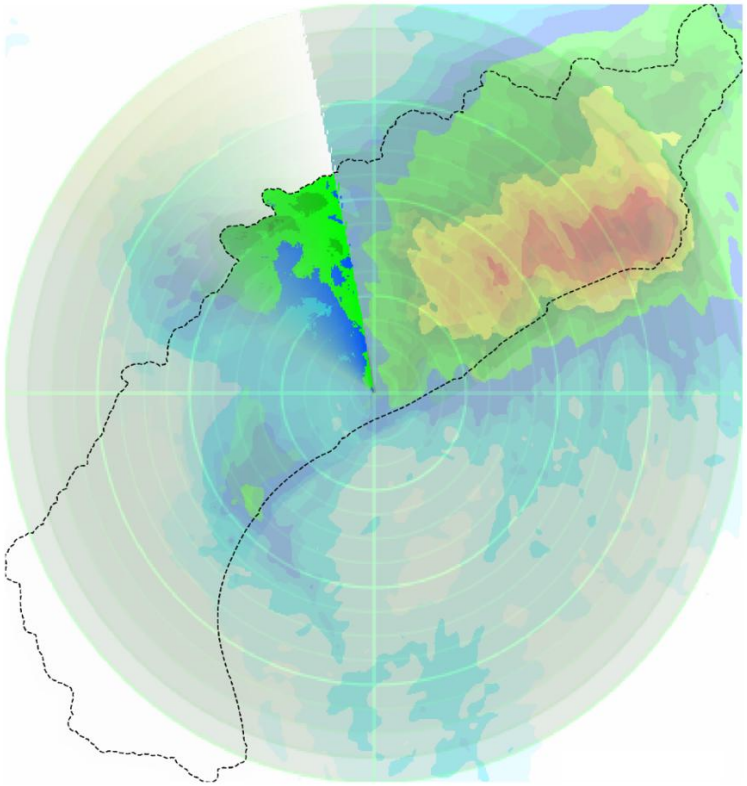
ENHANCED RADAR PRECIPITATION PRODUCT FOR THE TEXAS COASTAL AREA

Contract No. 2301792723



Prepared For:

Coastal Science Team

Texas Water Development Board (TWDB)

1700 N. Congress Avenue, Austin, TX 78711

Submitted By:

Nick Fang, Ph.D., P.E.

Robert S. Gooch Endowed Professor

Director, Water Engineering Research Center

Civil Engineering Department

The University of Texas at Arlington

Box 19308, Room 431, Nedderman Hall, 416 Yates St.,

Arlington TX 76019-0308

UTA

WERC Water Engineering Research Center

February 2026

ACKNOWLEDGEMENTS

This research project is supported by the Texas Water Development Board (TWDB) Contract No. 2301792723. We would like to acknowledge the Coastal Science Team at TWDB: Caimee Schoenbaechler, Amin Kiaghadi, Ram Neupane, and Kim Karimi for their management and close coordination with the research team in ensuring the smooth execution of this project.

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Texas Water Development Board (TWDB). This report does not constitute a standard, specification, or regulation.

The authors reserve the right to revise, update, or expand upon the material presented in this report as new data, methods, or interpretations become available. Furthermore, the authors retain full intellectual property rights and exclusive authority to publish this material whether in part or in full in academic journals, technical reports, conference proceedings, or any other appropriate medium. They may also adapt the findings for educational purposes, derivative works, or future research efforts. Any reuse, distribution, or reproduction of this material by third parties must include proper attribution to the original authors and work.

The authors declare that they have no conflicts of interest related to the content, data, or outcomes presented in this report.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Full Form
AQC	Accumulation Quality Control
AQC	Accumulation-based Quality Control
ASOS/AWOS	Automated Surface/Weather Observing Systems
DOD	Department of Defense
FAA	Federal Aviation Administration
FEWS	Flood Early Warning System
GHCNd	Global Historical Climatology Network Daily
GPM	Global Precipitation Measurement
HADS	Hydrometeorological Automated Data System
HCFCD	Harris County Flood Control District
HEC-HMS	Hydrologic Engineering Center – Hydrologic Modelling System
IDEW	Inverse Distance and Elevation Weighting
IDW	Inverse Distance Weighting
IRNC	Information Rich Non-Critical
MADIS	Meteorological Assimilation Data Ingest System
MPE	Multisensor Precipitation Estimates
MRMS	Multi-Radar Multi-Sensor
MRMS_GageCorr	MRMS Gage Corrected
MRMS_ROnly	MRMS Radar Only
NCC	Normalized Cross-Correlation
NCEP	National Centers for Environmental Prediction
NEXRAD	Next Generation Weather Radar
NMQ	National Mosaic and Multi-Sensor QPE
NOAA	National Oceanic and Atmospheric Administration
NSE	Nash-Sutcliffe model Efficiency coefficient
NWS	National Weather Service
OLS	Ordinary Least Square
PCA	Principle Component Analysis
QC	Quality Control
QPE	Quantitative Precipitation Estimates
R^2	Coefficient of Determination
RAIDW	Regression-residuals density-Attenuated IDW
RBQC	Radar Based Quality Control
RFC	River Forecast Centers
RMSE	Root Mean Square Error
RQI	Radar Quality Index
SSIM	Structural Similarity Index Measure
TAQC	Temporal Alignment Quality Control
TWDB	Texas Water Development Board
WERC*	Water Engineering Research Center

**Used in report to reference this study's products*

Executive Summary

Texas is one of the most economically vital and disaster-prone states in the nation. If considered independently, it would rank as the world's eighth-largest economy, with critical energy, petrochemical, agricultural, and shipping assets. Yet Texas also leads the U.S. in billion-dollar disasters, suffering more than $400 billion in losses and dozens of fatalities annually since 1980. Extreme rainfall events, particularly from hurricanes and tropical storms, are especially destructive, driving over $250 billion in damage and repeatedly overwhelming infrastructure and emergency systems. Accurate measurement of rainfall extremes is therefore an economic and public safety imperative. Existing radar-based products like such as the Multi-Radar/Multi-Sensor (MRMS) provide broad coverage but suffer from systematic errors and sparse local calibration. This project addresses these shortcomings by combining radar data with available rain gage networks across Texas, applying rigorous quality control and advanced bias-correction methods to produce rainfall fields that are more reliable and regionally representative.

Coastal Texas depends on accurate rainfall information for flood warning, ecological monitoring, reservoir operations, and other long-term planning purposes. Yet rain gages are unevenly distributed, and radar, while providing extensive geographic coverage often over or under estimates rainfall. The study had four primary objectives: (1) create a higher-quality rainfall dataset for the 2015 to 2024 period, (2) include more gages by enhancing quality control, (3) document a transparent and repeatable method, and (4) identify remaining gaps to guide future work. The project produced an enhanced gridded rainfall product for Coastal Texas, ready-to-use data packages (GeoTIFFs, DSS files, and PCP files), a practical toolkit with documentation, and summaries of the approach and findings.

Expanding the rain gage network used to correct MRMS rainfall estimates reduced errors and increased agreement with ground observations, resulting in a 67 percent improvement in rainfall estimates. The enhanced rainfall product was further validated by applying the corrected rainfall to a hydrologic rainfall-runoff model for the Hurricane Harvey Event. Model generated hydrographs matched observed hydrographs (Nash–Sutcliffe Efficiency of 0.91), outperforming alternative datasets such as MRMS or Stage IV. This validation further supported the need for increased MRMS accuracy using additional gage networks.

Despite its improvements demonstrated, the developed product has several inherent limitations tied primarily to data availability, measurement uncertainty, and intended operational scope. The bias-correction approach depends on the quality and spatial distribution of available gages, which remain uneven across much of coastal and rural Texas. While the inclusion of additional local and private gages can greatly enhance spatial coverage, tipping-bucket instruments introduce known measurement uncertainties, particularly under high-intensity rainfall conditions that can lead to underestimation during extreme events. Moreover, the adopted regression–interpolation framework (RAIDW) cannot fully correct missed radar detections in zero-rainfall zones and may underrepresent variability in regions with sparse gage density. Temporal inconsistencies in gage reporting and uncorrected radar artifacts can also contribute to localized

discrepancies. Finally, the workflow was designed primarily for retrospective analysis; extension to near-real-time applications will require further optimization of computational efficiency and automated data quality screening.

Future research should focus on operationalizing the workflow for real-time or near-real-time applications to support active flood forecasting and emergency response. Integrating additional data sources such as satellite precipitation estimates, and soil moisture observations could further strengthen correction performance across diverse storm types. Machine learning techniques also present a promising avenue for improving spatial bias correction, dynamically identifying faulty gages, and detecting radar anomalies. Expanding the framework to other climatic regions would help test its generalizability and support refinement of key parameters under varying meteorological regimes. Establishing continuous validation using streamflow observations and hydrologic model outputs could further quantify real-world performance, guide adaptive recalibration, and improve detection of gage failures. Finally, sustained collaboration among state, federal, and local agencies to enhance gage density in critical areas and standardize quality control (QC) protocols will be essential for sustaining and extending the reliability of gage-corrected rainfall products statewide. Taken together, these limitations do not diminish the utility of the current product for regional analysis and planning, but they highlight clear pathways for targeted enhancement and operational deployment.

1 Introduction

Texas is one of the most flood-prone and economically vital regions in the United States, where accurate measurement of rainfall is critical to infrastructure design, reservoir management, and emergency response. The Multi-Radar Multi-Sensor (MRMS) gage-corrected product developed by the National Weather Service (NWS) is one of the standard gridded radar precipitation datasets used for these purposes. MRMS integrates radar observations with surface measurements predominantly from the Hydrometeorological Automated Data System (HADS) networks (prior to 2018) and later from the Meteorological Assimilation Data Ingest System (MADIS) to produce bias-corrected rainfall estimates at high spatial and temporal resolutions (Martinaitis et al., 2021). However, gage networks used by MRMS are limited in station density and uneven across Texas, leaving large portions of the state, especially coastal and rural areas, underrepresented in the bias-correction process.

This project was commissioned by the Texas Water Development Board (TWDB) Coastal Science team to advance methods that provide consistent, continuous, and high-quality precipitation data for the Texas Gulf Coast. It improves the accuracy of rainfall estimates by expanding and refining the ground-truth gage foundation used for bias correction and validation. Building on the MRMS framework, the study integrates a larger network of rain gages compiled through Synoptic Data by the Water Engineering Research Center (WERC) at The University of Texas at Arlington, supported by rigorous multi-stage quality control and dynamic validation. The resulting product is an enhanced, region-specific rainfall dataset that produces radar fields using improved statistical and spatial methods tailored to Texas' hydroclimatic conditions. By combining regionally optimized bias-correction method with denser, quality-screened gage observations, the project delivers a corrected and enhanced product with higher spatial fidelity and reliability for flood forecasting, hydrologic modeling, and long-term water management applications.

1.1 Research Goals and Objectives

The need for this work was identified through the Texas Integrated Flooding Framework (TIFF), which emphasizes the importance of high-quality and consistent precipitation data to support statewide flood modeling, risk assessment, and early-warning systems. This study builds upon those findings by developing and testing a region-specific bias-correction and validation framework for radar rainfall products, enhancing their performance in coastal environments where existing ground-truth data are limited.

The primary goals of this project are to improve the precision, reliability, and usability of gridded rainfall fields and to establish a transferable framework that can be expanded to other regions of Texas.

Specific objectives of this study are to:

- Compile and quality-control precipitation observations from multiple existing networks to create a unified, validated dataset for coastal Texas;
- Develop and apply regionally tailored bias-correction algorithms that incorporate dynamic validation to improve temporal and spatial consistency in rainfall fields;
- Evaluate the effectiveness of the bias-corrected product relative to uncorrected datasets and prepare documentation to support its operational and research use; and
- Demonstrate a reproducible framework that can be extended to other regions of Texas and integrated into statewide modeling and early-warning programs.

This study was designed as a multi-year proof of concept demonstrating how advanced rainfall bias-correction and validation methods can be strengthened when applied regionally using a scalable system. The initial phase focused on the Texas Gulf Coast, where the diversity of rainfall mechanisms and historical data limitations provide an effective test environment for evaluating and refining the approach. Lessons learned from this effort will guide future deployment across Texas and will inform continued enhancements to the data and modeling infrastructure developed under the TWDB.

The results also serve as a prototype for potential statewide implementation under the TWDB's TIFF, contributing to its goals of reducing data gaps and advancing the scientific basis for flood planning across Texas. By establishing a corrected MRMS product for the Texas coast, this work provides a foundation to improve future coastal flood forecasting, model calibration, and infrastructure design.

1.2 Scientific and Operational Rationale for a Region-Specific Rainfall Product

The development of a region-specific rainfall product is grounded in the complex physical nature of precipitation and the operational demands of hydrologically vulnerable regions. Along the Texas Gulf Coast, land-sea interactions, warm-rain processes, and tropical influences create rainfall patterns that differ from continental regimes. The scientific, economic, and operational implications of these regional characteristics are discussed in the following subsections.

1.2.1 Regional and National Importance of Accurate Rainfall Estimation

Texas holds a uniquely important position in the United States, supporting a rapidly expanding population, major industrial centers, and several of the nation's most critical ports and energy corridors. If considered independently, Texas would rank as the world's eighth-largest economy. Because of this concentration of people and infrastructure, disruptions originating in Texas can reverberate across national supply chains, energy markets, and commercial activity. These sensitivities are most acute along the Texas Gulf Coast, where assets such as the Port of Houston and the petrochemical corridors from Freeport to Port Arthur form the backbone of U.S. energy and trade.

The United States has sustained nearly $3 trillion in damages from weather and climate-related disasters since 1980, with precipitation-driven hazards dominating both the frequency and cost

of those events. Severe storms, hurricanes, and related flooding account for more than 80 percent of all billion-dollar disasters, while hurricanes alone are responsible for more than half of the total economic losses. From 1980 to 2024, approximately 15 percent of all billion-dollar disasters occurred in Texas, the highest of any state, with tropical storms accounting for roughly $250 billion in damages across 14 major events (Smith, 2025). Texas ranks first in the U.S. in terms of human safety toll from hydrometeorological disasters, with an average of 42 fatalities per year (Paul et al., 2018). These losses underscore how sensitive the Texas economy is to extreme precipitation and reinforce the economic and societal imperative for accurate rainfall information across Texas, especially in hydrologically sensitive coastal regions.

1.2.2 Coastal Rainfall Characteristics and Limitations of Existing Products

The Texas coastal zone is repeatedly exposed to tropical and post-tropical systems whose rainfall characteristics challenge conventional radar-correction techniques. Events such as Hurricane Harvey (2017), Tropical Storm Imelda (2019), and Hurricane Nicholas (2021) illustrate the compound flood hazards that occur when slow-moving or stalled systems produce prolonged, high-intensity rainfall over low-relief terrain. During Harvey (2017), rainfall totals exceeded 60 inches in parts of the upper Texas coast, overwhelming both radar and gage-based estimation systems and exposing major spatial inconsistencies between MRMS fields and observed accumulations (Gao et al., 2021). Imelda revealed similar issues on shorter timescales (Li et al., 2023). These recurring challenges show that rainfall-estimation errors are not isolated anomalies but systematic regional deficiencies requiring targeted correction and validation strategies.

Accurate rainfall estimation is further complicated by the physical and observational characteristics of coastal Texas. Radar-to-gage relationships vary markedly between the coastal plain and inland regions due to shallow warm-rain processes, sea-breeze convection, and frequent storm translation from the Gulf (Birch et al., 2015; Suh et al., 2021).

1.2.3 The Need for Regionally Optimized Interpolation

The Texas coast presents unique observational and analytical challenges that justify the need for a region-specific rainfall product. National products rely on interpolation and bias-correction approaches optimized for continental uniformity but do not fully reflect local factors that influence spatial rainfall-field bias. For example, the MRMS system provides dynamic reflectivity-to-rainfall adjustments but does not explicitly incorporate coastal-specific microphysical and atmospheric effects nor does it apply regionally optimized radar-gage validation within its bias-correction framework. Thus, these national methods often underperform in regions with distinctive hydroclimatic signatures and uneven gage density conditions that define the Texas Gulf Coast.

By implementing regionally optimized bias-correction procedures specifically calibrated to regional meteorological and observational patterns, this study addresses this gap for the Texas coast. A central driver of this work is the need for an interpolation framework that is unique to the spatial density, clustering, and distribution patterns of the Texas Gulf Coast gage network

rather than relying on broad continental assumptions. Since regional gage clusters, spacing irregularities, and gage distributional biases cannot be adequately addressed by nationwide smoothing techniques, this study introduces a smooth, continuous, and differentiable gage-density index that systematically attenuates gage influence with distance. This innovative procedure eliminates abrupt transitions common in other interpolation methods, improves spatial coherence, and yields a more stable basis for hydrologic modeling. The result is a rainfall-correction method that more closely aligns with the processes that dominate rainfall variability in this region.

1.2.4 Relevance to TWDB

Rainfall-estimation errors hinder calibration of hydrologic models used for reservoir operations, evacuation planning, and long-term flood-mitigation design. Even modest uncertainty in rainfall estimates can translate into disproportionate uncertainty in modeled discharge and inundation, complicating emergency response decisions and cost–benefit evaluations that guide public investment. Enhancing rainfall accuracy therefore yields measurable value: it supports more reliable forecasts, defensible design standards, and better prioritization of mitigation spending.

Collectively, these scientific, observational, and methodological drivers establish the need for a region-specific rainfall product tailored to the Texas Gulf Coast. Integrating dense gage networks with robust quality control, regionally optimized interpolation, and an information-based validation framework produces a higher-fidelity rainfall dataset that more accurately represents storm structure, timing, and magnitude.

The Texas Gulf Coast provides an ideal proof-of-concept environment for this work. Due to its complex hydroclimatic conditions, high exposure to flood risk, and economic importance to the state, advancements demonstrated in this region can strengthen flood-risk modeling, real-time forecasting, and long-term planning while paving the way for broader application of regionally tailored rainfall products across Texas.

1.3 Geographic Extent and Coastal Relevance

The study area encompasses the Texas Gulf Coast and its contributing watersheds, extending inland to include the upstream river basins. This region includes several of the state's most hydrologically complex and economically important coastal counties, characterized by low-relief terrain, tidal influence, and dense urbanization. The project's analysis and validation focus on the coastal hydrologic units where rainfall uncertainty has the greatest influence on flood modeling and emergency-response decisions. **Figure 1** illustrates the geographic extent of the study area , which includes a 5-km buffer to reduce interpolation edge effects while maintaining computational efficiency.

Figure 1: *Coastal Texas area of interest (dashed lines), with a 5 km buffer region (solid line)*

The coastal region is particularly vulnerable to compound flooding driven by coincident rainfall, storm surge, and riverine inflows during tropical cyclones and other high-impact weather events (Ali et al., 2025; Dykstra & Dzwonkowski, 2021). The tools used to assess and forecast these risks, hydrologic and hydraulic models, depend directly on accurate, continuous, and spatially

representative rainfall data (Morbidelli et al., 2021). When rainfall estimates lack sufficient spatial detail or when gage density is low, flood models can either underpredict or overpredict inundation depths and extents, creating uncertainty in both real-time forecasting and long-term planning. Thus, high-quality rainfall information is essential to mitigate these risks and to support proactive resilience strategies (Jayawardene et al., 2021; So, 2023; Yu et al., 2018).

The expanded gage network and dynamic bias-correction methods developed in this study enable better calibration of radar estimates to these coastal conditions, improving rainfall representation in basins where flood risk is strongly influenced by short-duration, high-intensity localized events.

1.4 Report Organization

This report presents the scientific motivation, data foundations, methodological development, and validation results supporting the creation of a region-specific rainfall product for the Texas Gulf Coast. **Section 2** describes the precipitation datasets used in this study and outlines the associated acquisition and quality-control procedures. **Section 3** details the bias-correction methodology, including the regression-residual density-attenuated interpolation framework, and dynamic validation components. **Section 4** provides statistical and hydrologic validation results comparing the corrected and uncorrected rainfall fields. **Section 5** summarizes key findings and discusses implications for operational use. **Section 6** provides concluding remarks, and lastly, **Section 1.1**outlines limitations and recommendations for future development, and potential statewide expansion.

2 Data Sources

Rainfall information in the United States is obtained from three primary observational systems: surface rain gages, weather radar networks such as Next Generation Weather Radar (NEXRAD), and satellite-based remote sensing missions such as the Global Precipitation Measurement (GPM) program (Sun et al., 2018). Radar data provides superior spatial distribution and temporal resolution of rainfall when available, but gages provide more accurate point measurements of precipitation. To obtain reliable rainfall datasets, radar products must therefore be bias-corrected using gage observations (Gourley et al., 2010; Sexton et al., 2010; Shakti et al., 2019; Tan et al., 2021; Wang et al., 2021; Yoon et al., 2012).

This study addresses data needs for the region-specific framework by expanding gage networks that can be used for bias-correction. When compared to legacy HADS coverage, the expanded gage network substantially increased spatial density across the study region.

The following sections describe the specific data sources used in this study.

2.1 Rain Gage Networks

Rain gage observations form the foundation for quantitative precipitation analysis and provide the reference against which radar and satellite products are calibrated. For this study, rainfall gage data were obtained through the Synoptic Data API, which aggregates real-time and historical measurements from multiple public and institutional networks operating across Texas (Synoptic Data Program, 2024). The Synoptic platform enables uniform access to otherwise independent datasets and provides consistent formats, basic quality flags, and metadata.

In addition to major sources such as HADS, Automated Surface/Weather Observing System (ASOS/AWOS), and TexMesoNet, a diverse set of smaller or locally maintained systems contribute precipitation observations across Texas. These include gage networks operated by river authorities, drainage and flood control districts, municipalities, universities, and private entities. While these systems vary in quality and reporting frequency, collectively they improve spatial coverage in otherwise ungaged areas and enrich the composite dataset available for further research and flood-risk applications.

All gage data acquired through Synoptic Data were subjected to the quality control procedures described in **Section 2.4**, including accumulation screening, temporal alignment, and radar-based consistency checks. Gages passing these tests were retained to form the WERC Gage Network, a curated and quality-controlled subset of all available records tailored for this study's bias correction and validation framework. The resulting dataset includes approximately 2,048 active precipitation stations across the study domain, representing a substantial increase in observation density relative to legacy HADS coverage, particularly in coastal and rural counties.

The primary observation networks contributing to this dataset are summarized in **Table 1**. The spatial distribution of contributing gages is shown in **Figure 2**, representing a snapshot of network availability at the time of analysis. The data shown is a simple visualization of Synoptic metadata,

where periods of record were estimated by the earliest and latest record of any data type in the gage, disregarding gaps in between.

Table 1: *Rain gage count per source as of 2024*

All Sources:	**Count**
AIRNow	78
ASOS/AWOS	52
Automatic Position Reporting System WX NET/Citizen Weather Observer Program	463
Climate Reference Network	2
Coastal Carolina University	1
Community Collaborative Rain, Hail and Snow Network	687
DataWise Environmental Monitoring Inc.	36
DTN	3
Harris County Flood Control District	333
Hydrometeorological Automated Data System	91
Intellisense Systems Inc.	1
Interagency Remote Automatic Weather Stations	24
Jefferson County Drainage District No. 6	76
Johnson Space Center NASA	2
Lower Colorado River Authority	31
National Estuarine Research Reserve System	2
National Ocean Service Physical Oceanographic Real-Time System	2
National Ocean Service Water Level Observation Network	16
NWS Cooperative Station Climate	15
NWS Cooperative Station Climate - Hydro	40
NWS Cooperative Station Hydro	37
NWS Cooperative Station Local or Met	7
NWS Unofficial Cooperative Station	2
San Jacinto River Authority	13
Soil Climate Analysis Network	4
Texas Soil Observation Network	2
Texas Water Development Board	22
Weather For You	3
WeatherSTEM	1
Unknown	2
Total:	**2,048**

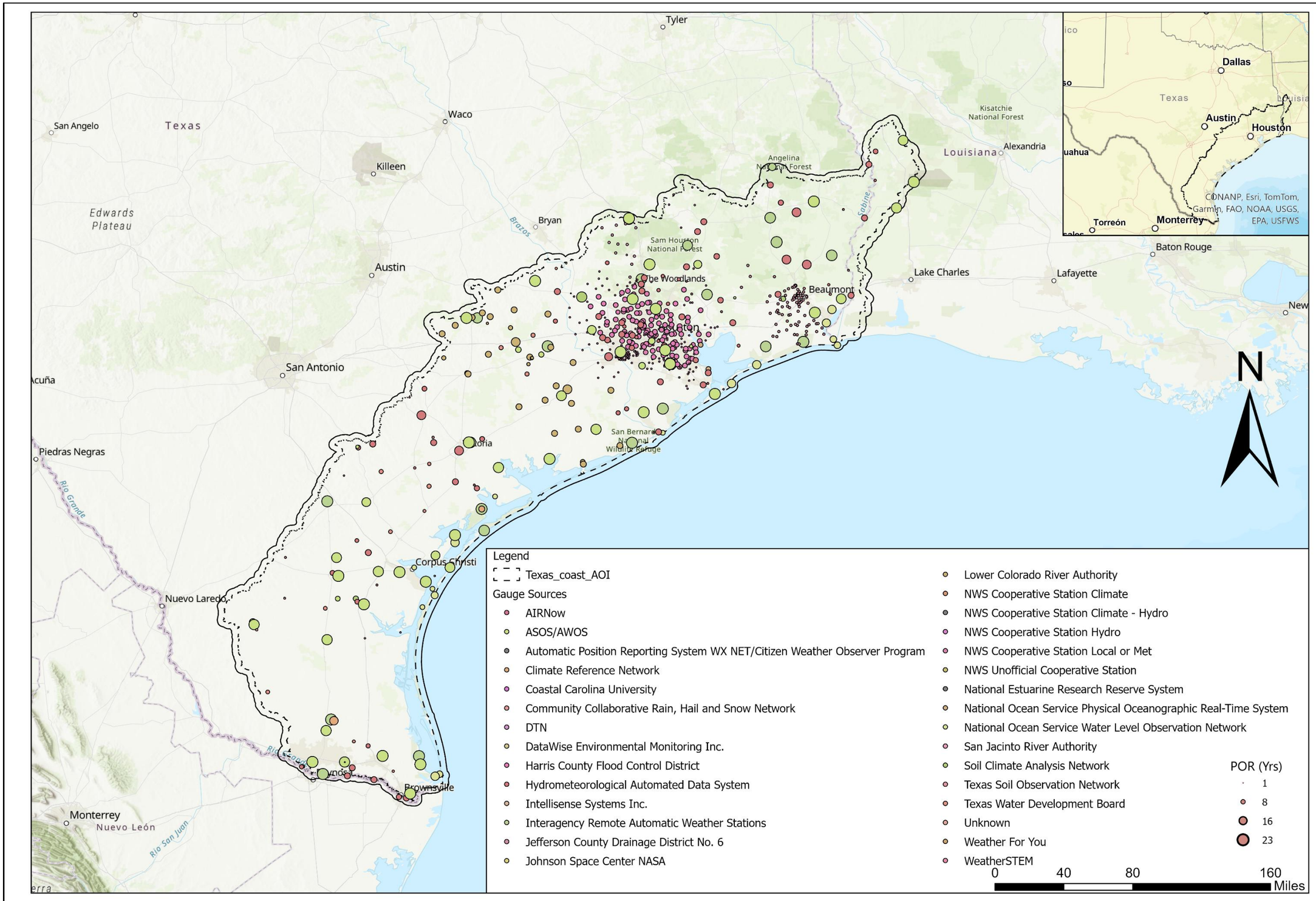


Figure 2: *All rain gage sources and their periods of record across the Texas Gulf Coast, showing spatial coverage as of late 2024*

2.2 Gridded Precipitation Products

The coastal Texas area has a high radar quality index (RQI ≈ 1), indicating excellent coverage and reliable reflectivity measurements (Zhang et al., 2012). Thus, radar is widely recognized as a powerful tool for capturing the spatial variability of precipitation events.

The MRMS system aggregates radar data from multiple WSR-88D sites, which are quality-controlled and merged into a three-dimensional mosaic. This forms the basis for calculating surface precipitation rates for each radar. Surface precipitation rates are processed into MRMS continental mosaics (Zhang et al., 2016, 2020). This mosaic is acquired and used as the primary gridded input in this study.

2.3 Data Acquisition

The data-acquisition framework translates the diverse precipitation observations into standardized, analysis-ready datasets suitable for bias-correction modeling. Because the contributing radar and gage records originate from numerous agencies, file formats, and temporal resolutions, the acquisition stage required coordinated retrieval, harmonization, and spatial alignment steps that extend well beyond simple data download. The retrieval period spans January 1, 2015, through December 31, 2024, encompassing hurricanes, convective storms, and drought periods representative of the coastal hydroclimate in Texas. All raster operations were conducted on a uniform 0.01° × 0.01° (≈1 km) grid using the Geographic Coordinate System (GCS_WGS 1984), chosen to balance computational efficiency with the spatial details needed to resolve rainfall gradients. This consistent framework ensures that radar, gage, and satellite inputs can be merged without projection-related distortions.

2.3.1 Point Precipitation Data Acquisition

Rain gage data for the area of interest (AOI) were obtained through the Synoptic Data API service (Synoptic Data API, https://api.synopticdata.com/v2/stations/timeseries). This source consolidates precipitation observations from the networks listed in **Table 1**. In total, 2,048 gages were retrieved for the AOI, as shown by source and periods of record in **Figure 2**. Data from Synoptic was obtained in JSON format and converted to a comma-separated (*.csv*) structure for each station, containing two columns: datetime and hourly accumulation. A separate *.csv* file stores metadata for all stations in the AOI, including latitude, longitude, elevation, period of record, and record type. Each accumulation value represents total rainfall over a one-hour period, recorded at the timestamp marking the end of that interval (i.e., period-cumulative accumulation). As expected, some data exhibited quality issues, which are addressed in the following sections.

2.3.2 Gridded Precipitation Data Acquisition

The MRMS radar-only "Q3RAD" product, which provides hourly surface precipitation accumulations, was selected as the primary source of radar-derived precipitation data. These data are primarily retrieved from the Iowa State Mesonet Archive as *.grib2* files[1]. In cases where Q3RAD data are missing, a supplementary retrieval process is implemented. Specifically, the 2-minute surface precipitation rate product (mm/hr) is used as an alternative source[2]. To reconstruct hourly accumulations, all 2-minute precipitation rate grids corresponding to the hour of interest (ideally 30 values) are retrieved. The hourly accumulation is estimated by averaging the available 2-minute rates (mm/hr) and multiplying by one hour (hr). This approach yields accurate precipitation accumulation when all 30 data are available. When fewer than 30 values are available due to missing data, missing 2-min precipitation rates are assumed to equal the mean of the available precipitation rates (**Equation i**).

$$Accumulation_{1hr} = \frac{\sum_{i=1}^{k} Precip\ Rates_i(\frac{mm}{hr})}{k} \times Time(hr) \Bigg\} k \leq 30$$

$$Accumulation_{1hr} = \frac{\sum_{i=1}^{k} Precip\ Rates_i}{k}\ mm \Bigg\} \begin{matrix} k \leq 30, \quad Time = 1hr, \\ Precip\ rates\ in\ mm/hr \end{matrix} \quad \text{i}$$

Where k is the number of available 2-minute precipitation rates.

When neither primary sources provide usable data, the Iowa State Mesonet GIS archive serves as the tertiary fallback[3]. In this archive, MRMS products are stored as 8-bit *.png* files with pixel values ranging from 1 to 255. These values are decoded into precipitation accumulations using the provided scaling keys[4]. The primary fallback product used is the 1-hour accumulation layer (mrms_p1h). If the 1-hour accumulation is unavailable, the 2-minute accumulation product (mrms_a2m) is used instead of the 2-min precipitation rate product applied in **Equation i**. In this case, the hourly accumulation is reconstructed by averaging available 2-minute accumulations and scaling the result to a 60-minute total **(Equation ii)**:

$$Accumulation_{1hr} = \frac{\sum_{i=1}^{k} 2Min_{Accumulation_i}}{k} \left(\frac{mm}{2min}\right) \times 60min \Bigg\} k \leq 30$$

$$Accumulation_{1hr} = \frac{\sum_{i=1}^{k} 2Min_{Accumulation_i}}{k} \times 30 \Bigg\} mm \qquad k \leq 30 \quad \text{ii}$$

Where k is the number of available 2-minute accumulations.

This approach ensures that when full data are available, the hourly accumulation reflects the true sum of observations. When gaps exist, missing values are filled using the mean of the observed 2-minute accumulations, minimizing the uncertainty introduced by missing data. This approach enables the reconstruction of uninterrupted radar-only precipitation data for 99.98% of the study period. For the remaining 0.02% of missing data, temporally adjacent hourly radar products are

used as substitutes. As these gaps predominantly correspond to no-event conditions (i.e., negligible or zero rainfall), their addition introduces no meaningful uncertainty into the dataset.

Gridded precipitation data sources are listed below in order of priority:

[1] https://mtarchive.geol.iastate.edu/2020/01/01/mrms/ncep/RadarOnly_QPE_01H/

[2] https://mtarchive.geol.iastate.edu/2020/01/01/mrms/ncep/PrecipRate/

[3] https://mesonet.agron.iastate.edu/archive/data/

[4] https://mesonet.agron.iastate.edu/GIS/rasters.php?rid=5

2.4 Gage Data Quality Control

This section outlines the sequential procedures applied to ensure the reliability, consistency, and temporal alignment of rain gage observations prior to their integration with radar datasets. The quality-control process addresses common issues such as accumulation resets, irregular reporting intervals, and discrepancies between radar and gage measurements.

2.4.1 Accumulation quality control (AQC)

Many gages experience accumulation resets at midnight, at the end of the year, or when manually flushed. This necessitates a preliminary quality correction of the precipitation accumulation record. To address this, accumulation vector (*A)* is shifted by one step (A_{-1}) and differenced from the original accumulation values to obtain interval precipitation vector *(I)* at time *(t)*. All negative interval values are set as zero to account for sudden drop in accumulated precipitation resulting from resets (**Equation iii**). Accumulation is recalculated from the adjusted interval values (**Equation iv**).

$$I = A - A_{-1} \qquad \text{iii}$$

$$For\ all\ \ i\ \in I, \qquad if\ i < 0\ set\ i = 0$$

$$a_t = \sum_{0}^{t} i_t\,,\ \ a_t \in A_{AQC}, \qquad i_0 = 0 \qquad \text{iv}$$

Here, a_t and i_t denote the accumulation and interval (precipitation increment) values, respectively, at time t.

This process is referred to as accumulation quality control (AQC) for the remainder of this report. This corrected accumulation forms the basis for all subsequent quality control described in the following sections.

2.4.2 Temporal alignment quality control (TAQC)

Temporal alignment is a critical step in reconciling gage and radar data, ensuring that precipitation values from gages are adjusted to consistent time intervals. This step is essential for

any meaningful comparison, correction, or data fusion effort between spatially distributed radar estimates and temporally irregular gage observations.

Gage data present several challenges related to reporting times. A notable issue arises with the tipping bucket gage reporting mechanism: although data may be reported frequently, no precipitation is recorded until the bucket fills and tips, introducing a minimum detection threshold. For instance, if one tip corresponds to 0.04 inches of rainfall, the gage will report zero in the intervals between tips even if precipitation is occurring continuously. Thus, depending on the number of reporting intervals between tips, the actual temporal distribution of rainfall within those intervals is unknown and could represent many possible combinations that sum to observed accumulation. As a result, a reported sequence such as 0.04, 0, 0, 0, 0.04 must be redistributed to preserve the cumulated total while better reflecting temporal variability. One possible redistribution is 0.04, 0.01, 0.01, 0.01, 0.01.

Another challenge is the lack of time synchronization among gages. For example, three hourly-reporting gages might log data at 2:02, 2:09, and 2:32, respectively. Furthermore, some gages also report at irregular intervals, for example, tipping bucket gages that record the exact timestamp of each tip or manually operated gages that log data whenever it is collected. To construct a coherent representation of rainfall across space and time, it is important to standardize reporting times such as aligning all data to the top of the hour. One approach is to use only gages that report within a narrow window of the target time (e.g., accepting 2:02 but excluding 2:09 and 2:32 based on a defined threshold). However, this strategy can result in the exclusion of valuable data and reduced spatial coverage. A more effective solution is to apply temporal alignment through linear interpolation to standardize reporting intervals.

By linearly interpolating accumulations between irregular reports, the interpolated accumulations can be used to infer increments at regular time steps. The regular increments obtained can be expressed as weighted combinations of adjacent observed intervals. Under approximately consistent spacing assumptions, each interpolated value is a convex combination of surrounding observations, which ensures that the estimated intensities remain within the bounds set by the observed data. Consequently, precipitation intensities derived from interpolated increments generally underestimate true peak intensities, while preserving total event accumulation. Furthermore, event durations can be artificially inflated because the total accumulation is conserved while intensity is distributed across adjacent time intervals. Intuitively, a burst of rainfall from 2:25 to 3:25, when temporally aligned to 2:00 to 3:00 and 3:00 to 4:00 will preserve the total rainfall amount but spread out the intensity over two (2) hours versus the actual 1 hour of rainfall duration.

Because this study enforces a fixed one-hour reporting interval (regular interval), only the spacing of the irregular gage observations can be constrained. This restriction on irregular interval is hereafter defined as temporal-alignment-quality (TAQ) threshold. To balance data inclusion with temporal consistency, TAQ threshold of 1.5 hours is applied. This threshold was identified during hyperparameter tuning phase of optimization methodology described in **Section 3.3**. If the time

gap between the two observations used in interpolation exceeds this threshold, the interpolated value is flagged as TAQ-FALSE. Moreover, any precipitation increment derived using a TAQ-FALSE accumulation is also marked as TAQ-FALSE. This process is referred to as temporal alignment quality control (TAQC) for the remainder of this report. Through this method, gage data are standardized to consistent one-hour intervals while maintaining a balance between temporal consistency and data quality.

The theoretical basis for TAQ threshold was evaluated using the source gage data. The underlying mathematical constraints that govern the reduction in rainfall intensity are detailed in **Appendix 7:** Temporal Alignment Methods with Equations. The reduction in rainfall intensity was found to be bounded by both lag time, i.e., the difference between irregular reporting times of source data and the required regular reporting time and by the ratio of irregular to regular spacing. Finer irregular intervals with low lag were shown to better retain peak rainfall intensities. This analysis provides a quantitative justification for the TAQ threshold selected in this study.

2.4.3 Radar-based quality control (RBQC)

The temporally-aligned gage reported precipitation values are synchronized with the hourly radar increments, enabling direct comparison between radar estimates and gage observations. For each gage and each hourly time step, radar-derived rainfall at the pixel corresponding to the gage location, is compared to the gage-reported increment. These comparisons reveal three primary quality issues.

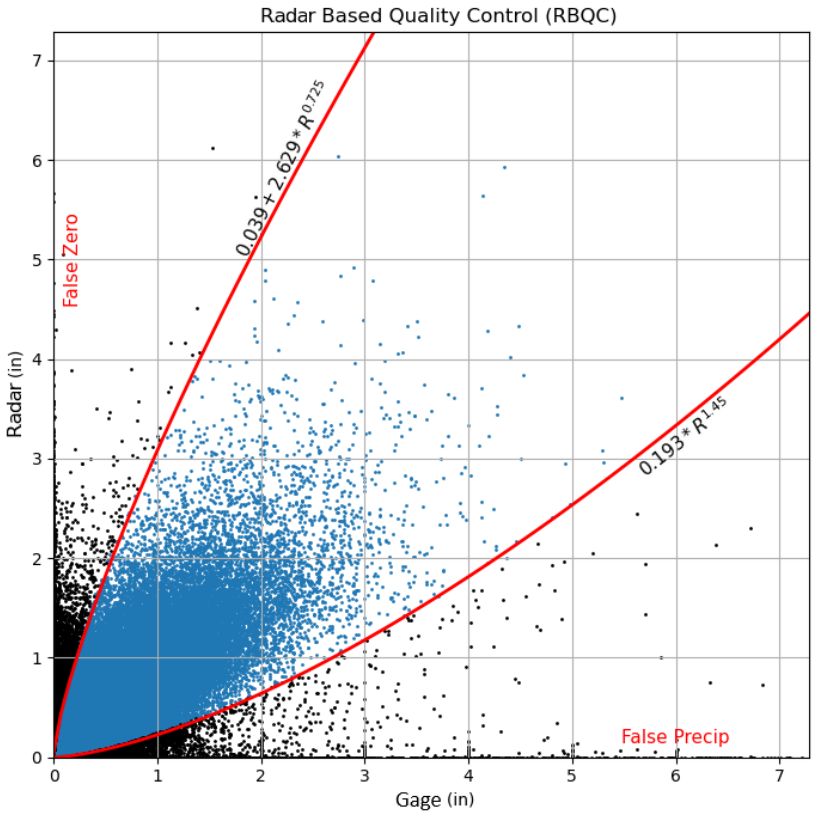


Figure 3: *Radar based quality control, including pass (blue) and fail (black) gage readings*

First, radar detects rainfall while the gage reports zero. This discrepancy may result from gage malfunctions, limited measurement resolution (i.e., rainfall amounts below gage precision), or the presence of frozen precipitation that is not immediately recorded.

Second, the opposite condition may occur where the gage registers rainfall while radar shows none. This typically occurs when previously frozen precipitation melts and is recorded by the gage but is not captured by radar.

Third, additional errors may arise from obstructions, wind effects, or siting issues that bias gage measurements. To mitigate these discrepancies, gage observations that deviate significantly from corresponding radar estimates are excluded using dynamic bounds as shown in **Figure 3** defined as $0.193R^{1.45}$ *to* $0.039 + 2.629R^{0.725}$, where R is radar-estimated rainfall in inches. Although this filtering removes approximately 4.2% of valid data, it successfully eliminates 74% of erroneous values. This radar-based quality control (RBQC) strategy closely follows the methodology used in the MRMS_GageCorr Product (Zhang et al., 2016).

After applying three layers of quality control correction (accumulation resets from AQC, temporal alignment from TAQC, and outlier-based filtering from RBQC), unusable gages are identified and dynamically filtered out. Moreover, because raw gage data availability and the quality control process are dynamic (time-dependent), the number of gages available for bias correction varies at each time step. **Figure 4** illustrates the number of quality-controlled gages (after AQC, TAQC, and RBQC) available within the AOI over the time-period as well as the percentage lost to RBQC as monthly averages.

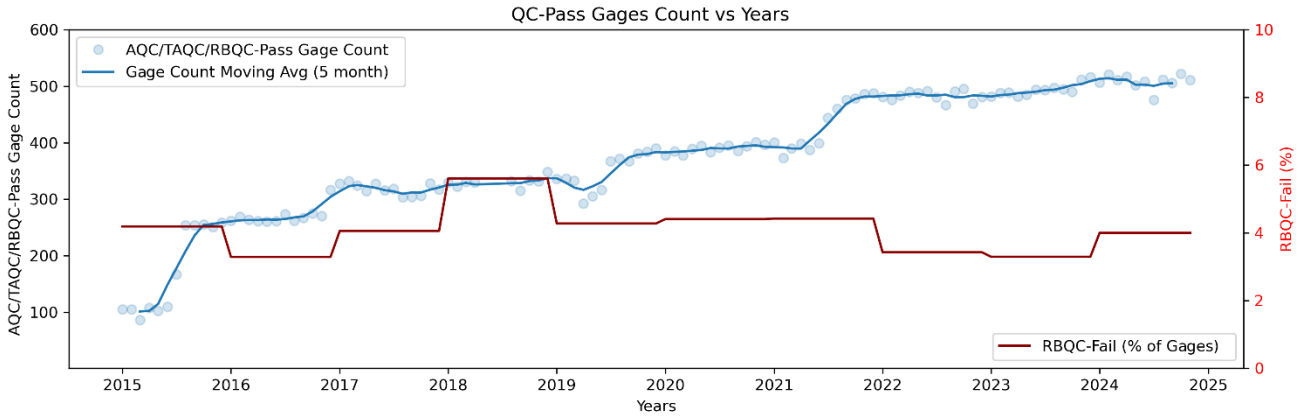


Figure 4: *Effect of the QC processes on gage count over time.*

2.5 Data Partitioning for Validation

Traditionally, bias-correction validity has been assessed by comparing errors between radar and gage measurements at gage locations reserved for testing (Kitzmiller et al., 2011; Li et al., 2023; Wootten & Boyles, 2014; Zhang et al., 2016). However, this approach has some caveats. Excluding certain gages for validation may omit critical information that cannot be reasonably inferred from surrounding gages. During most rainfall events, only a small fraction of gages record non-trivial rainfall, which means that gages excluded for validation often yield trivial comparisons (i.e., 0 vs. 0 rainfall). Furthermore, bias is introduced in the validation scheme based on the specifics of the gages chosen for testing.

Cross-validation techniques, which partition the available gages into subsets using all but one for calibration and the remaining one for validation, can help reduce bias but still suffer from inherent limitations. Although smaller subset sizes, such as leave-one-out cross-validation, further reduce bias, they are computationally expensive to implement over extended periods or large domains (Chen et al., 2019; Mapiam et al., 2022; Ware, 2005). Cross-validation is also affected by scaling issues, with differences in error measures observed as sample size varies (Ghasemzadeh et al., 2024). As gage density itself increases, the inclusion of predominantly trivial rainfall measurements can artificially reduce error metrics, even without genuine improvements in system performance. Consequently, direct comparisons between events or regions with different gage densities become unreliable. This limits such validation to specific event-based case-studies. Furthermore, it is computationally expensive as each *k*-fold requires approximately *k* times the processing effort, making it impractical for operational or low-latency applications.

To overcome these limitations, a novel gage selection algorithm was formulated that prioritizes gages according to event significance and spatial representativeness. Using this validation strategy, different interpolation algorithms for radar-gage bias correction were evaluated, identifying methods most suitable for the coastal Texas region. Furthermore, this approach makes it possible to evaluate the performance of the gage selection algorithm across the entire data record. By integrating event-relevant and spatially balanced gages, the validation framework ensures that uncertainty remains quantifiable without sacrificing critical information.

2.5.1 Classification methodology for validation dataset

The information provided by each gage can be regarded as a measure of how much it contributes beyond what neighboring gages already capture. In other words, a gage's importance depends not only on its own data, but also on its neighbors. To make use of this notion, the gages that pass the quality controls are further divided into three distinct categories based on the spatial information contained within them as critical, trivial, and information-rich non-critical (IRNC) gages, as defined below:

Critical Gages: Some gages hold unique information in relation to radar. Gages located in gage-scarce regions, where a single gage is the sole provider of discrepancy information, are critical for correcting radar grids. In gage-dense regions, if there is a large spatial variance in the discrepancy between radar and gage readings, the gage information becomes critical. In these cases, no other gages nearby can substitute for the critical one. Removing such a gage would degrade any bias-correction algorithm's ability to model local biases. Thus, it is detrimental to any analytical result to remove a critical gage which is defined here as the critical gage category.

Trivial Gages: Conversely, gages that provide little to no unique information can be defined as trivial gages. When there is no event ongoing over some gages, they report 0 inches. Zero values are trivial when evaluating any algorithm. Other trivial gages can be found in densely gaged areas, where most gage readings information matches with radar values, i.e., there is no discrepancy between gage and radar data for large swaths of areas. Even when discrepancies do exist

between gage and radar, but those discrepancies are consistent across many gages, only a few gages may be sufficient to correct the radar grids. In such situations, additional gages may provide trivial information; meaning that only a small number of gages are necessary to adjust the precipitation estimates. Since it is the nature of trivial information to be very easy to estimate, using these gages as input for any performance analytics inflates the metric, i.e., it yields a better than true performance of the model.

IRNC Gages: IRNC gages occupy the middle ground – neither trivial nor critical – providing meaningful information to be used in analytics without materially affecting the calibration product if withheld.

Critical gages must be retained in calibration, as their removal would degrade the quality of the final product. Trivial gages must be excluded from validation analytics, as their inclusion inflates performance metrics. To facilitate information-based classification, the spatial informativeness of each gage needs to be quantitatively assessed. The composite metric, defined in **Equation xxxix** of **Appendix 8:** Normalized Composite Information Index Definition, is designed to capture both linear and non-linear spatial variations by combining a simplified average difference index with modified forms of Local Moran's I and Geary's C (Anselin, 1995, 2019).

While the individual components of the index are well established in literature, their combination into a composite index is an original contribution of this study, designed to address the unique challenge posed by spatial rainfall distributions that exhibit both linear and non-linear variability. The combination of direct linear association measure from average difference, semi-linear association of Moran's I and non-linear association from Geary's C with no one factor having larger weight than the other ensures that this index captures both linear and non-linear variability in spatial rainfall distributions. The resulting information index *(I)* is normalized to range from 0 to 1 and provides a measure of the relative spatial uniqueness of each gage point.

2.5.2 Uncertainty analysis of information-based data partitioning

To identify thresholds delineating trivial and critical information zones, the empirical distribution of the information index was analyzed for the period of August 1st to September 30th, 2017, which includes the event Hurricane Harvey (2017), along with several smaller events. Several distinct paths can be identified corresponding to different events. The green curve in **Figure 5** illustrates this distribution. A distinct inflection point appears near the 90th percentile, corresponding to an information index of approximately 0.25. Beyond this threshold, a rapid increase in the index suggests the onset of critical information content.

Determining the lower threshold proved more challenging and involved iterative hyperparameter tuning and performance evaluation across multiple events. This is because, many gages remain trivial even during an active event due to limited spatial extent of most

storms. Ultimately, the lower threshold was set at 0.05, which aligns with roughly the *36th* percentile of the data.

The thresholds correspond to different percentile ranges at each timestep, and the prior analysis only represents average percentile correlation. For instance, during large, spatially extensive events, the lower threshold often aligns with a low percentile, activating nearly all gages for validation. In contrast, for smaller, locally concentrated storms, the same threshold corresponds to a higher percentile to help filter out many less relevant gages. Thus, gage selection for calibration-validation split is dynamic across events of varying spatial extent.

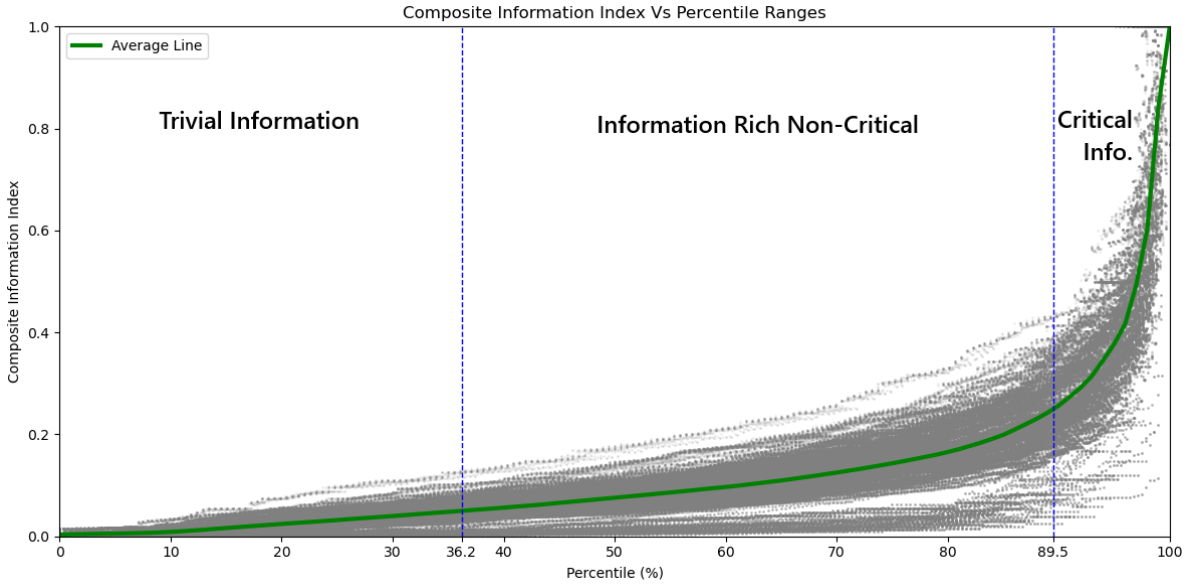


Figure 5: *Composite information index-based classification*

On average, this framework retains approximately 53.3% of the data (the IRNC data between the blue lines of 36.3% and 89.5% as shown in Figure 5) as suitable for validation. Probabilistic analysis revealed that around 4.17% of data classified as IRNC can still be trivial and 1.69 % of data classified as IRNC can be critical. While this introduces some instability in validation metric for short time-periods (5.86% of timesteps), metrics stabilize for events exceeding 5 hours (i.e. no rapid change in performance values is observed). Thus, analysis carried out over a long period of time the approach is found sufficiently robust.

The framework is designed to be adapted to different regions, with thresholds reselected based on the local gage distribution. The exact information index thresholds from this study should not be directly transferred. Instead, the information index values corresponding to the percentile values (36.2%, 89.5%) from the distribution of gages in the target region should be determined to establish initial thresholds, which can then be further optimized for the local context.

Importantly, this optimization is not performed with the objective of maximizing performance indices as such an approach would constitute overfitting, but rather as a form of hyperparameter

tuning to ensure that the performance indices remain robust, i.e., not overly sensitive to variations in random seed selection for validation gages. It was noted that IRNC gages identification for validation was paramount in stabilizing the Bayesian optimization detailed in subsequent sections.

2.5.3 Data classification for calibration and validation

Building on the information index percentile framework, observations of gage readings indicate that index values greater than 0.25 (typically around the 90^{th} percentile) correspond to critical information, while values below 0.05 are generally considered trivial. These thresholds were not predetermined but rather refined through iterative evaluation of the dataset and algorithm performance across multiple events, with particular attention to how different gage selection strategies influenced calibration and validation outcomes.

The intermediate range is defined as the Information-Rich Non-Critical (IRNC) region. A gage point is then randomly selected from the IRNC region. If a selected gage is excluded from any calibration efforts, the IRNC region must be recalculated after removing the selected gage. This recursive process is repeated to select a total of 10 validation gages.

A fixed number of gages is chosen rather than a percentage to ensure that the resulting validation set is an unbiased estimator, independent of the number of available gages (which may vary over time). If the IRNC region is too small to permit further selection, typically in trivial cases such as zero-rainfall events, the gage with the median index is used instead. In such cases, no further recursion is needed, as the available information is inherently trivial. Thus, the quality-controlled data is partitioned into two sets: a calibration set, and a validation set, as illustrated in **Figure 6**.

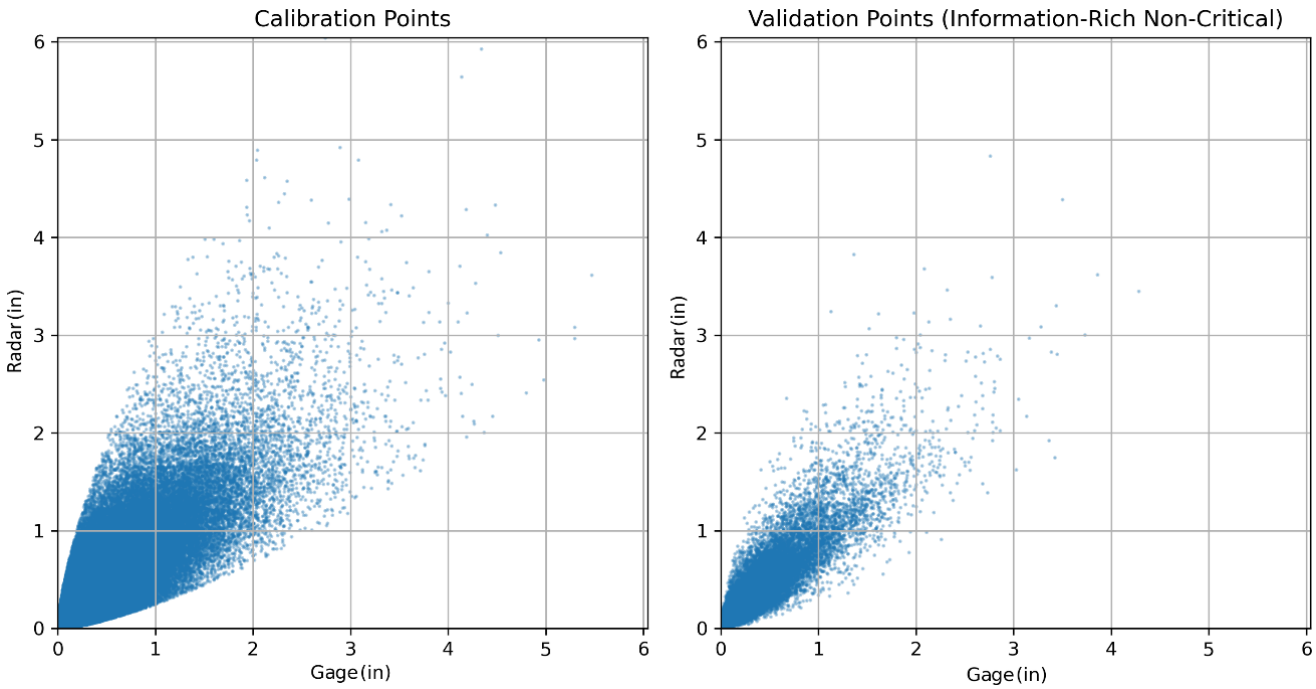


Figure 6: *Calibration and validation set*

The calibration set includes all the gage locations utilized in the bias-correction framework to produce a bias-corrected radar rainfall product. The validation set contains all the gage locations used for comparison with the bias-corrected product to evaluate performance metrics.

3 Methodology

This study aims to integrate weather radar rainfall data with the information from an expanded mix of public, private, and local rain-gage networks, safeguarded by a three-layer quality-control workflow (to detect counter resets, timing mismatches, and statistical outliers), and apply event-based radar bias correction anchored to trustworthy gages so totals and spatial patterns can better reflect what fell on the ground. To ensure stability in hydrologic models, the project team weighted gage influence using a smooth, continuous, and differentiable gage-density index that attenuates with distance, avoiding discontinuities. All workflow components were designed to be transparent and repeatable, and the project delivers modeling-ready products (GeoTIFFs, HEC-DSS, and HEC-HMS PCP) with concise user guidance. Performance is quantified across 2015–2024 events using independent gages and a modeling case study, with comparisons to radar-only and alternate gage-corrected gridded rainfall baselines.

3.1 Bias Correction Methods

Various methods have been developed over the years to fuse gridded and point-based precipitation data. These approaches typically assume that point observations (e.g., gage data) are free of error. The point data are then used either directly or represented as residuals or discrepancy values relative to the gridded data. These values are subsequently interpolated across the spatial domain using a range of algorithms. The resulting interpolated surface may serve as the final bias-corrected product or as an intermediate correction layer to adjust the original gridded dataset.

Common interpolation algorithms used to obtain gage-corrected rainfall are ordinary Kriging, regression Kriging, empirical Bayesian regression Kriging, multiquadratic interpolation, inverse distance and elevation weighting (IDEW), etc. (Abbas & Xuan, 2020; Li et al., 2023; Ware, 2005). Kriging is a family of interpolation methods in which linear combination of observation points is used for interpolation at every grid point, and the linear combination is dictated by the covariance between observation points. Multiquadratic interpolation uses radial basis functions to create a smooth, continuous and differentiable surface that balances fit and curvature while avoiding abrupt changes. On the other hand, inverse distance weighting (IDW) methods use some metric of distance between two points (usually Euclidean 2D distances) to weigh the influence of each observation point based on proximity to each grid point.

3.1.1 Existing bias-corrected gridded datasets

Two broadly-accepted gage-corrected rainfall products for Texas are local gage bias-corrected radar quantitative precipitation estimates (QPE) from Stage IV and Multi-Radar Multi-Sensor (MRMS) (Du, 2011; Gao et al., 2021; Habibi et al., 2021; Zhang et al., 2016). MRMS hourly product (MRMS_GageCorr) reports gage corrected data with an average latency of ~1.5 hours whereas Stage IV is updated monthly. MRMS uses the Hydrometeorological Automated Data System ([1]HADS) network of gages and is 0.01 x 0.01 degrees ~ 1km^2 in resolution. Stage IV is the output after the manual quality control on the Stage-III data by River Forecast Centers (RFCs) and is 4x4

km^2 in spatial resolution (Lin & Mitchell, 2005). Multisensor Precipitation Estimates (MPE) is a similar product to Stage-III published by National Centers for Environmental Prediction (NCEP) that undergoes additional quality control and is available as a national mosaic (Wootten & Boyles, 2014). Stage IV typically performs better than other datasets (Li et al., 2023), but its latency and coarser resolution limit its usefulness for flood warning and smaller watersheds. Users are therefore often left with either the less accurate Stage-III product, MPE, or the MRMS_GageCorr product.

MRMS uses an automated choice of Z-R relationships based on precipitation classification for its radar-only products, an improvement over NCEP Stage-II, which uses a single Z-R relationship per radar domain (Lin & Mitchell, 2005; Zhang et al., 2016). The MRMS bias-correction methodology is based on the approach proposed by Ware (2005), that corrects rainfall based on radar-gage difference errors (additive errors) calculated at gage stations corresponding to radar pixels, which are interpolated using an inverse-distance-weighted scheme. The interpolated values are then subtracted from radar pixels to obtain the bias-corrected product. MPE uses a similar method but with multiplicative errors. At a 1-hour interval, MRMS has been observed to have fewer biases than the MPE product (Habib et al., 2009; Habibi et al., 2021).

Both MRMS and MPE bias correction methods use direct gage-radar comparison errors, whether additive or multiplicative, requiring co-located gage observations, which are typically sparse relative to radar coverage. As a result, radar rainfall regions not collocated with gages depend on bias information from distant stations, whose influence weakens with distance. However, rainfall bias characteristics are often consistent across similar meteorological events, spatial domains, and times of day. It is uncommon for distinct rainfall regimes to occur simultaneously within a region. This makes global regression approaches, where radar is adjusted once across the region before interpolating residuals, viable for coherent domains like the Texas Coast. Such an approach is common in empirical Bayesian regression Kriging or regression Kriging methods (Krivoruchko & Gribov, 2019; Li et al., 2023; Wang et al., 2009; Yang & Ng, 2019). In these methods, regression is used to correct the radar globally to obtain a regression-adjusted radar rainfall, then this product is further compared with gages to obtain residuals (either additive or multiplicative) and interpolated using one of the many interpolation algorithms.

Each bias correction method (and thus product) has its own constraint. Global regression is not advisable over regions that are too large or meteorologically uncorrelated; for example, applying the same fit across simultaneous events in Texas and Florida would provide no meaningful bias information. In such cases, local discrepancies from direct gage-radar comparisons are more reliable. Thus, most national level products do not focus on global error corrections. However, along the Texas coast, where climate, storm types, and meteorological structures are generally coherent across space for a given time, global regression can be effective. Here, regression fitting can correct many systematic deviations, leaving only residuals to be interpolated. Residuals may be treated additively (with units) or multiplicatively (unitless). Additive interpolation can produce unrealistic negative rainfall or artifacts in areas without radar precipitation, whereas

multiplicative interpolation only scales rainfall, minimizing such artifacts. The drawback is that small discrepancies can yield large scaling factors (e.g., a 0.001 in radar vs. 0.002 in gage mismatch producing a 2× multiplier). Additionally, Kriging can introduce spurious extrapolations that introduce artifacts, while IDW is purely interpolative and may miss valuable extrapolation information. Ultimately, each method carries tradeoffs, and the most appropriate approach should be selected based on regional data itself.

[1]https://hads.ncep.noaa.gov/

3.2 Performance Metrics

Multiple metrics were used to evaluate bias correction performance:

- Coefficient of determination (R^2)
- Root mean square error (RMSE)
- Total Percent bias (Bias%)

To gain additional insights, the data was also stratified by rainfall intensity. Rainfall (pixel) values were categorized into three intensity classes based on percentile thresholds. These thresholds were adapted from commonly used percentile ranges in streamflow analysis and extended here to characterize rainfall intensity (Buda et al., 2022; Ley et al., 2016; Yilmaz et al., 2008).

- High intensity: Above 95th percentile
- Median intensity: 45th to 55th percentile
- Low intensity: Below the 30th percentile

Further evaluation included computing the median bias and estimating the slope of the bias in the mid-section range (between the 30th and 80th percentiles). While the mid-section slope does not represent overall goodness of fit, it provides insight into how biases vary across the distribution, indicating whether larger biases occur at higher rainfall intensities compared to lower intensities, or vice versa.

Each metric was calculated for both the raw (unprocessed) and processed radar rainfall datasets, yielding a total of four sets of performance evaluations. Additionally, for each set and each metric, a scaled percentage improvement was computed relative to the ideal (error-free) value of that metric, serving as a benchmark for expected performance improvements. The mathematical formulations for these metrics are provided below (**Equation v**):

$$R^2 = 1 - \frac{\sum\left(R_g - R_r\right)^2}{\sum\left(R_g - \bar{R}_g\right)^2} \qquad \text{v}$$

Coefficient of determination (R^2) defines the extent of variation of a dependent variable (R_r) that can be explained by independent variable (R_g). A higher value indicates larger goodness of fit of model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{n}(R_g - R_r)^2} \quad \text{vi}$$

Root mean square error (RMSE) gives the average magnitude of prediction errors without regards to direction (**Equation vi**). A lower value indicates better goodness of fit.

$$Percent\ Bias_{Total}\ (\%) = 100 \times \frac{\sum(R_r - R_g)}{\sum R_g} \quad \text{vii}$$

$$Percent\ Bias_{High\ Intensity}(\%) = 100 \times \frac{\sum(R_{r,0.95} - R_{g,0.95})}{\sum R_{g,0.95}} \quad \text{viii}$$

$$Percent\ Bias_{Median\ Intensity}(\%) = 100 \times \frac{\sum(R_{r,0.45-0.55} - R_{g,0.45-0.55})}{\sum R_{g,0.45-0.55}} \quad \text{ix}$$

$$Percent\ Bias_{Low\ Intensity}(\%) = 100 \times \frac{\sum(R_{r,0-0.3} - R_{g,0-0.3})}{\sum R_{g,0-0.3}} \quad \text{x}$$

$$Bias\ Midsection\ Slope(\%\ slope) = 100 \times \frac{[R_{r,0.8} - R_{r,0.3}] - [R_{g,0.8} - R_{g,0.3}]}{[R_{g,0.8} - R_{g,0.3}]} \quad \text{xi}$$

Where R_g is the rainfall increment for the gage location and R_r is the rainfall increment of the gridded product (either before or after processing) at corresponding locations.

Percent bias (**Equations vii**, **viii**, **ix**, **x**, and **xi**) describes the direction of discrepancy between two datasets i.e., it determines whether the model (R_r) tends to underpredict or overpredict relative to the reference dataset (R_g), calculated across percentile ranges ordered by R_g values. Previously, similar approach has been done for flow duration curves to analyze performance metrics at different ranges of data (Buda et al., 2022; Ley et al., 2016; Yilmaz et al., 2008). Bias midsection slope is added as an additional index which quantifies the rate and nature of changes in biases between low and high intensity values with large positive values indicating increasing bias at higher intensities.

The expected values of these metrics under ideal, error-free conditions, where radar estimates exactly match gage observations are as follows:

$$E(R^2) = 1$$
$$E(RMSE) = 0$$
$$E(Biases) = 0\%$$

Thus, percentage improvements are then estimated linearly by (**Equation xii)**:

$$Improvement\ (\%) = 100 \times \frac{Metric_{Processed\ radar} - Metric_{Raw\ radar}}{Expected\ Value_{Ideal} - Metric_{Raw\ radar}} \quad \text{xii}$$

Performance metrics were computed under the assumption that gage readings are infallible. While this assumption carries inherent limitations as previously discussed, it provides a practical basis for evaluating the performance of gridded precipitation data.

3.3 Spatial Interpolation Model Optimization (Regionalization)

Based on established literature and past related project experience, parameter optimization in spatial interpolation model was conducted by maximizing the percent improvement in the R^2 and RMSE of the validation set, while applying an equal-weighted penalty for computational runtime. Bayesian optimization using Gaussian processes was carried out using the scikit-optimize package (Pedregosa et al., 2011). The acquisition function was set to "*Expected Improvement*" with a noise level of 0.01. 3,000 iterations with 1,000 random initial samples.

The optimization was performed over August–September 2017, a period that includes Hurricane Harvey (Aug 25th to Sep 1st), and several smaller events before and after, as shown by AOI rainfall volume in **Figure 7**. The optimization process begins at the onset of rainfall and progresses step-by-step, encompassing the full range of intensities from very low (at the beginning and end) to very high (at peak hurricane intensity).

The information-rich nature of the validation dataset played a pivotal role in stabilizing the optimization algorithm by providing robust feedback on performance. The optimization framework considered a structured decision tree, which included the type of input values to be interpolated (direct gage values, gage–radar discrepancies, and regression residuals), the interpolation method (e.g., Kriging, Inverse Distance Weighting - IDW), and both static and dynamic parameter optimization at runtime (for Kriging) as shown in **Figure 8** (all methods tree) and **Figure 9** (regression residual interpolation branch).

Performance metrics at each time step are evaluated independently, and when combined with information-based validation gage selection, both extreme low and extreme high events are assessed with equal rigor when aggregated. This ensures coverage of the full spectrum of rainfall intensities, including an event which is among the most severe on record.

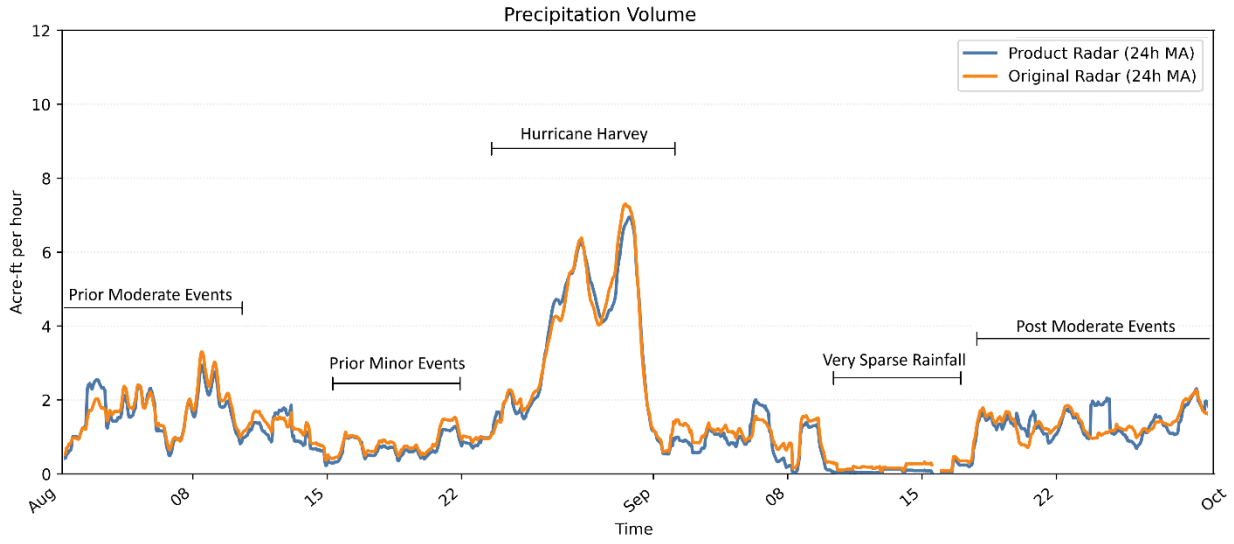


Figure 7: *Optimization period events (Moving average: precipitation volume (acre-ft) per hour)*

The optimization process converged on the following parameters from the tree (**Figure 9**):

- Using regression residuals
    - Linear (no-intercept) Regression: y=ax
    - Multiplicative residuals
    - IDW interpolation method
        - IDW Power = 2.969
        - IDW Max Distance = 1.954°
        - IDW Density Influence Factor = 0.091175

Another notable outcome was the identification of the best-performing Kriging configuration: ordinary Kriging with a spherical variogram model, applied to multiplicative discrepancy values. In this case, variogram parameters were optimized dynamically during runtime. While Kriging methods exhibited strong performance for small, localized events, several limitations emerged when applied at scale.

First, in large, spatially widespread events such as hurricanes, rainfall residuals can remain correlated over distances comparable to or exceeding the study area. This produces a strong spatial variability in distance (nearly flat semi-variogram without a clear sill), rendering parameter estimation unstable and undermining the basic assumption that local variance differs meaningfully from global variance. In such cases, Kriging tends to collapse toward the mean, reducing its effectiveness.

Furthermore, to the east, the coastal boundary introduces structural challenges since little to no gages exist offshore, thus requiring Kriging to extrapolate beyond observations, which increases uncertainty and leads to visible artifacts along the coast. The extrapolation effect is especially pronounced in elongated domains such as the AOI. The problem could be alleviated by incorporating gages much further away from the region of interest to the west of the AOI, though this comes with increased runtime overhead. Furthermore, the inherent computational overhead of ordinary Kriging is significant, especially when variogram fitting must be repeated dynamically for thousands of time steps.

IDW interpolation is computationally efficient and well suited to long-term, automated applications and outperformed Kriging consistently. Kriging also introduced occasional artifacts due to its extrapolative nature, whereas IDW, being limited to interpolation within the observation network, produced smoother and more stable grids. The project team selected IDW as the more reliable and efficient choice for long-term, automated gage correction.

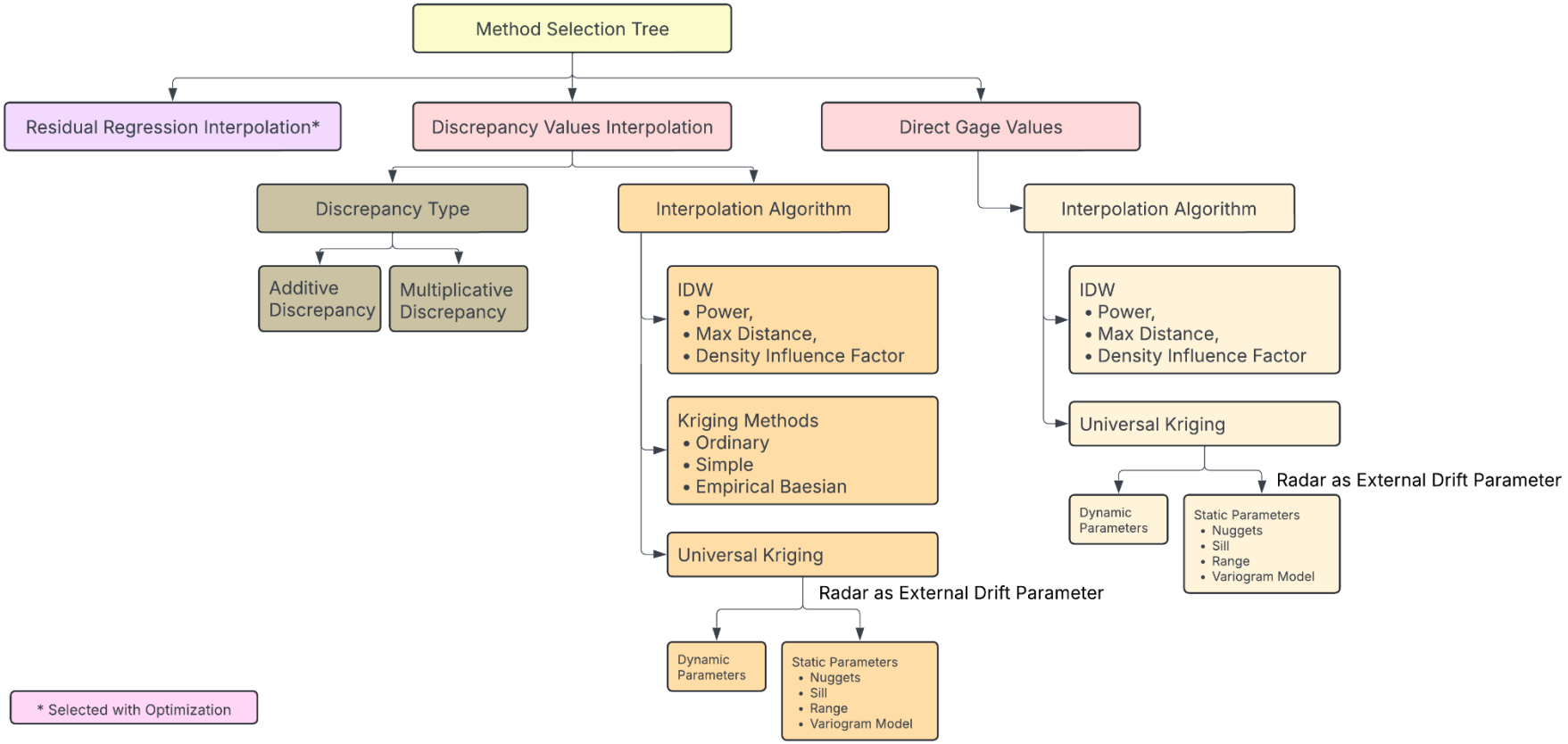


Figure 8: *Methodology selection tree for hyperparameter optimization (all methods considered)*

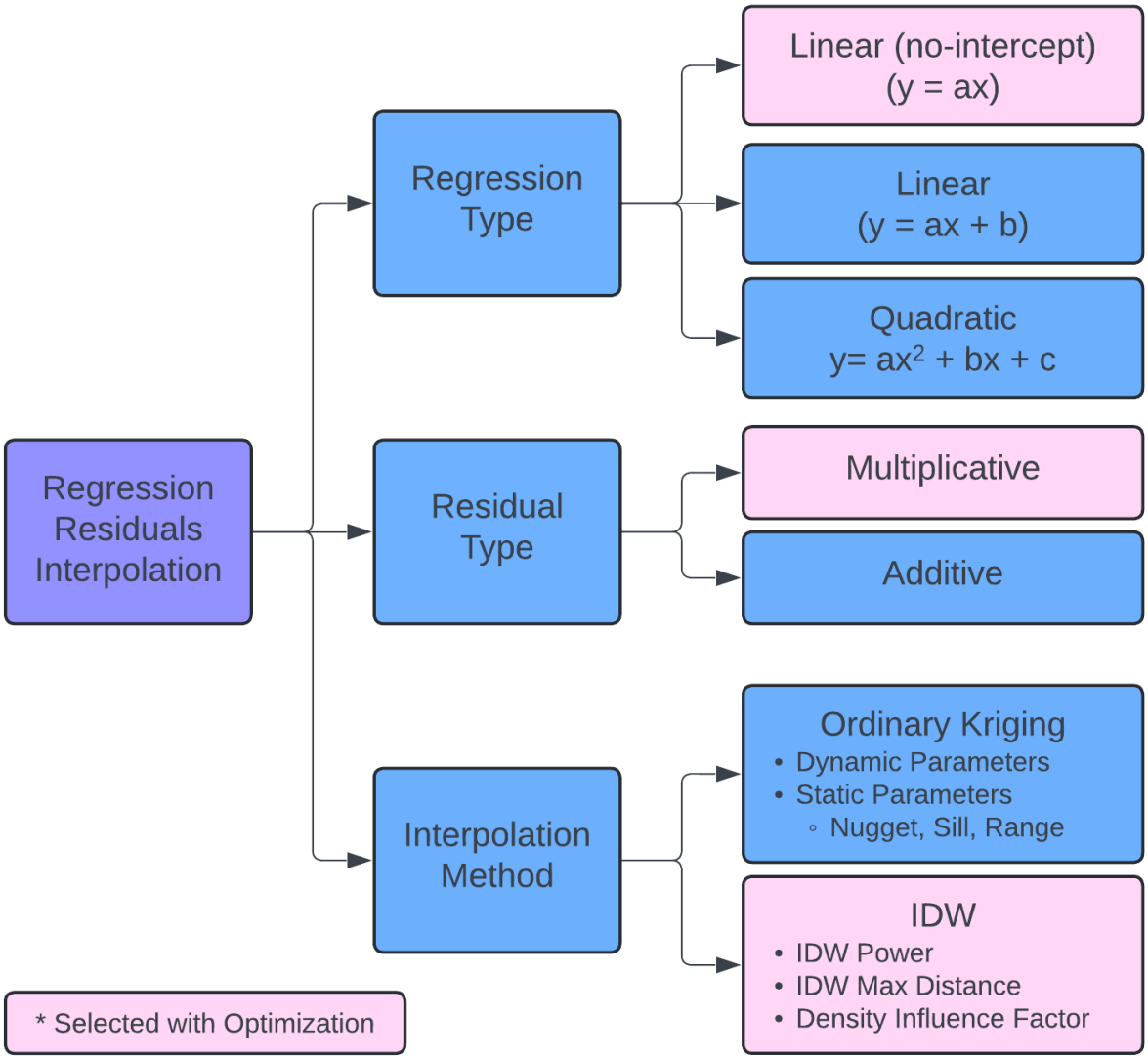


Figure 9: *Methodology selection tree for hyperparameter optimization (regression residuals interpolation method)*

3.4 Regression-Residual Density-Attenuated Inverse Distance Weighting (RAIDW)

Multiple bias correction methods were initially explored during this study, but most were subsequently excluded in favor of the optimized approach described below.

The chosen framework, Regression Residual Density-Attenuated Inverse Distance Weighting (RAIDW), was found to balance volumetric accuracy, spatial representation, and computational efficiency. The methodology includes enhanced validation strategies that account for both event-scale significance and spatial coverage, ensuring that bias corrections are informed by gages most representative of storm dynamics rather than trivial outliers. This dynamic validation approach strengthens the robustness of rainfall estimates by adjusting to the characteristics of each event.

This method forms the basis for all subsequent analyses and results presented in this report. For an overview of alternative methodologies evaluated but not adopted, readers are referred to: Foehn et al., 2018; Krivoruchko & Gribov, 2019; Ware, 2005.

For each time step, the raw radar data was processed into GeoTIFF (*.tif*) format in inches, clipped using the buffered AOI as shown in **Figure 1**, and reprojected to EPSG:4326 where necessary, utilizing cubic resampling. No-data values were replaced with zeros. Simultaneously, gage measurements and their spatial coordinates from the calibration set were extracted and reprojected to the same coordinate system (EPSG:4326), with snapping applied to (0, 0) for consistency. For each gage observation (*y*), the corresponding radar pixel value (*x*) was retrieved. A linear regression of the form $y = ax$, was then fitted using the ordinary least squares (OLS) method to determine the scaling coefficient (a). This regression model was applied across the entire radar raster to generate a regression-adjusted raster (Regression Raster). At each gage location, multiplicative residuals were computed (**Equation xiii**):

$$Residuals_i = \frac{Gauge\ Value_i}{Regression\ Raster_i} \quad \text{xiii}$$

These residuals are unitless, stored in a shapefile, and subsequently interpolated into a continuous residual grid using a modified Inverse Distance Weighting (IDW) technique.

For each grid cell, a neighborhood is defined as a circular buffer of radius $R = 1.954°$ which includes all residuals within this maximum distance. Let the set of residuals in the neighborhood be indexed by n, and the grid cell of interest by i. The value of residual i is denoted by R_i and the values of its neighboring point residuals are denoted by R_n. The distance between gage i and each neighboring gage is given by $D_{i,n}$. The total number of point residuals in the neighborhood is N. The interpolated value at the location of each grid center denoted by R_i is computed using the following weighted average (**Equation xv**): *(also see **Section 3.3** for hyperparameter values obtained from optimization, IDW Power = 2.969, IDW Max Distance = 1.954°)*

$$w_{i,n} = \frac{\frac{1}{(D_{i,n} + \varepsilon)^{2.969}}}{\sum_{n=1}^{N}\left(\frac{1}{(D_{i,n} + \varepsilon)^{2.969}}\right)} \quad \text{xiv}$$

$$R_i = \sum_{n} R_n w_{i,n} \quad \text{xv}$$

In trivial solutions R_i is set to 1. ε is a small positive value.

A density-based attenuation is introduced to ensure that the impact of interpolated residuals diminishes with increasing distance from the nearest gage points, preventing regions lacking nearby gages from being overly influenced by distant observations. The density term is derived from a Gaussian kernel applied to all gage distances at all center points of the interpolation grid (**Equation xvi**): *(See **Section 3.3** for hyperparameter values obtained from optimization, IDW Density Influence Factor = 0.091175)*

$$Decay\ Factor = \sum_{All\ Points} e^{-\frac{(D_{i,n})^2}{2\sigma^2}}\ , where\ \sigma = std.deviation\ of\ all\ Distances\ (D)$$

$$Density = 1 - e^{-0.091175 \times Decay\ Factor} \quad \text{xvi}$$

Equation xvi, is a novel density function providing a continuous, differentiable and tunable surface ranging from 0 to 1; representing the relative concentration of gage points across the domain.

An attenuation factor *K* is defined at each grid point to scale the influence of residuals based on proximity to the nearest gage (**Equation xvii**):

$$Attenuation\ Factor\ (K_i) = 1 - \left[\frac{min(D_{i,n})}{max(D_{i,n})}\right]^{\frac{1}{Density}} \quad \text{xvii}$$

Here, i represents each grid point and n represents all gage points.

This factor is equal to 1 at the location of the nearest gage and decreases with distance from that gage. The rate of decay is modulated by the local density: in regions of high gage density, the attenuation decreases gradually with distance, while in regions of low density, the attenuation falls off more rapidly. This ensures that in sparsely gaged areas, the influence of distant residuals is appropriately limited.

The final residual raster is then obtained by multiplying the interpolated residual values with the attenuation factor normalized around residual = 1, i.e., *Residual Raster = [1 + (R_i – 1) × K_i].*

A similar exponential attenuation factor α, **Equation xviii**,

$$\alpha = \begin{cases} \sum e^{-\frac{(D_{i,n})^2}{(\frac{D}{2})^2}}, & \alpha < 1 \\ 1, & \alpha \geq 1 \end{cases} \quad \text{xviii}$$

was used previously in the National Mosaic and Multi-Sensor QPE (NMQ) system (Zhang et al., 2011). The new attenuation factor *K* is preferred here due to its continuity, boundedness, and tunability via the inverse decay influence parameter (0.091175), which allows stable and flexible optimization and reduces spatial inconsistencies and overfitting in the final interpolated product.

The final bias-corrected radar rainfall product is obtained by multiplying the normalized residual raster with the Regression Raster: *Bias Corrected Product = Regression Raster × Residual Raster*. The methodology will be referred to as Regression Attenuated IDW (RAIDW) throughout the remainder of this report and is detailed in **Figure 10**. The validation set is then used to compute performance metrics for the final processed radar rainfall product.

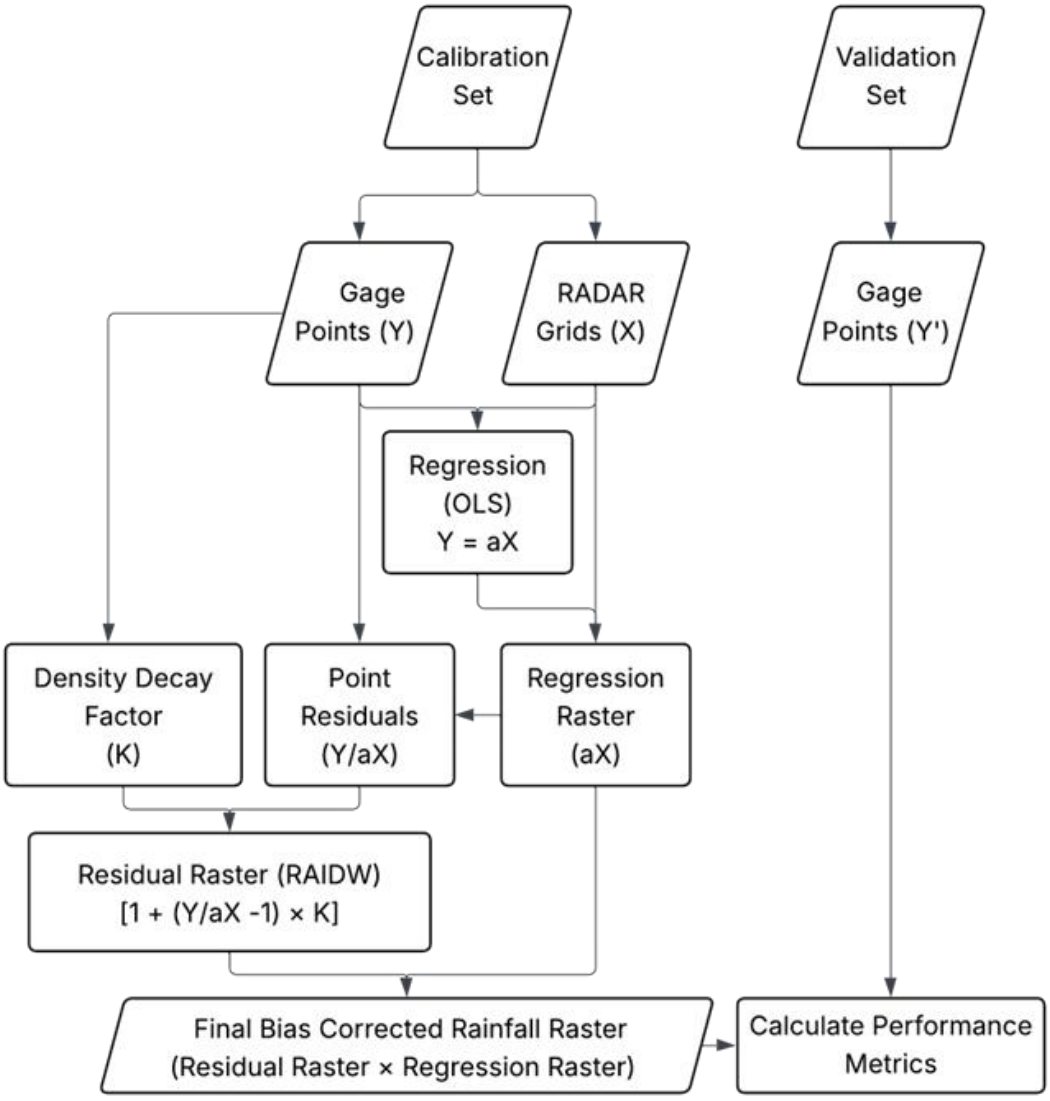


Figure 10: *RAIDW bias correction methodology flowchart*

3.5 Expected Density-Driven Improvements

Gage distribution was examined across varying gage networks by using the novel gage density function, a continuous, differentiable surface that characterizes the concentration of gages within the study area. Defined using a Gaussian kernel function and optimized for the region, this index provides an intuitive representation of spatial distributions of gages. For more details on the formulation of gage density, see **Section 3.4**.

3.5.1 HADS gage network

The distribution of gages for the year 2024 as obtained from exclusively the HADS network, following quality control defined in this study is shown in **Figure 11**.

3.5.2 WERC gage network

The distribution of gages for the year 2024 as obtained from Synoptic Data (rain gage data vendor for TWDB), following quality control as defined in this study (see **Section 2.4**), resulting in a selection of good-quality gages is referred to as WERC gages and is shown in **Figure 12**.

3.5.3 Comparative analysis of gage density differences between networks

Figure 13 illustrates the difference in gage coverage between the two networks. The additional gages were identified by applying strict quality control on the more numerous Synoptic network, resulting in a denser and more representative dataset. Examination of these differences motivates an important hypothesis:

Areas with higher gage density will exhibit greater improvements in radar bias-corrected rainfall estimates.

To demonstrate that greater gage density or greater observational support exerts a critical influence on radar bias correction and refined rainfall estimation, a strong correspondence should be observed between the gage density difference map and the spatial differences between this study's enhanced product and existing products (MRMS). Consequently, if this correspondence is confirmed, then it would indicate that the enhanced product provides a substantive improvement over existing alternatives. This hypothesis is examined and demonstrably verified in **Section 4.1**.

To evaluate this hypothesis, the spatial differences between the enhanced radar product and the MRMS_GageCorr dataset were later compared against the gage-density difference map. The strength of correspondence, quantified using normalized cross-correlation (NCC) and the Structural Similarity Index (SSIM) serves as a diagnostic of density-driven improvements.

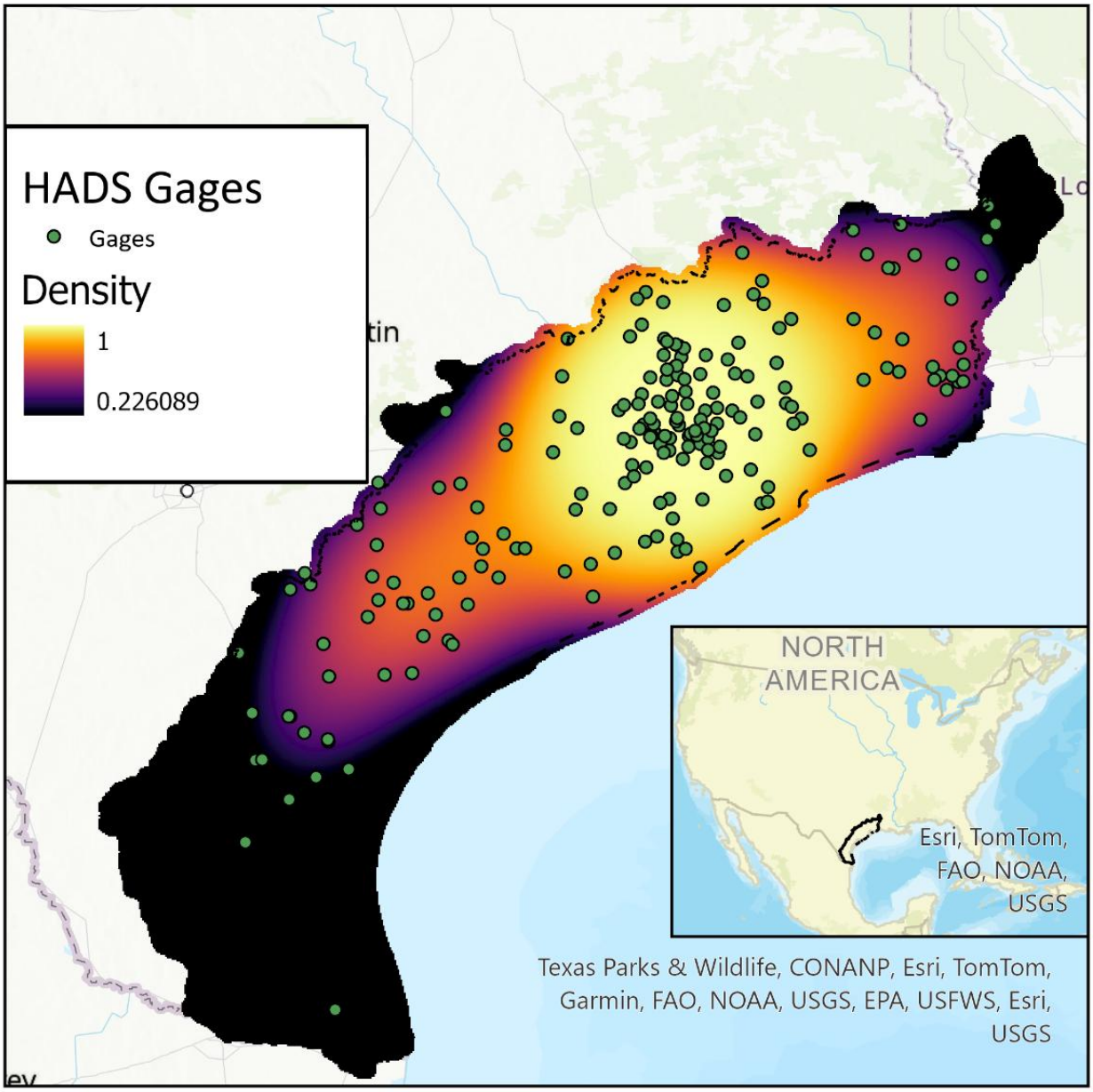


Figure 11: *HADS gages (after QC by WERC) with corresponding density*

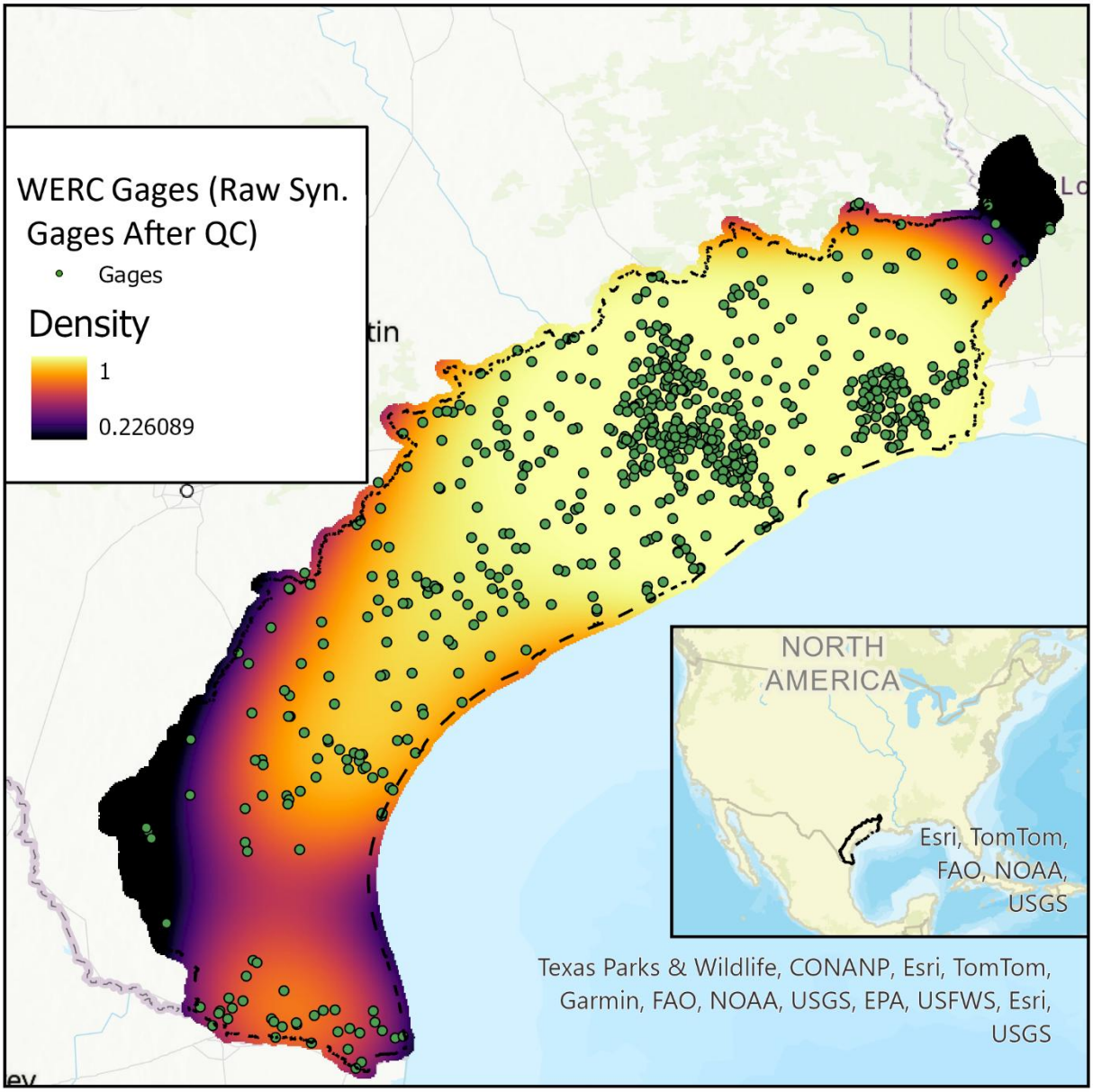


Figure 12: *WERC gages (raw Synoptic gage source- after QC by WERC) with corresponding density*

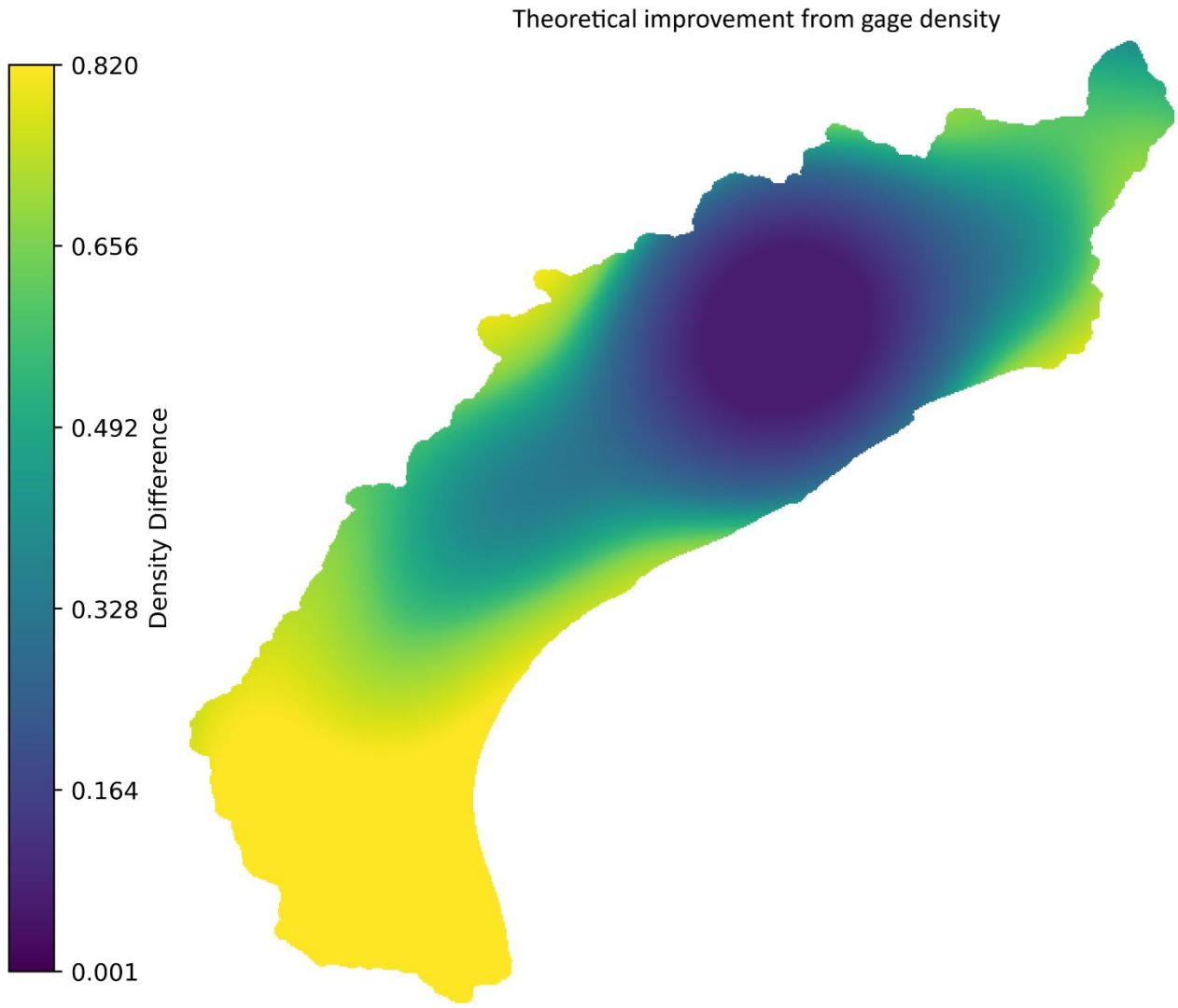


Figure 13: *Density difference between the HADS and WERC gage networks*

4 Implementation and Performance Analysis

The methodology outlined in this report was applied across a ten-year period (2015–2024) to evaluate the performance, spatial integrity, and hydrologic suitability of the enhanced radar precipitation product. The evaluation encompassed multiple complementary analyses, including comparison with the MRMS_GageCorr dataset, spatial assessment of rainfall patterns for representative events such as Hurricane Harvey (2017), and statistical performance evaluations for calibration, validation, and event-based datasets. Annual summaries with interannual trends were generated and uncertainty analyses were conducted to establish confidence bounds for key metrics such as bias, RMSE, and R^2. Temporal relationships between gage density and product performance were also investigated. Finally, independent verification using clustering analysis and hydrologic modeling with a pre-calibrated HEC-HMS model was also performed to assess the product's reliability for watershed-scale applications.

4.1 Evaluation of Product Improvements Relative to Gage Density

The product developed in this study was evaluated against the MRMS_GageCorr product, a similar 1-kilometer resolution dataset that primarily relies on the HADS gage network. RMSE between the two products was computed for the period 2015–2018, considering only cells corresponding to event durations and excluding trivial no-rainfall periods.

The RMSE was then spatially normalized to a 0–1 range, with the resulting distribution shown in **Figure 14**. This result was evaluated against the theoretical improvement expected from gage density (**Figure 13**), which reveals a low normalized cross-correlation (NCC) of 0.1208 at the pixel level (Lewis, 1995). Such a low correlation is expected, as differences between the two products arise from more than just gage density, including algorithmic approaches, spatial and temporal variability of events during the evaluation period, and the inherent differences in the spatial distribution of the density function.

The Structural Similarity Index Measure (SSIM) was also applied (Wang et al., 2004) to verify whether gage density influences rainfall product correction. An SSIM of 0.5015, indicates a moderate degree of pattern similarity, supporting the hypothesis that gage density plays an important role in improving gage-corrected rainfall products.

Overall, the comparison confirms that while gage density alone does not fully explain the spatial variability in product performance, it exerts a measurable and meaningful influence on radar bias correction accuracy. The moderate structural similarity observed suggests that areas of higher observational support correspond to regions of greater improvement, validating the theoretical expectation established in **Section 3.5**. These findings indicate that denser and better-quality gage networks, such as the WERC configuration, contribute significantly to the refinement of radar-based precipitation products, even when other algorithmic and environmental factors are considered.

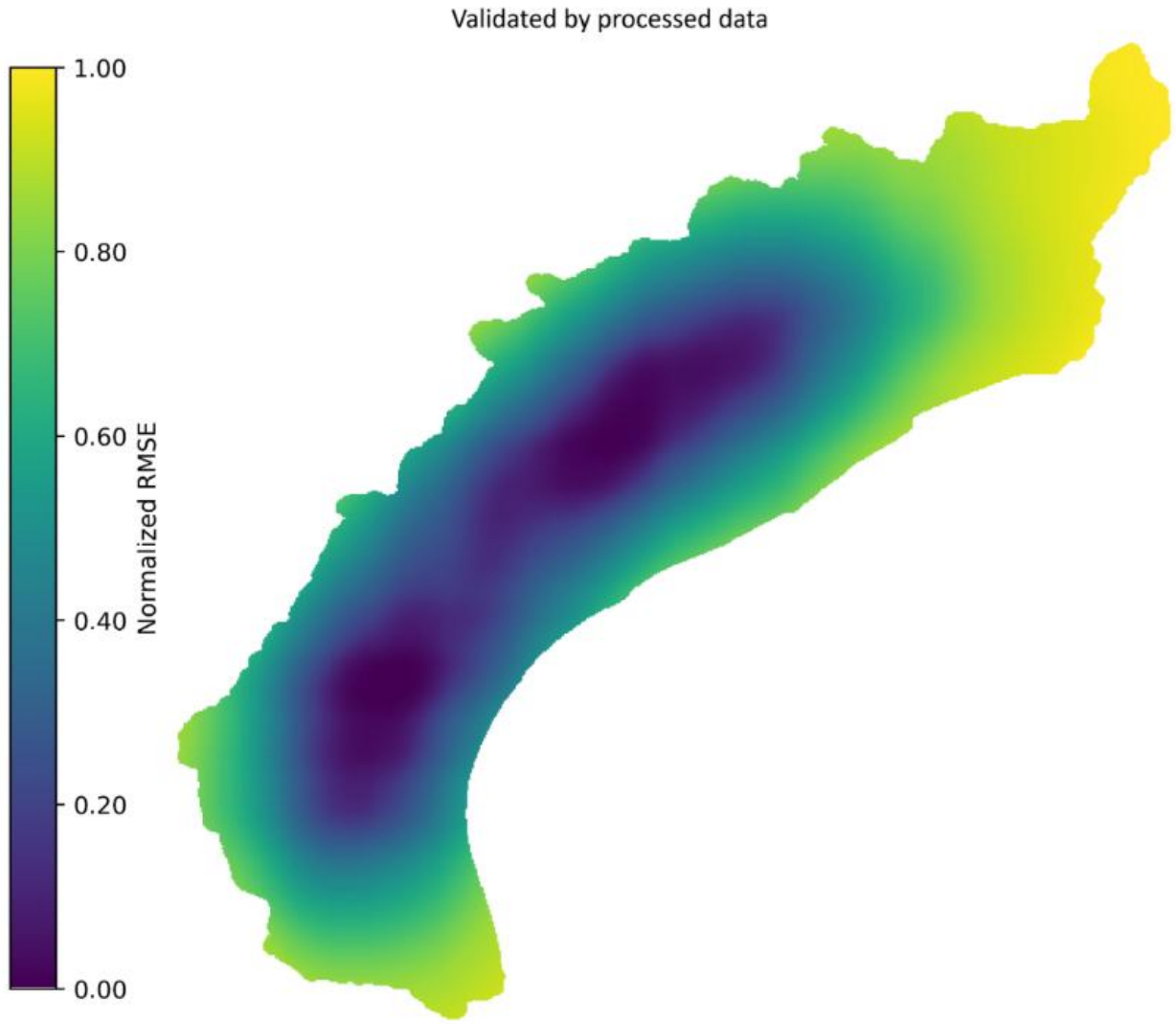


Figure 14: *RMSE differences (normalized to 0-1) between MRMS_GageCorr product and WERC product*

4.2 Spatial Variability

To assess the spatial consistency of the proposed product precipitation accumulation maps were examined. **Figure 15** compares the original radar data (left) and the enhanced product (right) for Hurricane Harvey (August 25th to September 1st, 2017), illustrating few expected radial dips in accumulated intensity, likely associated with beam blockage. Localized noise near radar stations, reflecting minimal ground clutter is limited in extent.

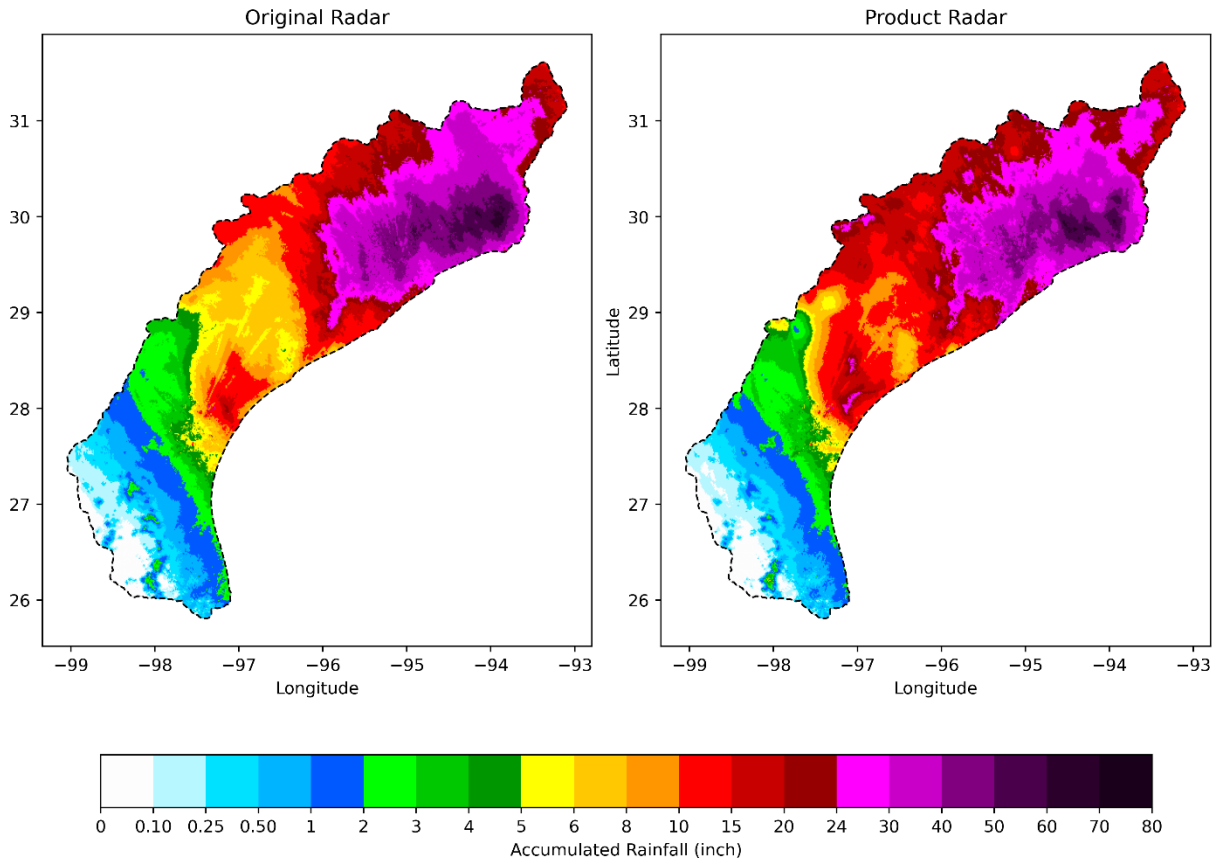


Figure 15: *Spatial comparison of gridded product (accumulation from Aug 25 to Sep 1, 2017)*

Figure 16 shows RMSE differences between the original radar and enhanced radar product. Increased spatial detail and localized variability become visible as textured or cloudy patterns on the product due to gage-informed alignment in gage-sparse areas. These patterns are expected as gage correction has more impact in areas with higher gage density. Locations farther from the radar (for example, at latitude 29.0 and longitude -96.5) also exhibit greater deviation consistent with well-documented decreases in radar data quality with distance from the station (Chumchean et al., 2003; Joss & Waldvogel, 1990; Steven, 1996).

Figure 16 also shows that, the data may exhibit both positive and negative biases across different areas within the same event. Boundary regions between the positive and negative bias zones appear as white lines in **Figure 16** representing locations where radar and gage values coincide by chance rather than by true agreement.

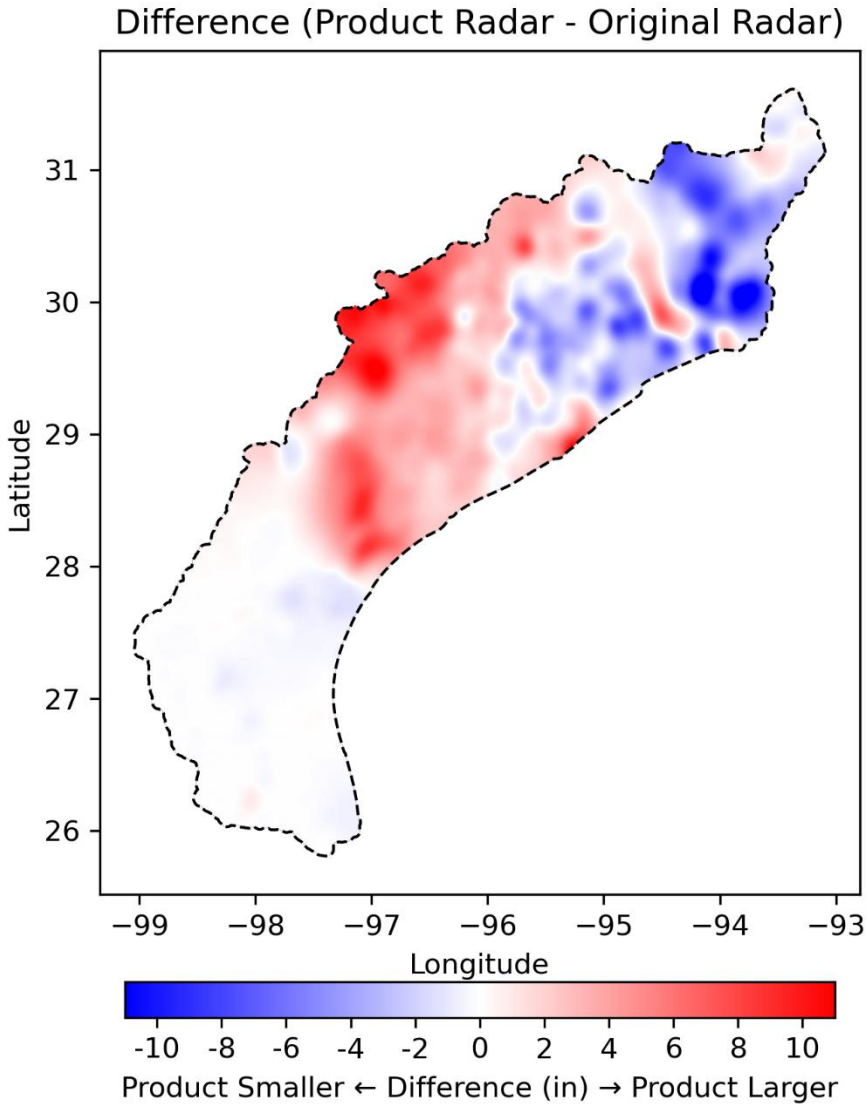


Figure 16: *Difference between product and original radar (based on total accumulation over Hurricane Harvey event)*

4.3 Product Performance Statistics

As stated in **Section 3.2**, rainfall values were stratified into low (<30th percentile), median (45th–55th percentile), and high (>95th percentile) intensity classes. Performance metrics across all three intensity classes were evaluated for all datasets (validation, calibration, and Harvey-validation) and are shown in subsequent figures and tables. Results are based on temporally-adjusted one-hour intervals for each unique location paired with the corresponding location's original or product radar value.

Figure 17, **Figure 18**, and **Figure 19** present performance metrics stratified by rainfall intensity. Gage readings across the entire coast are classified into percentile ranges and subtracted from 100 to obtain pseudo exceedance probability. The deviations of radar from each gage is shown as a line from the gage. With the line length representing the magnitude of the discrepancy. Due to the volume of points, the collection of lines forms a band that highlights large discrepancies more clearly than alternative visualizations with even a small number of outliers at different exceedance probabilities appearing as visible extensions of the band. By contrast, a confidence band would appear much narrower and smoother, while scatter plots tend to look sparse at high discrepancy values (see **Section 4.4** for confidence bands and scatter plots). These deviations, along with the points in the scatter plot, represent comparisons between gage and radar at individual one-hour time steps, such that each point corresponds to an hourly intensity value for a given gage across the full data record.

The methodology forces product radar estimates to closely match calibration gage values aside from minor noise from pixel-level gage clustering or rounding artifacts. Thus, high calibration fitness is expected, making the validation statistics the primary basis for performance assessment. Validation metrics across the full data range show an average improvement of 67.2%, summarized in **Table 2** and visualized in **Figure 17**. Calibration metrics across the full data range show an average improvement of 94.7% as summarized in **Table 3** and **Figure 18**. Event-based evaluation over Hurricane Harvey event (25th August to 1st September) shows an average improvement of 73.1%, as shown in **Table 4** and **Figure 19**. Performance metrics were also computed annually and are summarized in **Table 5**.

The mid-section slope and the median bias do not match, as expected, since the two metrics describe different aspects of product performance. The mid-section slope reflects how biases are distributed across the central portion of the rainfall intensity spectrum and is not intended for direct comparison between products, whereas bias values measure systematic deviation in magnitude. Biases remain relatively consistent across different rainfall intensities, which naturally produces a lower mid-section slope. The median bias often appears larger because it represents a range neither dominated by very small values, where trivial measurements can offset discrepancies, nor by very large values, which radar tends to estimate more accurately.

Table 2: *Validation set metrics (all data)*

Validation Dataset (All Data)			
Performance Metrics	**Original Radar**	**Product Radar**	**Percent Improvement**
Percent Bias (Median Intensity)	49.4	5.51	88.8 %
Percent Bias (High Intensity)	13.11	2.68	79.6 %
Percent Bias (Low Intensity)	11.04	4.21	61.9 %
Total Percent Bias	10.72	3.42	68.1 %
Coefficient of Determination (R^2)	0.83	0.94	64.7 %
Root Mean Square Error (RMSE)	0.1	0.06	40.0 %
Bias Midsection Slope (%)	-0.62	1.3	N/A
		Avg. Percent Improvement:	**67.2 %**

Table 3: *Calibration set metrics (all data)*

Calibration Dataset (All Data)			
Performance Metrics	**Original Radar**	**Product Radar**	**Percent Improvement**
Percent Bias (Median Intensity)	46.8	0.91	98.1 %
Percent Bias (High Intensity)	10.97	0.23	97.9 %
Percent Bias (Low Intensity)	2.14	-0.2	90.7 %
Total Percent Bias	6.05	-0.02	99.7 %
Coefficient of Determination (R^2)	0.72	0.99	96.4 %
Root Mean Square Error (RMSE)	0.122	0.018	85.2 %
Bias Midsection Slope (%)	-6.8	-0.2	N/A
		Avg. Percent Improvement:	**94.7 %**

Table 4: *Validation set metrics (Hurricane Harvey)*

Validation Dataset (25 August 2017 - 1 September 2017)			
Performance Metrics	**Original Radar**	**Product Radar**	**Percent Improvement**
Percent Bias (Median Intensity)	57.5	7.1	87.7 %
Percent Bias (High Intensity)	15.88	1.88	88.2 %
Percent Bias (Low Intensity)	17.8	5.59	68.6 %
Total Percent Bias	15.75	4.37	72.3 %
Coefficient of Determination (R^2)	0.81	0.95	73.7 %
Root Mean Square Error (RMSE)	0.114	0.059	48.3 %
Bias Midsection Slope (%)	3.78	1.88	N/A
		Avg. Percent Improvement:	**73.1 %**

Table 5: *Performance metrics of validation set from 2015-2024.*

Year / Performance Metrics	2015			2016		
	Original Radar	**Product Radar**	**% Improvement**	**Original Radar**	**Product Radar**	**% Improvement**
Percent Bias (Median Intensity)	63.49	4.32	93.2 %	72.03	12.55	82.6 %
Percent Bias (High Intensity)	21.23	2.23	89.5 %	27.44	4.32	84.3 %
Percent Bias (Low Intensity)	26.96	1.42	94.7 %	34.55	6.44	81.4 %
Total Percent Bias	23.09	1.8	92.2 %	28.49	5.33	81.3 %
Coefficient of Determination (R^2)	0.733	0.926	72.2 %	0.712	0.920	72.3 %
Root Mean Square Error (RMSE)	0.116	0.061	47.2 %	0.137	0.072	47.4 %
Bias Midsection Slope (%)	9.64	3.02	N/A	17.56	3.22	N/A
Average % Improvement			**81.5 %**			**74.9 %**
Year / Performance Metrics	**2017**			**2018**		
	Original Radar	**Product Radar**	**% Improvement**	**Original Radar**	**Product Radar**	**% Improvement**
Percent Bias (Median Intensity)	57.5	7.1	87.7 %	52.54	4.93	90.6 %
Percent Bias (High Intensity)	15.88	3.42	78.5 %	8.14	1.46	82.1 %
Percent Bias (Low Intensity)	17.8	5.59	68.6 %	-1.99	0.31	84.4 %
Total Percent Bias	15.75	4.37	72.3 %	1.61	0.99	38.5 %
Coefficient of Determination (R^2)	0.814	0.950	73.3 %	0.808	0.941	69.1 %
Root Mean Square Error (RMSE)	0.114	0.059	48.3 %	0.096	0.054	44.4 %
Bias Midsection Slope (%)	3.78	1.88	N/A	-7.57	-3.81	N/A
Average % Improvement			**71.4 %**			**68.2 %**

Table 6: *Performance metrics of validation set from 2015-2024, continued...*

Year / Performance Metrics	2019			2020		
	Original Radar	**Product Radar**	**% Improvement**	**Original Radar**	**Product Radar**	**% Improvement**
Percent Bias (Median Intensity)	55.47	12.19	78.0 %	44.74	5.03	88.8 %
Percent Bias (High Intensity)	16.12	2.51	84.4 %	11.14	2.69	75.9 %
Percent Bias (Low Intensity)	15.84	3.88	75.5 %	4.89	4.01	18.0 %
Total Percent Bias	14.02	3.27	76.7 %	6.91	3.38	51.1 %
Coefficient of Determination (R^2)	0.826	0.941	66.3 %	0.828	0.946	68.4 %
Root Mean Square Error (RMSE)	0.095	0.055	41.9 %	0.086	0.048	43.8 %
Bias Midsection Slope (%)	-0.16	-0.65	N/A	-3.52	-1.45	N/A
Average % Improvement			**70.5 %**			**57.6 %**
Year / Performance Metrics	2021			2022		
	Original Radar	**Product Radar**	**% Improvement**	**Original Radar**	**Product Radar**	**% Improvement**
Percent Bias (Median Intensity)	45.66	5.72	87.5 %	35.62	1.2	96.6 %
Percent Bias (High Intensity)	8.97	2.24	75.0 %	7.62	2.19	71.3 %
Percent Bias (Low Intensity)	5.35	5.07	5.2 %	2.2	4.3	-95.5 %
Total Percent Bias	5.97	3.74	37.4 %	3.67	3.08	16.1 %
Coefficient of Determination (R^2)	0.876	0.946	56.6 %	0.876	0.943	53.8 %
Root Mean Square Error (RMSE)	0.088	0.058	34.1 %	0.079	0.053	32.0 %
Bias Midsection Slope (%)	-3.12	1.52	N/A	-2.75	2.87	N/A
Average % Improvement			**49.3 %**			**29.1 %**

Table 7: *Performance metrics of validation set from 2015-2024, continued...*

Year / Performance Metrics	2023			2024		
	Original Radar	**Product Radar**	**% Improvement**	**Original Radar**	**Product Radar**	**% Improvement**
Percent Bias (Median Intensity)	41.67	3.72	91.1 %	40.46	4.49	88.9 %
Percent Bias (High Intensity)	10.04	2.89	71.2 %	8.7	2.42	72.2 %
Percent Bias (Low Intensity)	5.84	4.89	16.3 %	3.07	4.12	-34.2 %
Total Percent Bias	6.92	3.86	44.2 %	4.79	3.19	33.4 %
Coefficient of Determination (R^2)	0.867	0.957	67.3 %	0.876	0.956	64.2 %
Root Mean Square Error (RMSE)	0.085	0.049	42.9 %	0.095	0.057	40.2 %
Bias Midsection Slope (%)	-1.78	1.13	N/A	-4.31	4.38	N/A
Average % Improvement			**55.5 %**			**44.1 %**

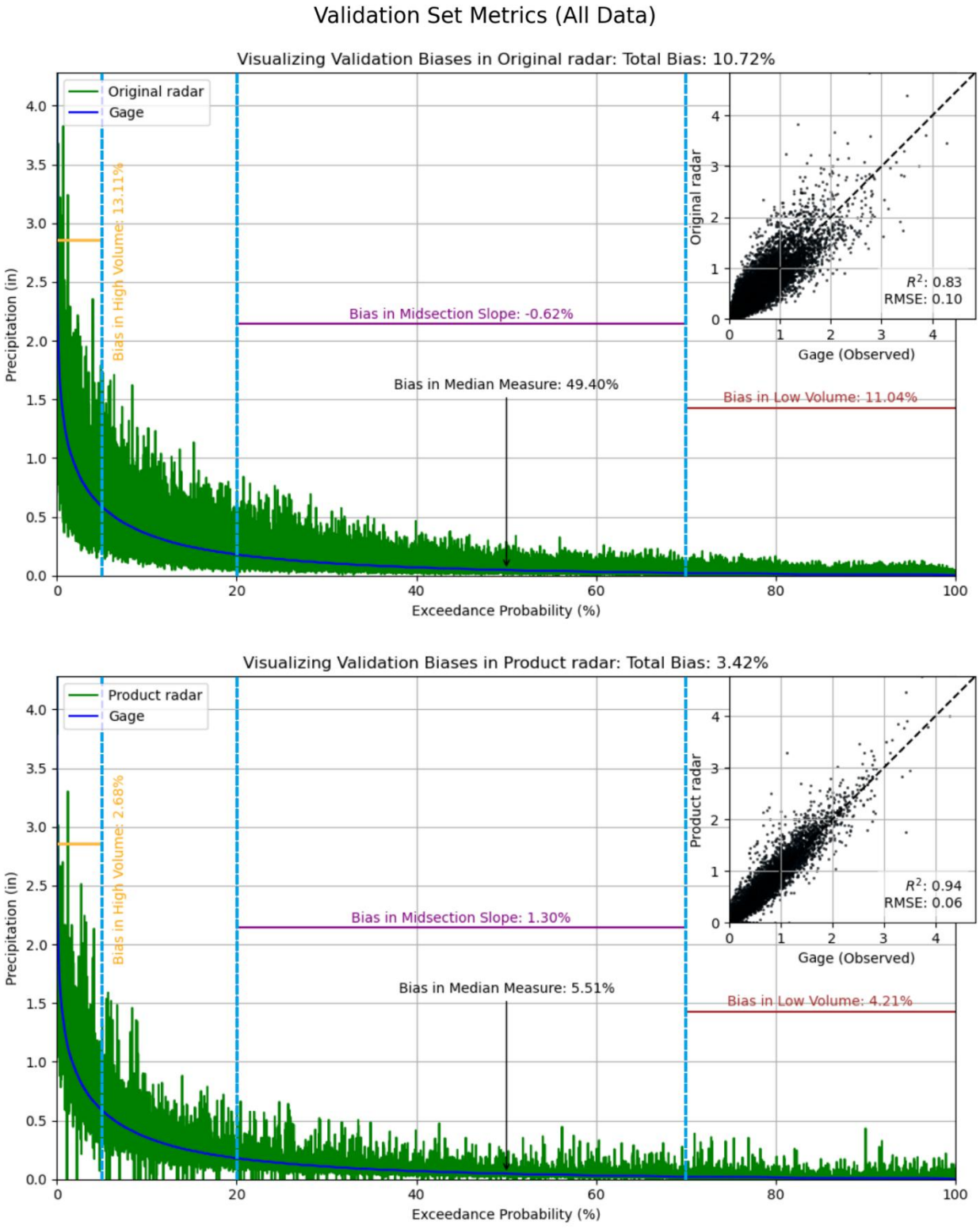


Figure 17: *Validation set metrics (all data)*

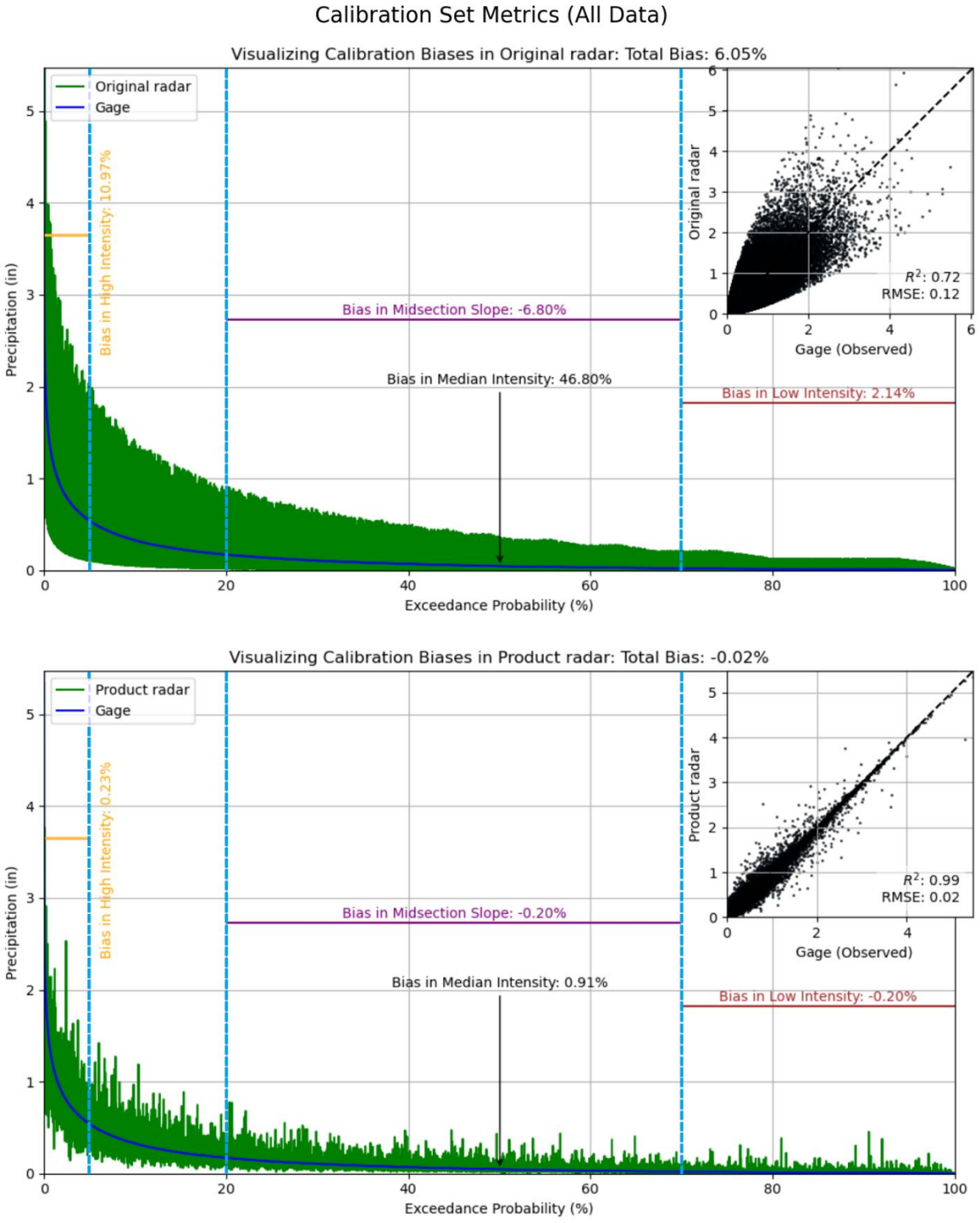


Figure 18: *Calibration set metrics (all data)*

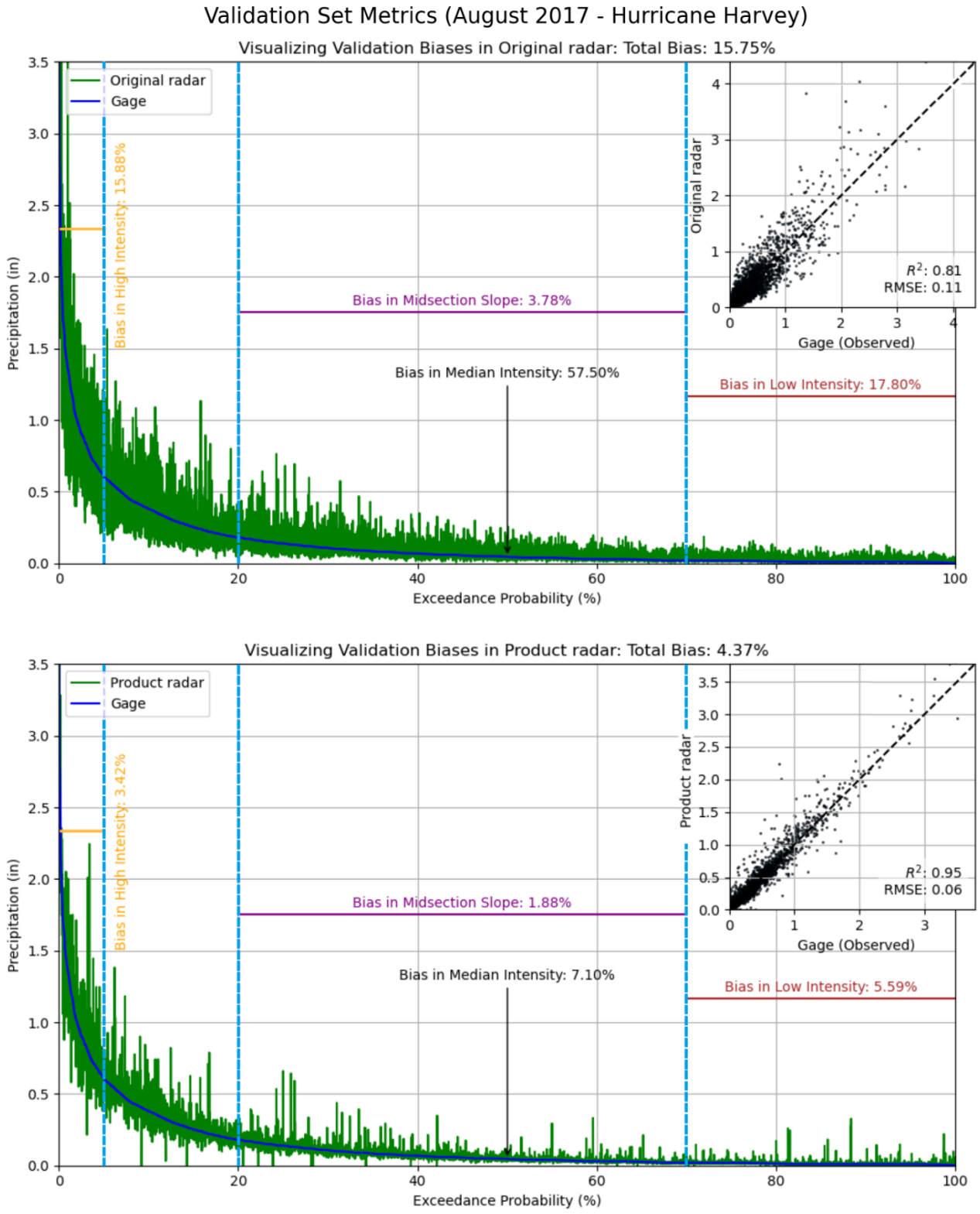


Figure 19: *Validation set metrics (Hurricane Harvey)*

4.4 Uncertainty Analysis of Performance Metrics

4.4.1 Global assessment across all events

Uncertainty analysis was carried out on the validation set for the full data range of 2015 to 2024. Although the overall gage-radar relationship exhibits heteroskedasticity (non-constant standard deviations), this does not bias coefficient estimates; thus, the performance metrics remain valid, particularly when expressed as percentage improvements. Optimization of the residual regression also favored a linear formulation, which is reasonable because event characteristics and radar bias tend to be relatively consistent within a region at the scale of a single time step producing an approximately linear relationship despite the broader heteroskedasticity. However, confidence intervals derived directly from ordinary least squares (OLS) assumptions would be unreliable under these conditions.

To address this, a bootstrapping approach with 10,000 iterations was used to quantify the uncertainty ranges for RMSE and R^2. For biases, the data were stratified into three rainfall categories (low, medium, and high) as defined in **Section 3.2**, and separate normal distributions were fitted to each category to minimize the influence of heteroskedasticity across regimes. The details of each are given in **Table 8** and **Figure 20**.

Table 8: *Confidence intervals for performance metrics*

Validation Dataset (All Data) Performance Metrics for 99% Confidence Interval			
Performance Metrics	**Original Radar**	**Product Radar**	**Percent Improvement**
Percent Bias (Median Intensity)	49.40 (46.66 – 52.13)	5.51 (4.20 – 6.81)	88.80% (85.41 – 91.94)
Percent Bias (High Intensity)	13.11 (12.70 – 13. 52)	2.68 (2.46 – 2.89)	79.60% (77.24 – 81.80)
Percent Bias (Low Intensity)	11.04 (10.20 – 11.87)	4.21 (3.74 – 4.69)	61.90% (54.02 – 68.49)
Total Percent Bias	10.72 (10.14 – 11.29)	3.42 (3.09 – 3.76)	68.10% (62.92 – 72.63)
Coefficient of Determination (R^2)	0.828 (0.82 – 0.836)	0.944 (0.939 – 0.948)	67.44% (62.80 – 71.11)
Root Mean Square Error (RMSE)	0.099 (0.097 – 0.102)	0.057 (0.054 – 0.059)	40.00% (39.18 – 47.06)
Containment Ratio	0.9664	0.9691	N/A
		Avg. Percent Improvement:	**67.20% (63.59 – 72.17)**

**No uncertainty thresholds were calculated for bias mid-section slope.*

For a nominal 99% confidence interval, the observed empirical coverage is approximately 97%, indicating good agreement between the theoretical interval and the actual data. Therefore, no further refinement of the distributional assumptions was pursued. The narrow confidence interval widths suggest that deterministic values are generally representative. The average percentage improvement was 67.20%, with a 99% confidence interval of 63.59%–72.17%, (span of 8.57%), derived by considering all possible combinations of upper and lower bounds for both the original and product radar. The narrow interval, together with the lower confidence bound of 63.59%, demonstrates the robustness of the results.

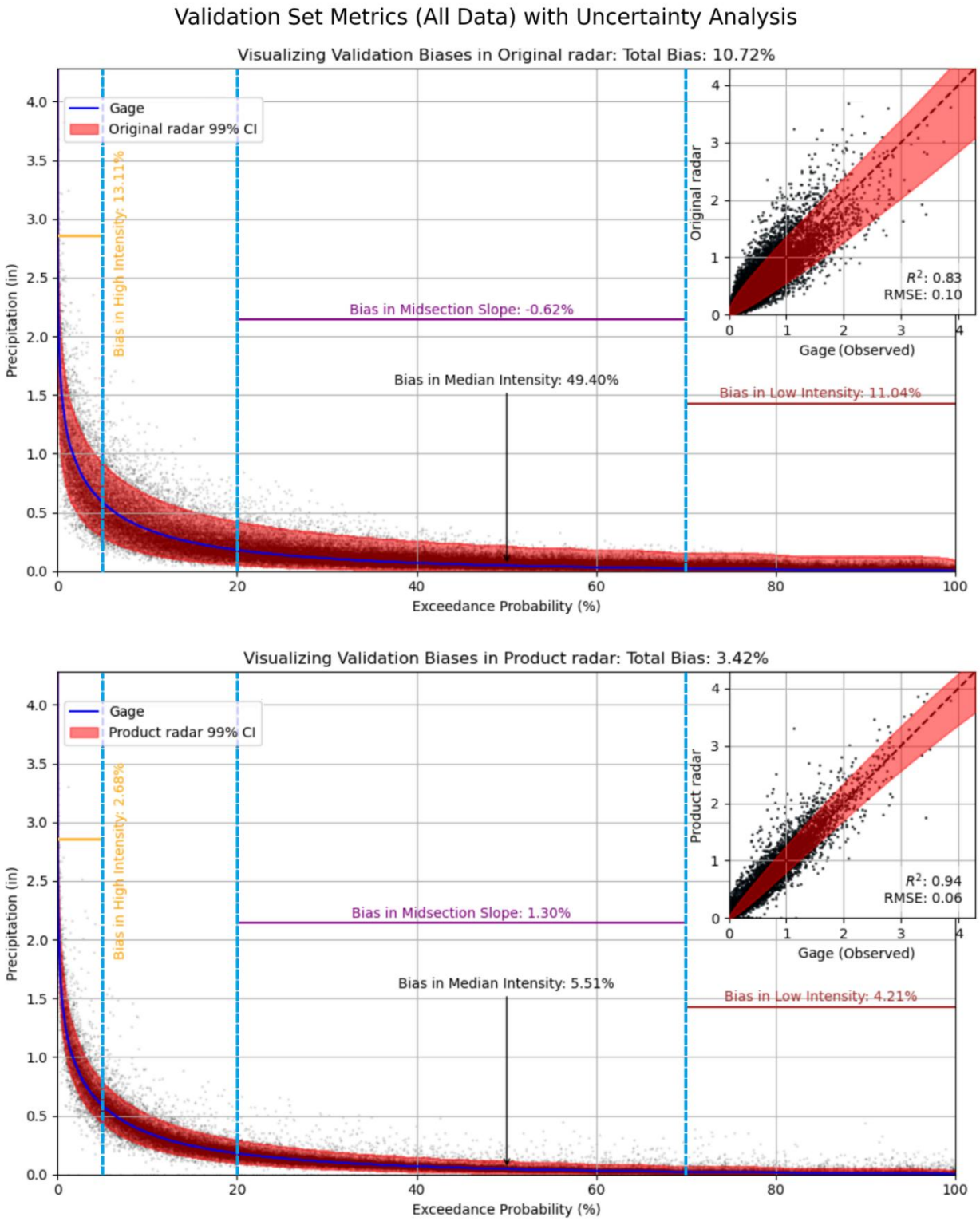


Figure 20: *Uncertainty analysis for validation set metrics for original (top) and product (bottom) radar with confidence bands*

4.4.2 Event-level assessments

To evaluate the product across varying event types, an uncertainty analysis was conducted on all rainfall events recorded between 2015 and 2024. Events were categorized based on the cumulative gage rainfall across all reporting stations during each event, which served as a robust proxy for both spatial spread and intensity. This method was selected over conventional return period analyses which are infeasible at the regional scale in this study. All available data was used, without splitting into calibration and validation sets to provide a holistic view of the final product's performance.

The analysis was based on computed average improvement in R^2 over the duration of each event. Events selected for analysis were restricted to those that lasted more than 5 hours, separated by at least 5 hours of no-rainfall and containing at least 5 valid data points. This arbitrary filtering yielded a total of 972 events. Of these, 52 events (5.34%) did not meet the lower-bound threshold of 62.80% average improvement in R^2. These were primarily low-volume events with fewer than 10 gages registering rainfall (e.g., 0.04 in. at the gage vs. 0.02 in. from the product). Each of these underperforming events was reprocessed using a different randomization seed, which was sufficient to resolve all deficiencies.

Following filtering and corrective steps, a 99% confidence interval was constructed using logistic regression under the assumption that residuals follow beta distribution, with predicted values capped at 100% to reflect theoretical limits. The empirical coverage rate of 99.15% confirms the distribution assumption. The resulting plot exhibits a slight downward slope (-0.028) in the regression line, with an overall average improvement of 96.6%, the product performs consistently well across event types with R^2 improvements averaging 98.06% for low-intensity events and 95.26% for high-intensity events. The lower bound of the 99% confidence interval remained at 70%, confirming the reliability of the approach. Results are illustrated in **Figure 21**.

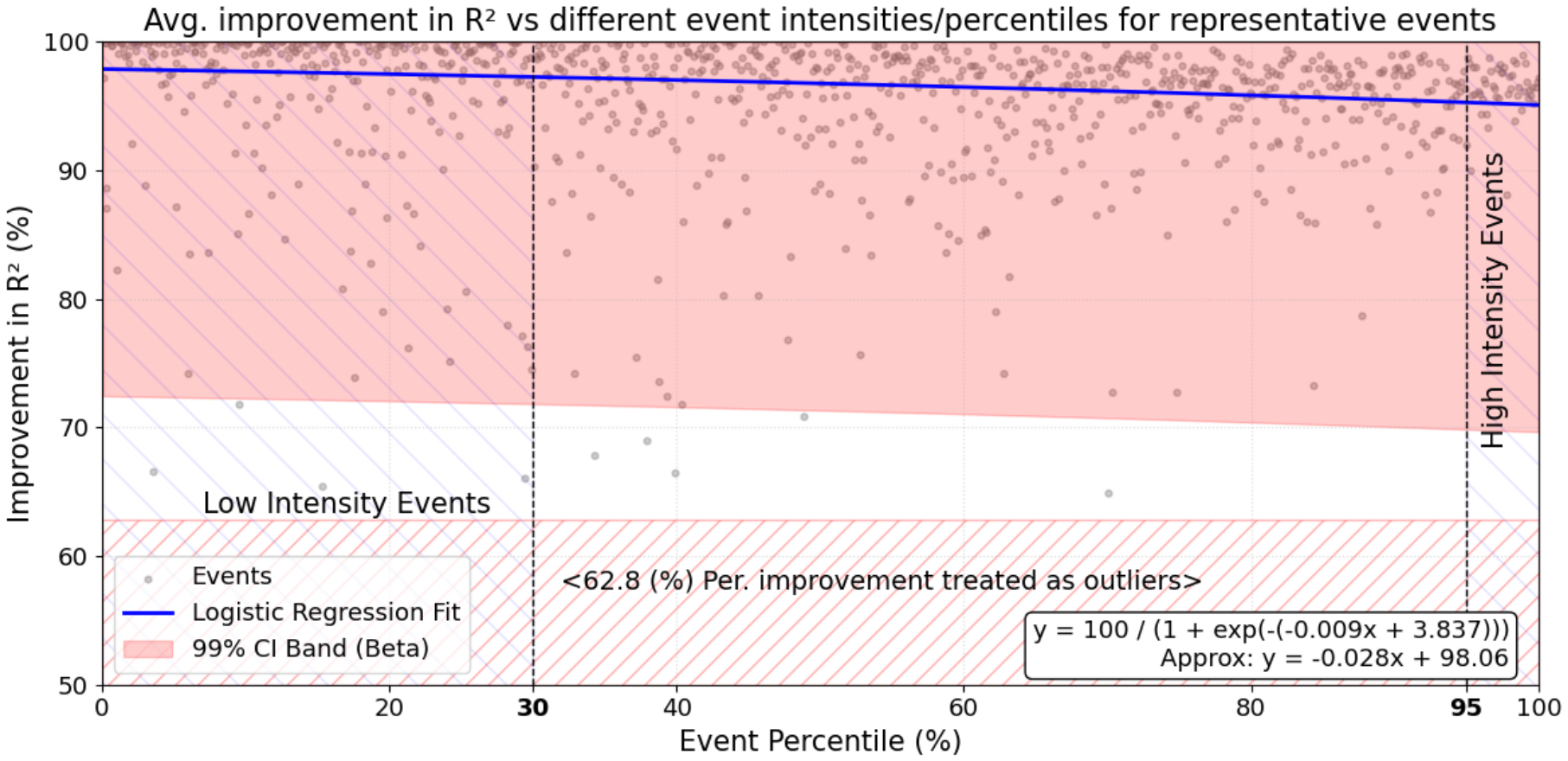


Figure 21: *Average improvement for different event types within the study period*

4.5 Temporal Trend in Performance Metrics

A clear temporal trend is observed in R^2 and RMSE, both of which show improvement over the study period. The number of gages available for calibration is one significant time-varying factor; as gage availability increases over time, performance improves correspondingly, suggesting a strong positive correlation. This is confirmed by a strong Pearson correlation coefficient of 0.794 between gage count and R^2. This relationship is illustrated in **Figure 22**.

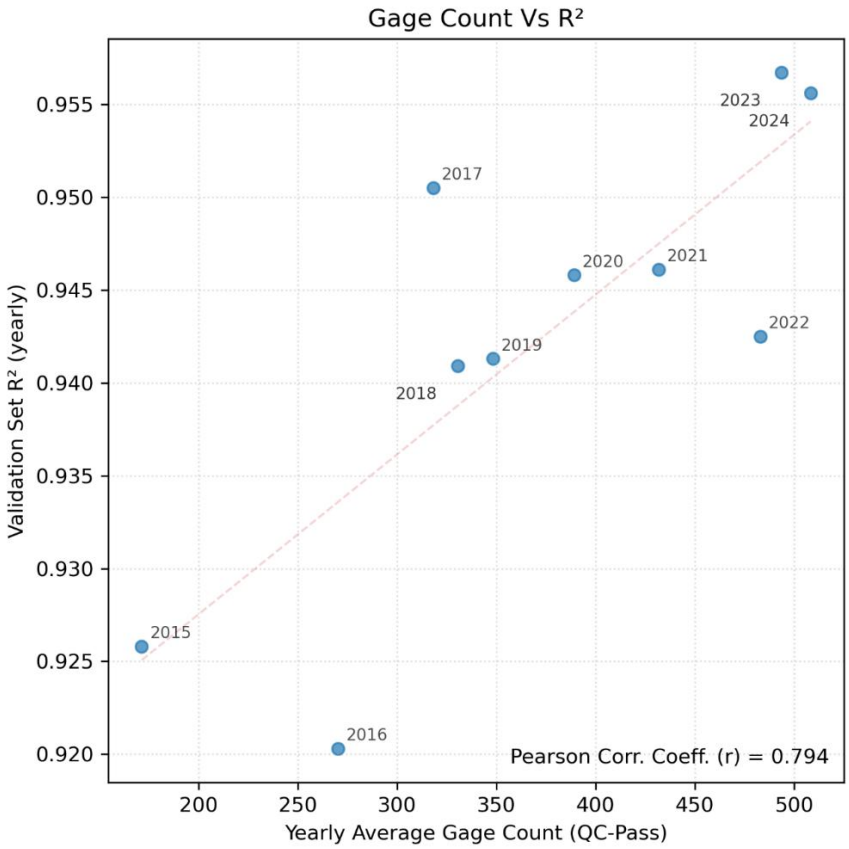


Figure 22: *Effect of the QC processes on gage count and R^2 over time (yearly averages)*

A moderate negative correlation is found between gage count and RMSE (–0.564), while the correlation between gage count and total bias is weak (0.160). These results suggest that increasing the gage count improves precision (as reflected in RMSE and R^2) but has limited effect on average bias which often arises from spatially and temporally uncorrelated noise that tends to cancel out over time or space. This implies that gage-corrected products, even with modest precision, can be suitable for long-term climatological applications where aggregate quantities are of interest while gage density becomes critical for applications requiring high-resolution accuracy, such as short-term hydrologic or hydraulic simulations.

Gage count may not be the only influencing temporal factor. Improvement of raw radar data quality over time has not been explicitly accounted for in this study, but its influence is suggested by the strong negative correlation (-0.883) between gage count and average percentage improvement. Correcting a poor-quality radar product from 2015 yields a large percentage improvement, while correcting a higher-quality product from 2023 yields a smaller percentage change even if the final corrected outputs are comparable in accuracy. In some cases, a negative percent improvement may be observed when the raw product itself is near the limits of correction (as seen in low intensity bias% of 2022 and 2024).

4.6 Evaluation with Preliminary Clustering Analysis

Up to this point, the evaluation of bias correction performance has relied exclusively on direct comparisons between gage observations and the WERC product. While gages are generally considered the ground truth in hydrometeorological studies, they are not immune to uncertainties. Issues such as instrument calibration drift, wind-induced undercatch, clogging, or localized anomalies can introduce measurement errors that compromise their reliability. Moreover, gage networks are sparse relative to the spatial resolution of radar, meaning that single-point errors may disproportionately influence regional assessments. These limitations highlight the need for independent, external validation to establish a more robust measure of confidence in the corrected precipitation estimates.

To address this, a preliminary clustering analysis and hydrologic modelling (the following section) were performed to evaluate the consistency of the WERC product with other independent data sources. The clustering approach allowed the project team to examine broader patterns of agreement across multiple datasets. By grouping similar behaviors and identifying outliers, the clustering framework provides an additional lens through which to assess the credibility of the corrected rainfall fields.

A representative basin was selected per project management team suggestion at Available gages located within or near the basin were compared with other established precipitation datasets, including the MRMS_Gagecorr product and 24-hour basin-aggregated rainfall files derived from the NOAA Global Historical Climatology Network Daily (GHCNd), obtained via personal communication with TWDB (NOAA NCEI, 2021; Zhang et al., 2016). Clustering was performed using the K-means clustering algorithm with Pearson correlation coefficients serving as the distance metric for the timespan of October 2017 to February 2018. The results were visualized in two dimensions using the two most significant principal components identified through Principal Component Analysis (PCA), as shown in **Figure 23**.

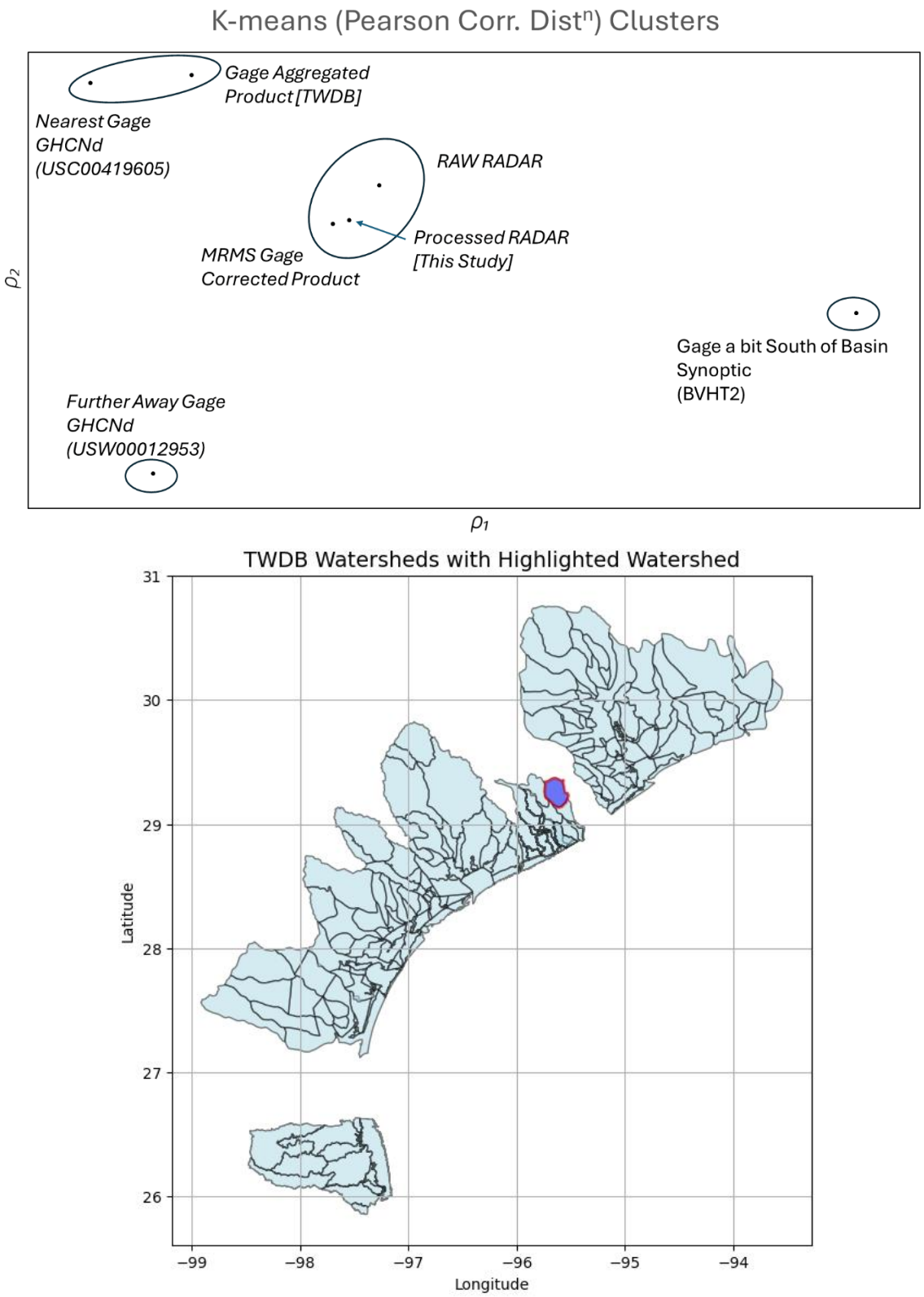


Figure 23: *Clustering alternative datasets (top) at Lower Brazos watershed location (bottom)*

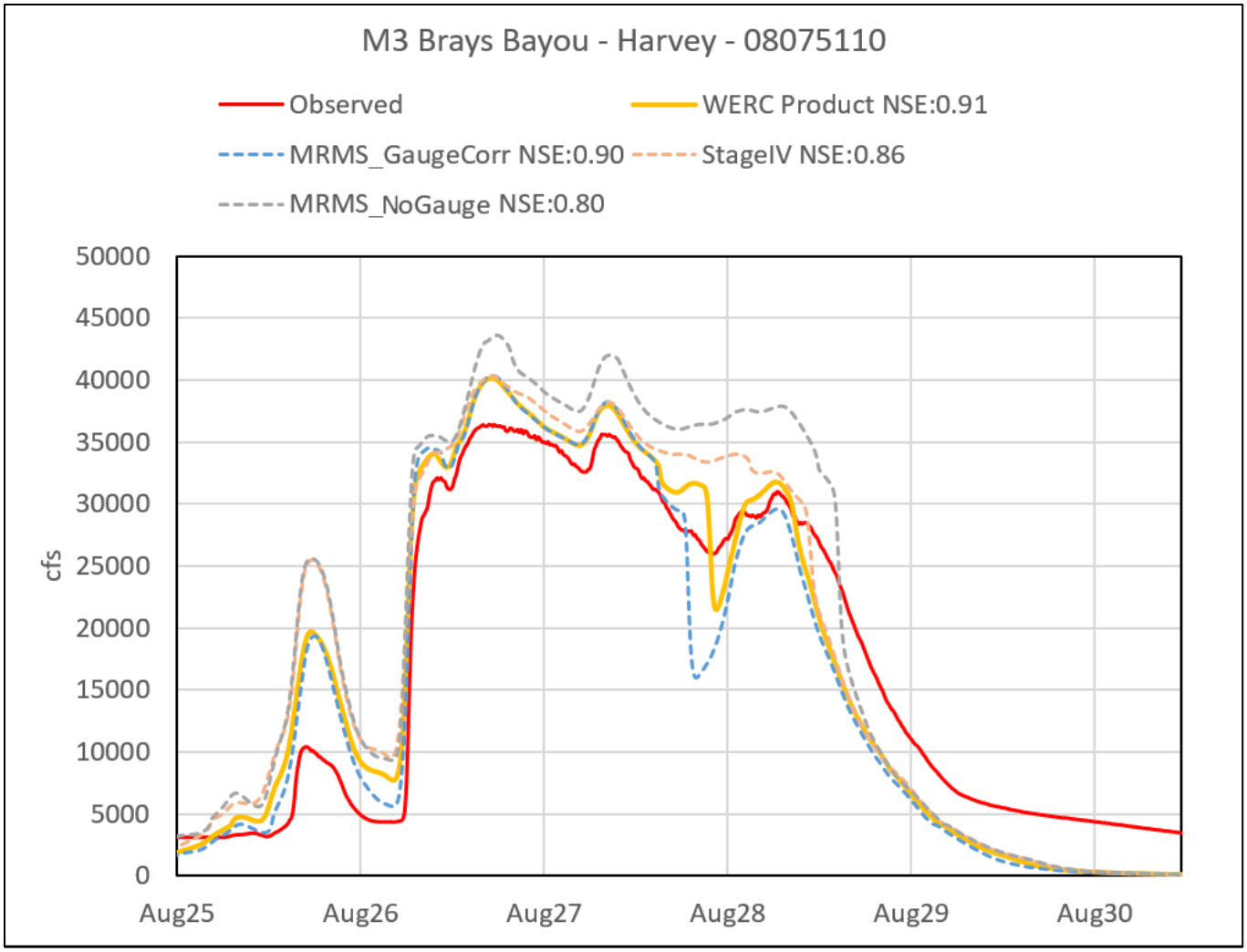
The clustering results indicate that the WERC product closely aligns with the MRMS_GageCorr product, with both falling within the same cluster. Based on intra-cluster distance, the WERC product is significantly closer to the MRMS gage-corrected data than to the raw radar data, as expected. In contrast, this cluster is distinct and relatively distant from those formed by raw gage data and the TWDB gage-derived product, suggesting that the WERC product is consistent with established datasets. Greater confidence in the absence of anomalous behavior is provided by the consistent clustering of the WERC product with the MRMS_GageCorr product, reinforcing its reliability for hydrological applications. At the same time, significant deviation from MRMS indicates that the WERC product represents an improved and distinct dataset.

The clear separation from gage-only products also highlights how single point uncertainties and sparse sampling can distort basin-scale precipitation patterns, whereas the WERC product preserves spatial uniqueness while not maintaining agreement with trusted benchmarks.

4.7 Hydrologic Modeling

To evaluate the hydrologic impact of different precipitation products, a pre-calibrated hydrologic model (i.e., HEC-HMS), for the Brays Bayou watershed, obtained from the Harris County Flood Control District (HCFCD)'s M3 system, was run for the Hurricane Harvey (Aug 2017) event. The hydrologic model was run for the new product radar, and the results compared with observed stage at USGS gage 08075110 (Brays Bayou at MLK Jr Blvd, Houston, TX) which corresponds to the junction "D1000000_0224_J" of the HEC-HMS model, shown in **Figure 24**. The same hydrologic model was rerun using alternative precipitation datasets to enable a comparative evaluation. These datasets included raw radar precipitation (i.e. MRMS mosaics without gage correction or radar-only data), the MRMS_GageCorr product, and the Stage IV product. For each case, simulated flows were compared against observed streamflow data from the USGS gage (08075110) to assess performance and consistency across different rainfall inputs.

The WERC product reproduces the observed hydrograph with greater accuracy, as shown in **Figure 24**, capturing subtle ridges and troughs more accurately, reflecting improved alignment in both timing and volume. The product achieves a Nash–Sutcliffe Efficiency (NSE) of 0.91, outperforming alternative datasets including the MRMS gage-corrected rainfall, demonstrating superior agreement in capturing rainfall-driven variations that shape hydrograph peaks (e.g., August 25) and recessions (e.g., August 28). This improvement is especially noteworthy given that the alternative datasets were directly used for model calibration, whereas the WERC product achieves higher accuracy without calibration, underscoring its robustness and reliability. The study area was intentionally selected to minimize differences in gage density relative to other products, thereby isolating improvements attributable primarily to the proposed methodology rather than changes in observational support. Consequently, regions experiencing larger increases in rain-gage density (green/yellow in **Figure 13**) are expected to exhibit even greater performance gains.



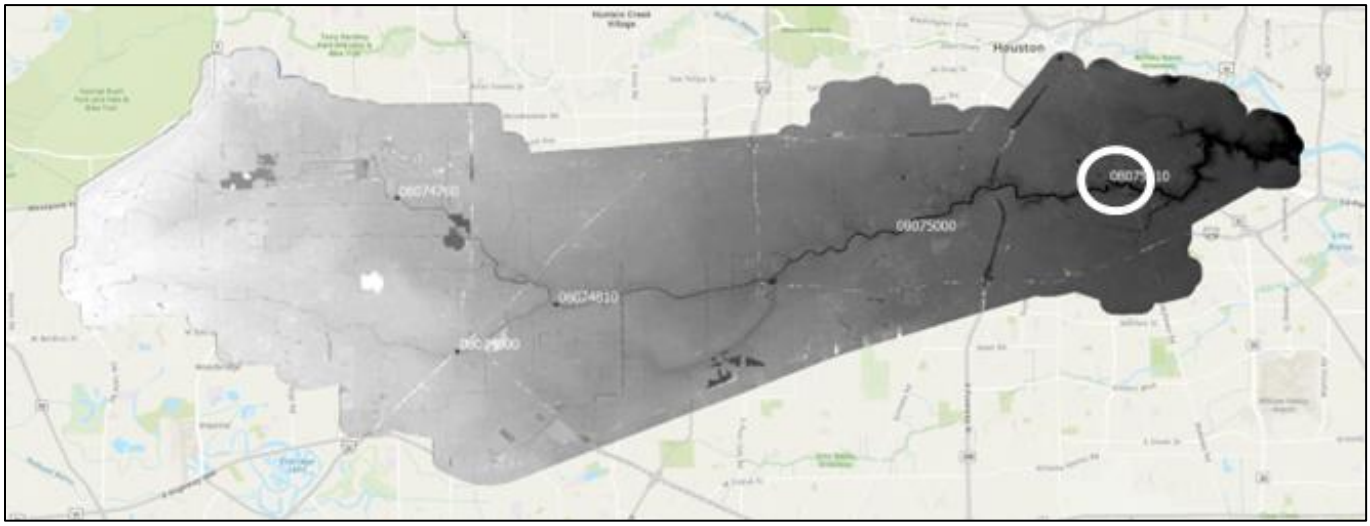


Figure 24: *HEC-HMS model comparisons (top) at the Brays Bayou watershed (USGS stream gage Location: 08075110) (bottom)*

5 Research Outcomes and Practical Recommendations

5.1 Key Findings

Overall, results confirm that higher gage density, rigorous quality control, and region-specific optimization substantially enhance rainfall accuracy and hydrologic reliability for coastal Texas. These key findings are summarized below:

1. The WERC gage-corrected rainfall product maintains spatial contiguity.
2. Average improvement in metrics is 67.2% for validation, 94.7% for calibration, and 73.1% for Hurricane Harvey validation. R^2 increased from ~0.83 to ~0.94 and RMSE decreased by 40% across datasets. Annual validation statistics from 2015–2024 show persistent accuracy gains, confirming model stability through time. Furthermore, 99% confidence intervals indicate narrow uncertainty bounds (containment≈97%), demonstrating high statistical robustness of the reported results.
3. Temporal trend analysis revealed a strong positive correlation between gage count and R^2 ($r = 0.794$) and a negative correlation with RMSE ($r = –0.564$). RMSE differences between the WERC and MRMS_GageCorr products, when compared with gage density differences, show moderate spatial correspondence (SSIM=0.5015). These results confirm that gage coverage influences correction accuracy.
4. A negative correlation ($r = –0.883$) observed between gage count and percent improvement reflects improved baseline radar quality over time.
5. Event-level analysis (972 events) showed an average R^2 improvement of 96.6%, with low-intensity events improving up to 98%. While 5.3% of events underperformed initially, all were corrected after reprocessing. Regression analysis across all rainfall events shows a slight downward trend (slope = –0.028) but sustains high overall performance with a lower 99% confidence bound of 70%, confirming the product's reliability across diverse event types.
6. Clustering analysis shows that the WERC product consistently grouped with the MRMS_GageCorr dataset, while maintaining distinctness stemming from independent refinement. This strong agreement with an established benchmark suggests no anomalous behavior in the enhanced data.
7. Hydrologic modeling using HEC-HMS for Hurricane Harvey (2017) achieved NSE = 0.91, outperforming all alternative rainfall inputs (Stage IV & MRMS) and more accurately reproducing peak and recession limb flows.

5.2 Implications for Engineering and Water Management Practice

The demonstrated gains in reliability and spatial coherence of the enhanced rainfall product developed in this study represent a significant advancement for hydrologic and hydraulic

modeling applications across coastal Texas. By providing spatially continuous, bias-corrected, and temporally consistent precipitation fields, the product improves the accuracy of rainfall–runoff simulations, peak-flow estimation, and flood-inundation mapping.

These improvements directly support more reliable assessments of storm impacts, drainage capacity, and watershed response under event conditions. They further enable greater confidence in gridded rainfall data used for regulatory and engineering design.

The integration of additional gages without compromising accuracy demonstrates the system's scalability and adaptability to evolving observational networks, a key attribute for long-term resilience planning.

5.3 Recommendations for Implementation

For practical adoption, it is recommended that the workflow and product be integrated into existing agency data pipelines and modeling platforms. For seamless integration into existing workflows, the enhanced rainfall product can be distributed through standardized data formats such as GeoTIFF, DSS, and PCP. These interoperable formats allow direct compatibility with widely used modeling platforms and GIS applications such as HEC-HMS, HEC-RAS, and TWDB TxRR model, among others.

The accompanying quality-control and bias-correction framework should be institutionalized as part of routine rainfall data processing across operational and planning divisions. Embedding these procedures within agency pipelines will maintain consistency in rainfall estimation, minimize systematic errors, and ensure that model calibration and design studies are based on the most accurate and validated precipitation inputs available.

To enhance operational readiness, the workflow can be expanded for near-real-time implementation, supporting active forecasting and flood early warning systems (FEWS). Establishing automated retrieval through the Synoptic API, coupled with scheduled recalibration and periodic validation, would ensure sustained accuracy as gage networks evolve. Such real-time integration improves situational awareness, providing critical lead time for flood forecasting, emergency response, and infrastructure management.

Collaboration among TWDB, regional partners, and local agencies is also recommended to systematically identify data-sparse regions and prioritize new gage installations in areas where increased density would produce the greatest improvements in model reliability. This strategic expansion of the gage network will promote more equitable spatial coverage, strengthen the observational foundation of the rainfall product, and enable a more cost-effective allocation of resources for gage deployment and maintenance.

6 Conclusions

This study developed regionally optimized radar-based precipitation estimates from an advanced rainfall bias-correction framework (RAIDW) for the Texas coast region. The approach integrated regression-based adjustment with residual interpolation using IDW that is especially valuable in areas with sparse gage networks, as regression applies a base correction to the entire raster based on average relationship between gage and radar rainfall data, thus not limiting corrections to areas that are near gages only. A dynamic quality-control and information-rich validation framework was designed and developed to address the limitations of traditional gage–radar comparisons. A novel gage-selection algorithm ensured that validation relied on spatially representative and event-significant gages, reducing limitations inherent in conventional cross-validation such as overrepresentation of trivial gages, exclusion of critical information, or large computational overhead. This information-rich validation design retained more than half of the data for evaluation and stabilized performance metrics across diverse rainfall regimes.

The generated WERC radar rainfall product demonstrates substantial improvements across calibration, validation, and event-based analyses. Average improvements over all performance metrics tested reached 94.7% for calibration and 67.2% for validation datasets, with a 73.1% gain observed for Hurricane Harvey (2017). The residual bias after correction is consistently positive between gage data and the WERC product. The temporal alignment of gages and gage uncertainty in recording large precipitations both suggest an inherent positive bias should exist; thus, the residual bias is permissible. The event-by-event evaluation confirmed stable improvements across 972 rainfall events from 2015 to 2024, with a 99% confidence interval indicating a consistently strong R^2 improvement of 94% per event. Importantly, the product exhibits robustness across both high- and low-intensity rainfall regimes without introducing spurious artifacts, as evidenced by only a slight deviation in performance across intensities: percentage improvements in R^2 averaged 98.06% for low-intensity events and 95.26% for high-intensity events, while the lower bound of the 99% confidence interval remained at 70%, underscoring the method's reliability for different event types.

A comparative spatial analysis against the MRMS_GageCorr dataset confirms that differences between the two products share moderate pattern similarity with gage density differences (SSIM= 0.5015). Building on this, analysis across the full study period revealed strong correlations between gage count and performance metrics (R^2 correlation coefficient = 0.794). Increased gage density enhanced precision (R^2 and RMSE), but had limited effect on bias, which often arises from uncorrelated noise. This suggests that for climatological applications, even moderate gage coverage may be sufficient, whereas short-term hydrologic applications benefit greatly from increased gage density. However, because radar quality also improved over time, future efforts should disentangle these overlapping temporal effects to isolate the contribution of gage count alone. This is evident in year-by-year performance metrics which confirmed consistent gains in

bias correction, with especially large improvements in earlier years when raw radar product quality was comparatively lower.

Comparative clustering analyses confirm that the WERC product aligns closely with the MRMS gage-corrected dataset while remaining clearly distinct from both raw radar and gage-only products. This consistent clustering reinforces the dataset's reliability, while deviations from MRMS indicate that the WERC product represents a distinct, improved rainfall field rather than a direct replication. Hydrologic validation using a pre-calibrated HEC-HMS model further confirmed its utility: without additional calibration, the WERC product achieved an NSE of 0.91, outperforming alternative datasets including MRMS and Stage IV in capturing both volumetric totals and temporal features with greater accuracy.

7 Limitations and Future Research

7.1 Limitations in Adopted Bias-Correction Methodology

The WERC product radar produced in this study relied on the RAIDW method, and the limitations of this approach must be acknowledged. The method is highly sensitive to small discrepancies, as even minor differences between gage and radar values can result in disproportionately large scaling factors that amplify rainfall estimates. Although multiplicative adjustments avoid the unrealistic negative values that can arise with additive corrections, they cannot correct for missed rainfall events in radar data, meaning zero-rainfall regions remain unadjusted even when gages record precipitation.

Additionally, IDW is strictly interpolative; while this prevents spurious extrapolations, it can underrepresent rainfall variability in areas with sparse gage coverage. Moreover, the regression framework assumes spatial coherence in bias patterns, which generally holds along the Texas coast, but may introduce inaccuracies when applied over larger or meteorologically diverse regions. Despite these limitations, the RAIDW approach has the advantage of reducing sporadic artifacts, leading to cleaner and more consistent precipitation fields compared to other methods.

7.2 Rain Gage Limitations and Challenges

Rain gages are used as point-based instruments that offer direct, ground-truth observations. Rain gages are generally considered more reliable than indirect radar estimates (Foehn et al., 2018). Nevertheless, rain gages are not without uncertainties. Sources of uncertainty include freezing precipitation which may cause undermeasurement during events and overmeasurement after thaw, high winds over gage which causes negative bias, undermeasurement by tipping bucket gages, and issues related to gage maintenance, etc. (Grossi et al., 2017; Habib et al., 2001; Muchan & Dixon, 2019; Pollock et al., 2018; Tokay et al., 2010; Villarini et al., 2008).

Each gage measures precipitation at a single point, while hydrologic and radar-based products represent spatially averaged conditions over grid cells or catchments. Uneven spatial distribution introduces representativeness error (Gao et al., 2021). In coastal Texas, dense clusters of stations

exist near major metropolitan areas such as Houston, but large portions of the middle and lower coast remain sparsely instrumented. As a result, localized convective cells, coastal rainbands, and inland gradients often go under-sampled, reducing the reliability of gage-based bias correction for radar rainfall fields (Villarini et al., 2008).

Instrumentation adds further uncertainty. Rain gage data from tipping-bucket instruments face a critical limitation during high-intensity storms: the finite time required for each bucket to fill and tip means that rapidly accumulating rainfall can outrun the instrument's counting mechanism, leading to significant under-measurement when rain rates are high. For example, at very high rainfall intensities the gage may miss the fraction of rainfall amounts during the tipping motion and record fewer tips than the true volume warrants. This under-catch effect grows non-linearly with increasing rain rates and becomes especially pronounced in convective or tropical-storm environments (Segovia-Cardozo et al., 2021). In addition, tipping-bucket gages remain subject to other error sources such as wind-induced under-catch, funnel splash-out, evaporation losses, and temporal resolution limitations, though these tend to be less impactful than the mechanical tip delay during intense storms (Duchon et al., 2013; Segovia-Cardozo et al., 2023; Wang et al., 2008).

Although quality-control and post-processing steps were applied in this study, this investigation did not explicitly correct for the severe-storm under-measurement inherent to the tipping-bucket mechanism.

7.3 Gridded Radar Rainfall Limitations and Uncertainties

Radar-based estimates are still subject to errors including signal attenuation, freezing precipitation, low topped convection, bright banding, beam blockage, and misinterpretation of reflectivity for rain rate (Berne & Krajewski, 2013; Germann et al., 2006; Villarini & Krajewski, 2010).

Differences in calibration among radar sites can also produce discontinuities at inter-radar boundaries, resulting in banding or abrupt spatial gradients in merged precipitation fields (Gourley et al., 2010). MRMS quality-control procedures mitigate, but cannot eliminate these artifacts.

Spatial smoothing inherent in gridded rainfall products introduces further limitations. The MRMS 1-kilometer grid averages rainfall over each pixel, filtering out small-scale variability that may be hydrologically significant for localized flooding or intense convective bursts (Habibi et al., 2021). As a result, radar products often fail to capture the localized extremes detected by point gages, reinforcing the need for gage-based bias correction.

Furthermore, some studies have identified a time lag of up to 8 minutes between radar-detected precipitation and corresponding ground-level gage measurements, due to the altitude at which radar observes rainfall before it reaches the surface (Ryu et al., 2025). Misalignment of

accumulation periods can therefore create apparent differences that are not strictly due to instrumentation.

7.4 Heteroskedasticity in Data

The overall gage-radar relationship was found to exhibit heteroskedasticity; however, this does not bias the coefficient estimates. Consequently, performance metrics remain valid, particularly when expressed as percentage improvements. The heteroskedasticity does, however, limit the statistical significance of R^2 values when compared in isolation. This limitation is mitigated when R^2 is interpreted in terms of percent improvement or relative to other R^2 values, where it still provides meaningful insight.

In practice, this ensures that higher values can be reliably interpreted as improvements over previous results, which is sufficient for the optimization phase of the methodology selection algorithm. The optimization algorithm's reliance on R^2 and RMSE also prioritized volumetric agreement over spatial contiguity. A potential refinement would be to assign greater weight to spatial consistency and artifact mitigation, even if this necessitates modest trade-offs in volumetric agreement metrics such as R^2 or RMSE.

7.5 Limitations in Quality Control

Other limitations arise from the reliance of radar-based quality control (RBQC) on radar data. Spatial artifacts within radar fields can cause false positives and negatives in gage selection, for example, good gages being excluded or poor-quality gages being retained since no additional corrections were applied to existing artifacts present in the raw radar. Some gages identified as critical by the information index were observed to report values inconsistent with nearby stations, despite passing RBQC checks. While these discrepancies are generally localized and detectable only in regions of high gage density, they point to vulnerabilities in the QC process that could distort precipitation estimates in areas with high gage density.

To further strengthen the QC process, temporal consistency checks could be incorporated to assess gage reliability across past and future observations, thereby improving the accuracy of real-time gage assessments. An additional QC step that flags gages with large deviations relative to nearby stations would also help reduce errors. It is also advisable to correct radar rainfall spatial artifacts through machine learning (ML) methods to identify faulty pixels and remove or rewrite them in future projects.

7.6 Limitations of Temporal Alignment

While Temporal Alignment Quality Control (TAQC) provides a systematic approach to harmonizing irregular gage reporting intervals with fixed radar accumulation windows, several limitations must be recognized. TAQC relies on linear interpolation between irregular gage reports to reconstruct hourly increments; however, this approach inherently smooths rainfall intensity. When interpolation spans large temporal gaps, peak intensities can be attenuated, and

storm durations are artificially lengthened due to the conservation of total accumulation. Although the TAQC threshold constrains allowable interpolation gaps, the threshold itself represents a compromise between data availability and quality: a strict threshold preserves peak intensity but excludes more stations, whereas a relaxed threshold increases spatial coverage at the cost of temporal fidelity.

A second limitation stems from the assumption of approximately uniform rainfall within the interpolation window. This assumption breaks down during high-intensity convective bursts, where rainfall can shift dramatically within minutes. Even when TAQ-TRUE, interpolated values may misrepresent the true timing or magnitude of sub-hourly fluctuations. This is particularly important in tropical or warm-rain environments along the Texas coast, where storm evolution is rapid and intense.

Additionally, TAQC treats each gage independently and does not incorporate information from nearby stations. As a result, isolated gages with irregular reporting may pass TAQC checks despite producing unrealistic increments when compared with neighboring sensors. A spatially informed TAQC incorporating neighborhood consistency checks or probabilistic temporal disaggregation—could help mitigate these artifacts.

7.7 Gap in Understanding of Temporal Relationship

A key research question emerging from this study concerns the correlation between gage count and performance metrics. Although a strong positive temporal correlation and clear spatial pattern similarity between increases in gage count and the performance of bias-correction frameworks were observed in this study, future research should aim to disentangle the temporal effects driven solely by gage density from those arising from concurrent improvements in radar data quality. This would help to more accurately quantify the impact of gage density on the effectiveness of gage-corrected precipitation products. Future efforts could focus on extending the methodology to other regions with different climatic conditions, incorporating machine learning techniques for further accuracy enhancement, and integrating real-time sub-hourly data assimilation for operational nowcasting improvements.

7.8 Regional Biases and Expected Improvement

This study clearly suggests that there is further potential for improvement of bias-corrected products, especially with regionalization. While the exact discrepancy was not quantified in this report, it can be evaluated in subsequent studies by comparing regional products with national level products throughout the United States.

Key recommendations include:

- Application and study on generalization potential of methods discussed in this study to other areas,
- Testing transformation techniques before fitting to address heteroskedasticity,

- Temporal consistency quality control scheme,
- Spatial consistency quality control scheme,
- Addressing artifacts in source radar,
- Differences from other analyses for relationship between gage count vs model performance, and
- Application to real time sub-hourly gridded rainfall correction.

7.9 Opportunities for Further Research

Future research can build upon this work by advancing both methodological refinement and operational applicability. A key opportunity lies in incorporating machine learning and data assimilation techniques to improve spatial interpolation, automate quality control, and dynamically correct radar bias in near real-time. Expanding the framework to other climatic regions would also help evaluate its generalizability and identify parameters requiring regional calibration. Future efforts should explore coupling the rainfall correction framework with hydrologic and hydraulic forecasting systems for predictive applications and early warning operations.

Data Availability

Data for this study is available in the following formats:

- .PCP Files (24-hour basin aggregates)
- .dss File (Hourly Increment)
- Raw .tif files

Credits

Principal Investigator:

Nick Z. Fang, Ph.D., P.E.

Researchers:

Daniel Li, Ph.D., P.E.

Asim Paudel

Reviewers:

Jerry L. Cotter, P.E.

Kelli M. Greenwood, E.I.T.

Matthew T. Lepinski, P.E.

References

Abbas, S. A., & Xuan, Y. (2020). Impact of Precipitation Pre-Processing Methods on Hydrological Model Performance using High-Resolution Gridded Dataset. *Water*, *12*(3), Article 3. https://doi.org/10.3390/w12030840

Ali, J., Wahl, T., Morim, J., Enriquez, A., Gall, M., & Emrich, C. T. (2025). Multivariate compound events drive historical floods and associated losses along the U.S. East and Gulf coasts. *Npj Natural Hazards*, *2*(1), 1–10. https://doi.org/10.1038/s44304-025-00076-5

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L. (2019). A Local Indicator of Multivariate Spatial Association: Extending Geary's c. *Geographical Analysis*, *51*(2), 133–150. https://doi.org/10.1111/gean.12164

Berne, A., & Krajewski, W. F. (2013). Radar for hydrology: Unfulfilled promise or unrecognized potential? *Advances in Water Resources, 35th Year Anniversary Issue*, *51*, 357–366. https://doi.org/10.1016/j.advwatres.2012.05.005

Birch, C. E., Roberts, M. J., Garcia-Carreras, L., Ackerley, D., Reeder, M. J., Lock, A. P., & Schiemann, R. (2015). *Sea-Breeze Dynamics and Convection Initiation: The Influence of Convective Parameterization in Weather and Climate Model Biases*. https://doi.org/10.1175/JCLI-D-14-00850.1

Buda, A. R., Reed, S. M., Folmar, G. J., Kennedy, C. D., Millar, D. J., Kleinman, P. J. A., Miller, D. A., & Drohan, P. J. (2022). *Applying the NWS's Distributed Hydrologic Model to Short-Range Forecasting of Quickflow in the Mahantango Creek Watershed*. https://doi.org/10.1175/JHM-D-21-0189.1

Chen, H., Cifelli, R., Chandrasekar, V., & Ma, Y. (2019). *A Flexible Bayesian Approach to Bias Correction of Radar-Derived Precipitation Estimates over Complex Terrain: Model Design and Initial Verification*. https://doi.org/10.1175/JHM-D-19-0136.1

Chumchean, S., Sharma, A., & Seed, A. (2003). Radar rainfall error variance and its impact on radar rainfall calibration. *Physics and Chemistry of the Earth, Parts A/B/C, Applications of Quantitative Remote Sensing to Hydrology*, *28*(1), 27–39. https://doi.org/10.1016/S1474-7065(03)00005-6

Du, J. (2011). *NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0* (Version 1.0, p. 343130 data files, 2 ancillary/documentation files, 24 GiB) [GRIB-2: Gridded Binary Edition 2 (GRIB)]. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.5065/D6PG1QDD

Duchon, C., Fiebrich, C., & Grimsley, D. (2013). *Using high-speed photography to study undercatch in tipping-bucket rain gauges*. EGU2013-2378. https://ui.adsabs.harvard.edu/abs/2013EGUGA..15.2378D

Dykstra, S. L., & Dzwonkowski, B. (2021). The Role of Intensifying Precipitation on Coastal River Flooding and Compound River-Storm Surge Events, Northeast Gulf of Mexico. *Water Resources Research*, *57*(11), e2020WR029363. https://doi.org/10.1029/2020WR029363

Foehn, A., García Hernández, J., Schaefli, B., & De Cesare, G. (2018). Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. *Journal of Hydrology*, *563*, 1092–1110. https://doi.org/10.1016/j.jhydrol.2018.05.027

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

Germann, U., Galli, G., Boscacci, M., & Bolliger, M. (2006). Radar precipitation measurement in a mountainous region. *Quarterly Journal of the Royal Meteorological Society*, *132*(618), 1669–1692. https://doi.org/10.1256/qj.05.190

Ghasemzadeh, H., Hillman, R. E., & Mehta, D. D. (2024). Toward Generalizable Machine Learning Models in Speech, Language, and Hearing Sciences: Estimating Sample Size and Reducing Overfitting. *Journal of Speech, Language, and Hearing Research*, *67*(3), 753–781. https://doi.org/10.1044/2023_JSLHR-23-00273

Gourley, J. J., Hong, Y., Flamig, Z. L., Li, L., & Wang, J. (2010). *Intercomparison of Rainfall Estimates from Radar, Satellite, Gauge, and Combinations for a Season of Record Rainfall*. https://doi.org/10.1175/2009JAMC2302.1

Grossi, G., Lendvai, A., Peretti, G., & Ranzi, R. (2017). Snow Precipitation Measured by Gauges: Systematic Error Estimation and Data Series Correction in the Central Italian Alps. *Water*, *9*(7), Article 7. https://doi.org/10.3390/w9070461

Habib, E., Krajewski, W. F., & Kruger, A. (2001). Sampling Errors of Tipping-Bucket Rain Gauge Measurements. *Journal of Hydrologic Engineering*, *6*(2), 159–166.

Habib, E., Larson, B. F., & Graschel, J. (2009). Validation of NEXRAD multisensor precipitation estimates using an experimental dense rain gauge network in south Louisiana. *Journal of Hydrology*, *373*(3), 463–478. https://doi.org/10.1016/j.jhydrol.2009.05.010

Habibi, H., Awal, R., Fares, A., & Temimi, M. (2021). Performance of Multi-Radar Multi-Sensor (MRMS) product in monitoring precipitation under extreme events in Harris County, Texas. *Journal of Hydrology*, *598*, 126385. https://doi.org/10.1016/j.jhydrol.2021.126385

Jayawardene, V., Huggins, T. J., Prasanna, R., & Fakhruddin, B. (2021). The role of data and information quality during disaster response decision-making. *Progress in Disaster Science*, *12*, 100202. https://doi.org/10.1016/j.pdisas.2021.100202

Joss, J., & Waldvogel, A. (1990). Precipitation Measurement and Hydrology. In *Radar in Meteorology: Battan Memorial and 40th Anniversary Radar Meteorology Conference* (pp. 577–606). American Meteorological Society. https://doi.org/10.1007/978-1-935704-15-7_39

Kitzmiller, D., Cooten, S. V., Ding, F., Howard, K., Langston, C., Zhang, J., Moser, H., Zhang, Y., Gourley, J. J., Kim, D., & Riley, D. (2011). *Evolving Multisensor Precipitation Estimation Methods: Their Impacts on Flow Prediction Using a Distributed Hydrologic Model*. https://doi.org/10.1175/JHM-D-10-05038.1

Krivoruchko, K., & Gribov, A. (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics*, *32*, 100368. https://doi.org/10.1016/j.spasta.2019.100368

Ley, R., Hellebrand, H., Casper, M. C., & Fenicia, F. (2016). Is Catchment Classification Possible by Means of Multiple Model Structures? A Case Study Based on 99 Catchments in Germany. *Hydrology*, *3*(2), Article 2. https://doi.org/10.3390/hydrology3020022

Li, W., Jiang, H., Li, D., Bedient, P. B., & Fang, Z. N. (2023). Evaluation of Radar Precipitation Products and Assessment of the Gauge-Radar Merging Methods in Southeast Texas for

Extreme Precipitation Events. *Remote Sensing*, *15*(8), Article 8. https://doi.org/10.3390/rs15082033

Lin, Y., & Mitchell, K. E. (2005). *The NCEP stage II/IV hourly precipitation analyses: Development and applications*. Conference on Hydrology, San Diego, CA. https://ams.confex.com/ams/pdfpapers/83847.pdf

Mapiam, P. P., Methaprayun, M., Bogaard, T., Schoups, G., & Ten Veldhuis, M.-C. (2022). Citizen rain gauges improve hourly radar rainfall bias correction using a two-step Kalman filter. *Hydrology and Earth System Sciences*, *26*(3), 775–794. https://doi.org/10.5194/hess-26-775-2022

Martinaitis, S. M., Cocks, S. B., Simpson, M. J., Osborne, A. P., Harkema, S. S., Grams, H. M., Zhang, J., & Howard, K. W. (2021). *Advancements and Characteristics of Gauge Ingest and Quality Control within the Multi-Radar Multi-Sensor System*. https://doi.org/10.1175/JHM-D-20-0234.1

Morbidelli, R., Saltalippi, C., Dari, J., & Flammini, A. (2021). A Review on Rainfall Data Resolution and Its Role in the Hydrological Practice. *Water*, *13*(8), Article 8. https://doi.org/10.3390/w13081012

Muchan, K., & Dixon, H. (2019). Insights into rainfall undercatch for differing raingauge rim heights. *Hydrology Research*, *50*(6), 1564–1576. https://doi.org/10.2166/nh.2019.024

NOAA NCEI. (2021). *Global Historical Climatology Network daily (GHCNd)* [Dataset]. https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily

Paul, S. H., Sharif, H. O., & Crawford, A. M. (2018). Fatalities Caused by Hydrometeorological Disasters in Texas. *Geosciences*, *8*(5), 186. https://doi.org/10.3390/geosciences8050186

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Pollock, M. D., O'Donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M. E., Colli, M., Stagnaro, M., Lanza, L. G., Lewis, E., Kilsby, C. G., & O'Connell, P. E. (2018). Quantifying and Mitigating Wind-Induced Undercatch in Rainfall Measurements. *Water Resources Research*, *54*(6), 3863–3875. https://doi.org/10.1029/2017WR022421

Ryu, S., Song, J. J., & Lee, G. (2025). Radar–Rain Gauge Merging for High-Spatiotemporal-Resolution Rainfall Estimation Using Radial Basis Function Interpolation. *Remote Sensing*, *17*(3), Article 3. https://doi.org/10.3390/rs17030530

Segovia-Cardozo, D. A., Bernal-Basurco, C., & Rodríguez-Sinobas, L. (2023). Tipping Bucket Rain Gauges in Hydrological Research: Summary on Measurement Uncertainties, Calibration, and Error Reduction Strategies. *Sensors*, *23*(12), 5385. https://doi.org/10.3390/s23125385

Segovia-Cardozo, D. A., Rodríguez-Sinobas, L., Díez-Herrero, A., Zubelzu, S., & Canales-Ide, F. (2021). Understanding the Mechanical Biases of Tipping-Bucket Rain Gauges: A Semi-Analytical Calibration Approach. *Water*, *13*(16), 2285. https://doi.org/10.3390/w13162285

Sexton, A. M., Sadeghi, A. M., Zhang, X., & Srinivasan, A. S. (2010). Using NEXRAD and Rain Gauge Precipitation Data for Hydrologic Calibration of SWAT in a Northeastern Watershed. *Transactions of the ASABE*, *53*(5), 1501–1510.

Shakti, P. C., Tsuyoshi, N., & Ryohei, M. (2019). The Role of the Spatial Distribution of Radar Rainfall on Hydrological Modeling for an Urbanized River Basin in Japan. *Water*, *11*(8), Article 8. https://doi.org/10.3390/w11081703

Smith, A. B. (2025). *U.S. Billion-dollar Weather and Climate Disasters, 1980—Present (NCEI Accession 0209268)* [Dataset]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/STKW-7W73

So, E. (2023). Data and its role in reducing the risk of disasters in the built environment. *Natural Hazards*, *119*(2), 1127–1130. https://doi.org/10.1007/s11069-022-05590-7

Steven, H. (1996). WSR-88D Radar Rainfall Estimation: Capabilities, Limitations and Potential Improvements. *Natl. Weather. Dig*, *20*, 26–38.

Suh, S.-H., Kim, H.-J., Lee, D.-I., & Kim, T.-H. (2021). *Geographical Characteristics of Raindrop Size Distribution in the Southern Parts of South Korea*. https://doi.org/10.1175/JAMC-D-20-0102.1

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Synoptic Data Program. (2024). *Synoptic Data API (rain gauge observations)* [Dataset]. https://synopticdata.com

Tan, M. L., Gassman, P. W., Liang, J., & Haywood, J. M. (2021). A review of alternative climate products for SWAT modelling: Sources, assessment and future directions. *Science of The Total Environment*, *795*, 148915. https://doi.org/10.1016/j.scitotenv.2021.148915

Tokay, A., Bashor, P. G., & McDowell, V. L. (2010). *Comparison of Rain Gauge Measurements in the Mid-Atlantic Region*. https://doi.org/10.1175/2009JHM1137.1

Villarini, G., & Krajewski, W. F. (2010). Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall. *Surveys in Geophysics*, *31*(1), 107–129. https://doi.org/10.1007/s10712-009-9079-x

Villarini, G., Mandapaka, P. V., Krajewski, W. F., & Moore, R. J. (2008). Rainfall and sampling uncertainties: A rain gauge perspective. *Journal of Geophysical Research: Atmospheres*, *113*(D11). https://doi.org/10.1029/2007JD009214

Wang, H., Ren, L. L., & Liu, G. H. (2009). A regression-kriging model for estimation of rainfall in the Laohahe basin. *International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining*, *7492*, 1321–1329. https://doi.org/10.1117/12.838036

Wang, J., Fisher, B. L., & Wolff, D. B. (2008). *Estimating Rain Rates from Tipping-Bucket Rain Gauge Measurements*. https://doi.org/10.1175/2007JTECHA895.1

Wang, S., Zhang, K., Chao, L., Li, D., Tian, X., Bao, H., Chen, G., & Xia, Y. (2021). Exploring the utility of radar and satellite-sensed precipitation and their dynamic bias correction for integrated prediction of flood and landslide hazards. *Journal of Hydrology*, *603*, 126964. https://doi.org/10.1016/j.jhydrol.2021.126964

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Ware, E. C. (2005). *Corrections to radar-estimated precipitation using observed rain gauge data.* [M.S. Thesis, Cornell University]. https://ecommons.cornell.edu/bitstream/1813/2115/1/pdfthesis.pdf

Wootten, A., & Boyles, R. P. (2014). *Comparison of NCEP Multisensor Precipitation Estimates with Independent Gauge Data over the Eastern United States*. https://doi.org/10.1175/JAMC-D-14-0034.1

Yang, P., & Ng, T. L. (2019). Fast Bayesian Regression Kriging Method for Real-Time Merging of Radar, Rain Gauge, and Crowdsourced Rainfall Data. *Water Resources Research*, *55*(4), 3194–3214. https://doi.org/10.1029/2018WR023857

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(W09417). https://doi.org/https://doi.org/10.1029/2007WR006716

Yoon, S.-S., Phuong, A. T., & Bae, D.-H. (2012). *Quantitative Comparison of the Spatial Distribution of Radar and Gauge Rainfall Data*. https://doi.org/10.1175/JHM-D-11-066.1

Yu, M., Yang, C., & Li, Y. (2018). Big Data in Natural Disaster Management: A Review. *Geosciences*, *8*(5), Article 5. https://doi.org/10.3390/geosciences8050165

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S., Arthur, A., Cooper, K., Brogden, J., & Kitzmiller, D. (2016). *Multi-Radar*

Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. https://doi.org/10.1175/BAMS-D-14-00174.1

Zhang, J., Howard, K., Langston, C., Vasiloff, S., Kaney, B., Arthur, A., Cooten, S. V., Kelleher, K., Kitzmiller, D., Ding, F., Seo, D.-J., Wells, E., & Dempsey, C. (2011). *National Mosaic and Multi-Sensor QPE (NMQ) System: Description, Results, and Future Plans*. https://doi.org/10.1175/2011BAMS-D-11-00047.1

Zhang, J., Qi, Y., Langston, C., & Kaney, B. (2012). Radar Quality Index (RQI)—A combined measure of beam blockage and VPR effects in a national network. *Weather Radar and Hydrology*, *351*, 388–393.

Zhang, J., Tang, L., Cocks, S., Zhang, P., Ryzhkov, A., Howard, K., Langston, C., & Kaney, B. (2020). *A Dual-Polarization Radar Synthetic QPE for Operations*. https://doi.org/10.1175/JHM-D-19-0194.1

Appendix 1: Rainfall Gage Corrector Toolkit, Demo and Code

Toolkit and a demo video have been attached externally and submitted with this document.

Appendix 2: Toolkit Documentation

This appendix provides documentation for the rainfall bias-correction toolkit developed for the Coastal Texas Radar Recalibrated Framework Technical Report (Contract No. #2301792723). The toolkit enables users to preprocess, correct, and analyze gridded precipitation products within the study domain. The following sections describe the program inputs, execution steps, outputs, requirements, restrictions, and licensing considerations.

A2.1 Input Data Requirements

Gridded Radar or Satellite Data:

- Input folder: Inputs\RadarOnly TIF\TIF\...
- Format: GeoTIFF (.tif)
- Projection: EPSG:4326 (WGS84)
- Temporal resolution: Hourly increments
- Units: Inches
- Naming convention: *YYYYMMDDHHMM.tif (UTC)

Area of Interest (AOI) Shapefile:

- Must consist of a single polygon.
- The attribute table should be empty (avoid unnecessary columns).
- Projection system units of the AOI must be known (used for extent buffers).
- The shapefile does not need to be copied into the input folder; it can be selected directly within the program.

A2.2 Program Execution

- Select program mode(s)
    - Download/QC Gages (Use the force option if you change AOI)
    - Gage Correct Rainfall
- Specify event period in UTC.
- Provide AOI shapefile path.
- Provide buffer to AOI in shapefile projection units.
- Click “Run Program.”

A2.3 Output Data

Major Outputs:

- Processed GeoTIFFs (bias-corrected rainfall grids). (Outputs/Processed TIF)

Intermediate files:

- Gage stations Lists: (Processing/myStationList.csv or Processing/downloadLog.csv)
- Raw downloaded CSVs: (Rainfall_Processor/RAW CSV/<stid>.csv)
- Temporally aligned processed files (incremental shapefiles, CSVs, and images): (Rainfall_Processor/Temporally Aligned Data/<stid>.csv)

Beyond the files stated here, other intermediate files are produced that the program requires for execution but have not been vetted for viewer quality.

A2.4 System Requirements

- Operating system: Windows 11
- Processor: Intel i5 or later
- Memory: Minimum 4 GB RAM

A2.5 Restrictions

- The toolkit will only operate within the spatial extents of the original study area (see Report).
- The toolkit can only be applied to rainfall events occurring up to five years prior to the current date.

A2.6 Debugging and Error Handling

The toolkit includes a basic debugger that provides descriptive exception messages for common issues. However, error handling is not exhaustive, and some algorithmic or data-specific issues may not be captured.

A2.7 Licensing and Proprietary Rights

This toolkit and all associated code, algorithms, documentation, and outputs are the intellectual property of Asim Paudel and the University of Texas at Arlington Water Engineering Research Center (WERC WERC).

Usage rights: Currently limited to the Texas Water Development Board (TWDB) for testing the procedures described in the technical report.

Restrictions: No copying, modification, redistribution, sublicensing, or commercial use is permitted without prior written consent.

The authors reserve the right to revise licensing terms, restrict access, or monetize the toolkit and its outputs in the future.

Users are strongly advised to prefer pre-generated rainfall datasets supplied with the report, which have undergone additional outlier identification and correction.

This program is intended strictly for research purposes. The authors disclaim responsibility for any unsanctioned use of the toolkit or its derived products.

A2.8 Author's Comments

The graphical user interface (GUI) was developed pro bono and was not included in the original contract scope.

Appendix 3: Monthly Progress Reports

Attached externally and submitted with this document.

Appendix 4: Product TIF Files

Attached externally and submitted with this document.

Appendix 5: DSS File

Attached externally and submitted with this document.

Appendix 6: PCP Files

Attached externally and submitted with this document.

Appendix 7: Temporal Alignment Methods with Equations

This approach converts irregularly timed (t_i) gage accumulations (a_i) into regularly timed (t_j) hourly values (a_j) while preserving the total observed accumulation. Here t_{i+1} corresponds to the first irregular timestep directly after required regular point.

$$a_j = \frac{a_{i+1}t_j - a_i t_j + a_i t_{i+1} - a_{i+1}t_j}{t_{i+1} - t_i} \quad \text{xix}$$

$$Where, \quad t_{i(Irregular\ Points)} < t_{j(Regular\ Points)} < t_{i+1(Irregular\ Points)}$$

For simplicity consider irregular points (e.g. reports at 32nd min mark every 2 hours) and regular points (e.g. reports at 0th min mark every hour) have consistent temporal spacing (e.g. 2hour and 1hour). Let a_{i+x} be accumulations after *x* regular timesteps where, $x \in \mathbb{Z}^+$ (non-negative integers). Let I_{i+x} be intervals, signifying rainfall increment at regular one unit of temporal spacing, then from equation xix.

$$a_{j+x} = \frac{a_{i+x+1}t_{j+x} - a_{i+x}t_{j+x} + a_{i+x}t_{i+x+1} - a_{i+x+1}t_{i+x}}{t_{i+x+1} - t_{i+x}} \quad \text{xx}$$

$$I_{j+x} = a_{j+x} - a_{j+x-1} \quad \text{xxi}$$

Similarly, let I_{i+x} and I_{i+x+1} be observed intervals before interpolations are carried out.

$$I_{i+x} = a_{i+x} - a_{i+x-1} \quad \text{xxii}$$

$$I_{i+x+1} = a_{i+x+1} - a_{i+x} \quad \text{xxiii}$$

Since a_{j+x} is linearly interpolated from a_{i+x} and a_{i+x+1}, it is bounded by the same. Also, since $a_{i+x} \leq a_{i+x+1}$, we can write $a_{j+x} \in [a_{i+x}, a_{i+x+1}]$. Continuing from equation xx:

$$a_{j+x} = a_{i+x} + (a_{i+x+1} - a_{i+x})\frac{t_{j+x} - t_{i+x}}{t_{i+x+1} - t_{i+x}}$$

Substituting equation xxiii into equation xx,

$$a_{j+x} = a_{i+x} + I_{i+x+1}\frac{t_{j+x} - t_{i+x}}{t_{i+x+1} - t_{i+x}}$$

$$a_{j+x} - a_{i+x} = I_{i+x+1}\frac{t_{j+x} - t_{i+x}}{t_{i+x+1} - t_{i+x}} \quad \text{xxiv}$$

Similarly for a prior index,

$$a_{j+x-1} - a_{i+x-1} = I_{i+x}\frac{t_{j+x-1} - t_{i+x-1}}{t_{i+x} - t_{i+x-1}} \quad \text{xxv}$$

Subtracting equation xxv from xxiv,

$$a_{j+x} - a_{i+x} - a_{j+x-1} + a_{i+x-1} = I_{i+x+1}\frac{t_{j+x} - t_{i+x}}{t_{i+x+1} - t_{i+x}} - I_{i+x}\frac{t_{j+x-1} - t_{i+x-1}}{t_{i+x} - t_{i+x-1}}$$

$$I_{j+x} - I_{i+x} = I_{i+x+1}\frac{t_{j+x} - t_{i+x}}{t_{i+x+1} - t_{i+x}} - I_{i+x}\frac{t_{j+x-1} - t_{i+x-1}}{t_{i+x} - t_{i+x-1}}$$

$$I_{j+x} = I_{i+x+1}\frac{t_{j+x} - t_{i+x}}{t_{i+x+1} - t_{i+x}} + I_{i+x}\left[1 - \frac{t_{j+x-1} - t_{i+x-1}}{t_{i+x} - t_{i+x-1}}\right]$$

$$I_{j+x} = I_{i+x+1}\frac{t_{j+x} - t_{i+x}}{t_{i+x+1} - t_{i+x}} + I_{i+x}\frac{t_{i+x} - t_{j+x-1}}{t_{i+x} - t_{i+x-1}} \quad \text{xxvi}$$

For simplicity, assume irregular intervals *X* are consistent, $t_{i+x} - t_{i+x-1} = t_{i+x+1} - t_{i+x} = X$

Similarly, assume regular intervals *Y* are consistent, $t_{j+x} - t_{j+x-1} = t_{j+x+1} - t_{j+x} = Y$

Then, lag *L* between two timeseries, $t_{j+x-1} - t_{i+x-1} = L_{j+x-1}$, $t_{j+x} - t_{i+x} = L_{j+x}$

Similarly, complement of lag L^c between two timeseries, $t_{i+x} - t_{j+x-1} = X - L_{j+x-1} = L^c_{j+x-1}$

Lag and its symmetric counterpart is constrained by irregular interval,

$$L \leq X \text{ and } Lc \leq X \Rightarrow O(L) = X \text{ , and } O(L^C) = X \quad \text{xxvii}$$

Substituting assumptions on equation xxvi,

$$I_{j+x} = I_{i+x+1}\frac{t_{j+x} - t_{i+x}}{X} + I_{i+x}\frac{t_{i+x} - t_{j+x-1}}{X}$$

$$I_{j+x} = I_{i+x+1}\frac{Y - X + L_{j+x-1}}{X} + I_{i+x}\frac{X - L_{j+x-1}}{X}$$

$$\frac{I_{j+x}}{Y} = \frac{I_{i+x+1}}{X}\frac{Y - X + L_{j+x-1}}{Y} + \frac{I_{i+x}}{X}\frac{X - L_{j+x-1}}{Y} \quad \text{xxviii}$$

Let *P* denote precipitation intensity i.e. interval per unit time, then from equation xxviii:

$$P_{j+x} = P_{i+x+1}\frac{Y - X + L_{j+x-1}}{Y} + P_{i+x}\frac{X - L_{j+x-1}}{Y}$$

$$P_{j+x} = P_{i+x+1}\left[1 - \left(\frac{X}{Y} - \frac{L_{j+x-1}}{Y}\right)\right] + P_{i+x}\left[\frac{X}{Y} - \frac{L_{j+x-1}}{Y}\right] \quad \text{xxix}$$

Calculating term $\frac{L_{j+x-1}}{Y}$,

$$\frac{L_{j+x-1}}{Y} = \frac{t_{j+x} - Y - t_{i+x} + X}{Y} = \frac{L_{j+x} - Y + X}{Y} = \frac{L_{j+x} - Y + X}{Y} = \frac{L_{j+x}}{Y} + \frac{X}{Y} - 1$$

Substituting term $\frac{L_{j+x-1}}{Y}$ to equation xxix,

$$P_{j+x} = P_{i+x+1}\left[1 - \left(\frac{X}{Y} - \frac{L_{j+x}}{Y} - \frac{X}{Y} + 1\right)\right] + P_{i+x}\left[\frac{X}{Y} - \frac{L_{j+x}}{Y} - \frac{X}{Y} + 1\right]$$

$$P_{j+x} = P_{i+x+1}\left[\frac{L_{j+x}}{Y}\right] + P_{i+x}\left[1 - \frac{L_{j+x}}{Y}\right] \quad \text{xxx}$$

Thus, from equation xxx it is observed that P_{j+x} is a convex linear combination of P_{i+x+1} and P_{i+x}, i.e. the coefficients of the linear combination add up to 1. Thus, for $L_{j+x} \neq 0$, if $P_{i+x} < P_{i+x+1}$ then $P_{j+x} < P_{i+x+1}$ and, vice versa if $P_{i+x+1} < P_{i+x}$ then $P_{j+x} < P_{i+x}$. If $P_{i+x} = P_{i+x+1}$, then $P_{j+x} = P_{i+x}$. Thus, it can be inferred that $P_j \leq Max(P_i)$.

For P_j to be as close to the maximum value as possible, one of the coefficients of the term in the linear combination should be close to 1 or, one of the coefficients should tend to 0. For case 1: $P_{j+x} = P_{i+x}$ the coefficient $\left[\frac{L_{j+x}}{Y}\right] \approx 0$ or simply, $L_{j+x} \approx 0$. Similarly for case 2: $P_{j+x} = P_{i+x+1}$ the coefficient $\left[1 - \frac{L_{j+x}}{Y}\right] \approx 0$ or, $Y - L_{j+x} \approx 0$ or, $L^c_{j+x-1} \approx 0$. Also, $\lim_{Y\to\infty}\left[1 - \frac{L_{j+x}}{Y}\right] = 0$ or $Y \to \infty$ is also a sufficient criterion. In simple terms, if regular spacing is large compared to irregular temporal spacing, the large intensity from irregular timestep is stored in the directly next regular timestep. Regardless of the criterion, to retain maximum intensity, regular timestep should be as close as possible with irregular timestep corresponding to maximum intensity.

Reduction to maximum intensity for one step can be obtained by dividing equation xxx by P_{i+x}:

$$\frac{P_{j+x}}{P_{i+x}} = \frac{P_{i+x+1}}{P_{i+x}}\left[\frac{L_{j+x}}{Y}\right] + \left[1 - \frac{L_{j+x}}{Y}\right] = \frac{P_{i+x+1}}{P_{i+x}}\left[\frac{L_{j+x}}{Y}\right] + \left[\frac{L^c_{j+x-1}}{Y}\right] \quad \text{xxxi}$$

Maximum reduction is given by upper bound to equation xxxi leveraging upper bound of lag from equation xxvii:

$$O\left(\frac{P_{j+x}}{P_{i+x}}\right) = \frac{X}{Y}\left[\frac{P_{i+x+1}}{P_{i+x}} + 1\right] = \frac{X}{Y} \times k$$

Assume the constant part "k" since those values cannot be altered. The symmetric condition where $P_{i+x} < P_{i+x+1}$ gives similar result the proof of which is left to the reader. xxxii

Appendix 8: Normalized Composite Information Index Definition

To facilitate information-based classification, first the spatial informativeness of each gage is quantitatively assessed. Discrepancy between gage and radar values are computed and scaled in a way that linearizes the differences. The scaled discrepancy *G* for each gage is unitless and is defined as:

$$G = \begin{cases} log_{10}\left(\frac{Gage\ value}{radar\ value}\right), & if\ Gage\ value \neq\ 0\ and\ radar\ \neq\ 0 \\ 0, & if\ Gage\ value = radar\ value = 0\ or\ radar\ value = 0 \\ -1, & if\ Gage\ value = 0\ and\ radar\ value \neq 0\ (edge\ case\ only) \end{cases} \quad \text{xxxiii}$$

For each gage, a neighborhood is defined as a circular buffer of radius 0.35° (optimized for computational efficiency), which includes all nearby gages, including the gage itself. Let the set of gages in the neighborhood be indexed by *n*, and the gage of interest by *i*. The value at gage *i* is denoted by G_i and the values of its neighboring gages are denoted by G_n. The distance between gage *i* and each neighboring gage is given by $D_{i,n}$. The total number of gages in the neighborhood is *N*. A custom information index is then defined based on the following components:

Weight $(w_{i,n})$ is calculated from each gage in neighborhood with respect to index gage:

$$w_{i,n} = \frac{\frac{1}{D_{i,n}^2}}{\sum_{n,\ n \neq i}\left(\frac{1}{D_{i,n}^2}\right)} \quad \text{xxxiv}$$

Neighborhood average gage value:

$$\bar{G}_n = \frac{\sum_n G_n}{N} \quad \text{xxxv}$$

Average difference information index:

$$H_{Avg.Diff_i} = \sum_{n,\ n \neq i} (G_i - G_n) \times w_{i,n}$$

$$H_{Avg.Diff.Norm_i} = \frac{\left|H_{Avg.Diff_i}\right|}{\max\left(\left|H_{Avg.Diff_i}\right|\right)} \quad \text{xxxvi}$$

Moran's I information index:

$$H_{Moran's\ I_i} = \frac{(G_i - \bar{G}_n) \times (N-1)}{\sum_{n,\ n \neq i}(G_i - \bar{G}_n)^2} \sum_{n,\ n \neq i}(G_n - \bar{G}_n) \times w_{i,n}$$

$$E(H_{Moran's\ I}) = \frac{-\sum_{n,\ n \neq i} w_{i,n}}{N-1}$$

$$H_{Moran's\ I.Norm_i} = \frac{\left|H_{Moran's\ I_i} - E(H_{Moran's\ I})\right|}{\max\left(\left|H_{Moran's\ I_i} - E(H_{Moran's\ I})\right|\right)} \quad \text{xxxvii}$$

Geary's C information index:

$$H_{Geary's\ C} = \frac{N-1}{\sum_{n,\ n\neq i}(G_i - \bar{G}_n)^2} \sum_{n,\ n\neq i} (G_i - G_n)^2 \times w_{i,n}$$

$$E(H_{Geary's\ C}) = \frac{2\sum_{n,\ n\neq i} w_{i,n}}{N-1}$$

$$H_{Geary's\ C.Norm_i} = \frac{\left|H_{Geary's\ C_i} - E(H_{Geary's\ C})\right|}{\max\left(\left|H_{Geary's\ C_i} - E(H_{Geary's\ C})\right|\right)} \quad \text{xxxviii}$$

Lastly composite information index is derived as the average of the three normalized indices from equation xxxvi, xxxvii, and xxxviii.

$$Comp.Information\ Index_i = \frac{\left(H_{Avg.Diff.Norm_i} + H_{Moran's\ I.Norm_i} + H_{Geary's\ C.Norm_i}\right)}{3} \quad \text{xxxix}$$

The composite index is a normalized quantifier, i.e. $Comp.Information\ Index_i \in [0,1]$. Its value represents the relative importance of one gage data point with respect to another at a given time-period.

Appendix 9: Technical White Paper

This study presents an enhanced radar precipitation product generated by an advanced bias-correction framework that integrates rain gage data retrieval, multi-stage quality control, radar-gage data fusion, information-based validation, and an optimal spatial interpolation method. The results indicate a significant reduction in errors, improved spatial consistency, and stronger correlation with ground observations, supporting the efficacy of the proposed approach. The regression-based approach is especially valuable in areas with sparse gage networks, as regression applies a base correction to the entire raster based on the average relationship between gage and radar, thus not limiting corrections to areas that are near gages only.

To ensure a larger volume of gage data for bias correction, the criteria for gage selection were relaxed, while the quality control methodology was simultaneously enhanced to ensure that increasing gage network sources would not come at the cost of overall quality. Three distinct quality control steps were implemented: accumulation reset adjustments (AQC), temporal alignment checks (TAQC), and radar-based outlier correction (RBQC).

A key part of the method was selecting validation points based on how much unique information each gage provided. By classifying gages based on an information index, this approach avoids the limitations of conventional cross-validation methods, such as overrepresentation of trivial gages, exclusion of critical information, or large computational overhead. The use of information-rich non-critical (IRNC) gages for validation ensures both spatial representativeness and event significance while enabling consistent and objective model evaluation.

Various interpolation methods and their variants were evaluated. The optimal model, reflecting the regional distribution of gages and validated using regional events, was selected using Bayesian optimization from among several competing bias-correction models, balancing R^2, root mean square error (RMSE), and computational efficiency. The final method chosen was Regression-Attenuated Inverse Distance Weighting (RAIDW), which interpolates regression residuals with a modified inverse distance weighting (IDW) approach. Method parameters were individually optimized before application to the full study period.

The generated rainfall product (2015 – 2024) for the Texas coastal region showed better agreement with gage data in both general, meaning all storms pooled across the full period, and event-specific tests, meaning performance evaluated for individual major storms such as Hurricane Harvey. Results indicate substantial improvement across all performance categories. The validation set showed an average improvement of 67.2%, while the calibration set showed an improvement of 94.7%. Event-based analysis for Hurricane Harvey yielded a 73.1% improvement. Year-by-year performance metrics confirmed consistent gains in bias correction, with especially high percent improvements in earlier years where raw (original) radar product quality was comparatively lower. While these results demonstrate clear advances, residual differences remain. The corrected product shows a consistently positive bias relative to gages,

which is reasonable given temporal alignment challenges and gage uncertainty at high precipitation levels.

Hydrologic validation using a pre-calibrated HEC-HMS model in Brays Bayou, Houston, during Hurricane Harvey further confirmed the utility of the corrected dataset. Without direct calibration, the WERC product achieved a Nash–Sutcliffe Efficiency (NSE) of 0.91, outperforming alternative datasets including MRMS and Stage IV in both volumetric agreement and temporal dynamics. The hydrograph derived using the product captured both volumetric totals and temporal features with greater accuracy. Additionally, the WERC product was shown to closely cluster with the Multi-Radar Multi-Sensor Gage Corrected (MRMS_GageCorr) product and clearly separate from the raw radar and unprocessed gage data, indicating that the corrections made are meaningful and consistent with trusted precipitation datasets.

The bias-correction framework's primary goal was to achieve volumetric agreement with gages while enabling scalable processing over large datasets. In doing so, modest concessions were made in spatial contiguity. For example, corrections occasionally appear as "cloudy blobs" centered around dense gaged clusters. Even though RAIDW was selected based on optimization results, the framework is flexible and can be adapted as needed. For instance, kriging methods produced smoother results during hyperparameter testing and can be incorporated in future studies if spatial contiguity is to be prioritized over volumetric agreement.

This study highlights a strong correlation between gage density and model performance (R^2 = 0.794), showing that while additional gages consistently reduce RMSE and improve R^2, their effect on bias is comparatively limited. For climatological applications, this suggests that moderate gage coverage may be sufficient, whereas short-term hydrologic applications benefit more substantially from higher gage density. Differences in spatial patterns between this study's product and the widely used MRMS_GageCorr product further underscore the importance of gage density, with the regionally optimized QC and expanded network applied here driving density variations that correspond to rainfall field differences. The resulting moderate spatial structural similarity (SSIM = 0.5) suggests that gage density improvements meaningfully enhance corrected precipitation products.

Looking ahead, opportunities exist to extend the methodology to regions with diverse climatic conditions, incorporate machine learning for improved spatial interpolation and quality control, and integrate real-time data assimilation for operational forecasting. Finally, additional QC steps, including temporal consistency checks and neighbor-based deviation filters, could further enhance robustness in dense gage networks by reducing the influence of outliers without compromising spatial representativeness.