# **Cross-evaluation of Uncertainties in Extreme Precipitation Events using Multiplicative Triple Collocation**

## RESULTS

### ***Overall behaviors***

The accumulative rainfall for all events, Harvey and non-Harvey in **Figure 2** shows that MRMS product always captures the most rainfall amount especially with larger portion inside Texas. The second largest is measured by IMERG satellite data and the last is NCEP gauge data. Harvey causes more intensive rainfall near the boundary of Texas and Louisiana state, while other events are scattered like a squall line. We can observe the lumped patches-like pattern for NCEP data, meaning the quality of gauge during extreme events is varying and will be investigated further in the Harvey events. But for MRMS and IMERG, they are stratified well which show the spatial continuity of rain field.

The first hand overview of applying TC to four events concatenated together, denoted as “All”, Harvey-only and non-Harvey that denoted as “other” are illustrated in **Figure 3** for CC**, Figure 4** for RMSE. Firstly, both CC and RMSE demonstrate consistent behaviour in which higher CC and lower RMSE. All three cases identified the same result that the ordered product uncertainties (from high to low) are MRMS, IMERG, NCEP according to the median value and metrics distribution at the whole domain. The median value is more insightful than mean value because there are noises that tend to shift up mean values especially in places where rainfall samples are limited. NCEP gauge data tends to have higher CC value in the western regions in which smaller amount of rainfall observed compared to core regions. MRMS is behaving more uniform that higher CC is wide spread over the domain. IMERG concentrates its higher value in the western region as well as NCEP, but cover more regions than NCEP. The violin plot inside each axes indicates the distribution of CC. They are more skewed to higher values (above 0.5) but still have some samples that close to zero, especially for NCEP data. These values normally appear outside of the core where we set non-rain samples to be . When the covariance of the estimated product e.g. NCEP with either other two products e.g. MRMS and NCEP becomes 0, it will impose estimated CC to be zero, meaning no correlation with the other two products. These low values will be eliminated when we only consider core regions. We would like to point out that, in event Harvey, CCs of three product is more divergent compared to either All or Other. This indicates the three products agree more with other for non-Harvey case, and also push us to investigate Harvey with special attention.

**Figure 4** depicted the RMSE value from TC results which reveals the other side of CC value. In **Figure 4**, we emphasized two remarkable consistent behaviours. One is for NCEP data that All and Harvey both experience higher RMSE value in Houston region. This needs our further scrutinization by inspecting its time series in the detailed Harvey analysis. The other for MRMS and IMERG shows anomalous signals at the boundary of Louisiana and Mississippi. The corresponding time series plot of rainfall at one location is selected and shown up in the small axes. The overall trend is well followed by three products except that MRMS radar has some anomalous impulse that disagrees with the other two data sources. This anomalous echo may not be coming from hydrometeors. It is probably caused by debris rolled up by strong wind that was recorded of 108 mph.

### ***Detailed Harvey Analysis***

**Table 3** listed both quantitative intensity difference and categorical difference in Harvey with three different percentiles. NCEP vs. MRMS, NCEP vs. IMERG, IMERG vs. MRMS are taken into account and the latter product is considered as “reference” because of lower uncertainty in the previous TC results. What this result conveys is with analogy to the TC results. IMERG vs. MRMS comparison has highest correlation coefficient for three percentiles, even though they are slightly lower than NCEP vs. IMERG in terms of RMS difference in lower percentiles. For categorical difference, all results suggest IMERG and MRMS behave relatively close while NCEP deviates from both MRMS and IMERG. It again validates the results that TC is capable to convey some consistent results with traditional methods.

The CC and RMSE from TC are further split based on accumulative rainfall by MRMS data at each interval of 50 mm in **Figure 5.** In doing so, we could identify the range that is suitable for each product to be having less uncertainty. As an overview, MRMS data is substantially robust and stable as rainfall increases, but NCEP gauge data generally acquire more uncertainties than IMERG and MRMS at all ranges. Besides, NCEP obviously aggravate its behavior at higher range of the rainfall i.e. above 1200 mm. IMERG data is behaving worse at lower tail i.e. below 150 mm. That indicates satellite data under-performs at lower tail probably due to the sensor sensitivity, signal attenuation and smoothing effect of large size of the footprint. It is been reported in literatures that IMERG data typically tends to overestimate light rain and underestimate heavy rainfall (Guo et al., 2016; O et al., 2017; Sharifi, Steinacker, & Saghafian, 2016). (Omranian et al., 2018) also concluded that IMERG final product has generally better performance with higher precipitation rates compared to lower rates in the case of hurricane Harvey.

**Figure 6** depicts TC metrics CC and RMSE for three grouped regions – whole, core and non-core. After thresholding, we clearly witnessed the improved performance for both MRMS and IMERG – higher CC and lower RMSE – in terms of median value and uncertainty bound. However, the RMSE for NCEP even remained unchanged. This points out the noise removal for NCEP data inside the core is superseded by the degradation of performance probably due to the impact of intensive wind or out-of-splash. While for the performance of each product, the results are still remained, which again validate the consistency of TC method. The whole region performance sits in between core and non-core because it neutralize the two tails. Looking at the distribution of rainfall for each product in **Figure 7**, It associates with the characteristics of the product itself. For NCEP gauges, It is likely to underestimate total rainfall because of splash-out-of-water, wind under-catching and not representative of rainfall variability. It is thus obvious for NCEP data to be concentrated in the range of 400 to 600 mm. Even though IMERG data is more wide spread than NCEP, it still cuts off at 1100 mm. The reasons behind are myriad e.g. the sensitivity of sensors, type of sensors (IR, PMW, DPR) but most importantly, since IMERG has resolution of 0.1 degree (around 10 km), it acts like a smoother that take the average of the cell. Hence, it is difficult to capture fine scale rainfall field. To be noted, MRMS data could also suffer from overestimation since we observed some anomalous impulses during this event, but we tend to believe that MRMS radar QPE is more or less close to “reference” as TC results suggested.

By inspecting the time series at selected representative points in **Figure 8**, we are able to unravel the reason that NCEP data behaved relatively worse. Point 1,3, 4 and 5 are picked based on the RMSE spatial map in which NCEP data had highest value or maximum rainfall are captured. In the corresponding time series plots, the grey window mentioned that in that certain period, NCEP data either recorded as zero value but the other two captured intensive rainfall (point 1,3) or stopped recording any data (point 4,5). This anomaly could be caused by wind that blows rain drops out of the range, or mechanical misfunctioning but more likely due to uncertainties introduced by interpolation. Because of the variability of rainfall field, sparse gauge network (e.g. 3000 in total around the U.S.) is not able to capture this variability. Point 2 is selected as IMERG data recorded maximum amount of rainfall in this event. The horizontal blue line marked the maximum rain rate that GMI (GPM Microwave Imager) can record due to sensor sensitivity (Skofronick-Jackson et al., 2017). In other words, IMERG product will cut off any rain rate larger than 60 mm/h except places swept by DPR (dual-frequency precipitation radar). Nevertheless, DPR has limited swath and coverage. Point 4 is where MRMS has relatively low CC value inside the core. We indeed found that MRMS sometimes jumped up instantaneously while NCEP and IMERG showed some agreements such as highlighted red window. It probably attributes to non-weather echoes or applying incorrect Z-R relations.

**Figure 9** has depicted the conditional inter-comparison results for each pair. As expected, those metrics are getting worse when larger percentile of data are exceeded because larger rain rate is associated with more difference. Except for RMS difference, all other statistics suggest IMERG and MRMS are more or less similar in whatever kinds of extreme rainfall conditions which is again the same signature in **Table 3**. It once again proved that NCEP data inside Harvey core has certain degree of degradation. Researchers need to be cautious when deciding to use this data set to evaluate extreme weather conditions, and additional justification should be provided.

## CONCLUSIONS AND FUTURE WORKS

In this experiment, we tried to explore the applicability of TC method in extreme events, and interpret the results from it. It is an ill-posed statement because we cannot find valuable reference data to absolutely validate the performance of TC method even though a great amount of previous works has proven that TC is a powerful statistical tool to analyse the uncertainties of three independent precipitation products. However, instead of directly compare TC with traditional methods for given reference data, we could evaluate the consistency of TC in the following ways:

1. Applying TC in the overall extreme events combined, and separately perform it with individuals. The results in **(Figure 3 and 4)** have suggested that MRMS data are always providing lowest uncertainties, and then follows IMERG satellite QPE. NCEP data are associated with largest amount of uncertainties, which may not be an appropriate product in evaluating rainfall at extremes.
2. The traditional evaluation metrics e.g. RMS difference, Correlation Coefficient, POD, FAR, CSI and also their conditioned values are well in line with the TC results. They indicate that NCEP data have larger difference with MRMS and IMERG.
3. The results to separate Harvey core and non-core **(Figure 5)** also demonstrated NCEP data performance is not substantially improved inside the core while the other two get different amount of increment. This points out that NCEP data may be subject to degradation in the Harvey core.
4. The time series for collocated pixels where gauge have more uncertainties inside the core showed that a large portion of NCEP data didn’t record rainfall while the other two products indeed observed some amount. We believe that gauge sparsity plays the first order role, and then it may also subject to other effects e.g. wind under-catching, splash-out-of-water, mal-functioning.
5. IMERG data has the second largest uncertainty because: 1. Its large resolution smooths out rainfall variability so that it does not capture enough rainfall **(Figure 8)**; 2. The sensitivity for IMERG data is not trivial in extreme conditions **(Figure 7)** as it cuts off rain rate beyond certain threshold; 3. Signal attenuation may also lower the rain rate captured by satellite QPEs; 4. Gauge may also smooth out extreme values for gauge adjusted satellite QPEs.

The TC results also provide us insightful considerations. Firstly, it can assess the quality of each product. In our extreme rainfall cases, we found NCEP gauge based QPE encountered some problems that may be due to the uncertainty introduced by interpolation, or systematic error e.g. wind under-catching, splash-out-of-water etc. IMERG satellite QPE sits in between NCEP and MRMS. MRMS radar QPE is proved to be the most reliable dataset in capturing extreme rainfall. Secondly, it tells the story about limiting boundaries of each product. NCEP gauge data increase its uncertainties as accumulative rainfall amount increases. One can infer that gauge-like QPE are more susceptible to higher rainfall rate due to its own deficiencies mentioned above. IMERG data have higher uncertainties at lower tail meaning light rainfall could cumber its performance. MRMS radar data are behaving more robust and stable at either high rain amount or low rain amount. It suggests that MRMS radar QPE is more appropriate to evaluate precipitation during extreme events.

This paper serves as the first order overview of the quality of each product during extreme events as each suffers from its own deficiencies. However, it is still unclear which systematic error plays the first role for radar and satellite QPE. It motivates us to explore more detailed error decomposition in the future researches. Beyond that, we believe that the gauge corrected products i.e. radar and satellite will not get improved performance in extreme events. Further comparisons regarding to this topic could also help developers to adopt more reasonable algorithms for quantitative precipitation estimation.

## ACKNOWLEDGEMENTS