy rainfall or flash flood event covered multiple counties and there were notable separations to the hardest-hit areas, the event was broken up into multiple cases.

Table 1. Criteria used to identify different types of flash flood impacts and the IBW levels that correlate to that impact.

Impact Type	Description	Possible IBW Levels
Roadway Flooding	Generally 6 inches or greater of swiftly moving water.	Base, Considerable
Roadway Flooding (major)	Any flooding of a major roadway (such as an interstate), or at least 3 feet of flooding from swiftly moving water.	Considerable, Catastrophic
Structure Flooding	Any ground floor flooding of a structure.	Considerable
Structure Flooding (Major)	Ground floor flooding of a structure exceeding 3-foot depth.	Considerable, Catastrophic
Water Rescues	Any situation where a person must be rescued from a vehicle or structure due to flooding.	Base, Considerable, Catastrophic
Other	Any flooding of normally dry areas that does not fit into the above categories.	Base, Considerable, Catastrophic
Miscellaneous, Minor Flooding	Roadway flooding or poor drainage area flooding not as severe as criteria indicated above	Advisory (non-FFW)

For further analysis and collective comparison of MRMS/FLASH values to flash flood severity, each severity level was assigned a numeric value ranging from 0.0 (no flooding) to 4.0 (flash flood - catastrophic). See Table 2 for a list of each severity level and the warning level equivalent. In some instances, it was somewhat ambiguous which flood severity should be assigned to a particular event. For those cases, the event was assigned a number in between severity levels (for example 3.5 instead of 3.0 or 4.0). This provided the ability to assess these uncertain cases differently, if desired.

Table 2. The numeric values used for classifying flash flood severity.

IBW Level	Severity (Numeric)
No Flooding	0.0
Advisory	1.0
Flash Flood Warning Base	2.0
Flash Flood Warning Considerable	3.0
Flash Flood Warning Catastrophic	4.0

2.2 Matching Events to MRMS and FLASH Data

Once the list of flash flood cases was created, each case was tied to the maximum value reached for MRMS and FLASH products in the 4-panel Technique. Peak values for 1-hr radar-only QPE, max QPE ARI, max QPE GFFG ratio, and unit streamflow, within several miles of the reports of flooding, were collected. Gridded values were averaged over four pixels to reduce the impact of single, isolated values skewing the results. Values over areas where no flooding could occur (such as Lake Michigan) were ignored.

Two approaches were taken to retrieve MRMS and FLASH data. The goal was to not just determine the typical values for known flash flood events, but to also determine the potential outcome of a situation where a given MRMS/FLASH value was indicated. The distinction, while subtle, is important for operations because null cases are excluded when you start with known events. First, data were retrieved for the documented flash flood cases (see sections 2.1 and 2.2). Second, data were retrieved for any instance where any MRMS/FLASH product exceeded the advisory level thresholds depicted in Table 3, regardless of whether a report of flash flooding was received. The next section provides more detail about finding potential null cases.

2.3 The 4-Panel Technique

At NWS Chicago, forecasters use four panels of MRMS and FLASH products together to assess needed flash flood products, known as the “4-Panel Technique.” The four panels include 1-hr Radar-Only QPE, Max QPE ARI, Max QPE GFFG Ratio, and Unit Streamflow. Each panel has custom color tables aligned to the expected magnitude of flash flooding. Forecasters use a consensus approach, looking for at least two to three panels (products) showing the same flash flood impact level, before issuing a particular hazard product. Although MRMS and FLASH products are a primary decision aid for making warning decisions, other information such as the spatial footprint of elevated MRMS/FLASH values, biases in MRMS data, trends in MRMS/FLASH values, meteorological considerations, and observed reports of flooding are also considered. The thresholds for each of the products were developed using a combination of journal articles (Gerard et al. 2021; Clark et al. 2014; Lincoln and Thomason 2018; Seo et al. 2013), Warning Decision Training Division (WDTD) training modules, and forecaster experience, with the assumption that they would be adjusted in the future as additional information was collected. A summary of the thresholds used by the 4-Panel Technique at NWS Chicago are shown in Table 3.

Table 3. Description of MRMS/FLASH value thresholds and colors currently used by the 4-Panel Technique at NWS Chicago.

Color	Assumed IBW/Impact Level	Value Range
Green	None	1-hr QPE: < 1.0 inches Max QPE ARI: < 1 year Max QPE GFFG Ratio: < 50% Unit Streamflow: < 100 cfs/mi^2
Yellow	Advisory	1-hr QPE: 1.00-1.99 inches Max QPE ARI: 1-1.9 years Max QPE GFFG Ratio: 50-124% Unit Streamflow: 100-199 cfs/mi^2
Red	Flash Flood: Base	1-hr QPE: 2.00-2.99 inches Max QPE ARI: 2-49 years Max QPE GFFG Ratio: 125-249% Unit Streamflow: 200-600 cfs/mi^2
Purple, White	Flash Flood: Considerable/Catastrophic	1-hr QPE: > 3.00 inches Max QPE ARI: > 50 years Max QPE GFFG Ratio: > 250% Unit Streamflow: > 600 cfs/mi^2

The flash flood events collected in section 2.1 were tested using the 4-Panel Technique to see how many panels (products) suggested flash flooding of either the advisory, flash flood base, or flash flood considerable/catastrophic levels. A script was written to download data from the flash server (flash.ou.edu), average over a four-pixel area (0.04 deg by 0.04 deg), and then count how many products exceeded the thresholds in Table 3.

2.4 Adding Null Events to List of Cases

It was determined early in this study that results could be biased if null events were not added to the list of cases. A null event would be a situation where at least one of the MRMS or FLASH products suggested an advisory level impact, but no flooding reports were received. To retrieve potential null cases, a script was written that downloaded gridded 1-hr radar-only QPE, max QPE ARI, max QPE GFFG ratio, and unit streamflow data from the FLASH archive server (flash.ou.edu) every 10 minutes covering late 2018 through 2020[1]. Downloaded data were averaged over a four-pixel area, then checked for values anywhere in the NWS Chicago and Milwaukee HSAs that exceeded the specified threshold (Table 3, advisory thresholds). This process was time and data intensive, requiring retrieval and processing of several gigabytes of data over a multi-month period.

The script suggested approximately 160 additional events for review. These potential cases were first checked against the existing list of cases and duplicates were removed. Each potential case was then reviewed manually, removing cases that were based upon erroneous data and assigning a subjective severity. After review, 100 cases were added, with just two cases having any coincident reports of flooding.

2.5 Assessing Skill for MRMS/FLASH Products and 4-Panel Technique

The probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) were calculated for each MRMS/FLASH product individually and collectively (as part of the 4-Panel Technique). The statistics were calculated for each severity level (0.0-4.0) using the values from Table 3. To count as a predicted flash flood event, at least three of the four products had to exceed the indicated threshold for that product and for the given severity level.

In an attempt to quantify the sensitivity of POD/FAR/CSI values to uncertainty associated with subjective assignment of severity values, statistics were also calculated two different ways for comparison. Using two of four panels (products) for a predicted flash flood forecast instead of three was tested, as well as including ambiguous severity levels (0.5, 1.5, 2.5, and 3.5) with both of the nearby levels. An example of the latter would be using severity levels 1.5, 2.0, and 2.5 for examining flash flood base events, rather than only severity level 2.

[1] Although limited amounts of archived MRMS/FLASH data were once available post-2020, the server that stored the data was shut down, making retrieval scripts no longer functional. This study only includes the data that could be retrieved while it was available.

2.6 Determining the Optimal Thresholds for MRMS/FLASH Products

A script was written to test numerous combinations of thresholds for 1-hr Radar-Only QPE, Max QPE ARI, Max QPE GFFG Ratio, and Unit Streamflow, against the collected flash flood cases to find the combination of thresholds with the highest CSI. This calibrated set of thresholds will guide possible adjustments to the 4-Panel Technique. Tables 4, 5, 6, and 7 show the threshold ranges and intervals used by the calibration script.

Table 4. Threshold ranges and intervals used for 1-hr Radar-Only QPE in calibration of the 4-Panel Technique. Values are in inches.

Flooding Severity	Minimum Value	Maximum Value	Interval
Advisory	0.5	2.5	0.05
Flash Flood base	1.0	3.0	0.05
Flash Flood cons.	1.0	4.5	0.05
Flash Flood cat.	1.0	5.0	0.05

Table 5. Threshold ranges and intervals used for Max QPE ARI in calibration of the 4-Panel Technique. Values are in years.

Flooding Severity	Minimum Value	Maximum Value	Interval
Advisory	1	10	1
Flash Flood base	1	20	1
Flash Flood cons.	5	200	1
Flash Flood cat.	40	200	1

Table 6. Threshold ranges and intervals used for Max QPE-to-GFFG Ratio in calibration of the 4-Panel Technique. Values are percent.

Flooding Severity	Minimum Value	Maximum Value	Interval
Advisory	40	125	5
Flash Flood base	75	190	5
Flash Flood cons.	75	300	5
Flash Flood cat.	125	500	5

Table 7. Threshold ranges and intervals used for Unit Streamflow in calibration of the 4-Panel Technique. Values are in cfs/mi^2.

Flooding Severity	Minimum Value	Maximum Value	Interval
Advisory	100	250	10
Flash Flood base	140	600	10
Flash Flood cons.	200	1000	10
Flash Flood cat.	400	2000	10

For comparison, the script also allows the user to calibrate a threshold for a single component of the 4-Panel Technique to maximize CSI. The script was also written to allow easy calibration of any list of cases provided, which would allow for possible subdividing of cases by land cover type (urban vs. rural) or soil moisture condition (wet vs. dry). The script was not written to be location or NWS WFO specific; it should work for any location given it is provided the properly formatted list of flash flood cases.

3.0 Results and Discussion

3.1 Known Flash Flood Cases

In total, 190 cases (both flash flood cases and null events) were analyzed. Of these cases, approximately half (100) were deemed null (no flood) events. Cases deemed advisory level or flash flood base level were the next most common, with flash flood considerable and flash flood catastrophic events the rarest (Figure 1).

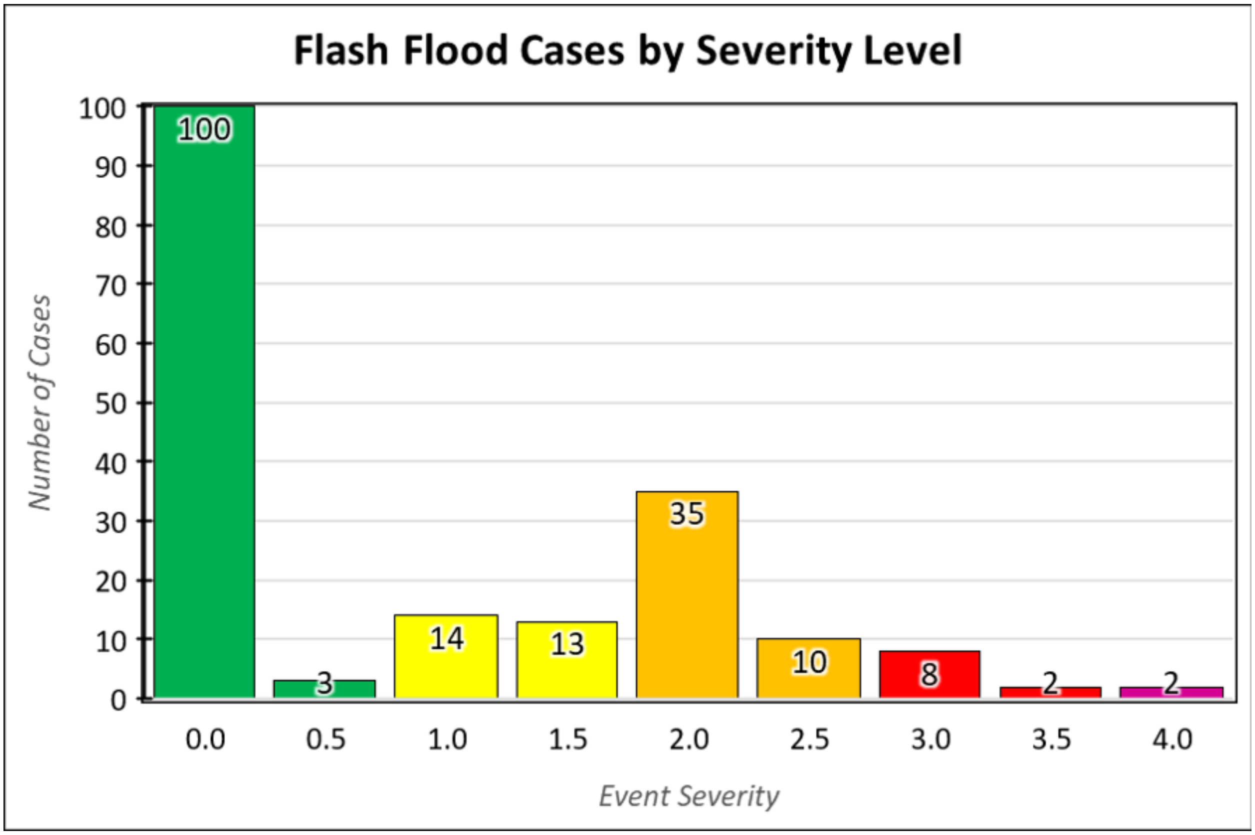

Figure 1. Numeric severity levels for the collected flash flood cases analyzed in this study. A null event (no flooding) corresponds to a value of 0, advisory level corresponds to a value of 1, flash flood base corresponds to a value of 2, flash flood considerable corresponds to a value of 3, and flash flood catastrophic corresponds to a value of 4.

Values for 1-hr radar-only QPE ranged from 0.2 inches to 5.8 inches with a median of 1.6 inches. Values for max QPE ARI ranged from 0 years to 200 years with a median of 15 years. Values for max QPE GFFG ratio ranged from 30% to 450% with a median of 130%. Values for unit streamflow ranged from 10 cfs/mi^2 to 1900 cfs/mi^2 with a median of 295 cfs/mi^2. The range of values for each MRMS/FLASH product broken up by event severity is shown by Figures 2, 3, 4, and 5. Note that in these figures the severity bins are inclusive of ambiguous cases, for example the bin for flash flood considerable (severity value 3) includes cases with severity 2.5, 3.0, and 3.5, to account for uncertainty and to increase the number of available cases for binning. A strong correlation was noted between increasing severity and increasing product values for max QPE ARI, max QPE GFFG ratio, and unit streamflow. Although there was a correlation between increasing severity and increasing product value for 1-hr radar-only QPE, it was not as strong as with the other products and there was significant overlap.

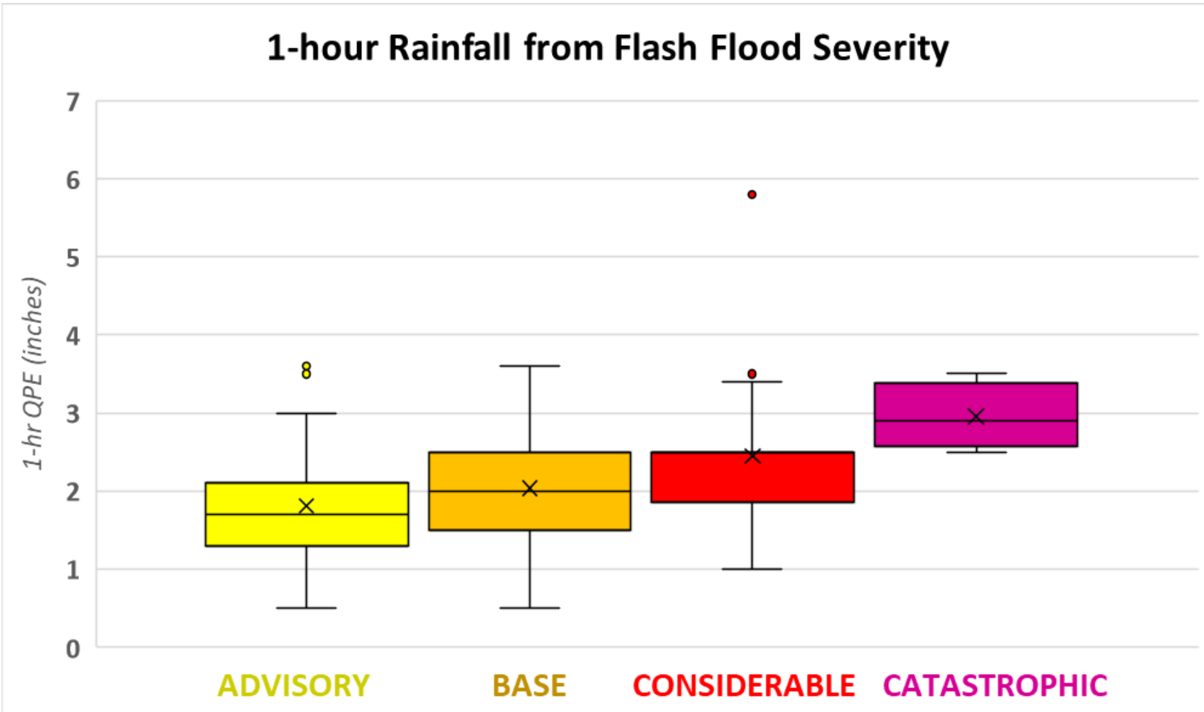

Figure 2. The range of 1-hr Radar-Only QPE values for collected flash flood cases at the Flood Advisory (1), Flash Flood Warning Base (2), Flash Flood Warning Considerable (3), and Flash Flood Warning Catastrophic (4) severity levels. Note the mean values depicted with "x" marks.

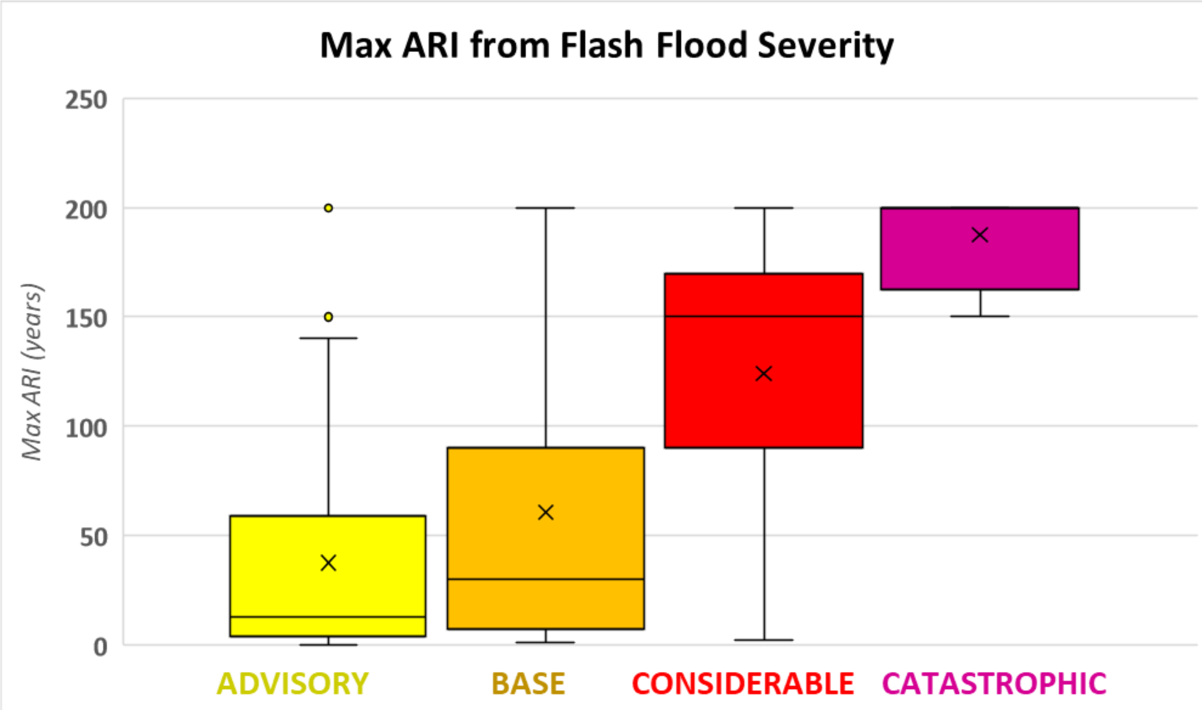

Figure 3. The range of Max QPE ARI values for collected flash flood cases at the Flood Advisory (1), Flash Flood Warning Base (2), Flash Flood Warning Considerable (3), and Flash Flood Warning Catastrophic (4) severity levels. Note the mean values depicted with "x" marks.

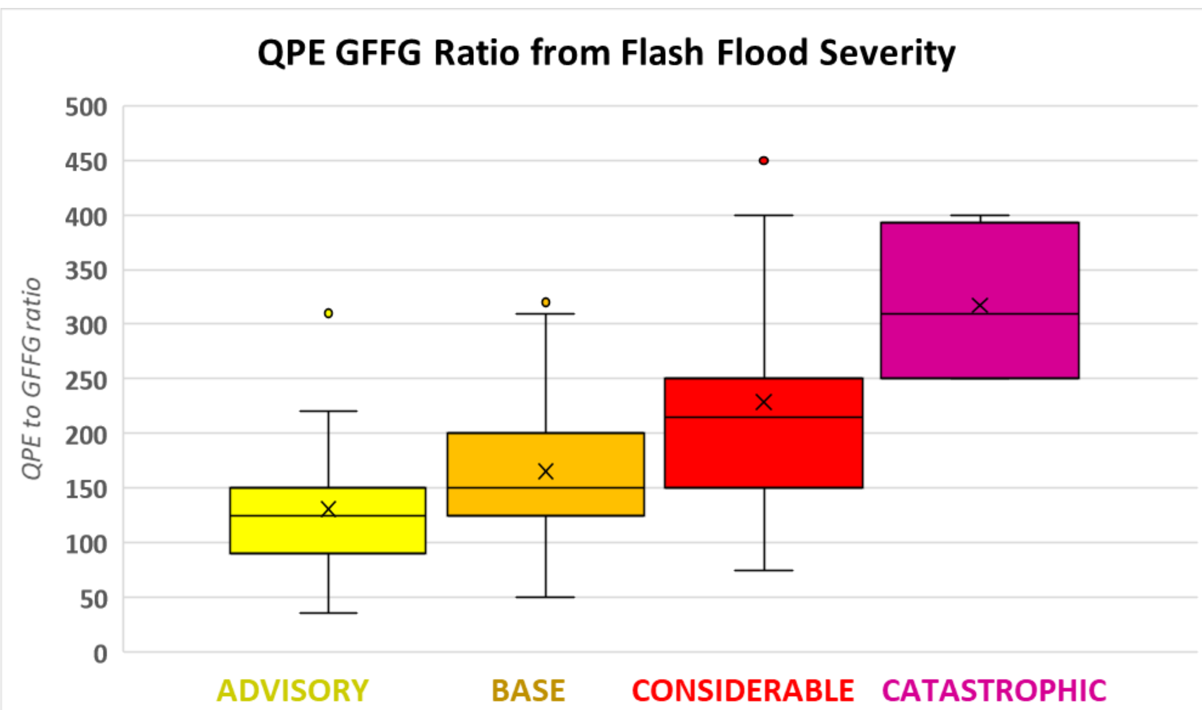

Figure 4. The range of Max QPE to GFFG Ratio values for collected flash flood cases at the Flood Advisory (1), Flash Flood Warning Base (2), Flash Flood Warning Considerable (3), and Flash Flood Warning Catastrophic (4) severity levels. Note the mean values depicted with "x" marks.

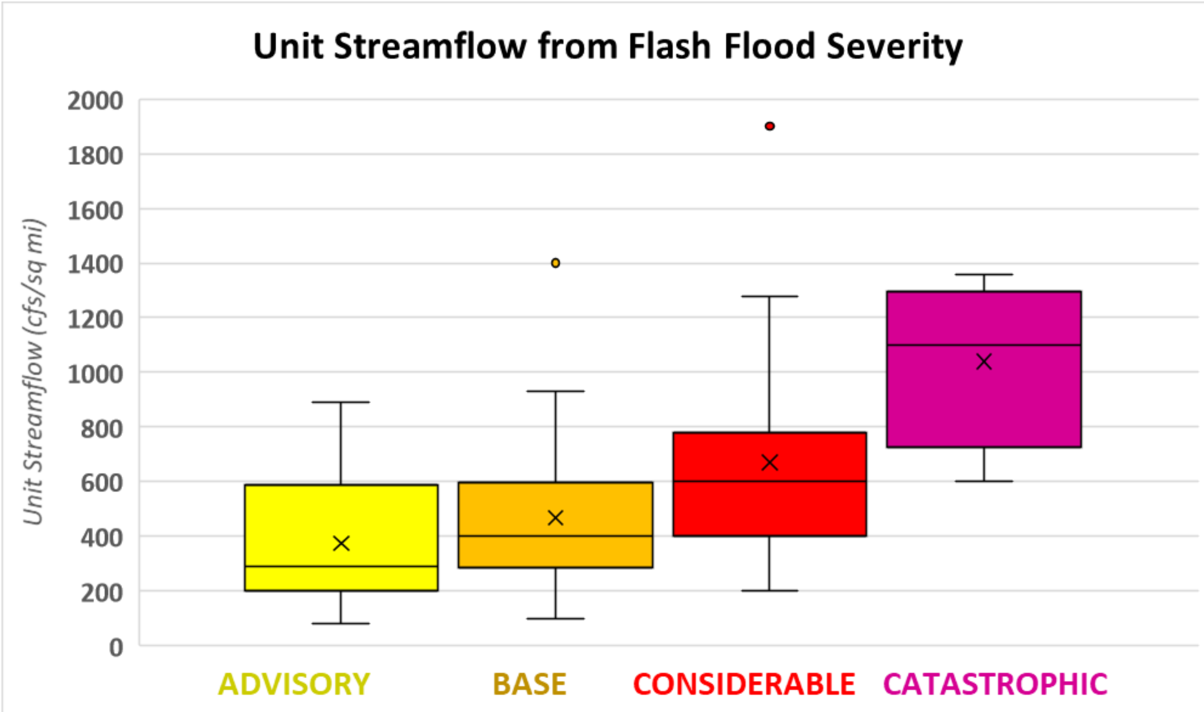

Figure 5. The range of Unit Streamflow values for collected flash flood cases at the Flood Advisory (1), Flash Flood Warning Base (2), Flash Flood Warning Considerable (3), and Flash Flood Warning Catastrophic (4) severity levels. Note the mean values depicted with "x" marks.

The data for each of the flash flood cases were then looked at in a different way; the values for each product were binned and the range of severity values was determined for each. Due to the weak correlation, 1-hr radar-only QPE was excluded. The range of resulting flash flood severities notably increased with higher max QPE ARI bins, although there were enough null events with large ARI values to make the 25th-to-75th percentile range of severity values include "no flooding" even with the highest bin (Figure 6). The range of resulting flash flood severities notably increased with higher max QPE GFFG ratio bins (Figure 7). Although a few null events with large GFFG ratio values were evident in the data, the largest two bins excluded "no flooding" from the 25th-to-75th percentile range. The range of resulting flash flood severities notably increased with higher unit streamflow bins (Figure 8). Although a few null events with large unit streamflow values were evident in the data, the largest bin excluded "no flooding" from the 25th-to-75th percentile range. It was also noted that almost no flash flood events were observed with unit streamflow values less than 125 cfs/mi^2, potentially yielding a useful minimum for operations.

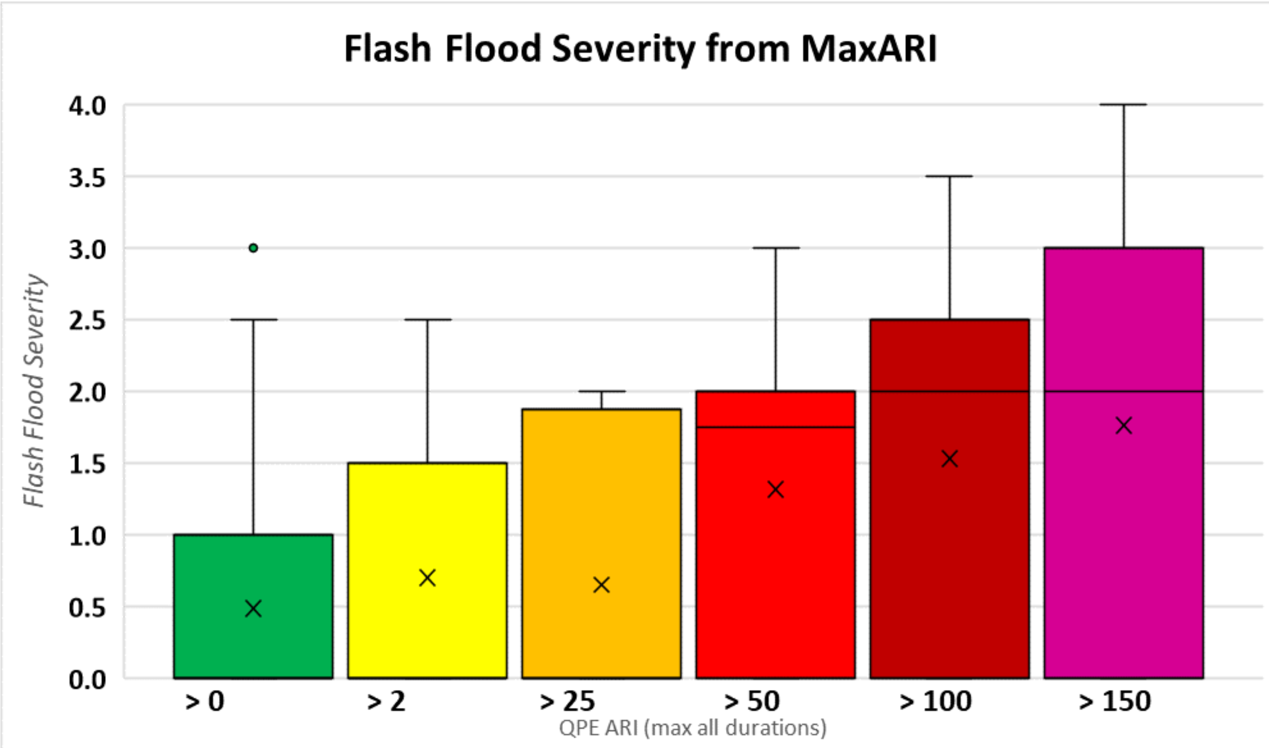

Figure 6. The range of flash flood severity values associated with six bins of Max QPE ARI values.

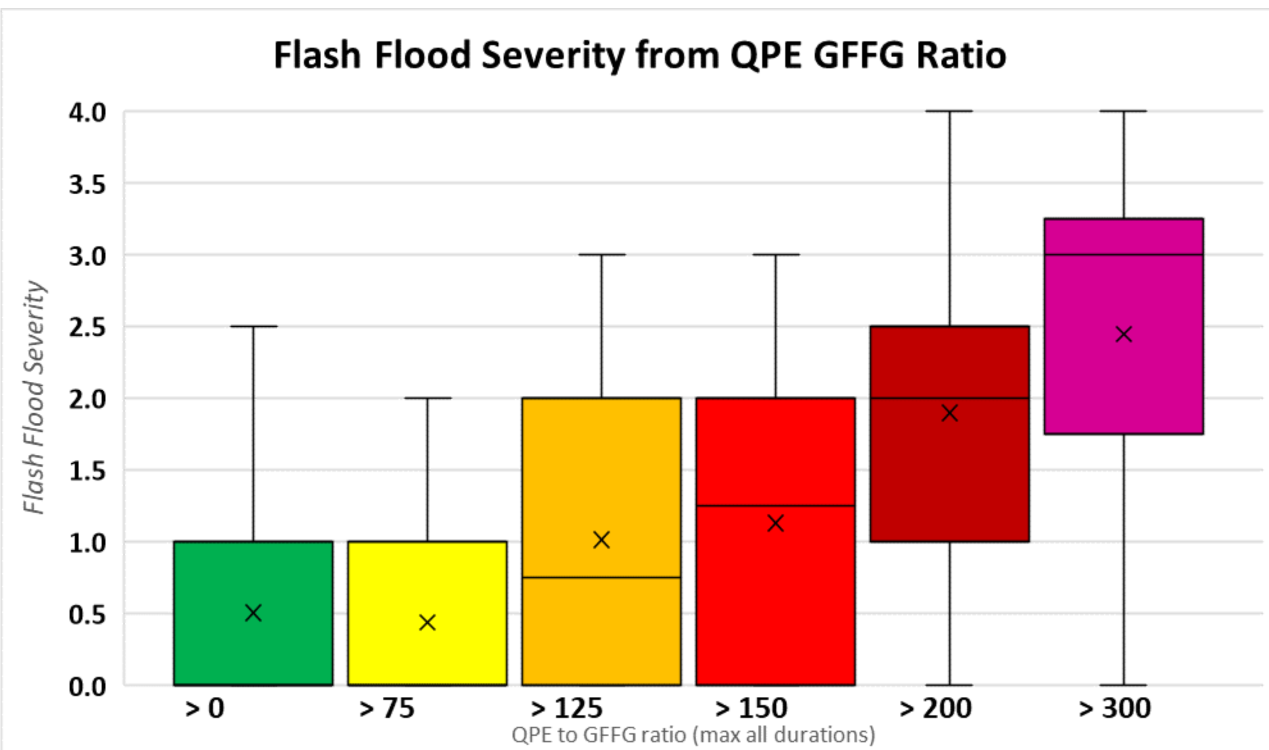

Figure 7. The range of flash flood severity values associated with six bins of Max QPE ARI values.

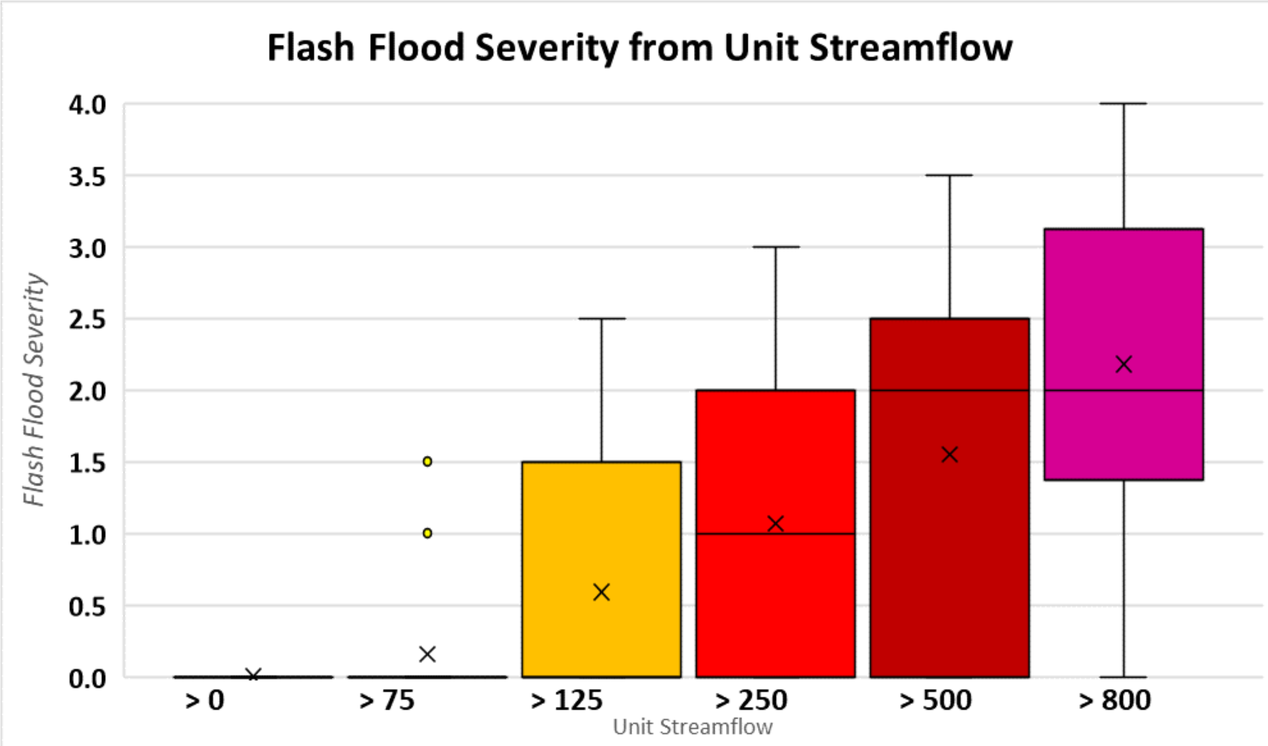

Figure 8. The range of flash flood severity values associated with six bins of Unit Streamflow values.

The higher-value bins that include a severity of 0 (no flooding) are a likely source of concern for flash flood operations. Possible explanations include lack of flash flood reports due to an event's rural location, lack of flash flooding due to very dry antecedent conditions, and uncertainty in the radar-only products that drive MRMS and FLASH. The exact contributing factor or factors is beyond the scope of this report, but should be considered for future work.

3.2 POD/FAR/CSI for MRMS/FLASH Products and the 4-Panel Technique

Tables 8, 9, 10, and 11 indicate the POD, FAR, and CSI when using the 4-Panel Technique as well as each of the MRMS/FLASH products individually. These statistics are based upon the currently used thresholds in the 4-Panel Technique provided in table 3. The 4-Panel Technique, as currently implemented, generally provides similar or better results than any individual MRMS/FLASH product. For example, using the QPE-to-GFFG product for base warning (severity level 1) decisions would provide a similar CSI to the 4-Panel Technique, while the other three products used individually would provide a lower CSI.

The alternative methods used for POD/FAR/CSI calculation yielded generally similar results, although there was a consistent dip in CSI values (due to increase in FAR) when using two panels instead of three for a forecasted event. Interestingly, there was a consistent dip in both POD and FAR values noted when ambiguous severity events (0.5, 1.5, 2.5, and 3.5) were included for neighboring severity levels, but the drop in FAR was larger and caused a slight CSI increase. This suggests that warning forecasters should be cognizant of values *nearing* a threshold when using the 4-Panel Technique, and not treat them as hard rules.

Table 8. POD, FAR, and CSI values for Advisory (1) level flooding using the 4-Panel Technique and for each of the components of the 4-Panel Technique individually.

	4-panel Technique	1-hr QPE Only	Max QPE ARI Only	Max QPE-to-GFFG Ratio Only	Unit Streamflow Only
POD	0.97	0.86	0.98	0.89	0.98
FAR	0.48	0.40	0.50	0.50	0.53
CSI	0.51	0.55	0.50	0.47	0.56

Table 9. POD, FAR, and CSI values for Flash Flood Warning Base (2) level flooding using the 4-Panel Technique and for each of the components of the 4-Panel Technique individually.

	4-panel Technique	1-hr QPE Only	Max QPE ARI Only	Max QPE-to-GFFG Ratio Only	Unit Streamflow Only
POD	0.85	0.73	0.92	0.76	0.85
FAR	0.45	0.48	0.63	0.38	0.47
CSI	0.50	0.44	0.36	0.52	0.48

Table 10. POD, FAR, and CSI values for Flash Flood Warning Considerable (3) level flooding using the 4-Panel Technique and for each of the components of the 4-Panel Technique individually.

	4-panel Technique	1-hr QPE Only	Max QPE ARI Only	Max QPE-to-GFFG Ratio Only	Unit Streamflow Only
POD	0.67	0.67	0.92	0.58	0.75
FAR	0.47	0.74	0.73	0.59	0.76
CSI	0.42	0.23	0.26	0.32	0.23

Table 11. POD, FAR, and CSI values for Flash Flood Warning Catastrophic (4) level flooding using the 4-Panel Technique and for each of the components of the 4-Panel Technique individually. Note that a very small sample size of cases was available for this severity level.

	4-panel Technique	1-hr QPE Only	Max QPE ARI Only	Max QPE-to-GFFG Ratio Only	Unit Streamflow Only
POD	1.00	1.00	1.00	0.50	1.00
FAR	0.60	0.85	0.82	0.80	0.67
CSI	0.40	0.15	0.18	0.17	0.33

3.3 Optimized Thresholds for MRMS/FLASH Products

The calibration script was run to determine the best combination of MRMS/FLASH values for use with the 4-Panel Technique, as determined by highest CSI value for severity 2 (flash flood base). Based upon the provided flash flood cases and their estimated severity levels, the best CSI would occur when using 2.1-2.2 inches for 1-hr radar-only QPE, 1-2 years for max QPE ARI, 140-145% for max QPE-to-GFFG ratio, and 280-290 cfs/mi^2 for unit streamflow (some values have ranges because 26 combinations yielded the same CSI value). This combination yielded a POD, FAR, and CSI of 0.81, 0.34, and 0.57, respectively. The highest POD and the lowest FAR were found to have similar values for most MRMS/FLASH products, although lower thresholds for 1-hr QPE and QPE-to-GFFG ratio were correlated to the best POD. For a comparison of statistics for different combinations of threshold values, see Table 12.

Table 12. Calibrated threshold values for 1-hr Radar-Only QPE, Max QPE ARI, Max QPE-to-GFFG Ratio, and Unit Streamflow that yield the best CSI, best POD, and best FAR for severity level 2 (flash flood base), based upon the provided list of cases. For comparison purposes, the POD/FAR/CSI values for each "best" are provided.

	Best CSI	**Best POD**	**Best FAR**
1-hr Radar-Only QPE	2.0-2.2	1.9	2.0-2.2
Max QPE ARI	1-2	1-2	1-2
Max QPE GFFG Ratio	140-145	115-150	140-145
Unit Streamflow	230-290	300-400	230-290
POD	0.81	0.86	0.81
FAR	0.34	0.39	0.34
CSI	0.57	0.55	0.57

The MRMS/FLASH products were also calibrated to best CSI value individually for comparison with the 4-Panel Technique (Table 13). The 4-Panel Technique was again found to have the highest POD, lowest FAR, and highest CSI when compared to individually calibrated MRMS/FLASH products. The 1-hr radar-only QPE and max QPE-to-GFFG ratio were both found to have similar threshold values when calibrated individually compared to when calibrated as part of the 4-Panel Technique. In contrast, individually calibrated threshold values for max QPE ARI and unit streamflow were both found to be notably higher than when used with the 4-Panel Technique.

Table 13. POD, FAR, and CSI values for 1-hr Radar-Only QPE, Max QPE ARI, Max QPE-to-GFFG Ratio, and Unit Streamflow calibrated individually for best CSI at severity level 2.0 (flash flood base). The statistics for 4-Panel Technique (each MRMS/FLASH product calibrated together) and threshold values associated with the best CSI are also indicated for comparison purposes.

	POD	FAR	CSI	1-hr QPE	ARI	GFFG	Unit Flow
4-Panel Technique	0.81	0.34	0.57	2.0-2.2	1-2	140-145	230-290
1-hr Radar-Only QPE only	0.64	0.39	0.46	1.90-1.95			
Max QPE ARI only	0.68	0.55	0.37		20		
Max QPE-to-GFFG Ratio only	0.78	0.39	0.52			145	
Unit Streamflow only	0.76	0.39	0.51				400

3.4 Discussion

The presented results illustrate the utility of the various MRMS and FLASH products for flash flood nowcasting, both individually and together (such as with 4-Panel Technique) but also illustrate remaining uncertainties and challenges. Relative similarity between threshold values determined based upon best POD, best FAR, and best CSI suggest reasonable confidence in operational use. The number of null cases associated with relatively large MRMS/FLASH values remains a source of concern, however.

Each product making up the 4-Panel Technique has strengths and weaknesses for flash flood monitoring. The QPE-to-GFFG ratio has its basis in some of the earliest flash flood monitoring techniques used by the NWS, beginning as simple county-level look-up tables several decades ago. A method to provide guidance values on a higher-resolution, gridded scale was developed in 2007 based upon physical land cover and terrain parameters and a set climatological rainfall amount (“design storm”). In theory, a major strength of GFFG is that seasonality and antecedent soil moisture conditions are taken into account. Despite improvements in recent decades, GFFG has been found to still have considerable uncertainty. Some of this uncertainty is tied to the same design storm (5-year ARI) being applied across the country, despite questionable assumptions (Lincoln 2017). Clark et al. (2014) found that the CSI of using GFFG was only 0.07 CONUS-wide (assuming QPE-to-GFFG ratio of 150-175%) and 0.14 for the area covered by the North Central River Forecast Center. It is unclear why CSI values reported by Clark were so much lower than those calculated for this study (0.52), but part of the discrepancy could be related to implementation; here we look at the max value for any time duration at the peak of an event, while Clark looked at specific durations to calculate CSI values. Despite some issues, GFFG has become a mature product with long-standing use among warning forecasters. Usage of rainfall ARI and unit streamflow, however, is much more recent. Although ARI should be correlated to streamflow and potential flash flood impacts as it is the basis for engineering designs of infrastructure, it does not take into account seasonality or antecedent conditions. Usage of ARI also cannot account for varying infrastructure design requirements across the country or even the same metropolitan area. WDTD recommendations and conclusions from Lincoln and Thomason (2018) were used to create the first thresholds for ARI, and it remains a useful product that puts rainfall estimates into climatological context. Unit streamflow is based upon a hydrologic model and thus seasonality and antecedent soil moisture conditions should be taken into account. Because it is providing an actual estimate of flow on the earth's surface, it is conceptually the product most closely related to flood impacts. Another strength is that unit streamflow is normalized based upon the upstream contributing area, which allows better comparison between different-sized streams. Unit streamflow does suffer from a similar limitation to that of ARI and GFFG, however, in that it cannot account for difference in capacity of stormwater infrastructure. The hypothesis behind the 4-Panel Technique is that the limitations of each product are minimized when using them together.

A review of the calibration output suggests that to get the best CSI value (severity 2, flash flood base), there is a small trade-off between POD and FAR. POD can be increased from 0.81 (best CSI) to 0.86, but this comes at the expense of FAR increasing from 0.34 to 0.39. As would be expected, a detailed look at the calibration results for individual MRMS/FLASH products suggests that POD values of 1.00 are technically possible, but they come at the expense of FARs greater than 0.60 and CSI values of 0.30 to 0.40. Because 1-hr radar-only QPE and max QPE-to-GFFG ratio were both found to have similar threshold values when calibrated individually compared to calibration as part of the 4-Panel Technique, recommended thresholds for operational use will have higher confidence. For max QPE ARI and unit streamflow, which each had notably higher thresholds when calibrated individually compared to when calibrated collectively, recommended values for operational use will have lower confidence. For max QPE ARI and unit streamflow, the best recommended thresholds will be a blend between both approaches.

Although the large number of cases presented (190) and the inclusion of null events likely yields relatively high confidence in the results of this study, some important caveats remain. The calibrated threshold values are based entirely upon cases occurring from central Wisconsin into northeast Illinois and into far northwest Indiana (NWS Chicago and Milwaukee HSAs), which means caution should be used when applying these values elsewhere. The possibility of some flood events - especially those in rural areas - going unreported may contribute to the collection of cases with no reported flooding but very high MRMS/FLASH values. Issues with the quality of flash flood reports in NWS databases have been discussed in multiple studies over many years (Gourley et al. 2013). This study also does not take into account land cover (urban vs. rural) or antecedent soil moisture conditions, which may suggest different thresholds in different situations. The exact contributing factor or factors and their relative contribution is beyond the scope of this report. It is possible, for example, that in urban environments, the unit streamflow threshold may need to be higher, and in rural environments, the 1-hr radar-only QPE, max QPE

ARI, and max QPE-to-GFFG thresholds higher. These possibilities can be explored in future work using the calibration scripts already developed for this purpose.

3.5 Recommended MRMS/FLASH Thresholds for 4-Panel Technique

Based upon all of the information collected and analyzed for this study, recommended thresholds are provided for MRMS and FLASH products used as part of the 4-Panel Technique (Table 14). The recommendation is mostly based upon the calibrated thresholds (section 3.3), but adjusted slightly to mitigate the drop in POD. These new recommended thresholds are very similar to thresholds already in use at NWS Chicago. For the flash flood warning base severity level, slightly reduced POD from current implementation (0.85 to 0.83) is exchanged for a slight drop in FAR (0.45 to 0.42) and improved CSI (0.50 to 0.52). This differs from the best-calibrated CSI where POD would drop to 0.81 in exchange for a FAR drop to 0.34. Although using all four MRMS and FLASH products together as part of the 4-Panel Technique yields the best POD, FAR, and CSI values, they can be used individually with varying degrees of usefulness. Table 15 provides recommended values for MRMS and FLASH products when used individually. Especially for the flash flood catastrophic severity level, warning forecasters should be cautioned to not use radar-derived products alone when making warning decisions.

Table 14. Recommended threshold values for 1-hr Radar-Only QPE, Max QPE ARI, Max QPE-to-GFFG Ratio, and Unit Streamflow when used as part of the 4-Panel Technique, based upon calibration and minor adjustments to improve POD. Note that due to the very small sample size for "flash flood catastrophic" (severity level 4) cases, borderline events (severity level 3.5) were also included; values should still be used with caution.

	Advisory	Flash Flood Base	Flash Flood Considerable	Flash Flood Catastrophic
1-hr Radar-Only QPE	1.5	2.0	2.5	2.7
Max QPE ARI	1	5	125	175
Max QPE GFFG Ratio	125	140	325	375
Unit Streamflow	200	230	850	1100
POD	0.91	0.83	0.58	1.0
FAR	0.35	0.42	0.30	0.60
CSI	0.61	0.52	0.47	0.40
Number of Cases Exceeding Severity	87 / 190	59 / 190	12 / 190	4 / 190

Table 15. Recommended threshold values for 1-hr Radar-Only QPE, Max QPE ARI, Max QPE-to-GFFG Ratio, and Unit Streamflow when used individually, based upon calibration and minor adjustments to improve POD.

	Advisory	**Flash Flood Base**	**Flash Flood Considerable**	**Flash Flood Catastrophic**
1-hr Radar-Only QPE	1.5	2.0	2.5	2.5
Max QPE ARI	2	8	150	200
Max QPE GFFG Ratio	125	145	350	375
Unit Streamflow	190	400	1100	1350

4.0 Conclusions and Future Work

The MRMS and FLASH products 1-hr radar-only QPE, max QPE ARI, max QPE-to-GFFG ratio, and unit streamflow have been used for at least the last few years to provide useful insights into possible flash flooding. NWS Chicago currently uses all of these products together as part of a 4-Panel Technique, where three out of four products indicating the same severity level is typically used as the basis for a flood hazard, assuming no biases in MRMS QPE. Because of limited guidance on the best thresholds to use for this method, NWS Chicago and NWS Milwaukee collected information on 190 flash flood cases - of which 100 were null cases - and assigned a flash flood severity value to each based upon documented impacts. POD, FAR, and CSI values were then calculated for the current implementation of the 4-Panel Technique, and thresholds for MRMS and FLASH products were calibrated. Calibrated thresholds provide improved values for FAR and CSI, and these thresholds were used as the basis for new recommended values. These thresholds are likely to be most useful in areas with similar terrain, land cover, and infrastructure standards to the area from where the cases were selected, roughly central Wisconsin southeast to northern Illinois and northwest Indiana. Warning forecasters should be cautioned to not use radar-derived products alone when making warning decisions, especially for higher-end floods. Although not explicitly discussed in this study, the spatial footprint of MRMS/FLASH values exceeding given thresholds, trends in MRMS/FLASH values, meteorological considerations, and observed reports of flash flooding could all be used as “nudgers” for issuing higher-severity products.

Planned future work using MRMS/FLASH products for flash flood nowcasting includes breaking up collected cases by land use type (rural, suburban, urban), soil moisture conditions (dry, moist, near average), and additional QPE durations (3 hour, 6 hour). Future work may also include breaking up collected cases by WFO to check for applicability of results from one area to another nearby area.

The authors would like to thank NWS Chicago Science and Operations Officer Kevin Donofrio for his helpful guidance and comments, NWS Milwaukee forecaster Aidan Kuroski, who assisted with the collection of cases, and Jeff Manion for his helpful review.

5.0 References

Clark, R.A., J.J. Gourley, Z.L. Flamig, Y. Hong, and E. Clark, 2014: CONUS-Wide Evaluation of National Weather Service Flash Flood Guidance Products. *Wea. Forecasting*, 29, 377-392.

Gerard, A., S.M. Martinaitis, J.J. Gourley, K.W. Howard, and J. Zhang, 2021: An Overview of the Performance and Operational Applications of the MRMS and FLASH Systems in Recent Significant Urban Flash Flood Events. *Bull. Amer. Meteor. Soc.*, 102, E2165–E2176.

Gourley, J.J., Y. Hong, Z.L. Flamig, A. Arthur, R. Clark, M. Caliano, I. Ruin, T. Ortel, M.E. Wieczorek, P.-E. Kirstetter, 2013: A Unified Flash Flood Database over the US. *Bull. Amer. Meteor. Soc*., 94, 799-805.

Lincoln, W. S., and R. F. L. Thomason, 2018: A preliminary look at using rainfall average recurrence interval to characterize flash flood events for real-time warning forecasting. *J. Oper. Meteor.*, 6, 13–22.

Seo, D., T. Lakhankar, J. Mejia, B. Cosgrove, R. Khanbilvardi, 2013: Evaluation of Operational National Weather Service Gridded Flash Flood Guidance over the Arkansas Red River Basin. *J. Am. Water Resour. Assoc.*, 49, 1296-1307.

Vergara, H.J., J.J. Gourley, and A. Arthur, 2019: Lessons learned in the application of the MRMS-FLASH system at scales ranging from the river reach to the Conterminous United States for Impact-based Flood Prediction. American Geophysical Union, Fall Meeting 2019, abstract #H43I-2108.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, 97, 621–637.

Appendix A

Table A1.All cases (both flash flood events and null events) that were used in this study, along with their assigned severity values and peak MRMS/FLASH values. See main text for description of severity values.

Date	Location (County/Region)	Severity	Max 1-hr QPE	ARI	QPE GFFG Ratio	Unit Streamflow
7/23/2016	Winnebago, Boone, and McHenry counties	2.5	2.5	150	150	400
7/23/2016	Lake and Cook counties	1.5	3	5	125	400
7/29/2016	DeKalb, Kane, DuPage, Will, and Cook counties	2	2.5	75	150	500
8/15/2016	Livingston, Kankakee, Lake (IN), and Porter counties	1.5	2.5	2	75	100
8/15/2016	Kankakee, Lake (IN), and Porter counties	1.5	3.5	10	125	200
8/28/2016	Grundy and Livingston counties	2	2	5	125	200
8/29/2016	Winnebago, Boone, McHenry, DeKalb, Cook, and DuPage counties	2	3	5	125	200
8/29/2016	Lee and LaSalle counties	1.5	2	2	75	200
6/28/2017	Winnebago and Boone counties	2.5	2.5	150	250	400
6/28/2017	McHenry, Lake, Kane, and Cook counties	2	2	10	150	400
7/12/2017	Lake County (IL)	2	2.5	150	250	900
7/12/2017	Kane, Cook, and DuPage counties	1	3	150	150	600
7/12/2017	Lake County (IL)	2	1.5	5	100	300
7/22/2017	Ogle County	1	2	5	125	200
7/23/2017	Cook, DuPage, and Will counties	1	1.5	1.5	75	200
10/14/2017	LaSalle, DeKalb, Kane, Kendall, DuPage, and Cook counties	2	1.5	5	75	200
5/30/2018	Cook and DuPage counties	2.5	2	75	150	600
6/15/2018	Winnebago and Ogle counties	2	2.5	75	150	400
6/18/2018	Winnebago County	3.5	2.5	150	250	600
6/26/2018	Winnebago and Boone counties	1	1	1.5	75	200
6/26/2018	McHenry, Lake (IL), and Kane counties	1.5	1.5	5	125	250
8/7/2018	Ogle and Lee counties	1	1	5	75	150
8/7/2018	Cook and DuPage counties	2.5	1.5	2	75	400
8/17/2018	Rock County	1	2.5	90	150	330
8/17/2018	Dodge and Jefferson counties	2	2.1	200	250	550
8/17/2018	Green and Rock counties	0	1.9	140	140	210

Date	Location (County/Region)	Severity	Max 1-hr QPE	ARI	QPE GFFG Ratio	Unit Streamflow
8/20/2018	Dane County	4	2.8	200	370	1360
8/20/2018	Milwaukee County	0	1.1	37	200	710
8/27/2018	Milwaukee County	1	1.3	32	220	850
8/27/2018	Ozaukee County	2	2.2	200	200	840
8/27/2018	Sauk County	2.5	1.8	200	250	830
8/28/2018	Central Wisconsin	2	2.2	200	320	420
9/3/2018	Sauk and Columbia counties	2	1.8	27	160	930
10/3/2018	South-central Wisconsin	0	1	1	60	200
2/7/2019	Morris and Grundy counties	0	0.2	160	190	320
2/24/2019	DuPage and Cook counties	0	0.3	1	56	260
3/9/2019	Ogle and Lee counties	0	0.4	1	50	30
3/14/2019	DuPage and Cook counties	0	1.1	1	90	490
3/20/2019	Cook County	0	0.3	1	30	70
3/26/2019	Lee and Livingston counties	0	2.3	25	120	80
3/27/2019	Livingston County	0	0.8	1	50	10
4/7/2019	Cook County	0	1.2	2	80	210
4/8/2019	Brown County	0	0.5	1	40	10
4/10/2019	Dane and Green counties	0	0.9	1	50	60
4/14/2019	Lake County (IN)	0	0.5	1	100	140
4/17/2019	Dane and Columbia counties	0	0.7	1	60	20
4/18/2019	Lake County (IN)	0	0.4	1	110	110
4/22/2019	Lee and DeKalb counties	0	1.1	7	105	30
4/27/2019	Will County	0	1.3	2	80	130
4/29/2019	Kendall, Kane, DuPage, and Will counties	0	0.4	2	85	40
4/29/2019	Cook County	0	0.4	1	70	210
4/30/2019	Livingston and Kendall counties	2	1	3	110	210
5/1/2019	Will, Kankakee, and Lake (IN) counties	2	1.2	10	150	290
5/9/2019	DuPage County	0	0.3	1	70	120
5/16/2019	Cook County	0	1.3	8	150	240
5/17/2019	Green and Dane counties	0	1.7	13	130	150
5/19/2019	DuPage County	0	0.9	1	70	210
5/22/2019	Cook County	0	0.4	0	70	110

Date	Location (County/Region)	Severity	Max 1-hr QPE	ARI	QPE GFFG Ratio	Unit Streamflow
5/23/2019	Livingston, Ford, and Iroquois counties	0	1.7	15	150	120
5/23/2019	Milwaukee County	0	0.7	2	95	330
5/24/2019	Green and Rock counties	0	0.8	4	75	180
5/25/2019	Dane, Columbia, and Dodge counties	0	1.1	14	150	380
5/27/2019	Kane and Kendall counties	0.5	1.5	59	141	470
5/27/2019	DuPage and Cook counties	1.5	1.2	50	130	770
5/28/2019	Cook, Will, and Lake (IL) counties	0	1.1	9	210	225
5/30/2019	Lee County	0.5	1	4	117	140
5/31/2019	LaSalle County	0	1.8	79	192	210
6/1/2019	Rock and Walworth counties	0	1.7	137	180	290
6/1/2019	Cook County	0	1.4	12	125	380
6/4/2019	Winnebago County	0	0.9	1	105	195
6/4/2019	Cook County	1	0.5	0	36	80
6/4/2019	Iowa and Dane counties	0	1.3	10	103	100
6/5/2019	LaSalle County	0	0.6	10	123	70
6/5/2019	Ogle County	1.5	1.6	13	141	190
6/12/2019	Walworth County	0	1.3	40	123	150
6/12/2019	Will and Cook counties	0	0.7	110	136	970
6/15/2019	LaSalle County	0	0.6	10	105	60
6/16/2019	Livingston and Kankakee counties	0	1.4	47	137	650
6/19/2019	Livingston County	0	1	34	162	140
6/20/2019	Lake County (IN)	1	1.5	50	125	200
6/24/2019	Rock and Dane counties	0	1.2	20	95	115
6/25/2019	Rock County	0	0.8	3	60	57
6/26/2019	Kane County	0	1.4	3	71	97
6/26/2019	DeKalb and Lee counties	0	1.8	8	100	20
6/27/2019	Cook, Will, and Lake (IN) counties	3	2	100	150	350
6/28/2019	Rock and Dodge counties	0	1.7	72	143	310
6/29/2019	Iroquois County	0	1.2	4	85	20
6/30/2019	Walworth County	0	1.8	25	116	180
6/30/2019	Kendall, Grundy, and Will counties	1	1.7	140	161	590
6/30/2019	Iroquois County	0	2.1	107	204	170
7/1/2019	Lee and Ogle counties	0	1.7	11	125	100

Date	Location (County/Region)	Severity	Max 1-hr QPE	ARI	QPE GFFG Ratio	Unit Streamflow
7/2/2019	Winnebago County	1	1.6	15	195	380
7/2/2019	Ogle County	0	1.6	35	126	110
7/2/2019	LaSalle County	0	1.7	154	195	710
7/3/2019	Sauk County	2	2.1	80	170	510
7/4/2019	Grundy County	0	1.3	3	90	70
7/5/2019	Rock County	0	2.1	59	155	430
7/5/2019	Stephenson and Winnebago counties	0	2.4	8	97	30
7/5/2019	Lee and Livingston counties	0	1.8	18	115	140
7/6/2019	Rock and Green counties	0	1.1	36	105	110
7/6/2019	Will County	0	1.5	18	128	210
7/14/2019	Jasper County	0	2.1	157	146	130
7/14/2019	Iroquois County	0	2.1	49	110	70
7/17/2019	Cook County	1	0.9	3	48	180
7/18/2019	Dane and Rock counties	0	1.7	13	96	360
7/18/2019	Livingston, Kendall, and DeKalb counties	0	1.9	37	130	140
7/18/2019	DuPage, Cook, Lake (IN) counties	0	1.1	15	113	440
7/19/2019	Columbia and Dane counties	0	1.6	64	144	440
7/20/2019	Milwaukee County	0	1.2	4	110	400
7/21/2019	Cook and Lake (IL) counties	0	1.4	7	170	500
7/21/2019	Will and Kankakee counties	0	2.2	171	151	550
7/21/2019	Lake (IN), Newton, and Jasper counties	0	1.4	185	126	95
7/29/2019	DuPage and Cook counties	0	0.5	19	98	430
8/3/2019	Milwaukee and /Waukesha counties	0	1.3	36	144	260
8/3/2019	Dane County	0	2	198	169	270
8/5/2019	Columbia County	0	1.2	8	74	65
8/6/2019	Winnebago County	0	1.4	4	73	240
8/11/2019	Dane, Green, and Rock counties	0	1.8	38	136	320
8/11/2019	Ogle, Lee, and DeKalb counties	0	2.4	68	90	170
8/12/2019	Cook County	0	0.8	7	95	240
8/13/2019	Dane County	0	0.9	9	94	270
8/14/2019	Waukesha County	0	1	79	133	280
8/16/2019	Benton County	0	3	172	134	250

Date	Location (County/Region)	Severity	Max 1-hr QPE	ARI	QPE GFFG Ratio	Unit Streamflow
8/17/2019	Cook County	0	0.8	75	111	770
8/18/2019	Cook and DuPage counties	1	0.9	4	99	450
8/21/2019	Ogle County	0	1.9	15	113	210
8/26/2019	Cook, Lake (IL), and Kenosha counties	0	0.5	4	109	220
8/26/2019	Kendall and Grundy counties	0	0.9	34	111	110
9/1/2019	Will and Cook counties	0	0.9	7	85	390
9/3/2019	Stephenson, Winnebago, and Ogle counties	1	2.1	61	130	660
9/10/2019	Green, Rock, Stephenson, and Winnebago counties	2	1.5	51	145	430
9/10/2019	Lake and Kenosha counties	0	1.9	26	203	350
9/11/2019	Rock, Walworth, Boone, and McHenry counties	0	1.9	141	180	620
9/11/2019	Kenosha County	1.5	1.8	80	210	790
9/12/2019	Lafayette County	2	2	41	160	640
9/12/2019	Iowa and Sauk counties	2	1.8	10	130	490
9/12/2019	Rock County	2	1.5	41	150	590
9/12/2019	Milwaukee County	2.5	1.6	133	230	760
9/12/2019	Walworth County	2	1.3	66	190	590
9/13/2019	Kane, DuPage, Cook, and Lake (IL) counties	2	2	10	250	300
9/13/2019	Winnebago, Boone, and McHenry counties	2.5	2.5	100	200	400
9/15/2019	Lee County	0	2	16	108	480
9/19/2019	Lafayette County	0.5	2.1	126	211	620
9/21/2019	Ogle and DeKalb counties	0	1.1	10	121	350
9/21/2019	LaSalle County	0	0.9	49	131	780
9/22/2019	Dodge County	0	0.7	9	113	180
9/27/2019	Kendall, Grundy, Will, and Cook counties	3	2.5	150	450	600
9/27/2019	Lake County (IN)	2	1.5	150	200	500
9/27/2019	Woodford and Livingston counties	2.5	3.4	185	243	800
9/29/2019	Ford and Iroquois counties	0	0.8	11	121	290
10/1/2019	Columbia County	2	1.1	13	160	370
10/1/2019	Ozaukee County	2	1.3	47	220	500
10/1/2019	Dodge County	2	2.1	30	310	650

Date	Location (County/Region)	Severity	Max 1-hr QPE	ARI	QPE GFFG Ratio	Unit Streamflow
10/1/2019	Dane County	2	1	5	140	470
10/2/2019	Dodge County	0	1.9	51	303	700
10/3/2019	DuPage and Cook counties	2	1.5	2	100	300
10/5/2019	Rock County	0	0.4	1	75	180
10/21/2019	DeKalb County	0	0.5	3	81	90
10/26/2019	Kendall, Grundy, Will, and DuPage counties	2	0.5	1	50	150
11/21/2019	Will and Cook counties	0	0.4	1	81	360
11/27/2019	Dane County	0	1	2	104	330
3/19/2020	Lafayette, Iowa, and Dane counties	0	0.6	0	65	70
3/28/2020	Lee County	0	1.9	40	139	230
3/28/2020	Ogle County	0	2	14	135	250
3/28/2020	Winnebago, Rock, Walworth counties	1.5	1.3	9	159	390
4/28/2020	Kenosha County	1.5	1.9	5	110	250
5/15/2020	Kane, DuPage, and Cook counties	3	2	100	150	550
5/15/2020	Livingston and Kankakee counties	2	2.5	75	150	400
5/17/2020	Kendall, Grundy, and Will counties	2.5	2.5	5	150	200
5/17/2020	DuPage and Cook counties	3	1	2	150	400
6/10/2020	Columbia, Dodge, and Fond du Lac coutnies	1.5	1.9	13	130	290
6/27/2020	Newton, Jasper, and Benton counties	3	2.5	150	350	600
6/29/2020	Rock County	2	2.5	68	190	280
6/29/2020	Dane County	2	2.7	10	160	450
7/7/2020	Dane County	1.5	2.1	13	90	230
7/7/2020	Waukesha County	1.5	2	13	120	370
7/9/2020	Racine County	0	1.6	40	180	920
8/2/2020	Milwaukee County	3	2.5	200	390	1900
8/9/2020	Sheboygan County	1.5	3.6	200	310	890
8/10/2020	Racine County	2	2.5	200	280	1400
8/27/2020	Columbia County	2.5	3.1	120	240	700
7/12/2021	LaSalle and Grundy counties	3.5	3.5	200	400	1100
8/12/2021	Ford, McLean, Champaign counties	4	3	200	250	1100
8/24/2021	Cook County	2	2.1	15	110	620
8/24/2021	McHenry County	2	2.85	160	190	480

Date	Location (County/Region)	Severity	Max 1-hr QPE	ARI	QPE GFFG Ratio	Unit Streamflow
7/5/2022	McHenry County	2	2.5	11	115	230
7/23/2022	Lake County (IL)	3	5.8	200	215	1280
8/7/2022	Cook County	2	2.2	31	130	560
8/8/2022	Ogle and Winnebago counties	2.5	1.8	80	200	480
9/11/2022	DuPage and Cook counties	3	1.9	154	160	720