

We thank the Reviewer for the comments. Our detailed replies are in blue below.

RC2: 'Comment on essd-2024-513', Anonymous Referee #2, 27 Feb 2025

The authors developed a framework based on their previous efforts on inverse rainfall-runoff modeling, and bias correction methods. Based on the framework, they developed a bias-corrected rainfall product that consists of 200+ storm events spreading across 20 basins in the Appalachian Mountains. The new rainfall product maintains a better basin-scale water budget and shows promising results for flash flood modeling in this region. As a manuscript submitted to ESSD, the uniqueness, usefulness, and completeness of the dataset should be the most important criteria. While I have no concerns about the uniqueness or completeness, the usefulness of the bias-corrected dataset is questionable. In addition, the dataset covers only a very limited region, only head watersheds in the Appalachians. Whether the proposed framework to other watersheds or storm events is another concern of mine. This is because there are many subjective choices in the framework. Justification of these routines is needed.

The first figure below is from Saharia et al. (2017) and it shows the alignment of flash flood climatology with the Appalachian Mountains east of 105° in the US. A recent example is Hurricane Helene in September 2024 which caused over 200 deaths and \$50 billion in property damage in the Southeast U.S. The second figure is a distribution of global flood hazards from the World Bank and Columbia University that illustrates the alignment of the most severe flooding with topographic features globally. Approximately 2 billion people live in such regions around the world, and apart from massive earthquakes, flashfloods are the deadliest natural hazard due to the combination of severity and frequency.

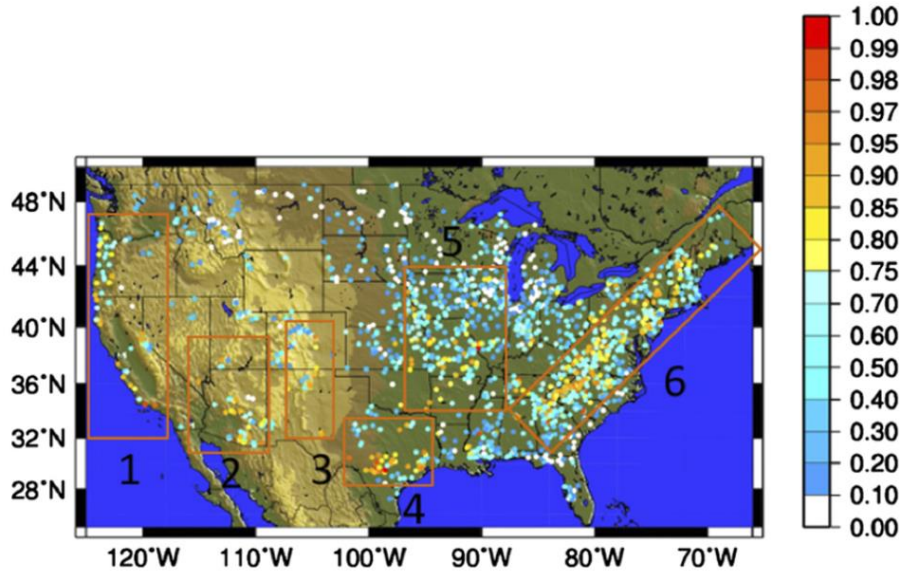
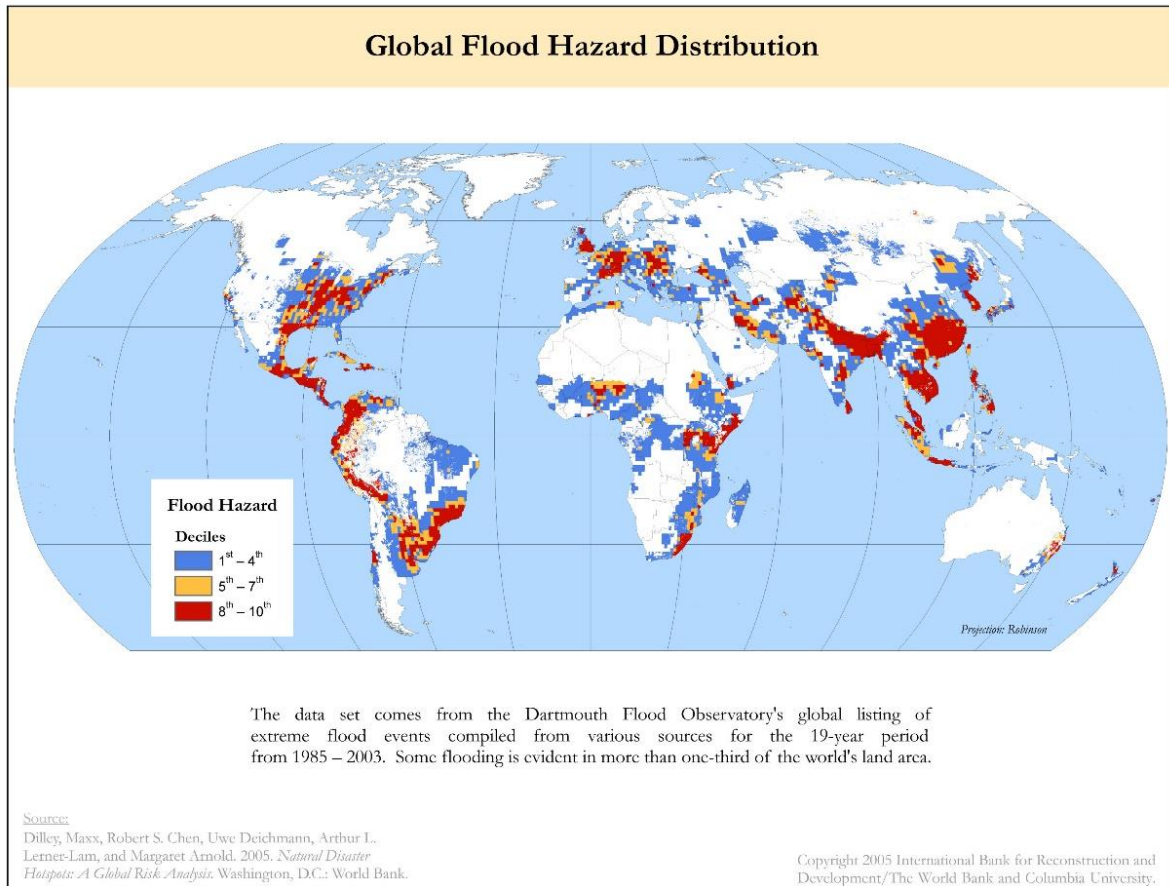
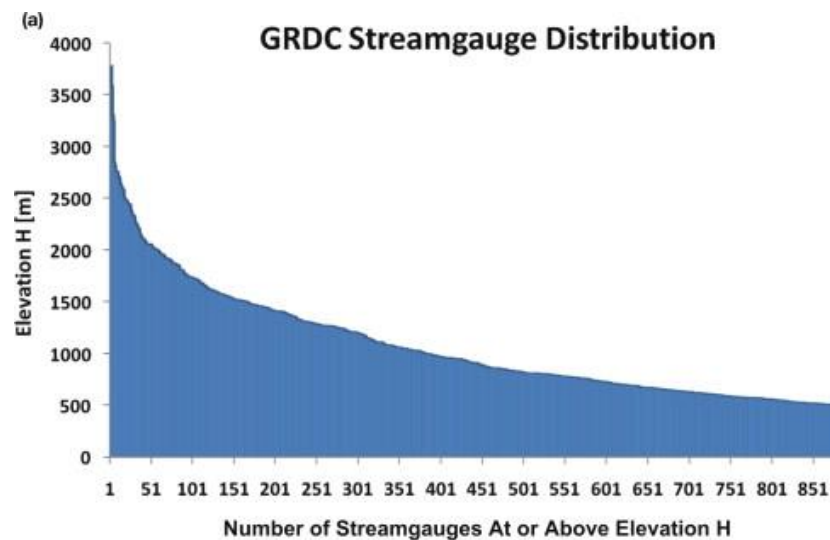


FIG. 2. Distribution of observed flashiness (0–1) over CONUS. The bounding boxes highlight known flash flood hotspots: 1) West Coast, 2) Arizona, 3) Front Range, 4) Flash Flood Alley, 5) Missouri Valley, and 6) the Appalachians.

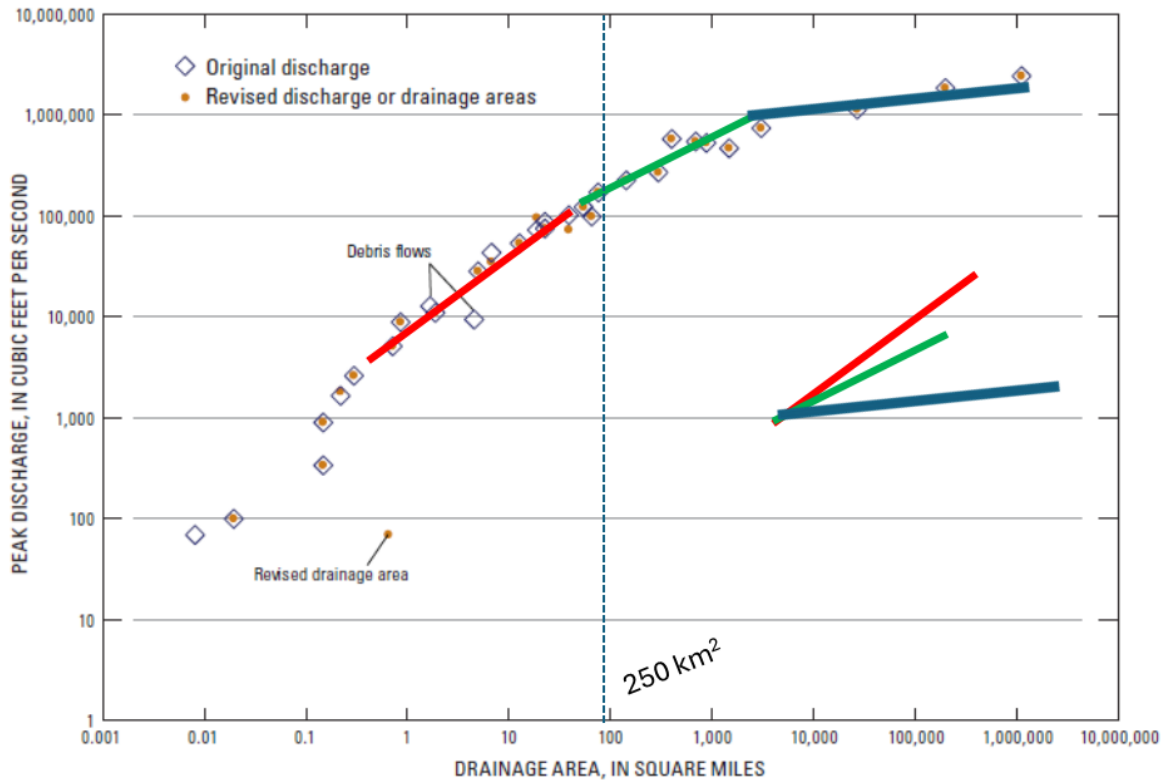
From Saharia et al. 2017



As shown by Barros (2013, see Fig. 19a below), the number of gauged basins in mountainous regions is very small globally and strongly decreases with elevation. Whereas the actual number changes in any given year, the situation has not improved in the last ten years, and indeed the number of well-maintained stream gauges that can be used reliably to evaluate and calibrate hydrological models and hydrological forecasts is even smaller and has decreased due to impacts of geopolitical instability and financial woes of regional agencies. The number of precipitation gauges is much higher but its density also strongly decreases with elevation. Studies of the historical record of exceptional floods (see figure below from Coast and Jarrett, 2008) show the steepest increase of flood magnitude with drainage area between 1 km<sup>2</sup> and 250 km<sup>2</sup>, the typical range of headwater basins in complex terrain. It is therefore critical that high quality data sets be available to support hydrologic (or AI based) model evaluation and calibration. Given the foundational importance of precipitation, we argue that having access to this high-quality data set is of great value and benefit to the community. This is the reason why we are making it available.



From Barros (2013).



Adapted from Costa and Jarrett (2008).

The Reviewer's point that there are approximations in the implementation of the IRC framework is well taken. These are well documented and justified, and the skill of the hydrologic metrics is very high. In the future, we will plan to rerun the full framework using more powerful computational resources; thus, we refer to the current data as version 1 (e.g., StageIV-IRC.1). These do not limit the utility of the dataset however. There are thousands of papers published in the peer-reviewed literature reporting on systematic evaluation of various precipitation products including optimal combinations of satellite, ground-based radar and rain gauges. None of such products meets the requirements needed at the small spatial scales of headwater basins, let alone capturing the high space-time variability of precipitation. Because of the readily available data sets for the continental US including streamflow and precipitation, and the quality of streamflow observations at gauges maintained by the USGS, the selected basins across the Appalachian Mountains provide a unique opportunity to generate a high-quality precipitation product that includes significantly different hydroclimatic regimes and physiographic settings suitable for research.

This is the first time that the previously established inverse precipitation correction approach (i.e. IRC) is systematically applied to a wide range of basins with drainage areas

ranging from 40km<sup>2</sup> to 500km<sup>2</sup>. Note Basin01 and Basin30 in this study are over 2,000 kilometers apart with drastically different weather regimes and topography. Another reason that this area is chosen is because these studied headwater basins are equipped with USGS streamgauges with 15-minute monitoring resolution, which allows us to study flash flood events in this region as rainfall runoff response can be as fast as 30 to 90 minutes in steep terrain. Similarly, the reasons for not studying larger basins (>1000 km<sup>2</sup>) are: 1: rainfall runoff response in bigger basins is relatively slower, therefore not meeting the criterion of flash floods timescales (<6 hours). 2: hydrological model resolution in this study is high (250m, 5 minutes) to capture flash floods, which makes it less suitable to run the same model configurations for bigger basins due to computational constraints. The authors also plan to use coarser resolutions and apply the same methods to larger basins in the future to study other flood events that last over 12 to 48 hours.

I agree that the golden rule is to test the performance of the dataset through hydrological modeling, but this does not mean that we should “force” it to happen. The core of the proposed bias-correction framework has been developed in the authors’ previous studies. I would thus suggest the authors to pick up either one route (framework or dataset) and resubmit to a more suitable journal. I have some other concerns, which are listed below (not necessarily in the order of importance).

We respectfully disagree, and this follows from the arguments stated above. The point of this manuscript is to provide this unique data set that was developed using the IRC framework along with detailed documentation how the data were obtained. This is not different from the publications that describe products like the ERA 5, or GPCP, or IMERG, and many others. Such products used to drive hydrologic models and for model calibration are “forced” to agree with rain gauge or radar observations in some statistical sense without meeting fundamental hydrologic criteria such as water budget closure besides exhibiting very large biases and lacking the sub-hourly space-time complexity of realistic storms. Whereas no doubt further improvements and bigger datasets will become available in the future, it is time to start a new generation of hydrology-centered precipitation products.

1. **Model uncertainty.** The authors emphasize that they are use a non-calibrated model in this region and model uncertainty is minimal. However, they attribute some of the poor performance in the new product due to model uncertainty (not capable of groundwater modeling). This is problematic, especially when transferring the framework to other regions with diverse land surface properties of runoff-generation mechanisms.

We appreciate the Reviewer's comment regarding model uncertainty, and we respectfully disagree with the statement above that is not accurate. The DCHM is capable of groundwater modeling. This model has been extensively used for the last 25 years in the Appalachians with parameters and parameterizations strongly tied to basin physics. What the model does not do is to simulate the complex subterranean hydraulics in karst terrain. Indeed, we are not aware of models of the same family as DCHM that do so. This is a question of model structural uncertainty as discussed in the manuscript. Consequently, the IRC precipitation for events in these basins are not expected to do well. We could have pretended that we did not select any basins in karst terrain. However, that would ignore an important element of the diversity of physiographic settings that surely can be found in other mountainous regions around the world as stated by the Reviewer. Therefore, it is critical that the challenges of karst hydrology and hydraulics be highlighted. This is like flagging a remote sensing product like for example SMAP soil moisture in the Amazon due to the limitations of soil moisture retrieval in densely forested areas.

This matter has no implications for transferability. In this paper we are making available a data set of extreme precipitation events that caused flash-floods for specific watersheds. We provide exhaustive description and documentation of how the data were obtained. ESSD is a journal with a well-defined mission to published data sets. This is why we submitted here.

**Presentation quality.** I would suggest the authors to substantially improve the presentation quality if resubmitted to other journals. This includes the structure of sentences which are too complicated to be understood, clear structure of the Introduction, more details for the IRC framework (right now it is only described in the figure), and less details in the equation and metrics of model evaluation. In addition, the manuscript lacks a map that clearly show the study region, including the watersheds, rain gauges, and different regions. This makes readers outside of US a miserable experience.

More details of the IRC are included in the revised manuscript along with new Figure A1-A3 in the Appendix. A new map showing the studied basins is included, and an improved version of Figure 4 that shows the regional perspective is added. In addition, links to the sites where all data are publicly available were added throughout. We regret causing frustration with the lack of these details. The manuscript was revised for clarity.

**Introduction.** The Introduction session should be reorganized in a more concise and logical way. The currently broad theme matter distracts from a focus on flash floods in mountainous areas. The second paragraph offers irrelevant context to the overall topic. It is vital to emphasize the significance of high-resolution precipitation estimates specifically in these regions. Instead of laboring extensively over the methods and details of this study, try to present a comprehensive overview of possible solutions to the challenges in QPE. The importance of Appalachian region should be highlighted as well. Why do the authors believe the new product should be able to contribute to earth system science?

Thank you for your suggestions for the introduction. The second paragraph is revised with a focus on flash floods. The importance of Appalachian region is explained according to the previous comment and is included in the revised manuscript. The importance of this work is to illustrate the uncertainties involved in the widely used radar QPE dataset (i.e. StageIV) and provide an improved dataset that can close the water budget at basin scale for flash floods.

2. More specific comments. I will not elaborate them all. They can wait till later rounds of reviews if applicable.

- Line 534-543: This paragraph is not appropriate in the Results session.

Thank you for your suggestions. This paragraph is therefore eliminated from the results session.

- Line 603-605: A lack of understanding in Karst-terrain physics does not justify the omission of model parameter calibration.

This sentence is meant to point out that model parameter calibration can hinder the understanding of physics and hide/compensate for errors in precipitation. This sentence is revised for clarity.

- Line 632-644: Discuss whether all flood-generating storms in mountainous regions align with terrain gradients and whether the greater consistency between precipitation spatial pattern and terrain gradients indicates better bias correction.

The spatial co-organization of orographic precipitation and topography across all mountain ranges (e.g. Konrad II, 1994; Smith et al., 2011; Forestia and Pozdnoukhov, 2012; Wolvin et al., 2024) is well documented in the literature including several publications on precipitation processes in the Southern Appalachians. In the Southern Appalachians in particular this is the case for heavy precipitation events tied to convective activity. A statement to this effect was added to the manuscript and references were included.

In addition, to further evaluate the effectiveness of bias correction, a comparison between StageIV and StageIV-IRC against MRMS, a radar-based product like StageIV (<https://www.nssl.noaa.gov/projects/mrms/>) using the same radar network but at higher spatial resolution (1 km) and different raingauge corrections, was conducted. MRMS is widely considered as the most advanced QPE data in the US, though large precipitation uncertainty still exists in the mountains due to low radar quality index and wide radar gaps. Note that there is no “true” reference product due to the lack of raingauges, and thus the comparison here first shows that StageIV and MRMS are well aligned as expected (see also Gao et al. 2021) while there is significantly more variance for StageIV-IRC with better hydrologic simulations.

Figure S1 shows overall good agreement between StageIV-IRC and MRMS QPE data, while StageIV-IRC produces much better hydrological responses as demonstrated in the manuscript. DEM maps are plotted as additional Figure A2 in the Appendix.

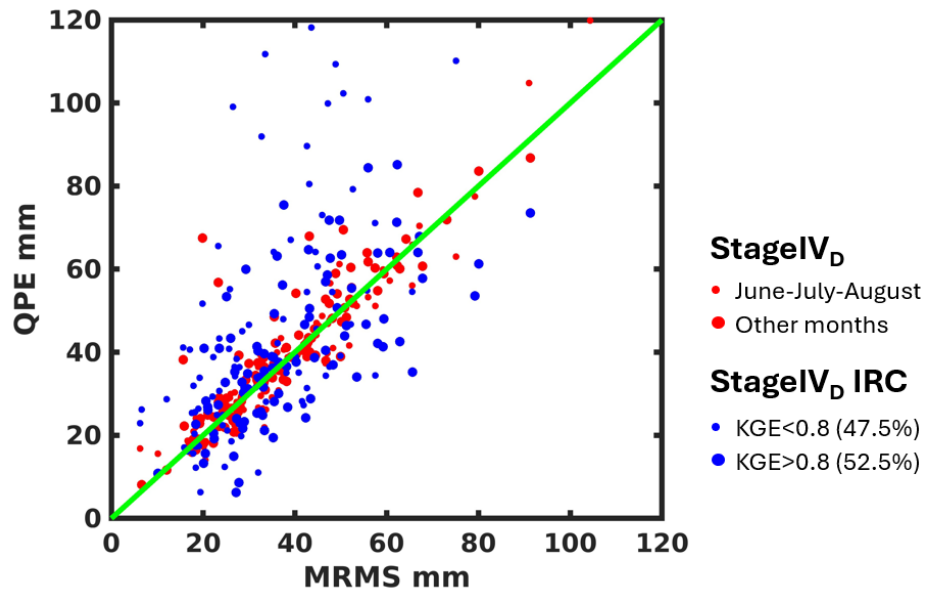


Figure S1 – Comparison of different QPE products for event total precipitation estimation (basin-averaged). Each point represents one event.

- Figure 5-7: Explain why results for specific seasons are shown.

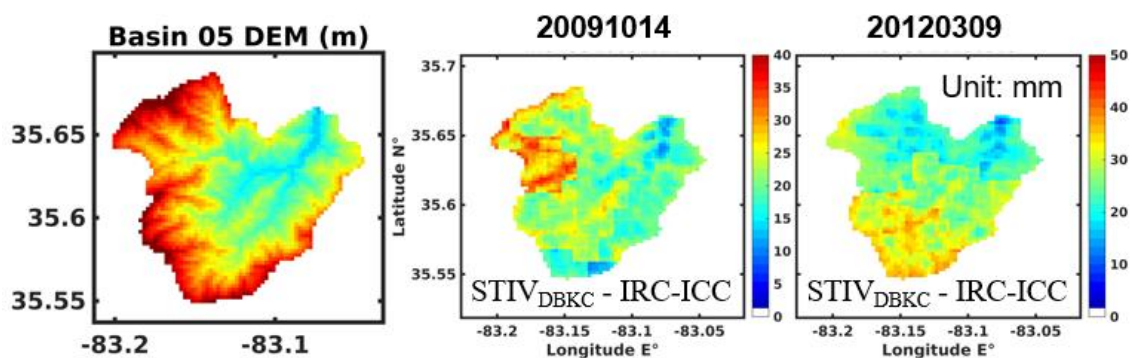
In this region, there are two primary precipitation regimes: long-lasting stratiform precipitation from shallow precipitation systems in the cold season (January, February and March), and high intensity thunderstorms in the warm season (July, August and September). Figure 5 shows the warm season and Figure 6 shows the cold season. A common issue is the missing detection of shallow precipitation systems



during late afternoon in the winter in this region which is highlighted in Figure 6. Figure 7 shows all metrics for the same season as in Figure 6.

- Figure 11: Show the terrain gradients, which makes it comparable to the spatial pattern of precipitation.

Based on previous comments, terrain gradients are plotted in the Appendix Figure A2 using the example of Basin 05. The slope map of Basin 05 matches well with rainfall corrections in Figure 10 using event 2012-03-09 and event 2009-10-14 as examples, and the figures are demonstrated below (event-total rainfall are plotted):



Reference:

Konrad II, C. E. (1994). Moisture trajectories associated with heavy rainfall in the Appalachian region of the United States. *Physical Geography*, 15(3), 227-248.

<https://doi.org/10.1080/02723646.1994.10642514>

Smith, J. A., Baek, M. L., Ntelekos, A. A., Villarini, G., & Steiner, M. (2011). Extreme rainfall and flooding from orographic thunderstorms in the central Appalachians. *Water Resources Research*, 47(4). <https://doi.org/10.1029/2010WR010190>

Foresti, L., & Pozdnoukhov, A. (2012). Exploration of alpine orographic precipitation patterns with radar image processing and clustering techniques. *Meteorological Applications*, 19(4), 407-419. <https://doi.org/10.1002/met.272>

Wolvin, S., Strong, C., Rupper, S., & Steenburgh, W. J. (2024). Climatology of orographic precipitation gradients over High Mountain Asia derived from dynamical downscaling. *Journal of Geophysical Research: Atmospheres*, 129(20), e2024JD041010.

<https://doi.org/10.1029/2024JD041010>

Gao, S., Zhang, J., Li, D., Jiang, H., & Fang, Z. N. (2021). Evaluation of multiradar multisensor and stage IV quantitative precipitation estimates during Hurricane Harvey.

*Natural Hazards Review*, 22(1), 04020057. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000435](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000435)