COSMIC HORIZONS CONFERENCE

Regression Scan Statistics for ALMA Platforming

Platforming anomaly detection for QA2 calibration spectra

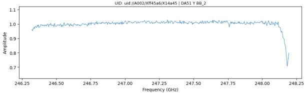
Gazi Abdur Rakib, Tristan Ashton, Ryan A. Loomis, Brian S. Mason, Eric J. Murphy, Ci Xue, Jeff M. Phillips

University of Utah, with NRAO and CosmicAI collaborators

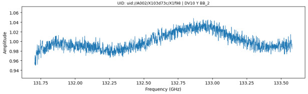
The astronomy problem

The scientific motivation

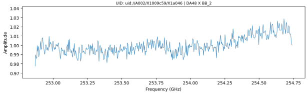
ALMA calibration spectra are an early checkpoint before science-ready hyperspectral data cubes are built. A localized instrumental offset can look small in one spectrum but propagate into non-astrophysical structure later.



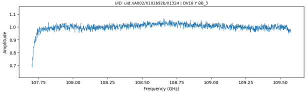
Normal signal with sharp dropoff at the end



Normal rich signal



Normal sparse signal

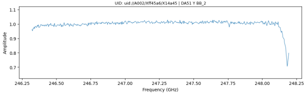


Regular signal with starting sharp dropoff

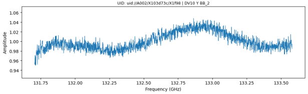
The astronomy problem

The scientific motivation

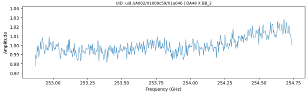
ALMA calibration spectra are an early checkpoint before science-ready hyperspectral data cubes are built. A localized instrumental offset can look small in one spectrum but propagate into non-astrophysical structure later.



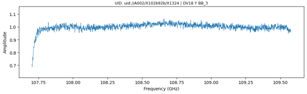
Normal signal with sharp dropoff at the end



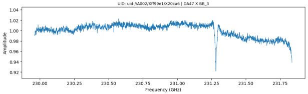
Normal rich signal



Normal sparse signal



Regular signal with starting sharp dropoff

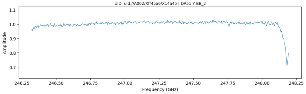


Regular signal with ATM interference

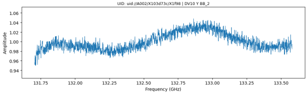
The astronomy problem

The scientific motivation

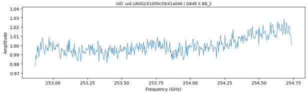
ALMA calibration spectra are an early checkpoint before science-ready hyperspectral data cubes are built. A localized instrumental offset can look small in one spectrum but propagate into non-astrophysical structure later.



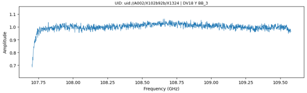
Normal signal with sharp dropoff at the end



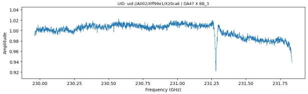
Normal rich signal



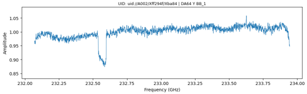
Normal sparse signal



Regular signal with starting sharp dropoff



Regular signal with ATM interference

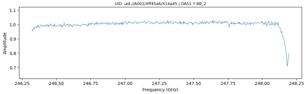


Anomalous signal with platforming

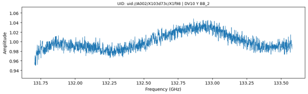
The astronomy problem

The scientific motivation

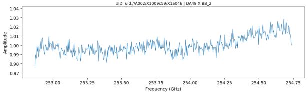
ALMA calibration spectra are an early checkpoint before science-ready hyperspectral data cubes are built. A localized instrumental offset can look small in one spectrum but propagate into non-astrophysical structure later.



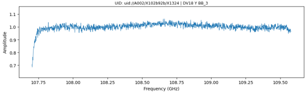
Normal signal with sharp dropoff at the end



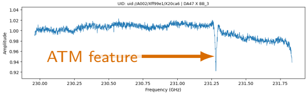
Normal rich signal



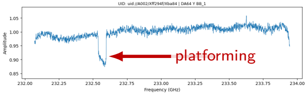
Normal sparse signal



Regular signal with starting sharp dropoff



Regular signal with ATM interference



Anomalous signal with platforming

From visibilities to science products

Where the detector sits

- **raw visibilities** antenna pairs →
- **QA2 spectra** platforming check →
- **calibrated data** bad rows flagged →
- **data cube** image × frequency

Operational question

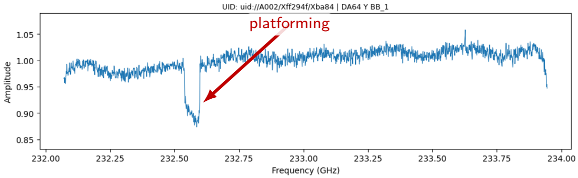
Which rows should be flagged before cube construction, and which frequency channels caused the flag?

Astronomy question

How do we preserve smooth spectral structure while removing calibration-induced discontinuities?

What platforming looks like

The morphology of the artifact

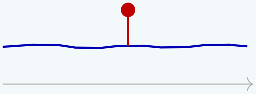


Platforming anomaly in ALMA Dataset

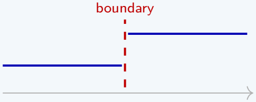
- Smooth background curvature is real and should not be flattened away.
- The anomaly is local in frequency, not a permanent regime change.
- It may overlap with noise, edge effects, or atmospheric structure.
- A good detector localizes the interval without hallucinating broad features.

What kind of anomaly is platforming?

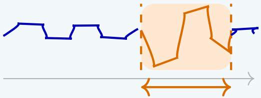
Anomaly detection landscape



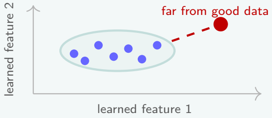
Point anomaly



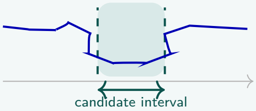
Changepoint



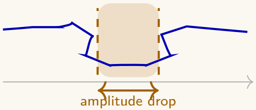
Discord subsequence



OOD anomaly



Scan-statistic interval



Platforming

Regression scan statistic

Score with a physical interpretation

Candidate interval $I_{a,b} = \{a, \ldots, b\}$. Each SSE is a sum of squared residuals from a fitted background.

observed data   global fit   inside fit   outside fit

$\mathrm{SSE}_{\mathrm{all}}$

one fit explains all channels

$\mathrm{SSE}_{\mathrm{in}}(I_{a,b})$

fit only the candidate interval

$\mathrm{SSE}_{\mathrm{out}}(I_{a,b})$

fit everything else

$$S(I_{a,b}) = 1 - \frac{\mathrm{SSE}_{\mathrm{in}}(I_{a,b}) + \mathrm{SSE}_{\mathrm{out}}(I_{a,b})}{\mathrm{SSE}_{\mathrm{all}}} \qquad \Phi(x) = \max_{I_{a,b}} S(I_{a,b})$$

Large score = one interval explains the spectrum much better when separated.

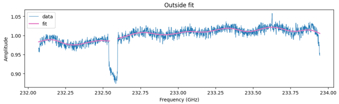
Why NWKR fits the spectra

Local smoothness for calibration spectra

ALMA bandpass spectra are smooth but non-flat. NWKR follows broad calibration curvature while allowing a compact platforming interval to be scored separately.

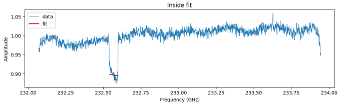
Outside fit

Background fit disregrading the candiate interval



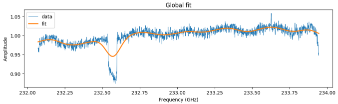
Inside fit

Candidate interval under consideration for score calculation



Global fit

Global fit considering all the channels of the bandpass solution



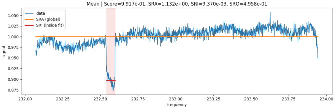
Why NWKR is the better fit

Comparison with Generic Scan Statistics

ALMA spectra are smooth but non-flat; the model should follow broad curvature without absorbing compact platforming.

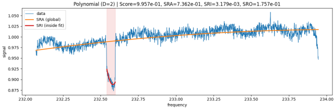
Mean scan

too rigid: assumes flat background



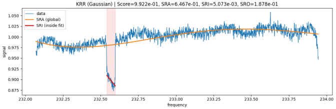
Polynomial scan

global curve can be pulled by local artifacts



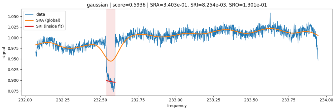
KRR scan

flexible, but less directly local/interpretable



NWKR scan

local smoothing matches bandpass geometry



NWKR tracks local bandpass curvature while exposing compact platforming.

The computational bottleneck

Why naive scanning is expensive

Outer search

try every interval $I_{a,b}$

$O(n^2)$

$\times$

Inner scoring

Refit NWKR $\times$ Evaluate mode

$O(n) \times O(n)$

Naive NWKR scan

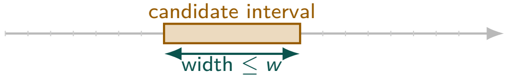
$$O(n^2) \times O(n^2) = O(n^4)$$

too slow for large ALMA screening

Key idea: same statistic, less recomputation.

Plausible platforming widths

Limit the scan range



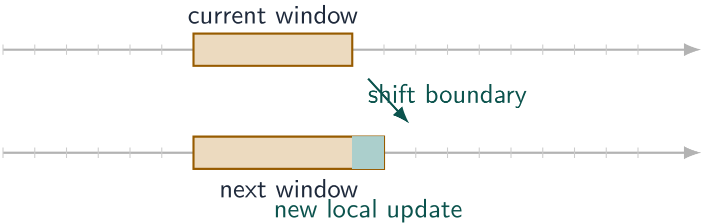
Restrict the scan to intervals whose length is plausible for platforming.

$$O(n^2) \text{ intervals} \quad \longrightarrow \quad O(nw) \text{ intervals}$$

Traditional scan-statistic reduction: fewer candidate windows before any fitting.

Reuse work as the window moves

Main algorithmic idea



Naive

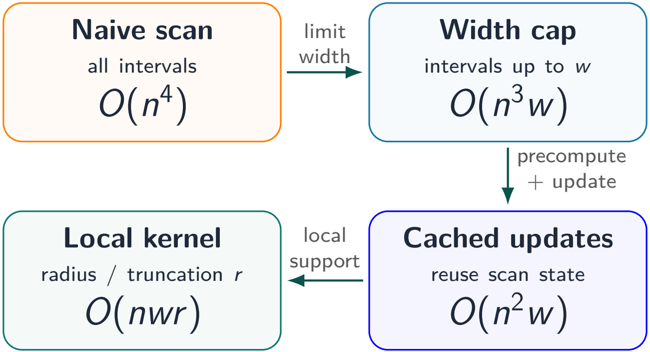
Rebuild the fit for every new interval.

Our insight

Precompute reusable pieces and update them as the scan boundary moves.

Complexity transition

Runtime story



Brute-force scan $\rightarrow$ streaming-style update.

We built a cleaned ALMA dataset

Evaluation dataset construction

Starting from raw QA2 calibration spectra, we cleaned the evaluation set in three stages before running the platforming scan.

1. Filter edge effected rows

Drop low-variation rows, edge plateaus, edge rolloff, and rows shorter than 128 channels.

2. Remove ATM rows

Match each spectrum to atmospheric transmission and discard rows overlapping atmospheric interference lines.

3. Trim flags

Remove channels marked by the ALMA flag array; discard rows with no valid channels left.

Output: cleaned spectra used as the evaluation set for interval localization.

ALMA QA2 dataset

The astronomy evaluation set

- Curated ALMA QA2 bandpass calibration spectra.
- Each row corresponds to one polarization of one spectral window observed by one antenna pair.
- Expert-identified platform intervals provide localization targets.
- Boundary buffers and atmospheric-interference cases are filtered for the core study.

38,881
spectra screened

231
confirmed platforms

0.6%
positive rate

128–3840
channels per row

Efficient enough for QA2-scale screening

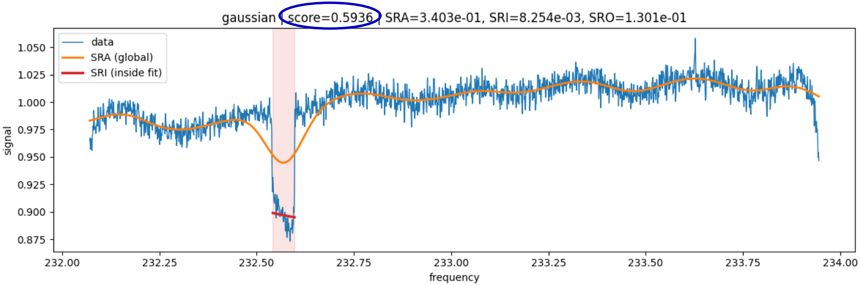
Runtime matters in calibration

- Brute-force scanning refits models for many candidate intervals.
- Optimized NWKR maintains numerator/denominator state for inside and outside fits.
- Moving a boundary changes only $O(r)$ local kernel contributions.
- Runtime is $O(nwr)$ for n channels, max interval width w, and kernel radius r.

method (N=1000)	time
$\mathcal{F}_0$ Mean	17 ms
$\mathcal{F}_{\mathrm{NWKR}}$ Laplace	69 ms
$\mathcal{F}_{\mathrm{NWKR}}$ Gaussian	70 ms
$\mathcal{F}_1$ Poly (deg 1)	392 ms
$\mathcal{F}_2$ Poly (deg 2)	472 ms
$\mathcal{F}_{\mathrm{NWKR}}$ Gaussian naive	759,396 ms
$\mathcal{F}_{\mathrm{NWKR}}$ Laplace naive	762,321 ms
CAPA	3 ms
LRT	37 ms
BOCPD	2,785 ms
STUMPY	2,776 ms
KernelCPD	11,279 ms

What the detector returns

Not a black-box alarm

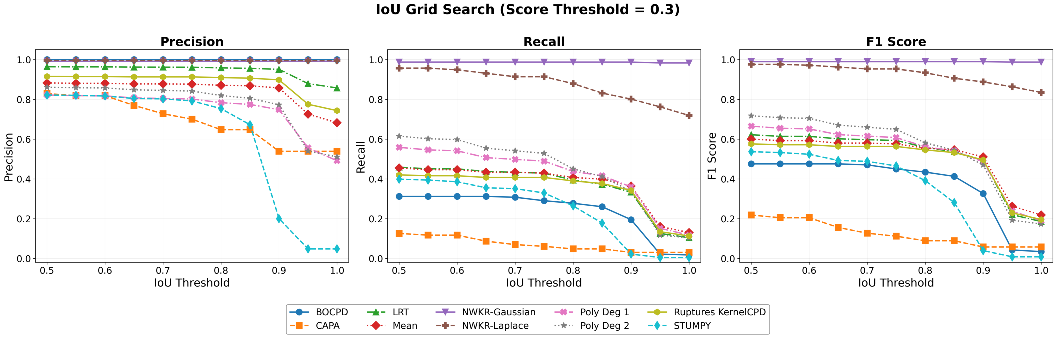


- **Score** $\Phi(x)$: how suspicious is this spectrum?
- **Window** $[a, b]$: which frequency channels are implicated?
- **Fitted background**: what smooth response was expected?
- **Review priority**: thresholded output for inspection or automatic flagging.

COSMIC HORIZONS / ALMA Platforming

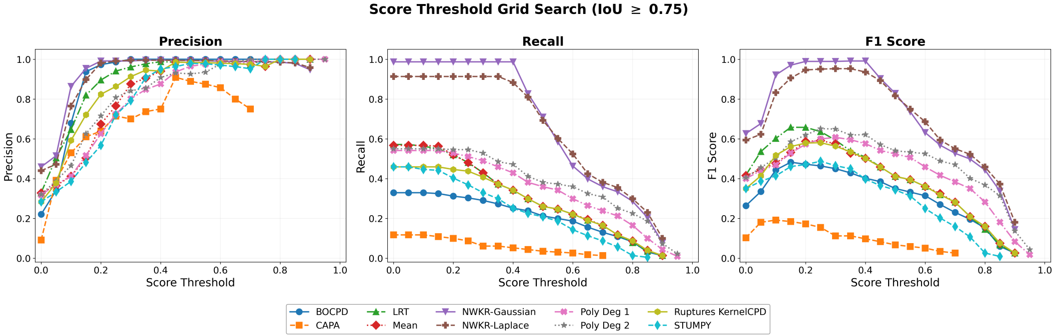
Interval localization on ALMA

Does the marked window match expert labels?



Interval localization on ALMA

Does the marked window match expert labels?

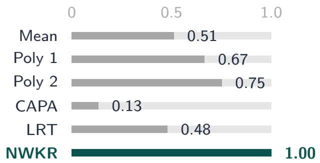


Screening comparison at $\tau_S = 0.3$

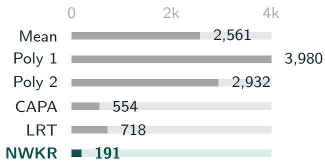
Model comparison

For QA2 screening, the best model should find almost all platforming rows while keeping the review burden small.

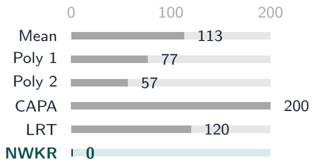
Known platforming rows found



False positives sent to review



False negatives missed



NWKR Gaussian is the only method here with perfect recall and zero false negatives.

Positives = 231, negatives = 38,650. Exact operating-point metrics are shown on the next slide.

NWKR Gaussian at production scale

NWKR operating point

Score-threshold screening on the cleaned ALMA evaluation set: $N = 38{,}881$, positives $= 231$, negatives $= 38{,}650$.

Confusion summary

	platforming	non-platforming
flagged	**231** TP	**191** FP
not flagged	**38,459** TN	**0** FN

1.000
recall
all known positives found

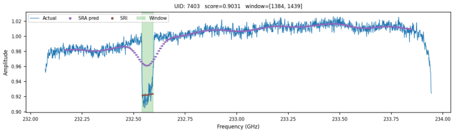
0
False Negatives
no false negative sample found

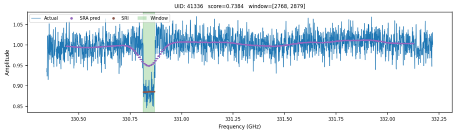
Operational reading

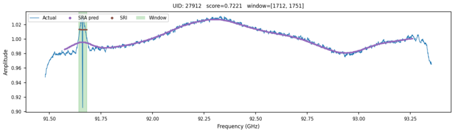
The screen prioritizes recall: no known platforming cases are missed, and the 191 extra rows are a manageable QA2 review/removal set.

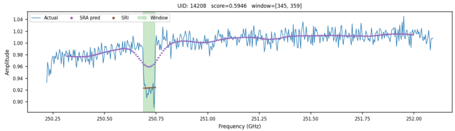
High-score ALMA rows

Examples that should be flagged





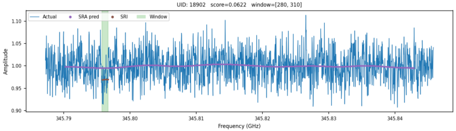


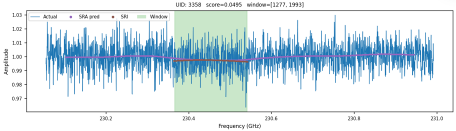


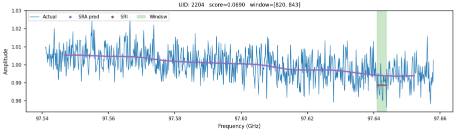
High-scoring rows show compact intervals whose amplitudes depart from a smooth local background. The highlighted windows provide interpretable evidence.

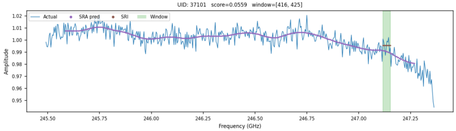
Low-score ALMA rows

Examples that should pass









Low-scoring rows may be noisy or curved, but no compact interval strongly improves the regression-scan fit.

Takeaway for Cosmic Horizons

Why this matters for astronomers

- Platforming is rare, localized, and scientifically consequential.
- Regression scan statistics match the geometry of the artifact.
- NWKR gives local adaptivity, efficient scanning, and auditable intervals.
- Output for QA2: anomaly score, frequency interval, and fitted background.
- Next: atmospheric masks, cycle-specific thresholds, broader ALMA workflow tests.

Comparison with current screening

method	FN rate	FP rate
human reducer	0.05	0.10
ALMA heuristic	0.585	0.052
our NWKR	**0.013**	**0.0049**

data pre-cleaned; rates are not a fully controlled head-to-head comparison

38,881

ALMA spectra

$O(nwr)$

NWKR scan

Main point: fewer missed platforming, fewer false alarms, and a localized interval for review.

Thank you

Project page

Project page and code

https://beardyman37.github.io/RegressionScanStats/

Questions?