Improving Micro Rain Radar Rain Rate Retrieval in Different Precipitation Regimes Using AI Methods

Jasmine Trevino, Kalyn Utenis, and Dr. Chuntao Liu
Texas A&M University - Corpus Christi
(jtrevino79@islander.tamucc.edu, kutenis@islander.tamucc.edu)

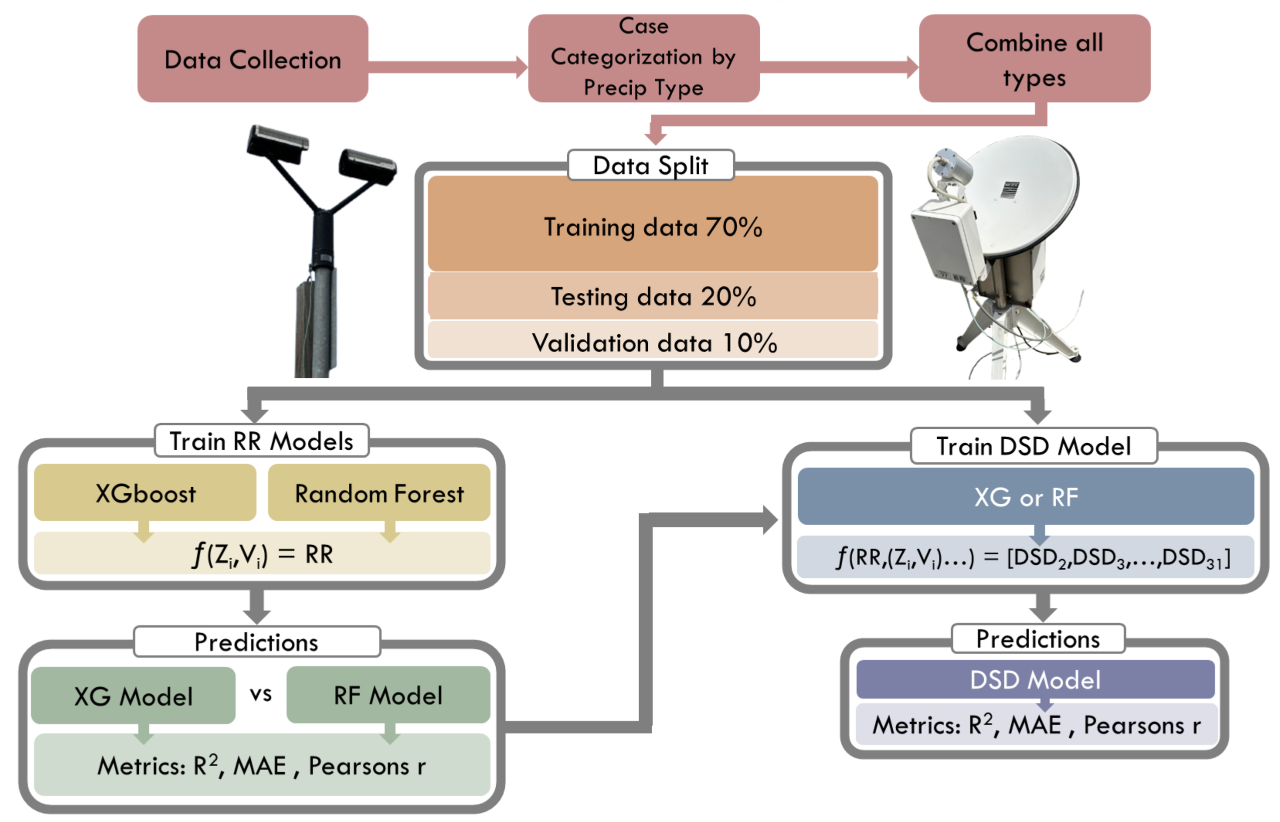
Research Questions/Objectives:

- Improve rain rate estimates across various precipitations regimes using AI
- Can one AI model predict different precipitation regimes accurately?
- Can AI model predict drop size distribution using MRR-Pro Observations?

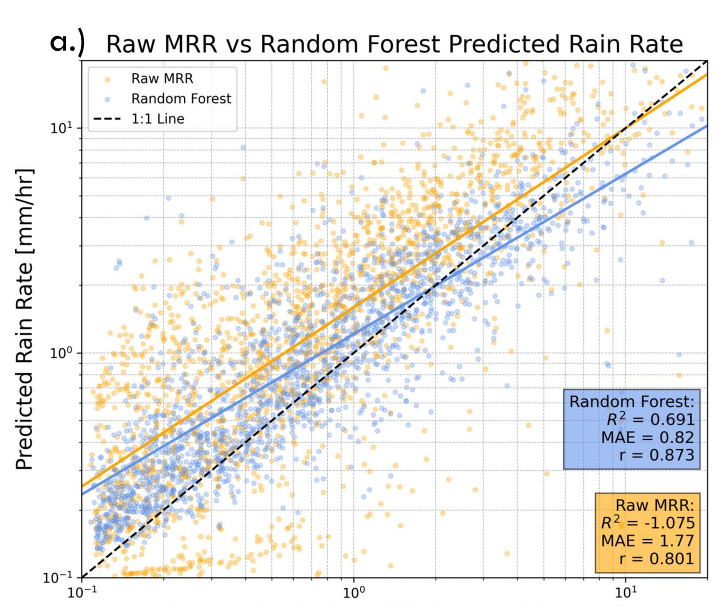
Instrumentation:

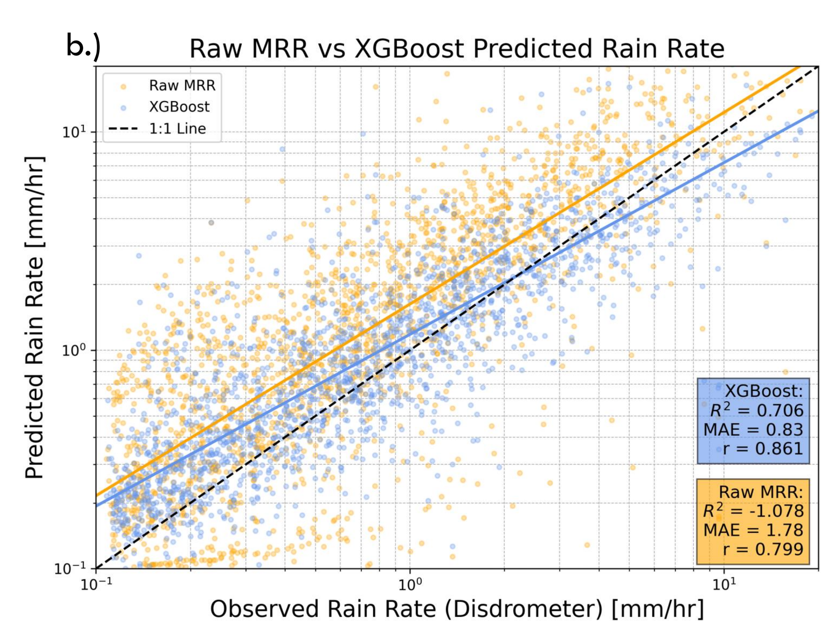
- MRR-Pro – A vertical pointing radar operating at K-Band (24 GHz) to observe precipitation at different altitudes
  - Vertical Profiles of precipitation properties such as reflectivity (R), doppler velocity (V), and derived rain rate (RR)
- Parsivel Disdrometer – Uses laser beam to measure the size and fall velocity of rain drops passing through
  - Provides drop diameter (0.2 mm to 24.5 mm), count, and derives rain rate

Methodology:



Random Forest VS XGboost:





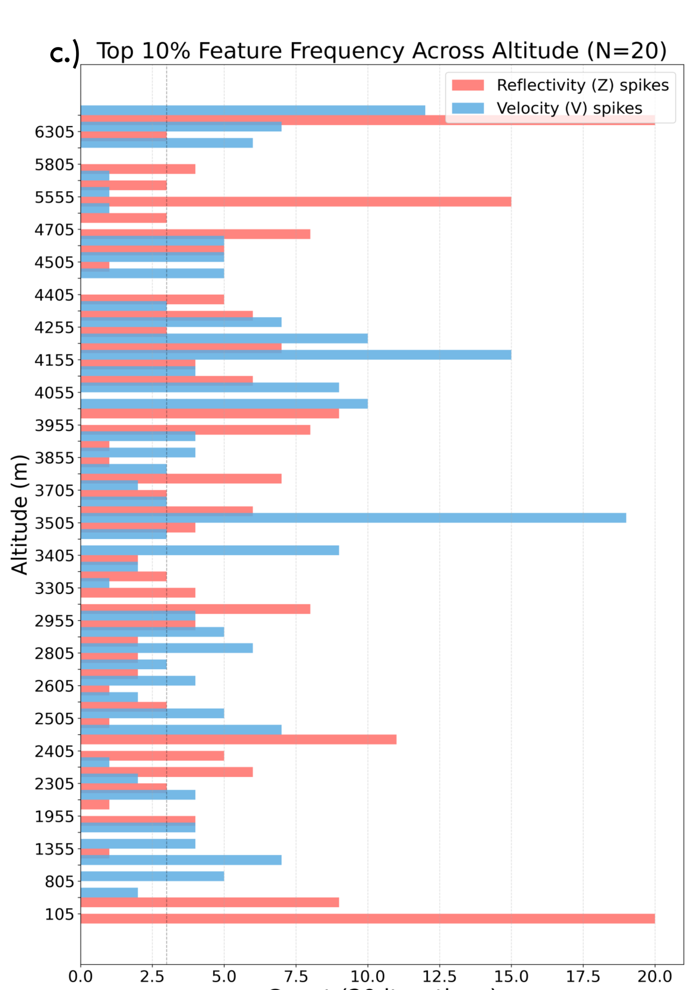
- Random Forest (RF, Fig. a) and XGboost (XG, Fig. b) models' performances are compared to the raw MRR-Pro RR retrievals across all precipitation types
- R squared = coefficient of determination, measures how accurately the model predicts vs the observed
- The orange dots represent the MRR-Pro and the blue dots represent the models
- The MRR-Pro, although linearly correlated with the disdometer (Pearson's r ≈ .8), consistently overestimates RR across most samples
- Both regression models demonstrate substantial improvement from the MRR-Pro raw RR
- Choosing the "best" model just from these two figures is difficult considering how similar they performed

Model Performance by Precipitation Regime

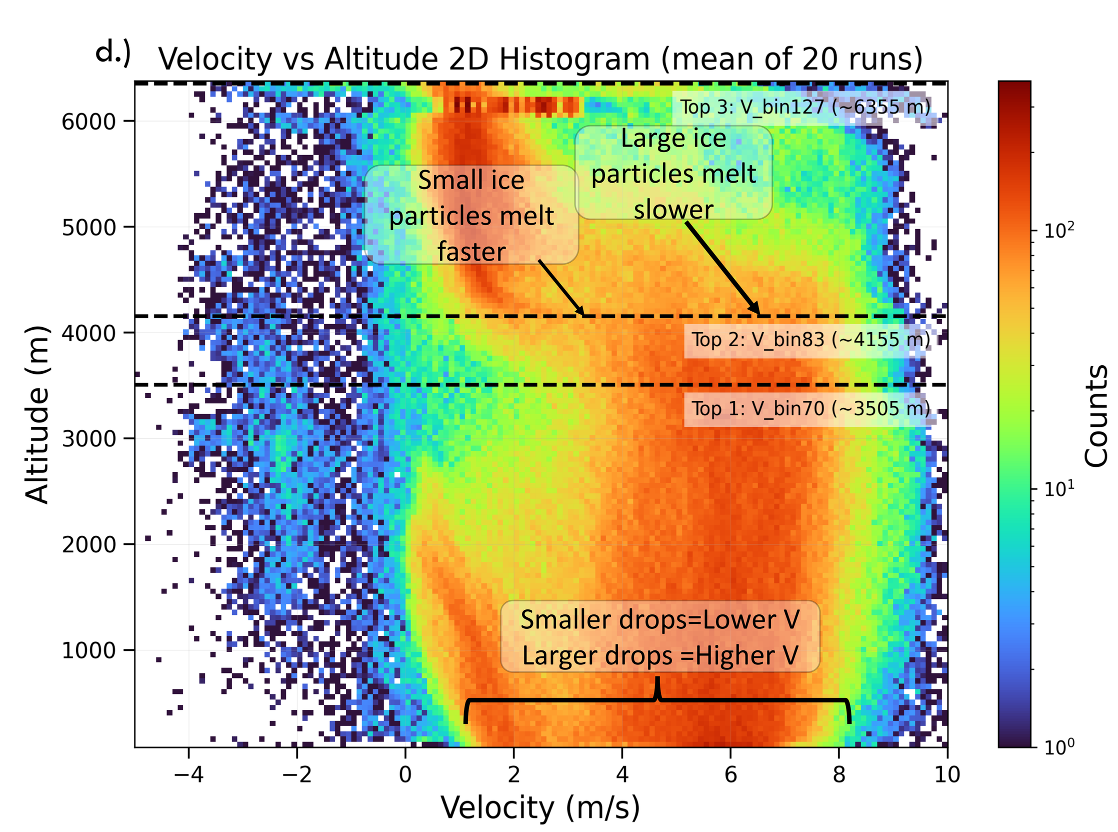
	Metric	All	Thunderstorm	Light/ Moderate	Tropical Storms	Cold Front	Heavy Rain	Drizzle	Warm Rain
MRR-Pro RR at 105m	R^2	-1.08	-11.16	-19.36	-4.01	-5.8	-3.65	-1.61	-3.44
	Pearson's r	0.79	0.9	0.87	0.69	0.82	0.77	0.76	0.85
	MAE (mm/hr)	1.78	4.81	1.79	7.82	1.54	5.48	0.51	7.69
XGboost	R^2	0.70	0.65	0.93	0.63	0.85	0.57	0.57	0.91
	Pearson's r	0.83	0.92	0.77	0.90	0.93	0.91	0.74	0.97
	MAE (mm/hr)	0.86	0.62	0.26	1.35	0.23	2.31	0.27	0.94
Random Forest	R^2	0.69	0.68	0.85	0.53	0.91	0.64	-1.53	0.58
	Pearson's r	0.87	0.94	0.84	0.89	0.96	0.91	0.81	0.87
	MAE (mm/hr)	0.82	0.64	0.26	1.67	0.16	2.31	0.44	3.34
Sample size	(minute)	9995	4411	3249	990	805	441	7	92

Table 1.) Each precipitation regime is evaluated independently, and model performance metrics are used to determine which model should advance to the next stage. Regimes with smaller sample sizes (e.g., warm rain and drizzle) show the largest differences between RF and XG performance, with XG providing notably better estimates. This improvement is likely due to XG boosted-tree learning structure, which captures more nonlinear patterns more effectively in limited datasets.

XG Feature Importance:



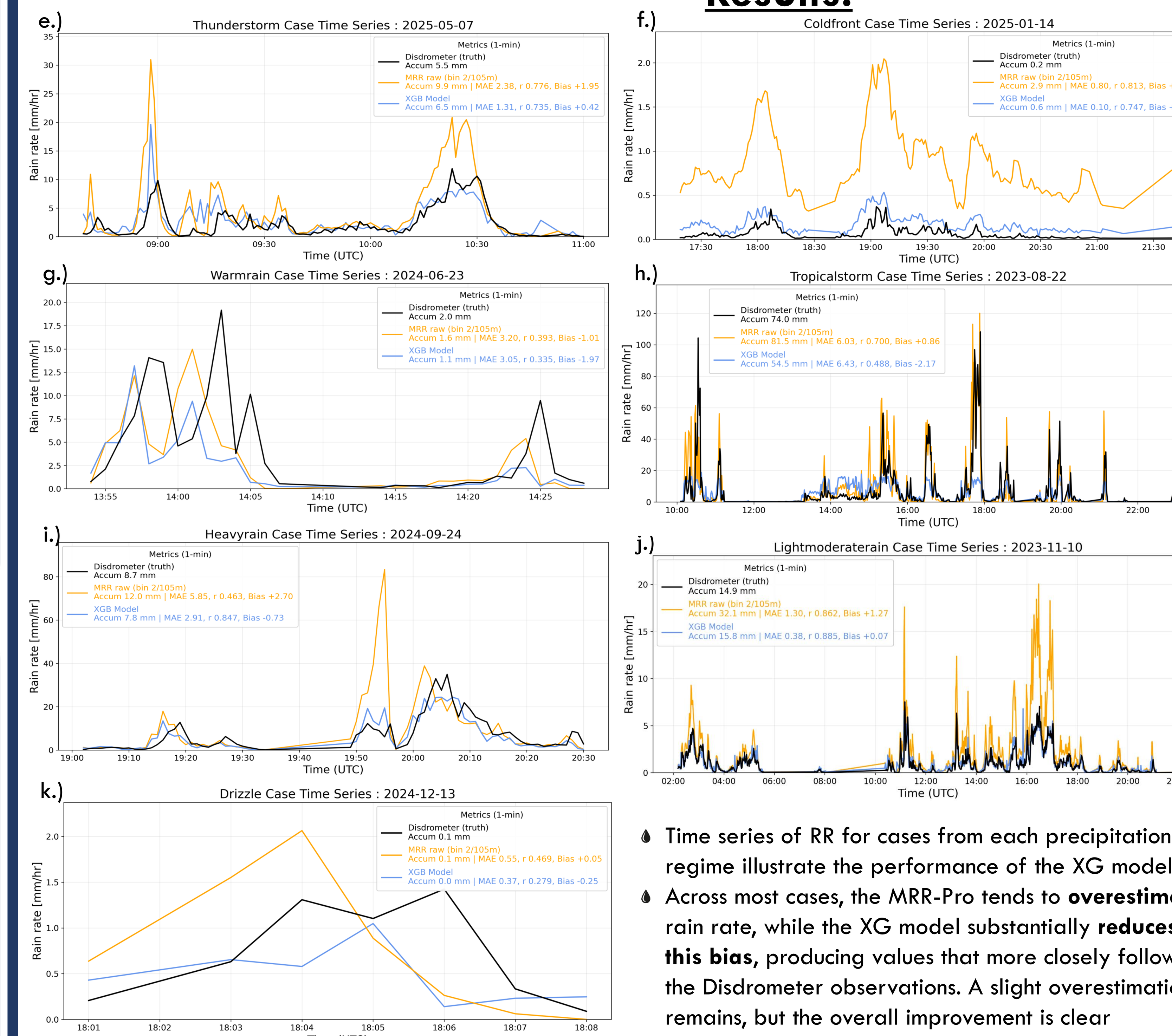
- In Fig. c, the feature importances across 20 iterations show how often each reflectivity or velocity bin appears in the top 10% of importance
  - Longer bars correspond to more reliable features
- Expected:
  - Z near surface (105 m) – reflects the hydrometeors reaching the ground
  - Z at 6355 m – indicates whether precipitation is from a deep cloud
- Unexpected:
  - Two doppler velocity features between 3500 – 4200 m stand out significantly
  - Their consistent importances warrant further examination of what XG is identifying at these levels



- To understand why the model picks mid-level velocity features, Fig. d examines the velocity-altitude structure seen by the XG model
- "Counts" represent the 2D histogram of doppler velocity vs altitudes of sample from all precipitation regimes
- Two most important velocity features are located around the melting layer, where ice particles transition into rain drops
- Smaller ice particles melt quickly, producing small drops with lower fall speeds, which correspond to lower rain rates with a possibly shallower melting band
- Larger ice particles melt more slowly, producing larger drops with higher fall speeds, which correspond to higher rain rates

Acknowledgements: NASA IPMSI grant: 80NSSC22K1795. Special thanks to program managers Drs. Emma Knowland and Barry Lefer, and to the NASA Precipitation Measurement Validation Network Technical POC, Dr. David B. Wolff and his team. Thank you to Pabla, Charanjit S. for reprocessing all the MRR data, and Dr. Christopher Williams for developing the new MRR data processing system.
This research was supported by the TAMUCC LSAMP Program (National Science Foundation Award No. HRD-1911375), and the TAMUCC McNair Scholars Program.

Results:



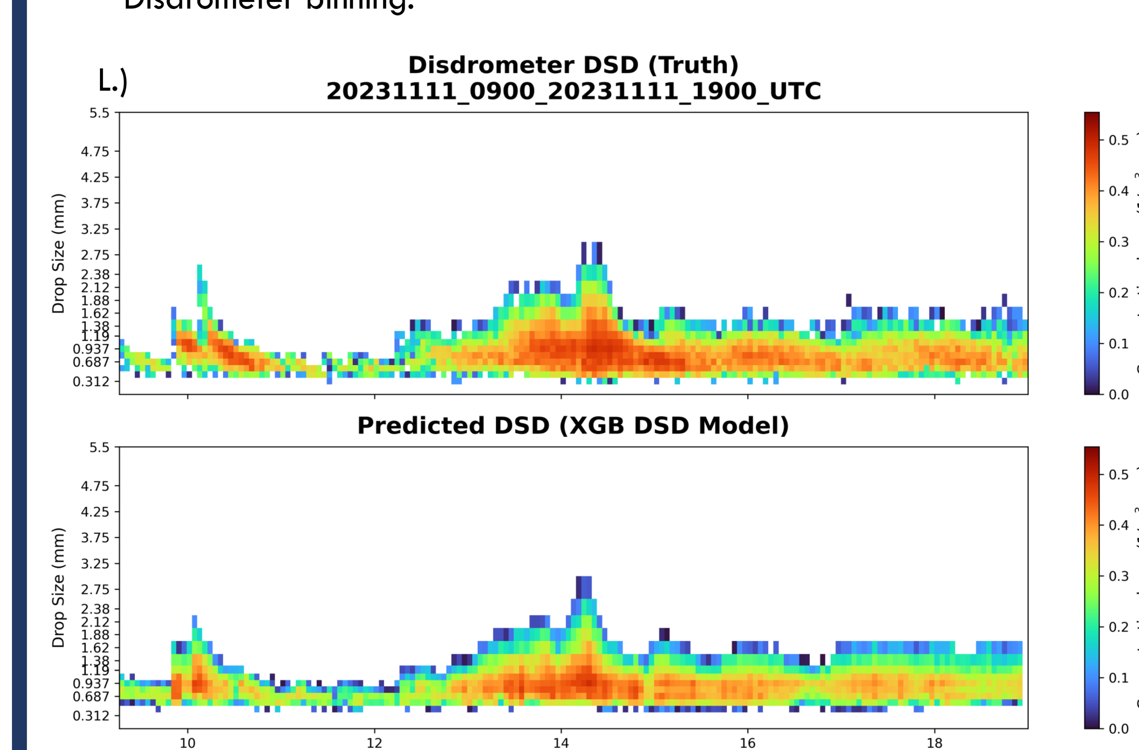
- Time series of RR for cases from each precipitation regime illustrate the performance of the XG model
- Across most cases, the MRR-Pro tends to **overestimate** rain rate, while the XG model substantially **reduces this bias**, producing values that more closely follow the Disdrometer observations. A slight overestimation remains, but the overall improvement is clear
- However, the XG model overestimates **in drizzle case**
- Both sensors have **intrinsic limitations** in drizzle:
  - The Disdrometer cannot reliably detect **very small drops (< 0.2 mm diameter)**, causing light-rain events to appear weaker or even absent
  - The MRR-Pro struggles to detect **weak signals from tiny droplets**, resulting in noisy or unreliable retrievals at low rain rates
  - Training data itself is **sparse and uncertain** in light-rain conditions, the model has limited ability to learn accurate patterns there
- The Tropical storm case (h.) shows a different challenge :
  - XG **underestimates the most intense peaks**, since extreme rain events are rare in the training set
  - Both sensors lose accuracy in very heavy rain (e.g., Disdrometer saturation), causing the models to underestimate the rain rate

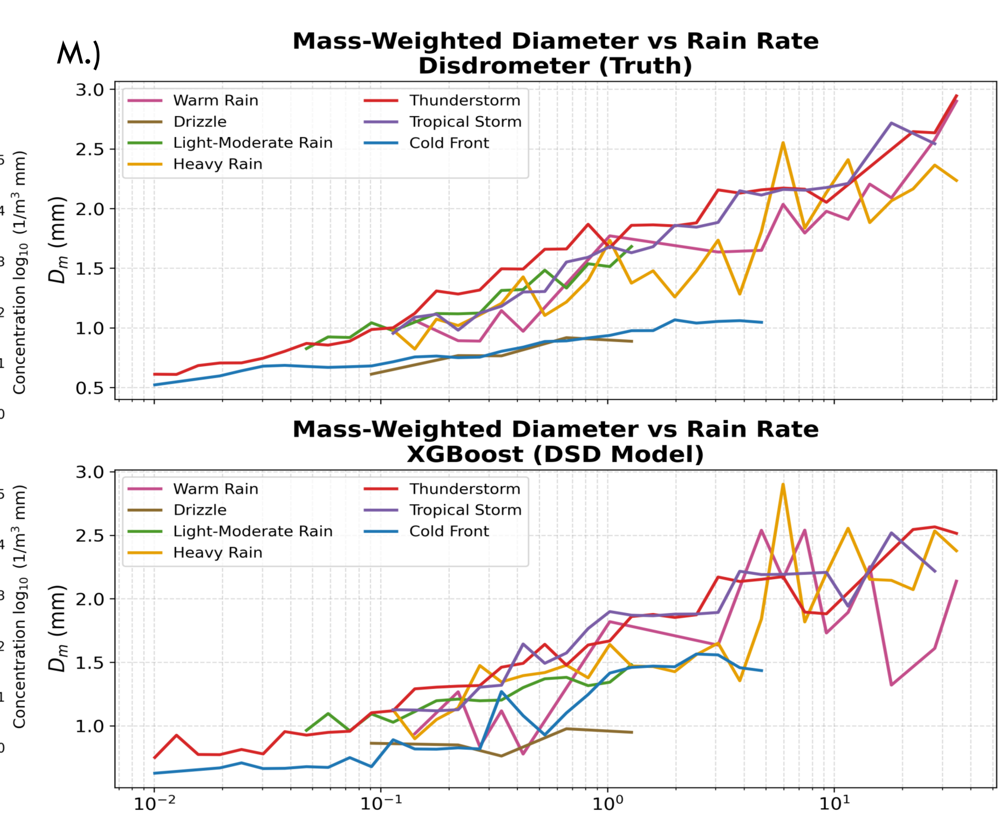
Summary:

- This study demonstrates that AI substantially improves rain-rate retrievals from the MRR-Pro across diverse precipitation regimes, reducing the raw MRR overestimation and producing results that align more closely with Disdrometer observations. The feature-importance and doppler velocity-altitude analyses show that the model captures physically meaningful patterns, including melting-layer processes, confirming that the XG is learning real microphysics rather than noise. Model performance decreases in very light rain, consistent with sensor limitations and sparse training examples in that regime. These findings complement the data-collection and observational analysis presented in poster #S249, which provides the detailed instrument context that supports the modeling results shown here.

Next steps [Drop Size Distribution Model]:

- The goal of the DSD model is to train on the MRR-Pro full vertical profile, using reflectivity (Z), Doppler velocity (V), and rain rate (RR) as inputs to predict the surface raindrop size distribution by producing drop diameters spanning 0.2–24.5 mm, consistent with the Parsivel Disdrometer binning.





- Fig. L) The XGboost model accurately tracks the timing and dominant drop sizes, confirming it has learned the underlying microphysics. However, it exhibits a systematic underestimation in concentration, especially during high RR periods.
- Fig. M) The model successfully preserves the regime-dependent relationship between mass-weighted diameter (D_m) and RR, capturing the distinct separation between precipitation types. Note that D_m is biased low at higher rain rates, consistent with the missing large drops
- Conclusion: The model captures the structural and regime behavior effectively but maintains a conservative low bias in both concentration and drop size.