

Tree-Based Supervised Classifier *of* ALMA Bandpass Calibration Anomalies

CosmicAI



Ci (Ceci) Xue, Brian Mason, Gazi Rakib, Tristan Ashton, Jeff Phillips,
Ryan Loomis, Ilsang Yoon, Ignacio Toledo, John Hibbard, Omkar Bait, Eric Murphy

National Radio Astronomy Observatory, Charlottesville, USA

NSF–Simons AI Institute for Cosmic Origins, Austin, USA

University of Utah, Salt Lake City, USA

Joint ALMA Observatory, Santiago, Chile

Interferometric Radio Observation Calibration



Observations

ALMA Calibration Pipeline
(Hunter et al. 2023)

Calibrated Data

Correct for system
temperature, water
vapor, antenna
position

Calibrate the
amplitude and
phase vs.
frequency of each
antenna

Calibrate the
amplitude and
phase vs. time of
each antenna

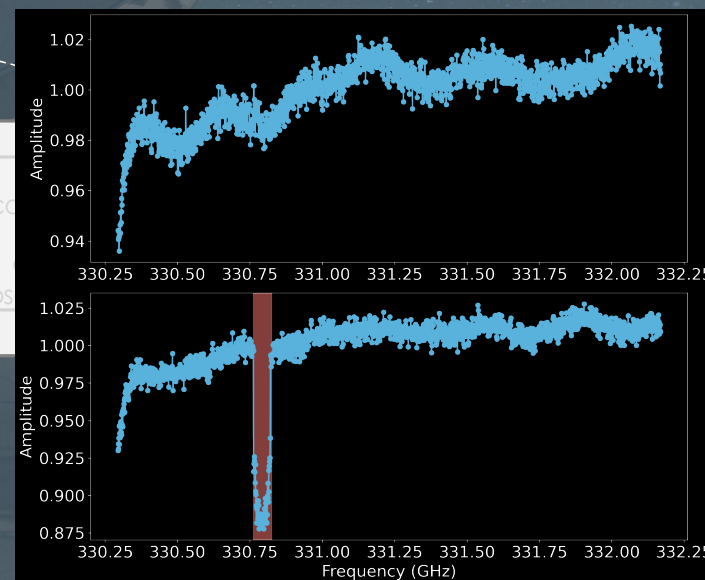
Calibrate the
absolute
amplitude scale

Calibrate the
bandpass

observed

Bandpass Calibration

- Observations of a point source (e.g. bright quasar) record antenna responses as a function of frequency → calibration solutions
- Anomalous data should be flagged to remove artificial distortions that can introduce artifacts into science products.

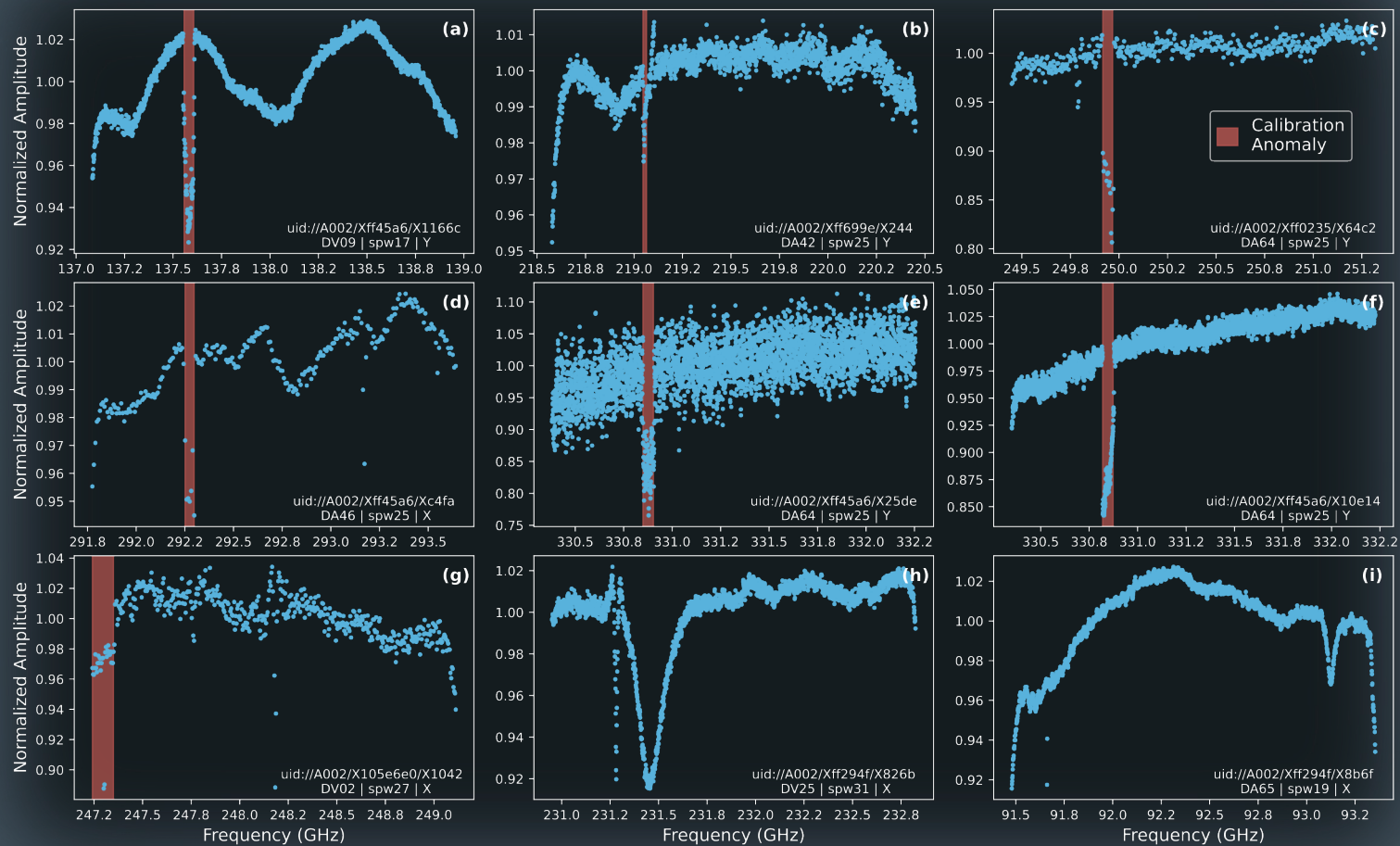


Top: Example normal, non-anomalous bandpass amplitude solution

Bottom: Example solution having an anomaly (shaded in red)

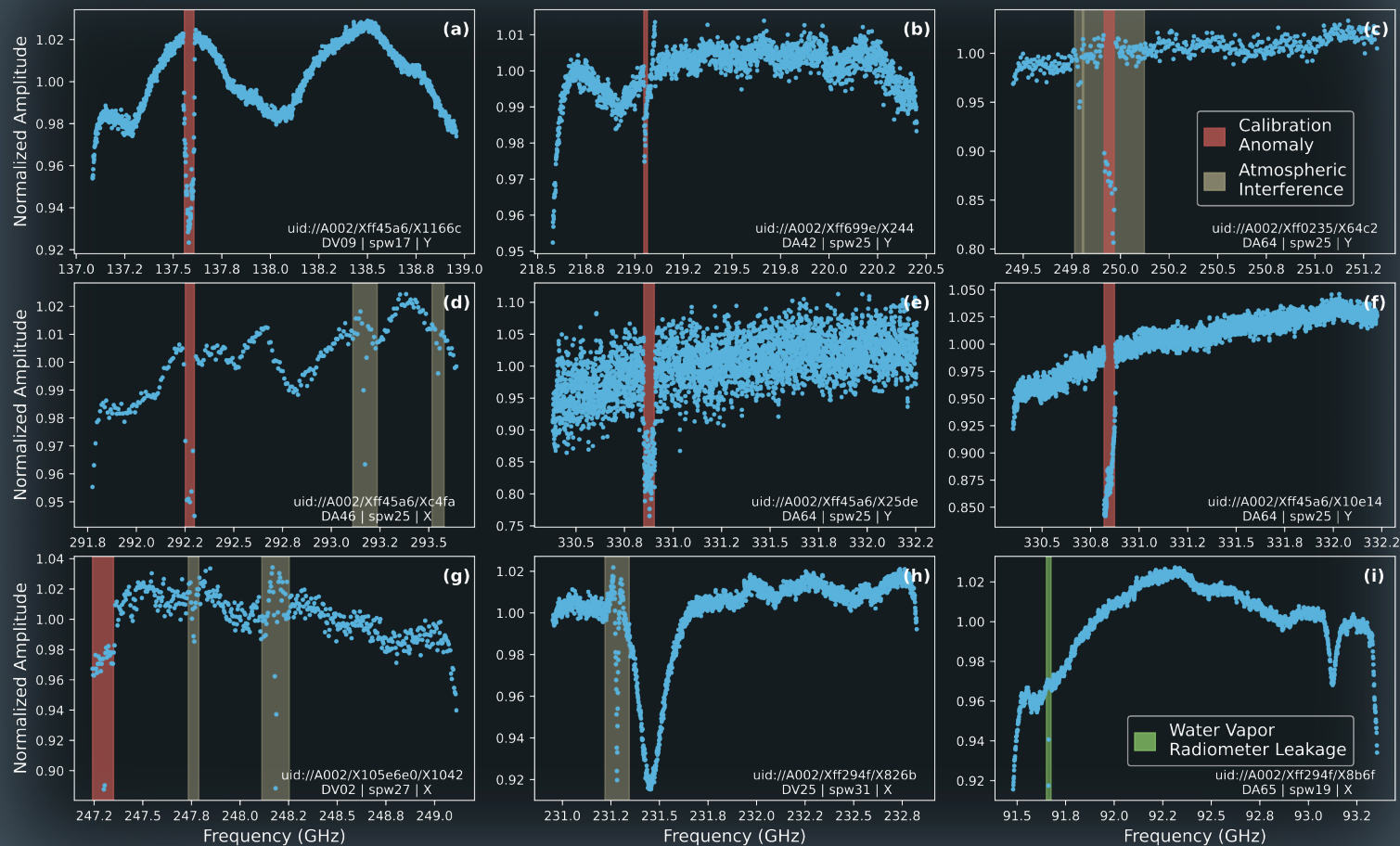
Bandpass Calibration Solutions

Example bandpass amplitude solutions from ALMA
Cycle 9 observations, project code starting with 2022.#.



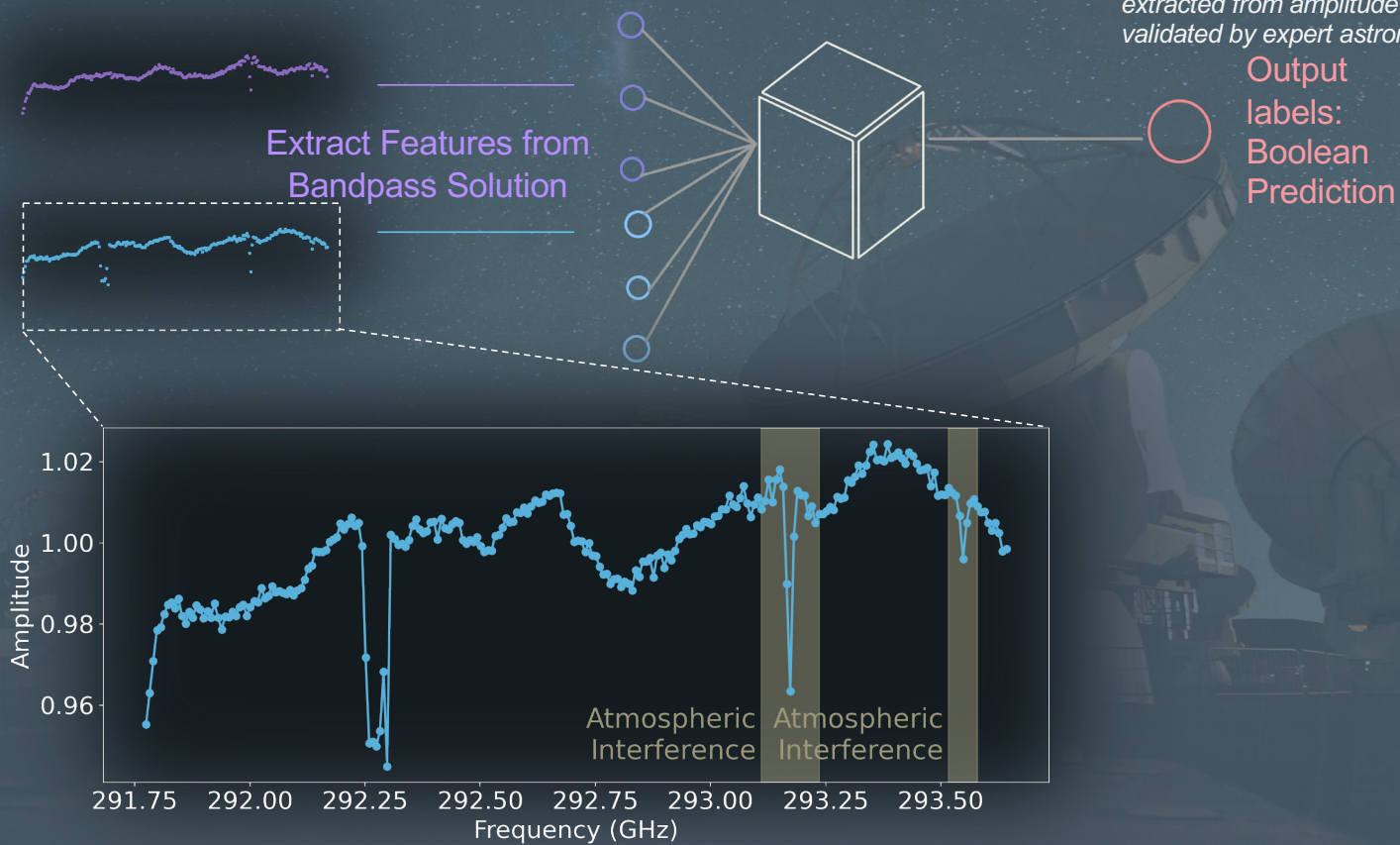
Bandpass Calibration Solutions

Example bandpass amplitude solutions from ALMA
Cycle 9 observations, project code starting with 2022.#.

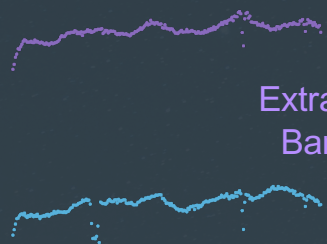


Decision Tree Based ML Classifier

Schematic of the ML classifier, with input features extracted from amplitude arrays and output labels validated by expert astronomers.



Decision Tree Based ML Classifier



Extract Features from
Bandpass Solution



Output
labels:
Boolean
Prediction

- Statistical features
 - **Scan statistics**
 - Amplitude dispersion
 - Channel spacing
- Categorical features
 - ALMA receiver band



Gazi Rakib
(U Utah)

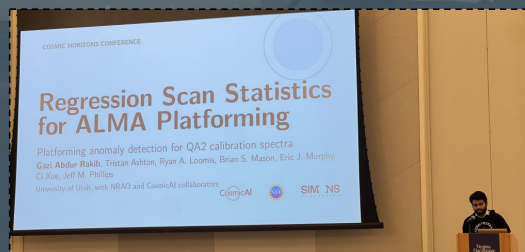


Jeff Phillips

Labels:
(Validated by expert astronomers,
ALMA QA2 flags)
True = bandpass calibration anomaly
False = normal, non-anomalous data

Model:

- Supervised learning
 - Labels from QA2 manual flags related to bandpass amplitude anomalies
- Gradient Boosting
 - Directly optimize predictive performance by combining weak models
- XGBoost
 - Chen and Guestrin 2014, standalone python library
 - Powerful for structured/tabular datasets



Data Preparation and Training Strategy

C. Xue (NRAO | CosmicAI)



Highly Imbalanced Data

- Inherently imbalanced between anomalies and normal data
- **~0.5%** data has bandpass amplitude anomaly

Per Pol Pair Flagging

- Flags are applied to pairs (polarizations), often resulting in one polarization being flagged alongside its anomalous counterpart

Potential Data Leakage

- Anomalies from the same EB might result in information leaks. Random splitting allows the solutions with the same EB id to be in both train and test
- Selected features may implicitly encode EB id and model learns EB-specific patterns

Mitigation Strategy

To account for the imbalanced class

- ✓ Use stratified sampling for training and testing splitting (80% | 20%)
- ✓ Adjust the weight of True and False weights during fitting, i.e. gives more weight to the minority True class

To account for the inconsistent pol labeling

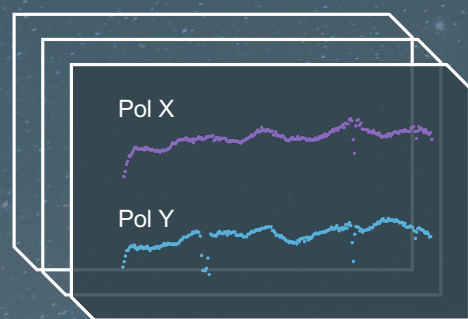
- ✓ Aggregate solutions at a specified level and combine labels into one, using the "or" logical operator

To account for potential data leakage

- ✓ Restrict features that may encode correlations with the labels
 - e.g. anomalies could be associated with specific antennas
- ✓ Split training and testing at EB levels

(Xue et al. in prep)

Model Performance Evaluation



Bandpass Amplitude Arrays

Feature
Extraction

Pol X	Feature	score ₁
	...	φ
	Feature	...
Pol Y	Feature	score ₂
	...	ζ
	Feature	...
...	...	ω
	...	...
	...	...

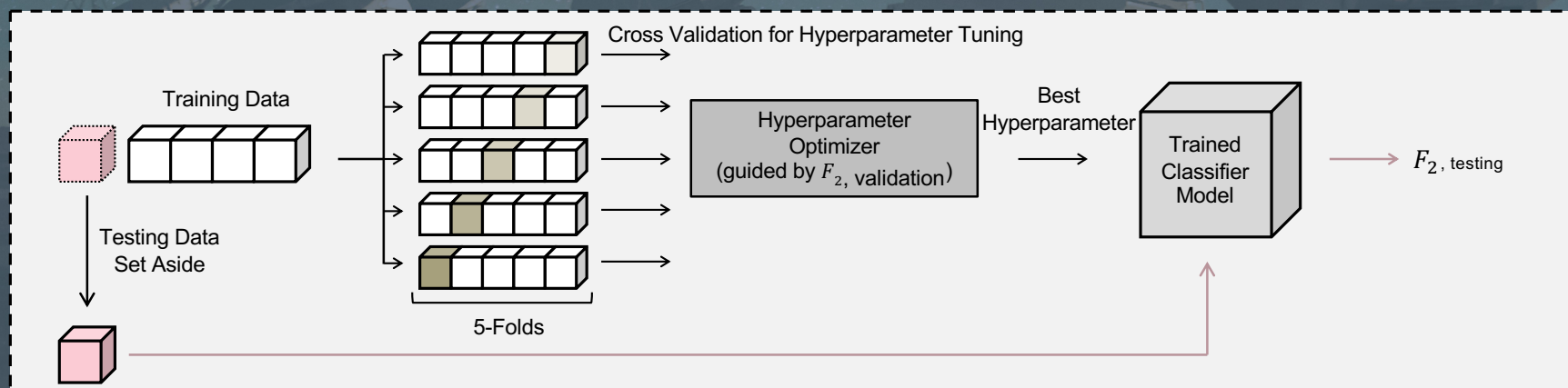
Features

Stratified
Splitting

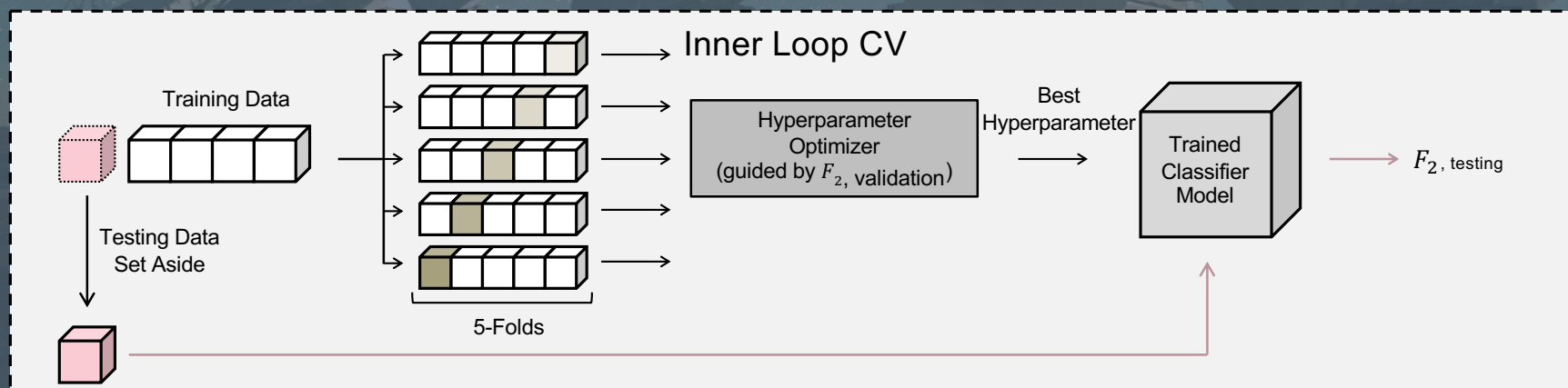
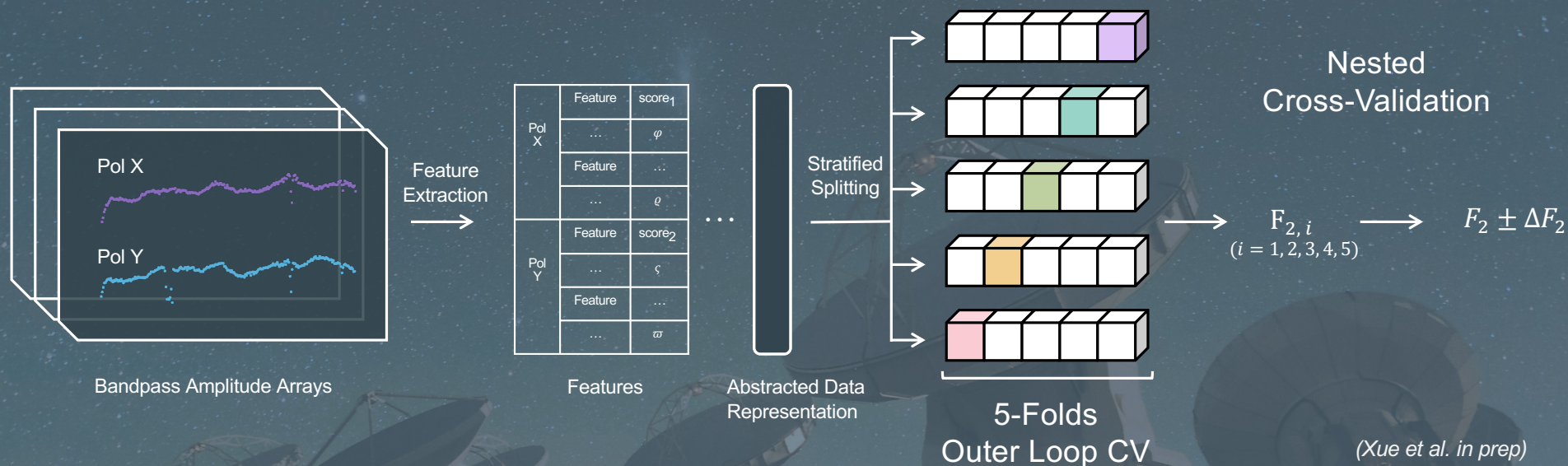


F_2

- F_2 is sensitive to the split as performance depends on the difficulty of the test set
- Only a point estimate, no uncertainty in F_2



Model Performance Evaluation



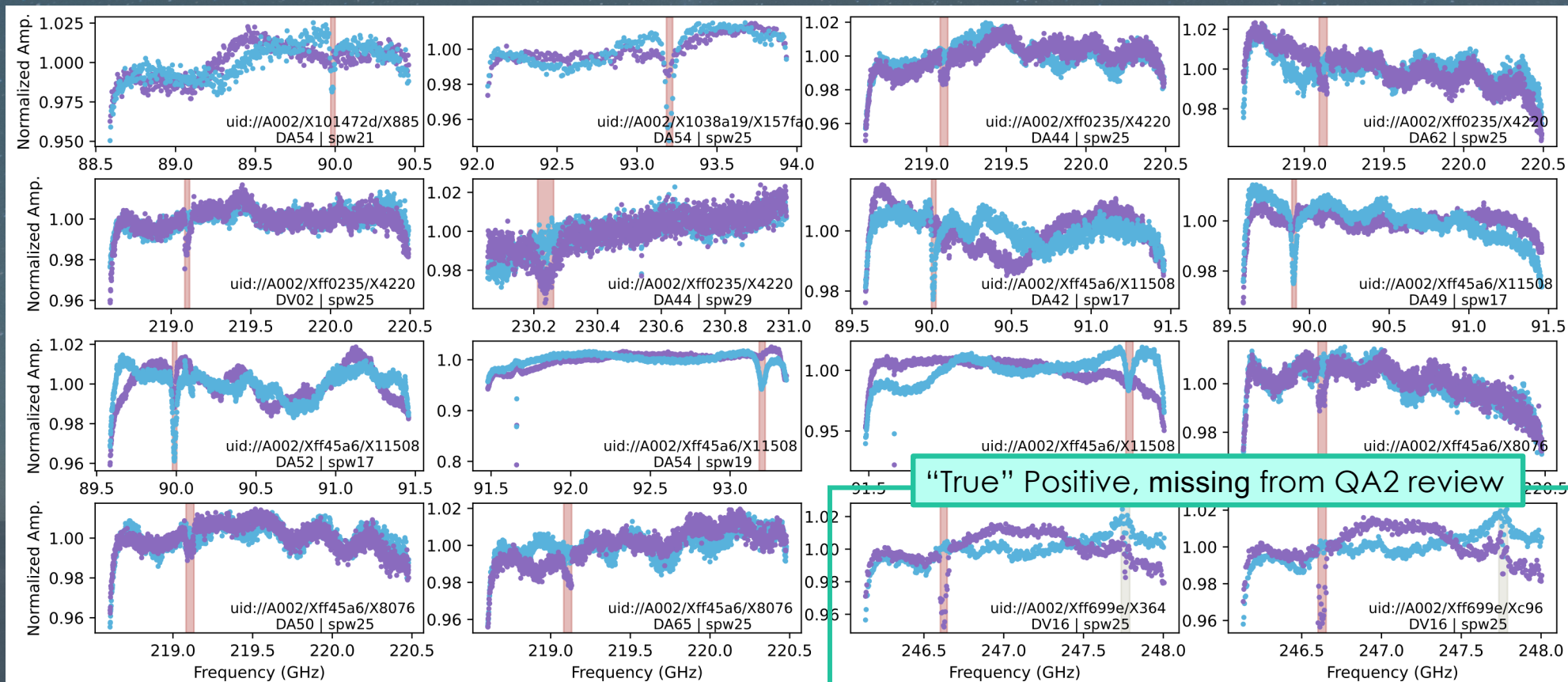
Model Performance Evaluation

~85,000 real observation data from the ALMA Cycle 9 observations

		FN	FP	TP	TN	FNR	$F_{2,i}$	$\overline{F_2} \pm \Delta F_2$
→		2 (0.02%)	4 (0.04%)	36 (0.4%)	8854 (99.5%)	5%	0.94	0.88 ± 0.05
→		6 (0.07%)	13 (0.15%)	40 (0.5%)	8500 (99.3%)	13%	0.84	
→		4 (0.05%)	0	34 (0.4%)	8556 (99.3%)	11%	0.91	
→		2 (0.03%)	4 (0.05%)	30 (0.4%)	7866 (99.5%)	6%	0.93	
→		9 (0.11%)	13 (0.16%)	39 (0.5%)	8274 (99.3%)	19%	0.80	

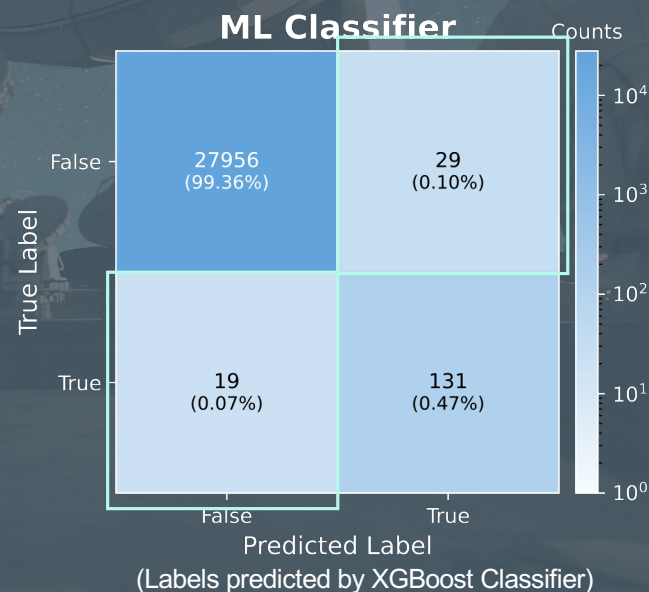
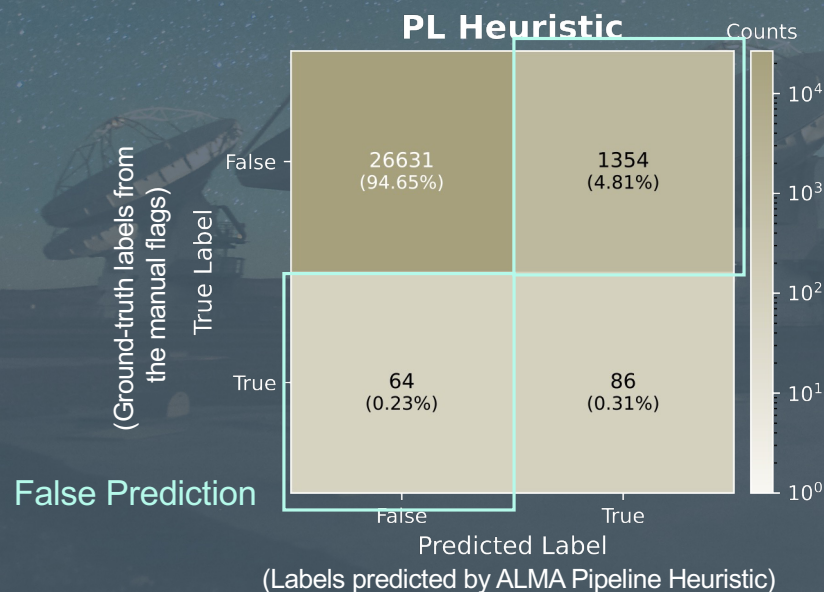
(Xue et al. in prep)

Example FP Cases Predicted by ML Classifier



ML Classifier vs. ALMA Pipeline Heuristic

- Existing ALMA pipeline heuristic
 - Uses 5 metrics designed to identify subband issues (including the “platforming” issues)
 - Applies defined thresholds for each metric
 - Evaluated only observations in ALMA BLC mode
- Cross checked with the ML training dataset
 - Evaluated ~59000 bandpass solutions from the ~85000 solutions training set
 - Aggregated two polarization labels using the “or” logical operator
 - ML classifier shows substantial performance improvement
 - ALMA Pipeline heuristic: $\text{FNR} = \frac{64}{64+86} = 43\%$, $F_2 = 0.21$



ML Classifier vs. ALMA Pipeline Heuristic

Possible reasons for behavior difference:

Modeling Approach / Methodology

- ML Classifier: supervised classification optimized for predictive performance
 - Balances false positives/negatives using cross-validation
 - Leverages multiple input features
- PL Heuristics: **threshold**-based decision
 - Rely on empirical distribution of the metric
 - Potentially fragile statistical assumptions to determine the threshold

Training Data Scales / Dataset Selection

- ML Classifier: used TACC to stage datasets on a per-cycle basis for a more representative distribution
- PL Heuristics: fine tuned on subset data (more tailored to specific cases)



Future work:

- ☐ Labels from the data reducer flag(s) might not be robust → define labels derived from visibility
- ☐ Integrate the ML classifier framework within the ALMA pipeline codebase
- ☐ Expand to other features

Summary

- Decision tree based gradient boosting model optimized for predictive performance
- Training strategies account for:
 - Highly imbalanced data
 - Observatory flagging practices
 - Potential data leakage from specific feature-label patterns
- Applied kernel-regression based scan statistics (Rakib et al. 2026) to extract statistical features
- Comparing against existing ALMA pipeline heuristic shows substantial performance improvement, achieving 70% reduction in false negative rates and 90% reduction in false positive rates

