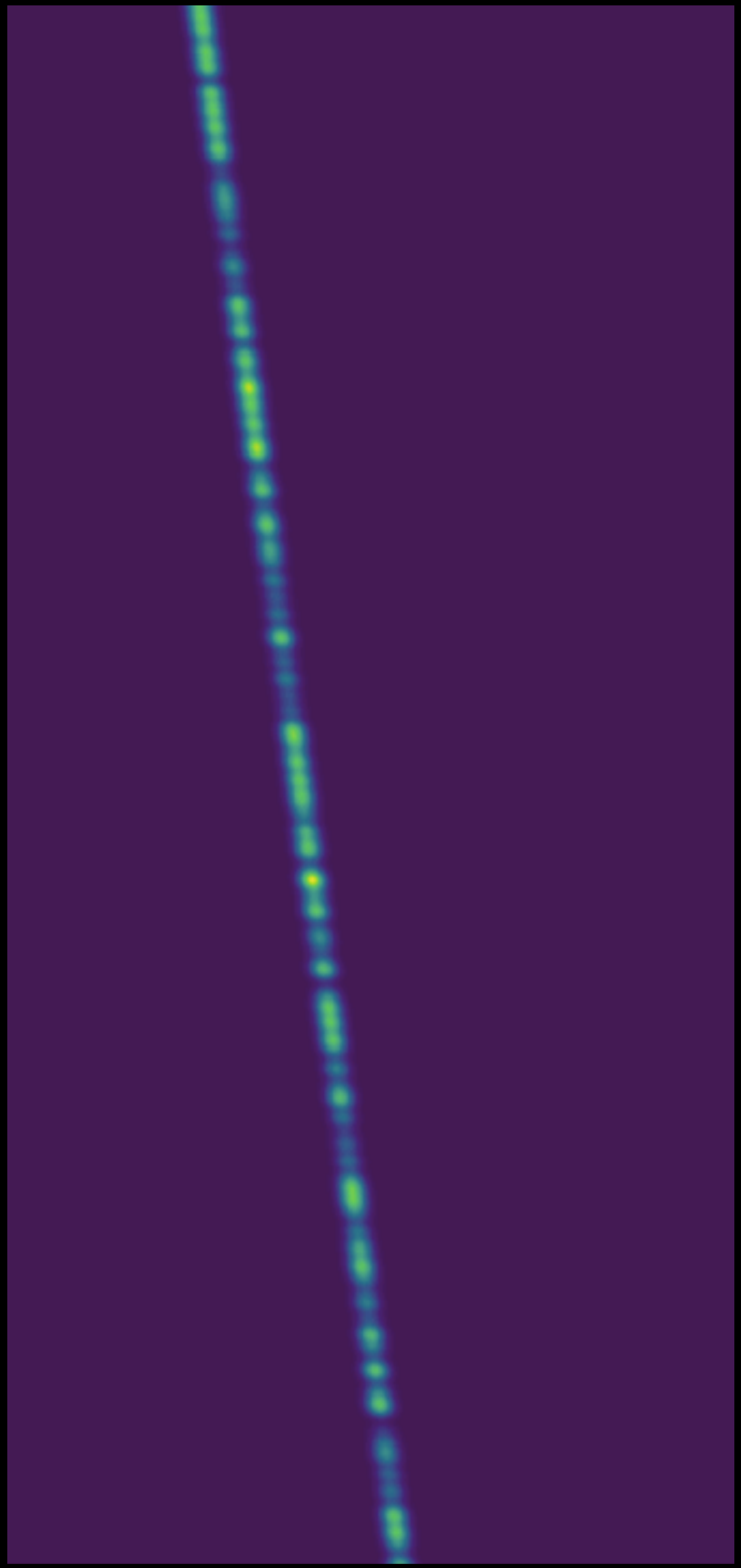


BREAKTHROUGH LISTEN

Detecting ISM Scintillation in Narrowband Signals: A New Filter for Radio SETI

Bryan Brzycki
University of California Berkeley

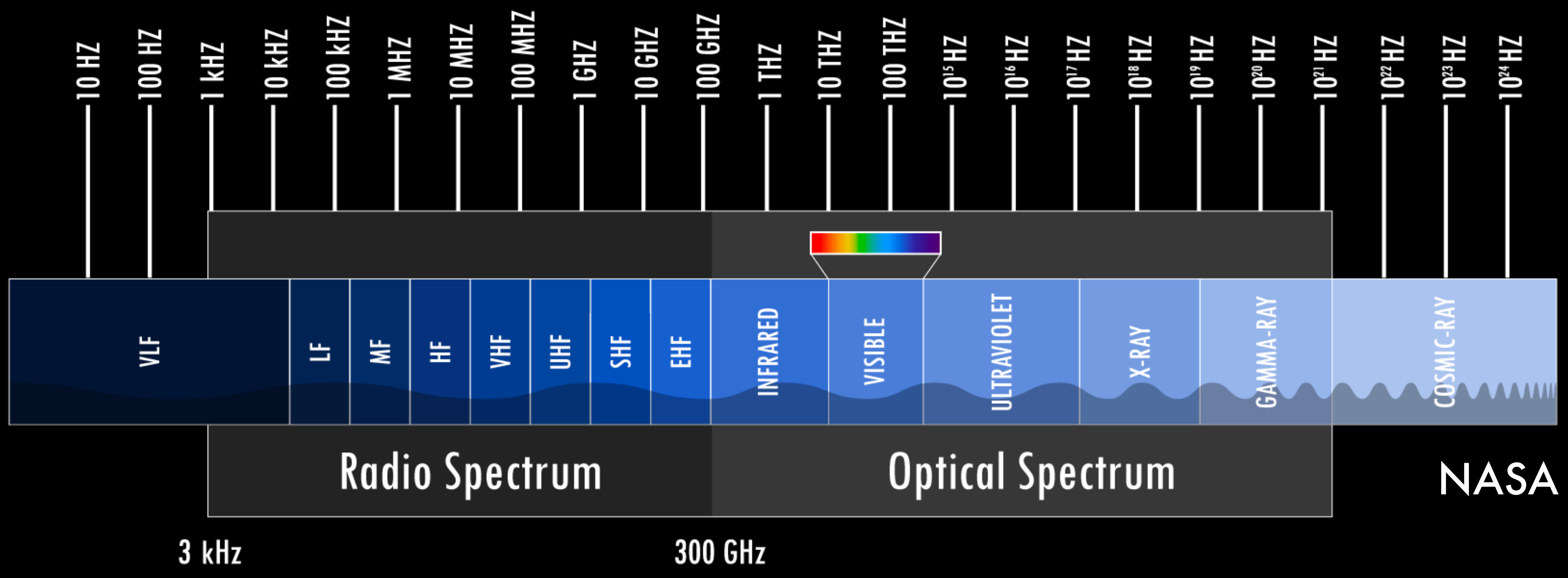
PSETI Seminar, September 21, 2023

BERKELEY SETI RESEARCH CENTER

The Search for Extraterrestrial Intelligence (SETI)

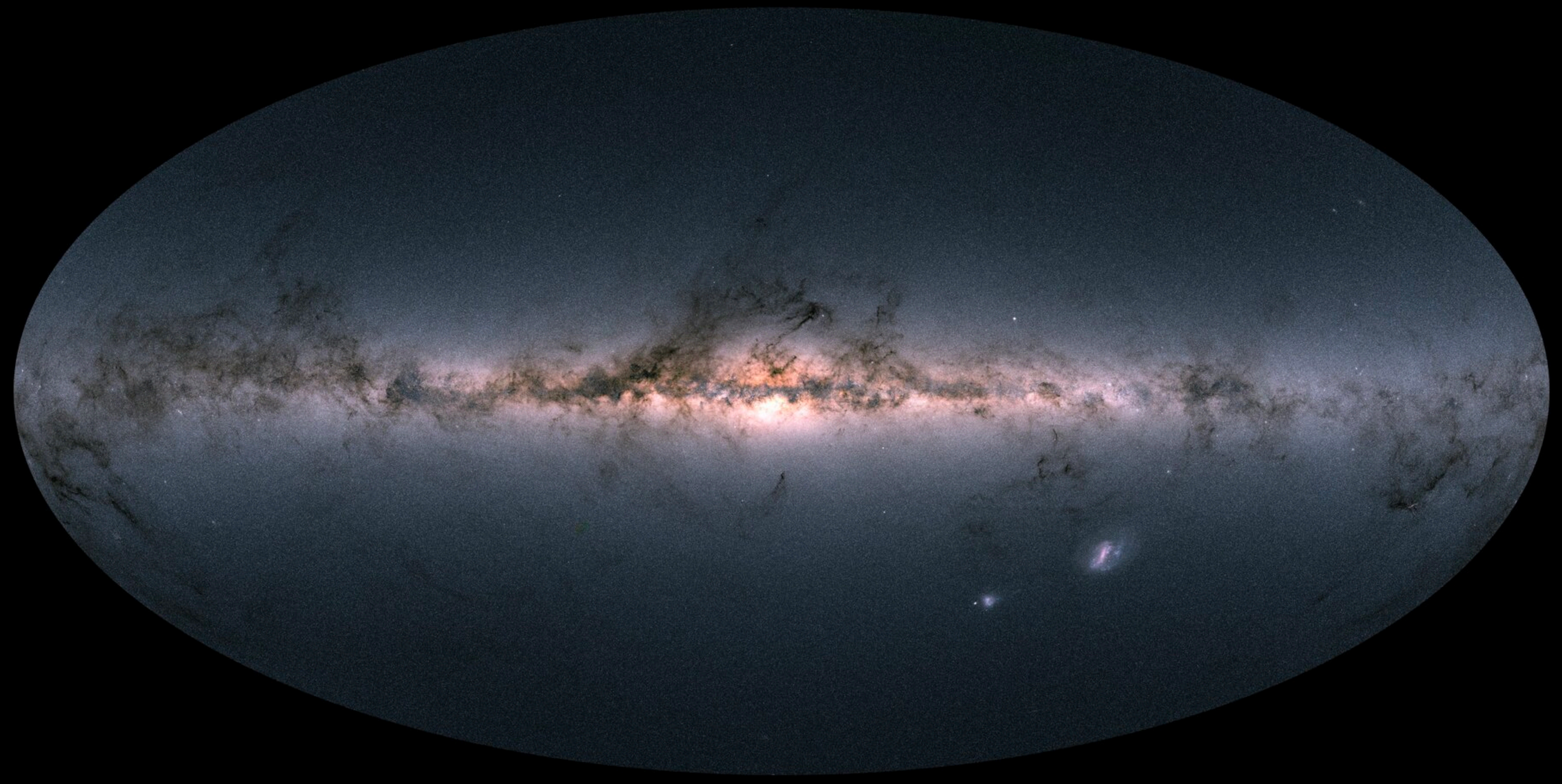
- Modern radio SETI began in the 1960s
- Vast improvements and expansion in:
    - Instantaneous bandwidth
    - Sensitivity
    - Survey size
    - Search strategies

Where should we look?

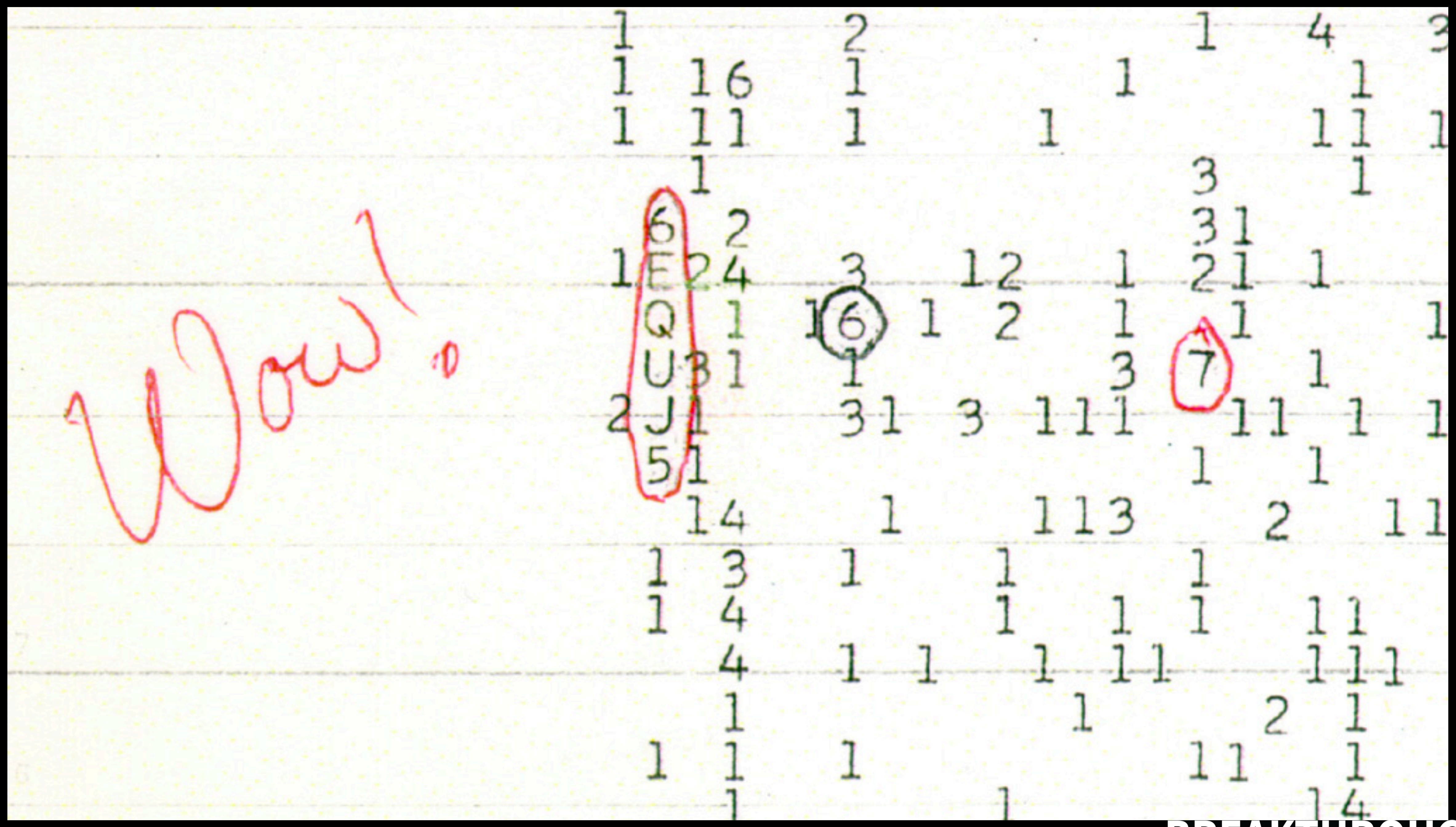


NASA

Big Ear Radio Observatory

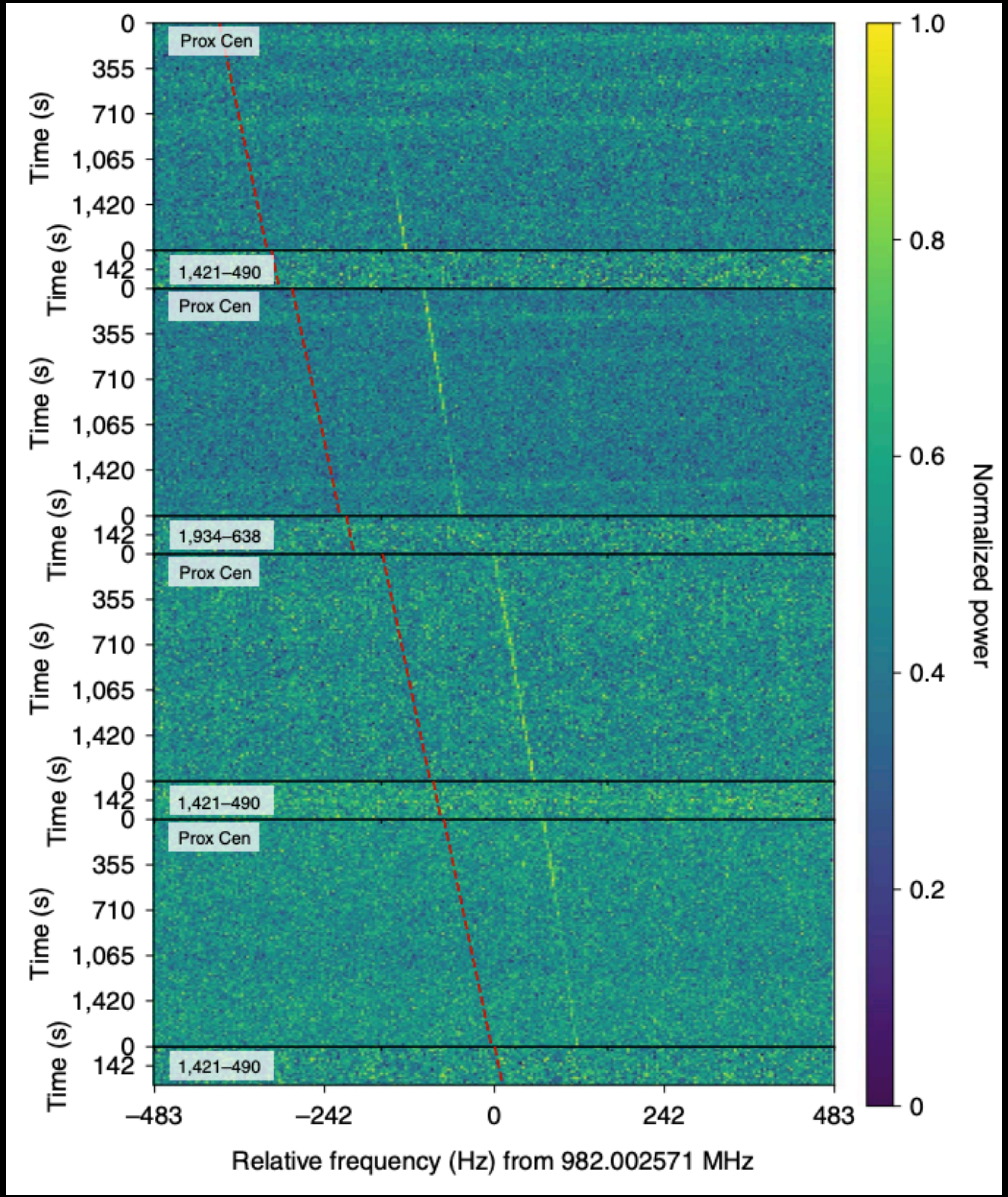


ESA



How should we look? What makes for a convincing candidate?

- **Narrowband** vs. astrophysical sources
- **Non-zero drift rate** vs. RFI
- **Sky localization** vs. RFI



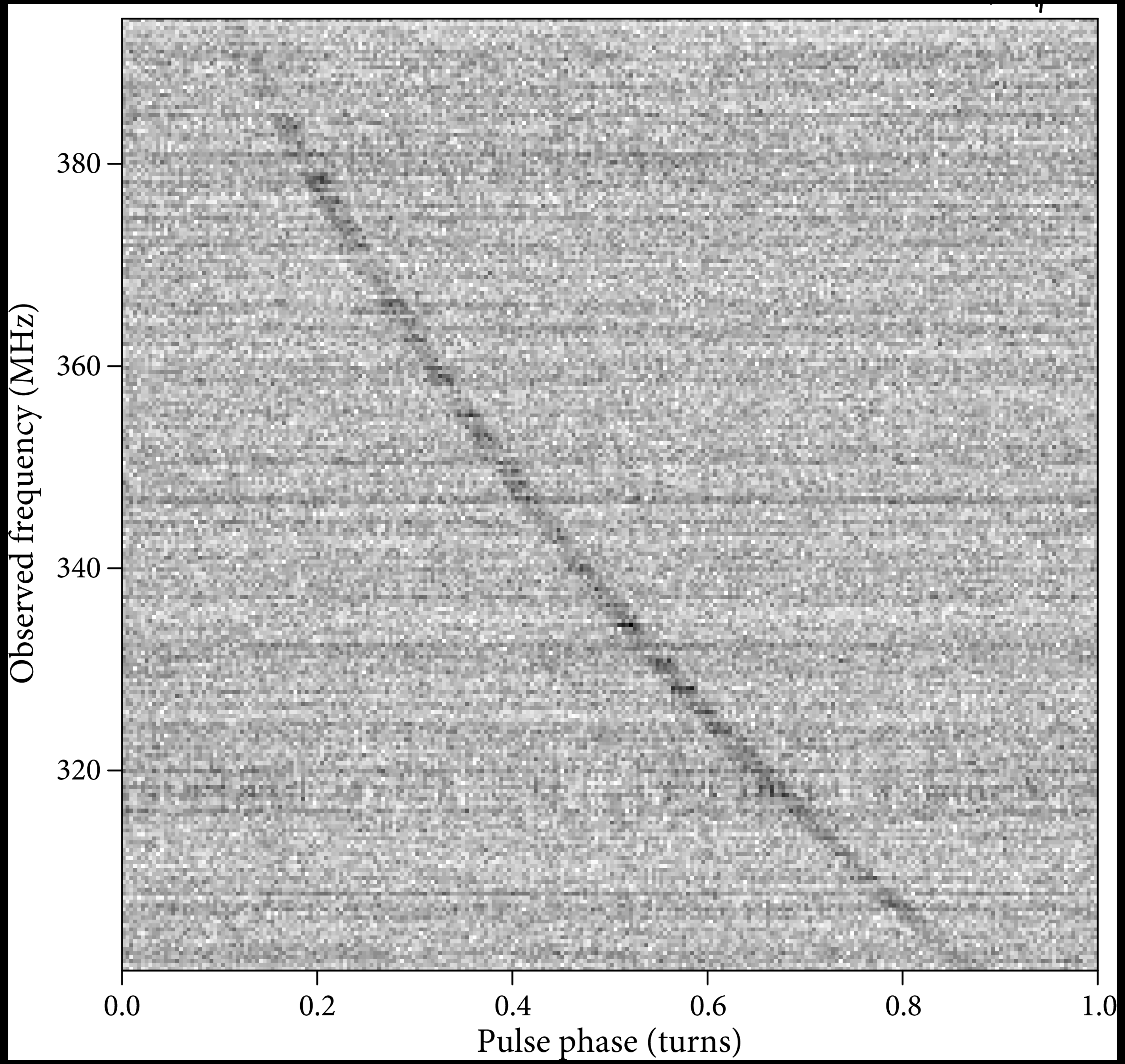
Smith et al. 2021

Can we use astrophysical phenomena as a filter to distinguish technosignatures from RFI?

ESA

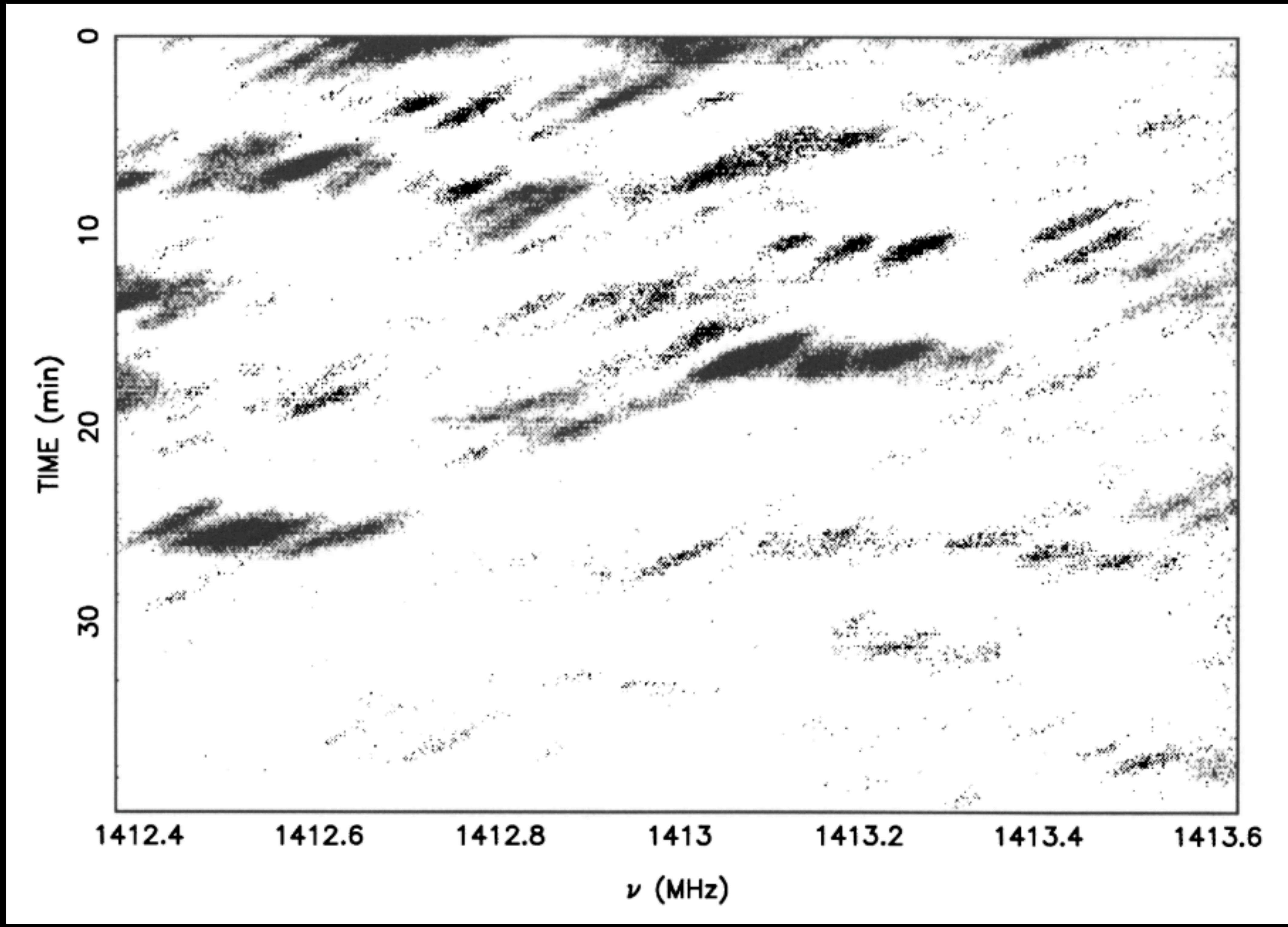
Pulsar observations probe radio plasma effects

Dispersion



Condon & Ransom 2016

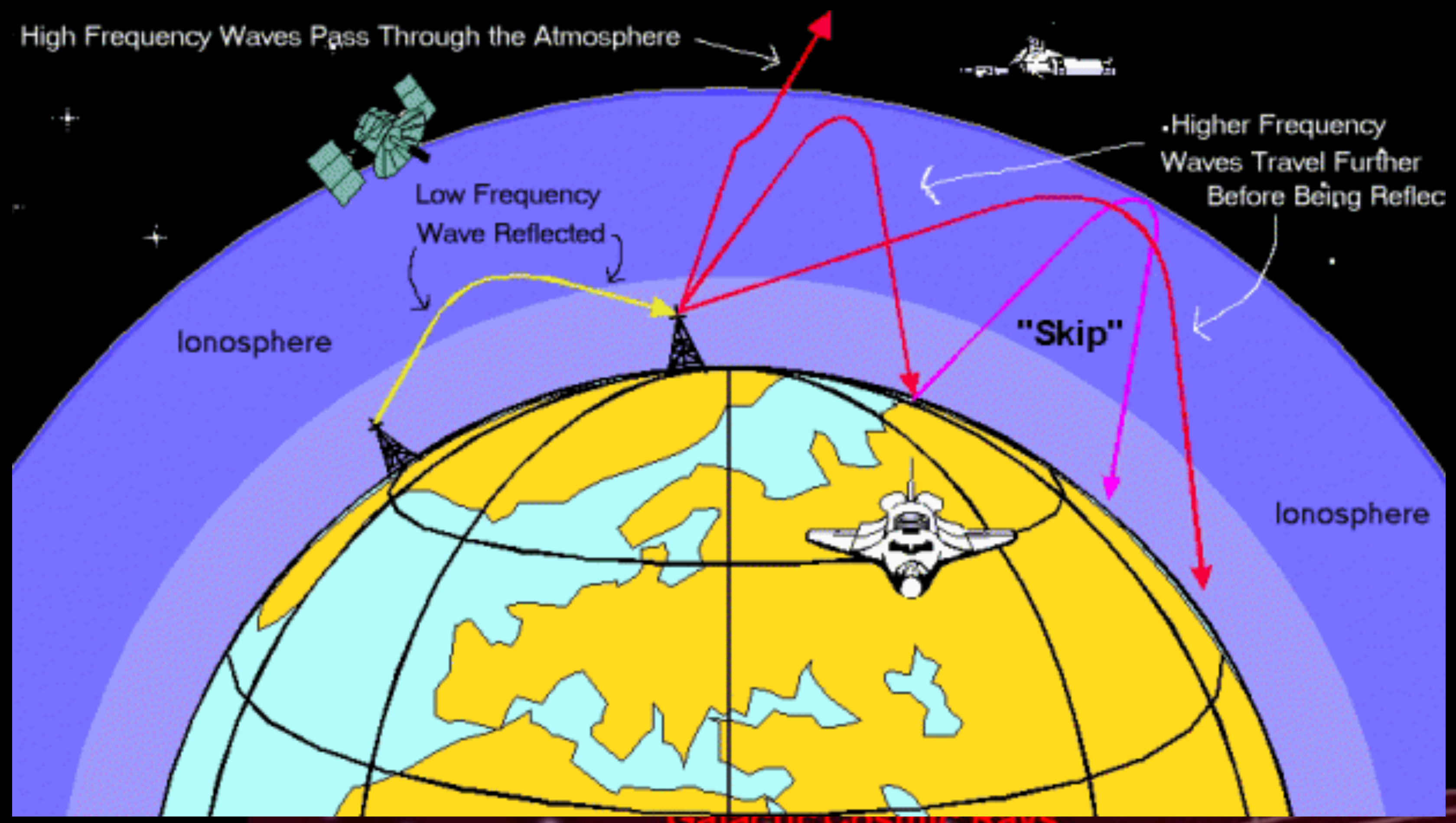
Scattering

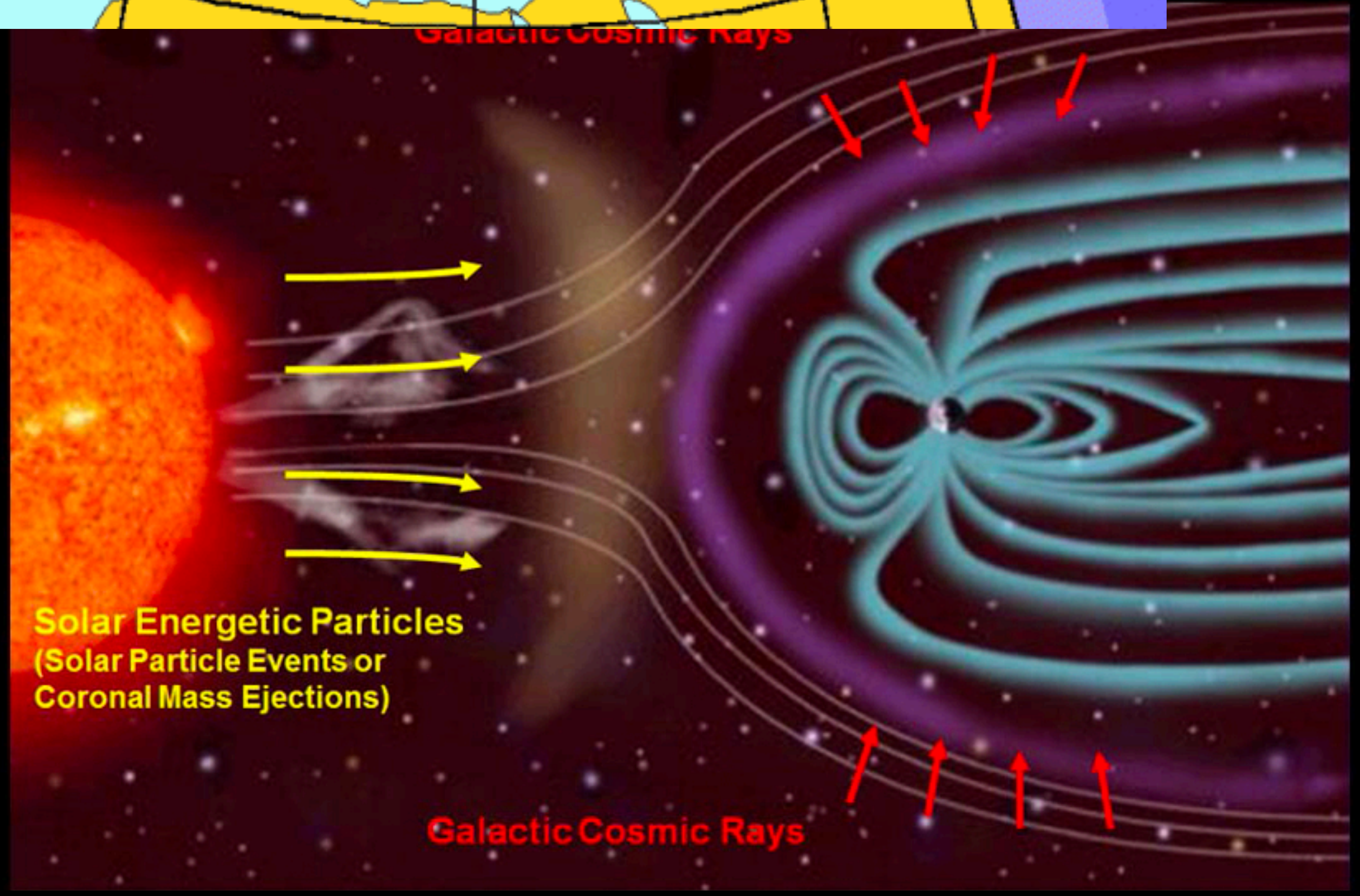


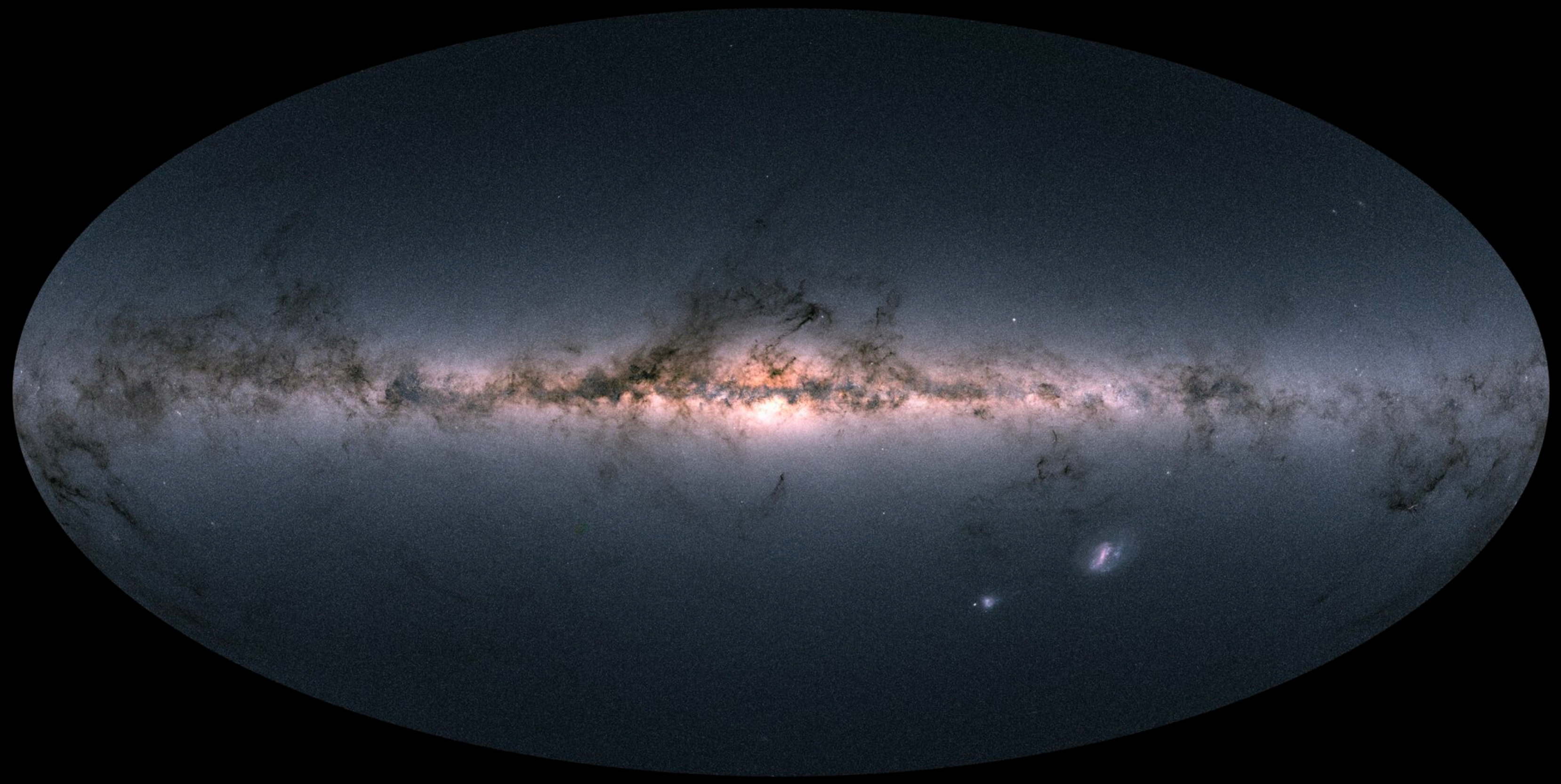
Cordes & Lazio 1991

Regions of ionized plasma

- Ionosphere
- Interplanetary Medium (IPM)
- Interstellar Medium (ISM)





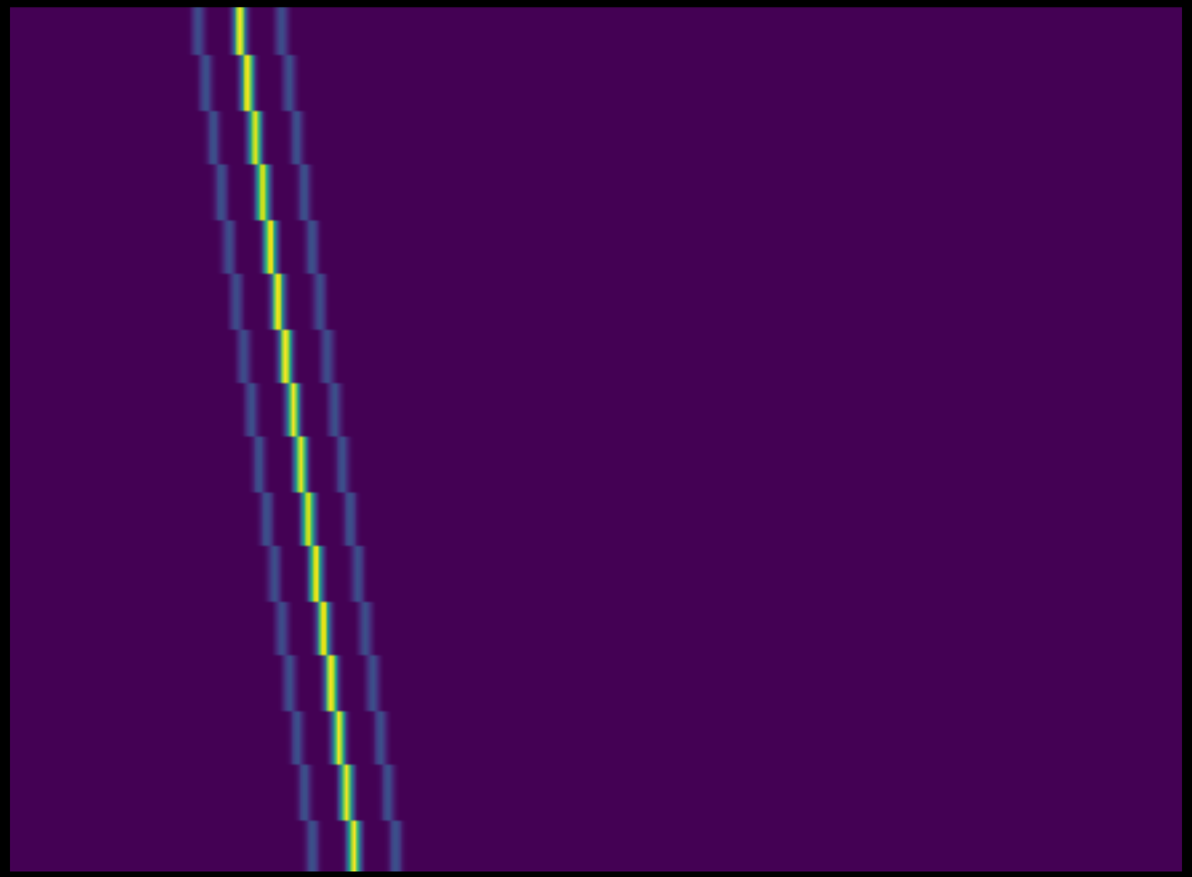


What would strongly scattered signals look like?

- Temporal scintillation
- Spectral broadening
- Pulse broadening
- Spectral de-correlation

What would strongly scattered signals look like?

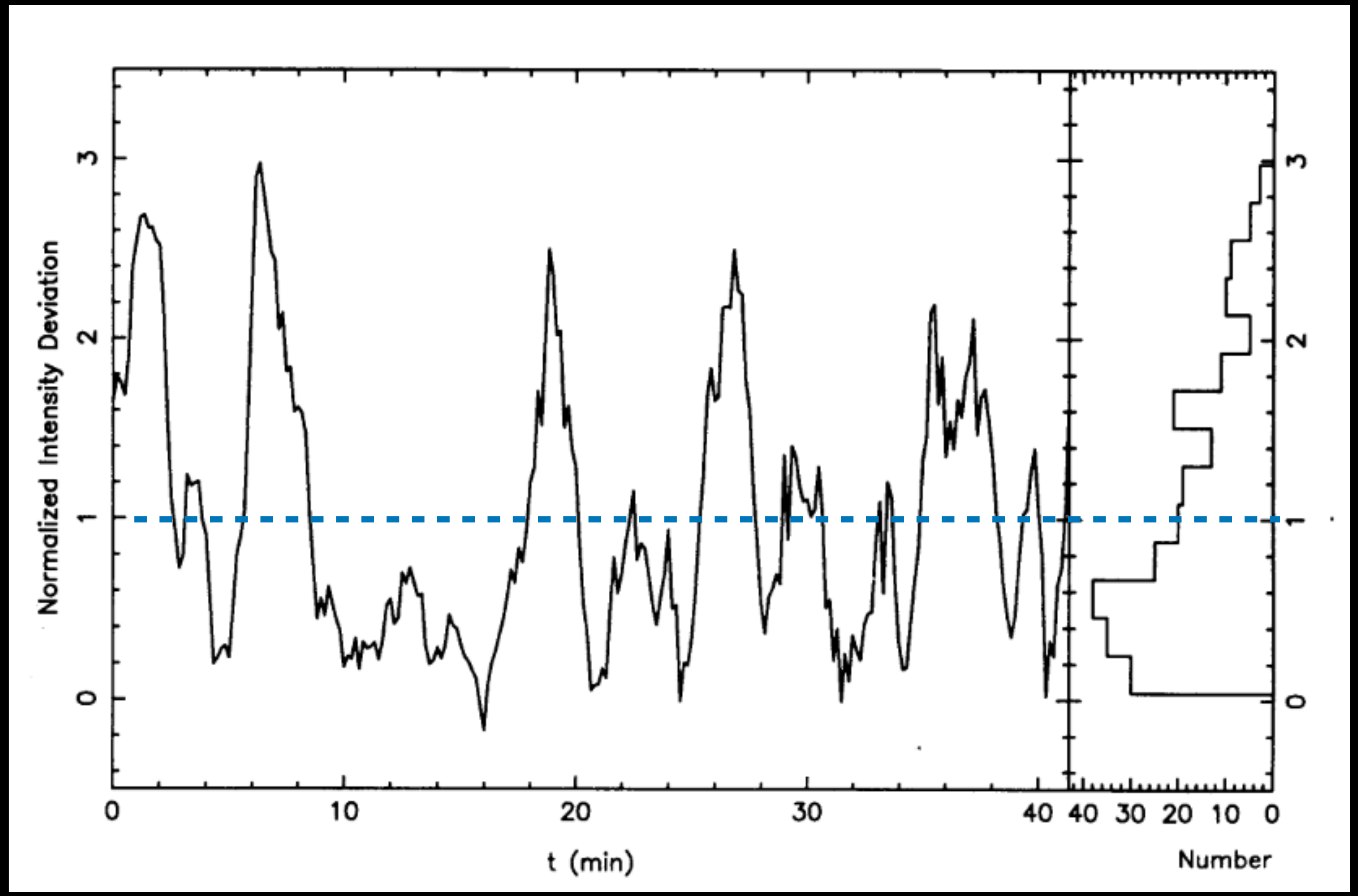
- Assuming a 100% duty-cycle narrowband transmitter



- Temporal scintillation
- Spectral broadening
- ~~Pulse broadening~~
- ~~Spectral de-correlation~~

What would strongly scattered signals look like?

- Assuming a 100% duty-cycle narrowband transmitter



Cordes & Lazio 1991

- **Temporal scintillation**
- ~~Spectral broadening~~
- ~~Pulse broadening~~
- ~~Spectral de-correlation~~

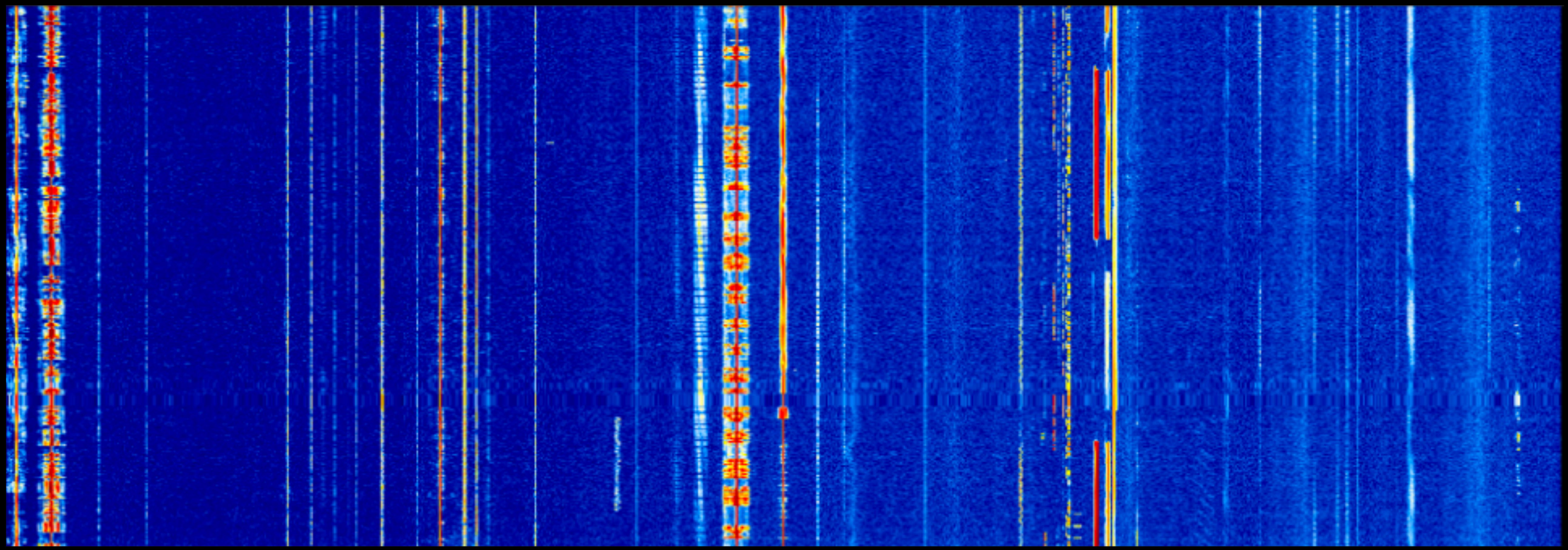
Prior SETI research on scattering

INTERSTELLAR SCATTERING EFFECTS ON THE DETECTION OF NARROW-BAND SIGNALS

JAMES M. CORDES AND T. JOSEPH LAZIO
National Astronomy and Ionosphere Center and Department of Astronomy, Cornell University, Ithaca, NY 14853

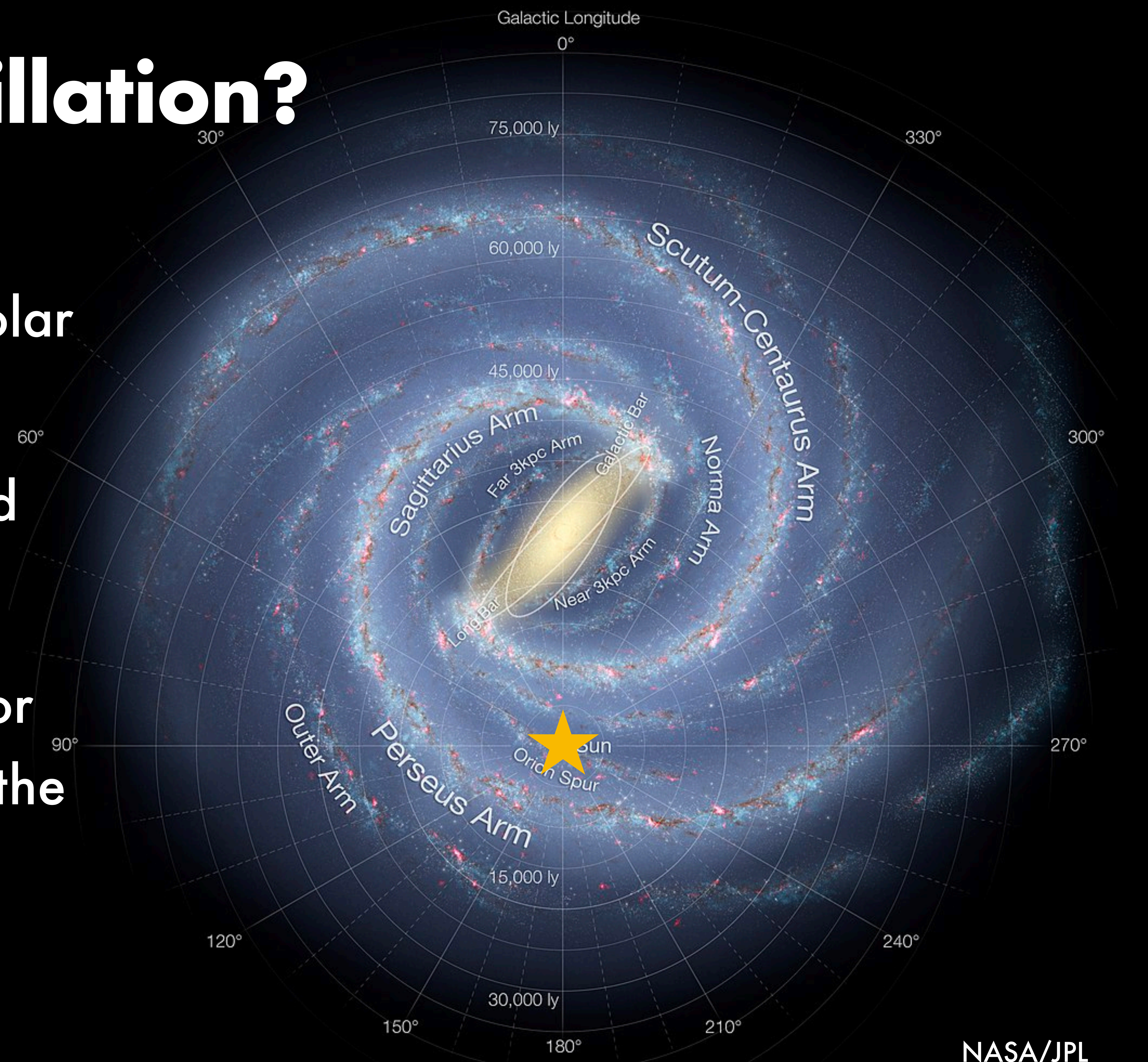
Received 1990 October 4; accepted 1991 January 15

- Many studies acknowledge scattering but attempt to avoid it
- Generally, SETI techniques aren't sensitive to detailed morphology
    - Noise, modulation, S/N
- Stochastic effects are hard to describe



Why search for scintillation?

- A filter that directly implies extra-solar origin
- Well-suited for continuous or pulsed *narrowband* signals
- One of the best places to search for scintillation corresponds to one of the best places to look for ETI – the Galactic Center

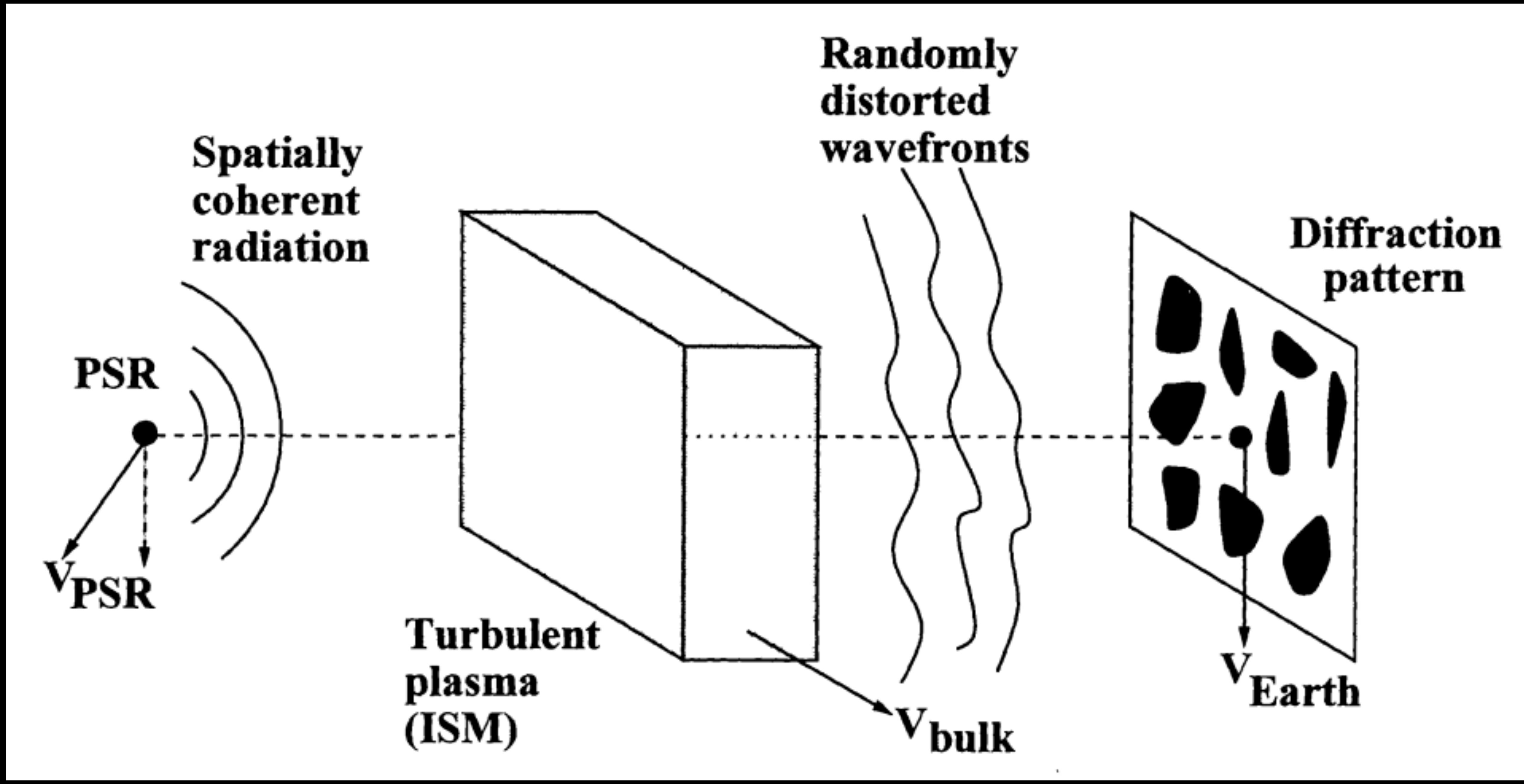


NASA/JPL

Diffractive scintillation in the ISM

- Electron density fluctuations in ionized plasma creates interference pattern

- Can lead to 100% intensity modulation, especially towards the Galactic center, with characteristic temporal scales Δt_d



Cordes 2002

INTERSTELLAR SCATTERING EFFECTS ON THE DETECTION OF NARROW-BAND SIGNALS

JAMES M. CORDES AND T. JOSEPH LAZIO
National Astronomy and Ionosphere Center and Department of Astronomy, Cornell University, Ithaca, NY 14853
Received 1990 October 4; accepted 1991 January 15

SCINTILLATION-INDUCED INTERMITTENCY IN SETI

JAMES M. CORDES,[1,2,3] T. JOSEPH W. LAZIO,[1,2] AND CARL SAGAN[1,3,4,5]
Received 1996 May 15; accepted 1997 May 9

- Showed that scattering can both help and hinder SETI

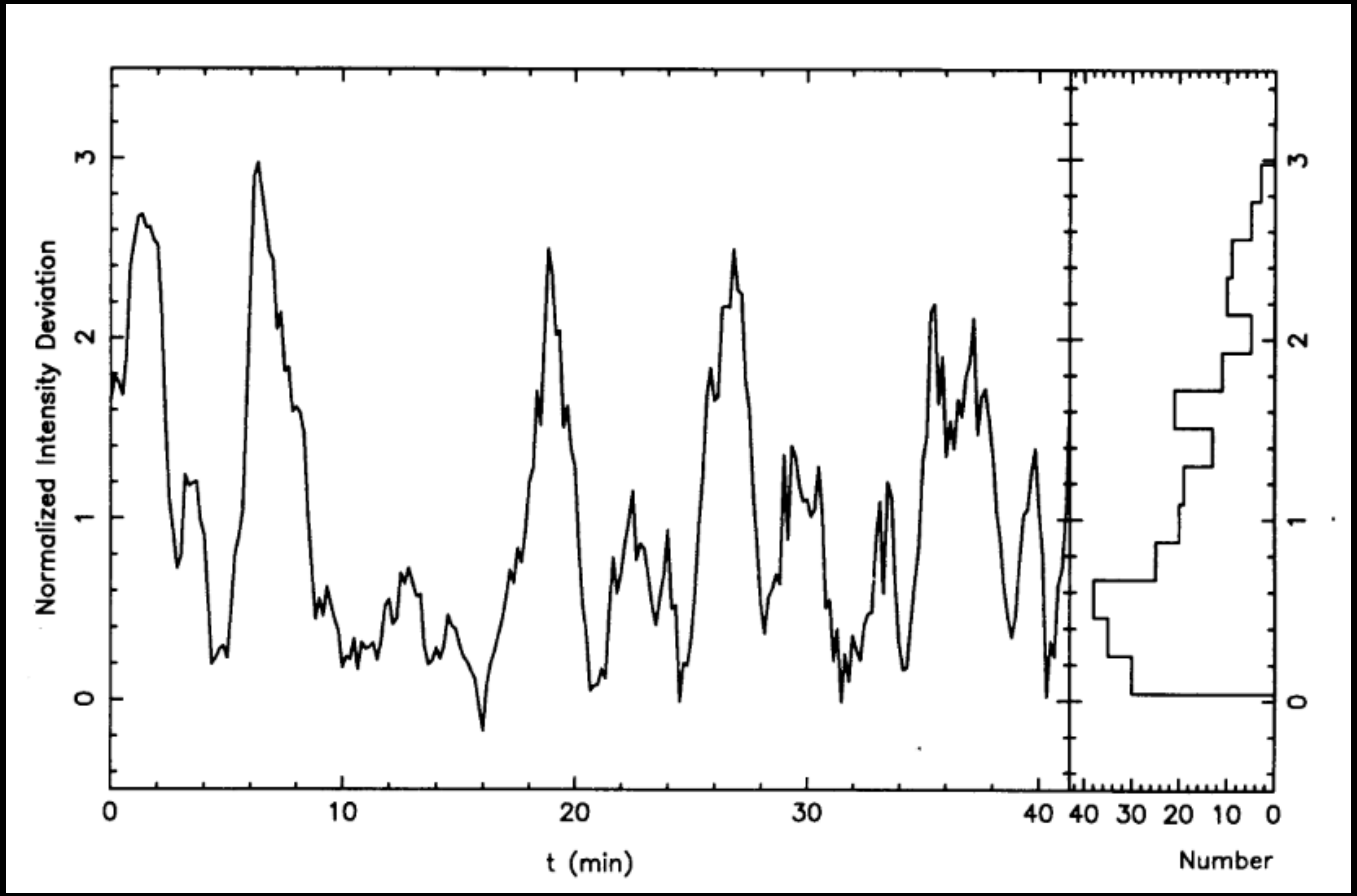
- Developed asymptotic expressions for detectability

What would strongly scintillated signals look like?

- Assuming a 100% duty-cycle narrowband transmitter
- Exponential intensity distribution

$$p(I) \propto e^{-I/\langle I \rangle}$$

- Near Gaussian auto-correlation (ACF)

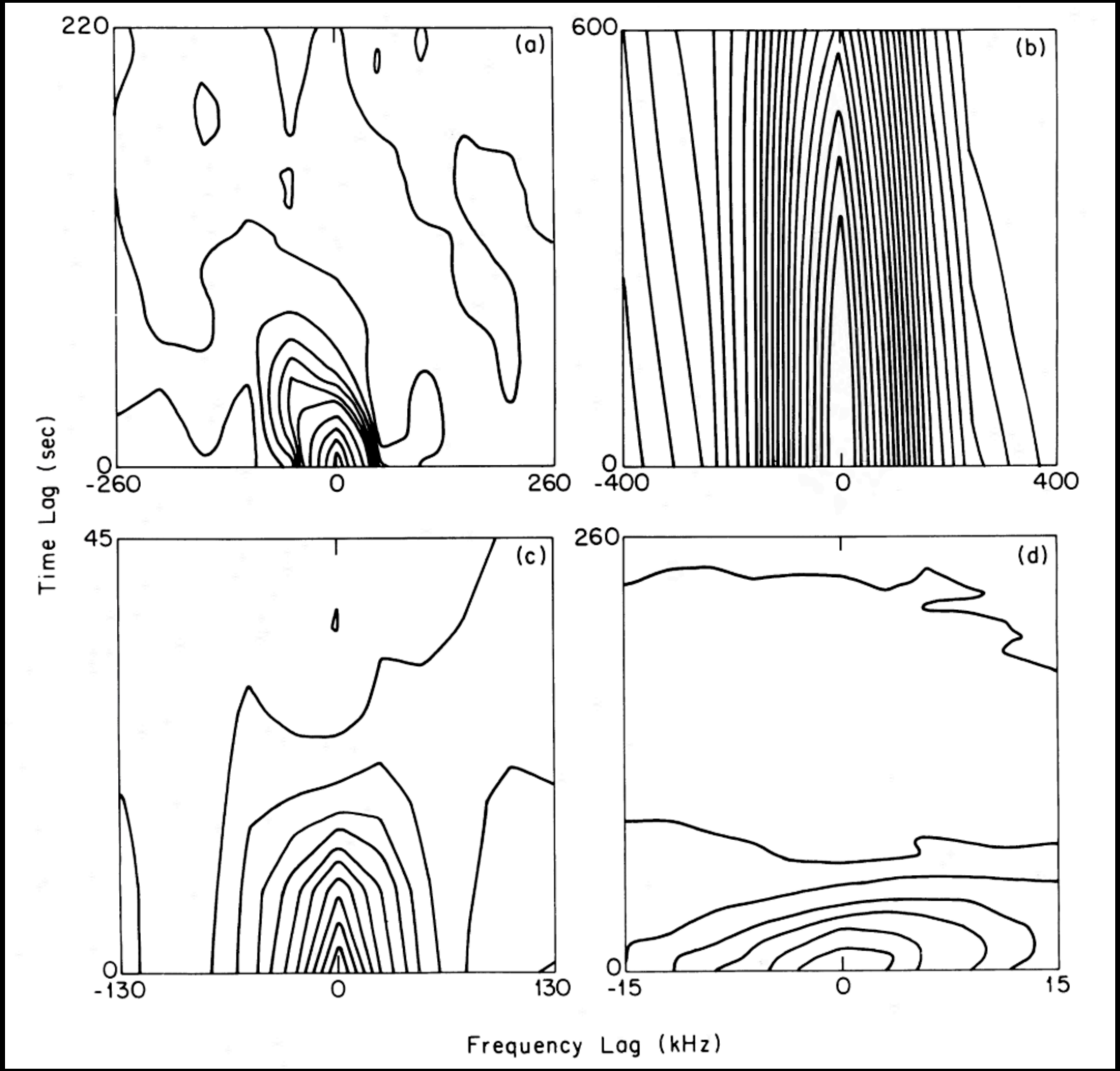


Cordes & Lazio 1991

What would strongly scintillated signals look like?

- Assuming a 100% duty-cycle narrowband transmitter
- Exponential intensity distribution
- Near Gaussian auto-correlation (ACF)

$$\rho(\tau) \sim e^{-(\tau/\Delta t_d)^{5/3}}$$



Cordes 1986

Can we detect scintillated narrowband technosignatures?

1. What scintillation timescales should we expect?
2. How can we probe asymptotic statistics?
3. Can we differentiate scintillated signals from existing RFI?
4. How can we design a survey to search for scintillated technosignatures?

Can we detect scintillated narrowband technosignatures?

1. **What scintillation timescales should we expect?**
2. How can we probe asymptotic statistics?
3. Can we differentiate scintillated signals from existing RFI?
4. How can we design a survey to search for scintillated technosignatures?

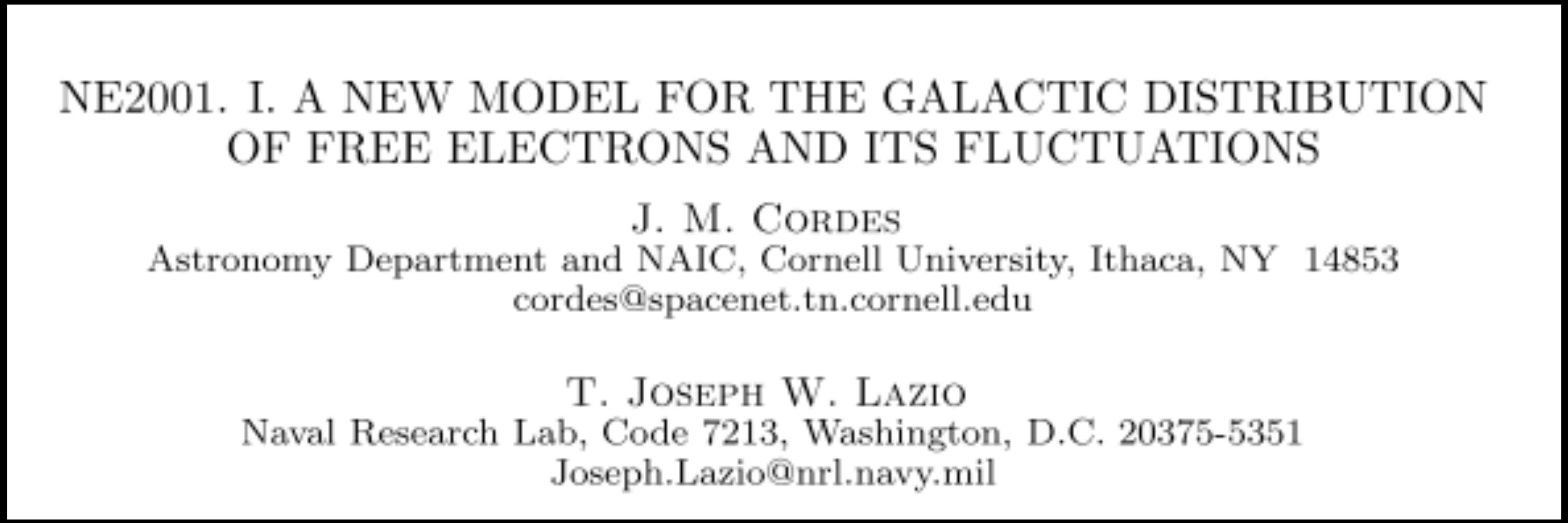
Estimating scattering strength

- NE2001 model: the standard for estimating pulsar distances for a while

- Estimates scattering parameters

- Computes values assuming defaults of 1 GHz and 100 km/s – requires scaling!

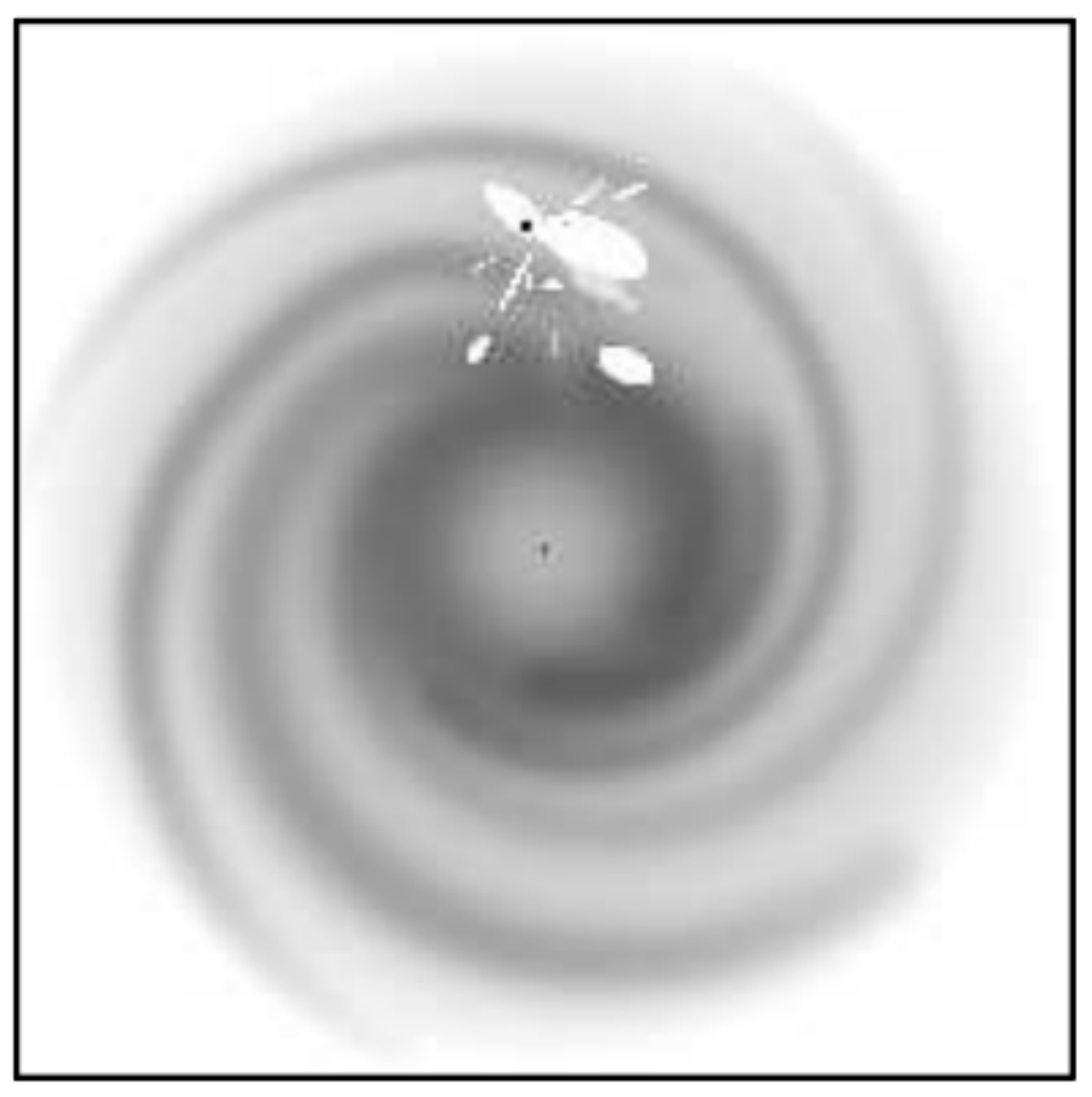
$$\Delta t_d \propto \nu^{6/5} v_T^{-1}$$



NE2001. I. A NEW MODEL FOR THE GALACTIC DISTRIBUTION OF FREE ELECTRONS AND ITS FLUCTUATIONS

J. M. Cordes
Astronomy Department and NAIC, Cornell University, Ithaca, NY 14853
cordes@spacenet.tn.cornell.edu

T. Joseph W. Lazio
Naval Research Lab, Code 7213, Washington, D.C. 20375-5351
Joseph.Lazio@nrl.navy.mil



Parameter space exploration of scattering parameters

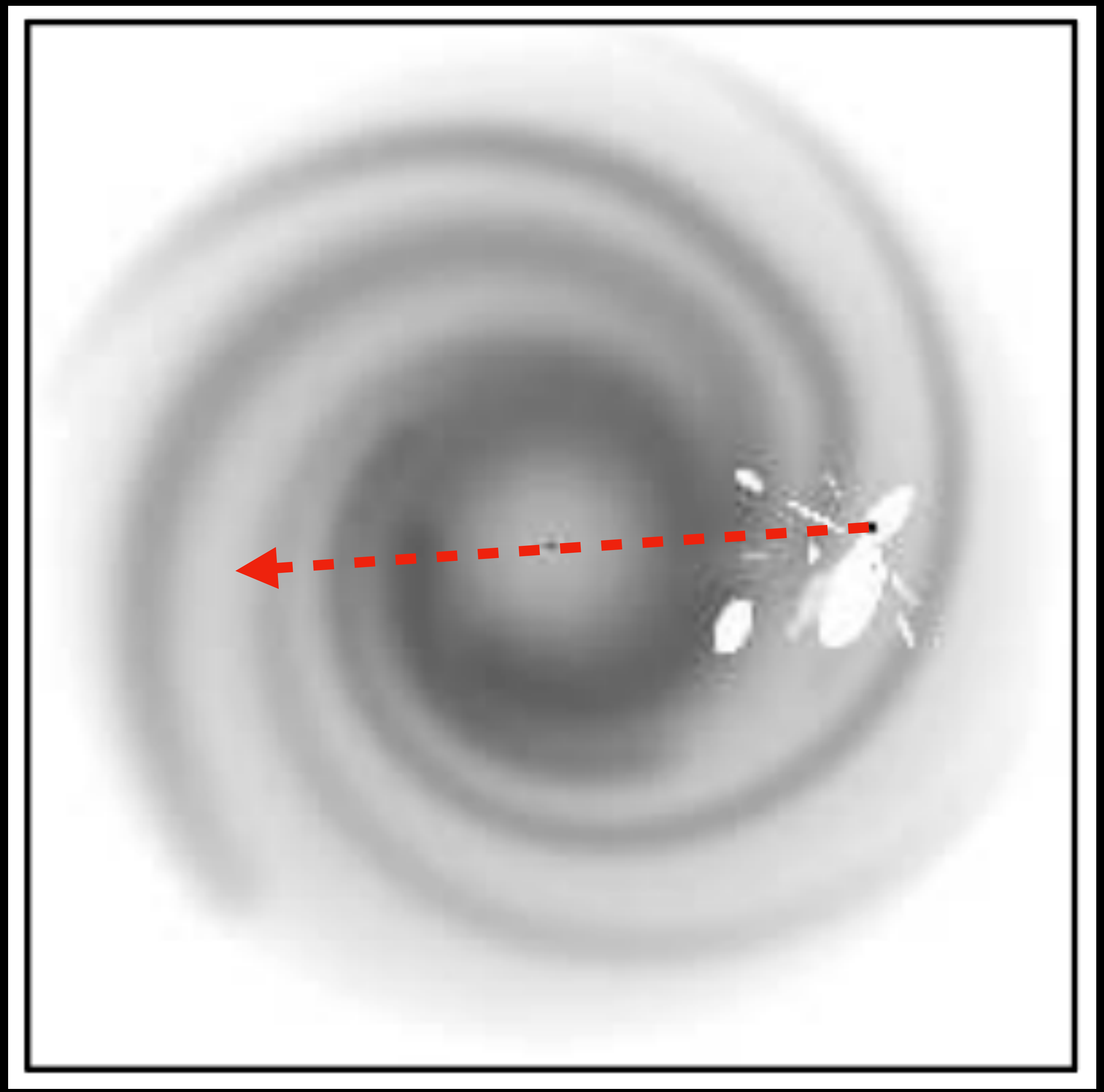
- A priori, we do not know:
  - Sky direction
  - Frequency
  - Distance
  - Transverse velocity

Monte Carlo sampling!

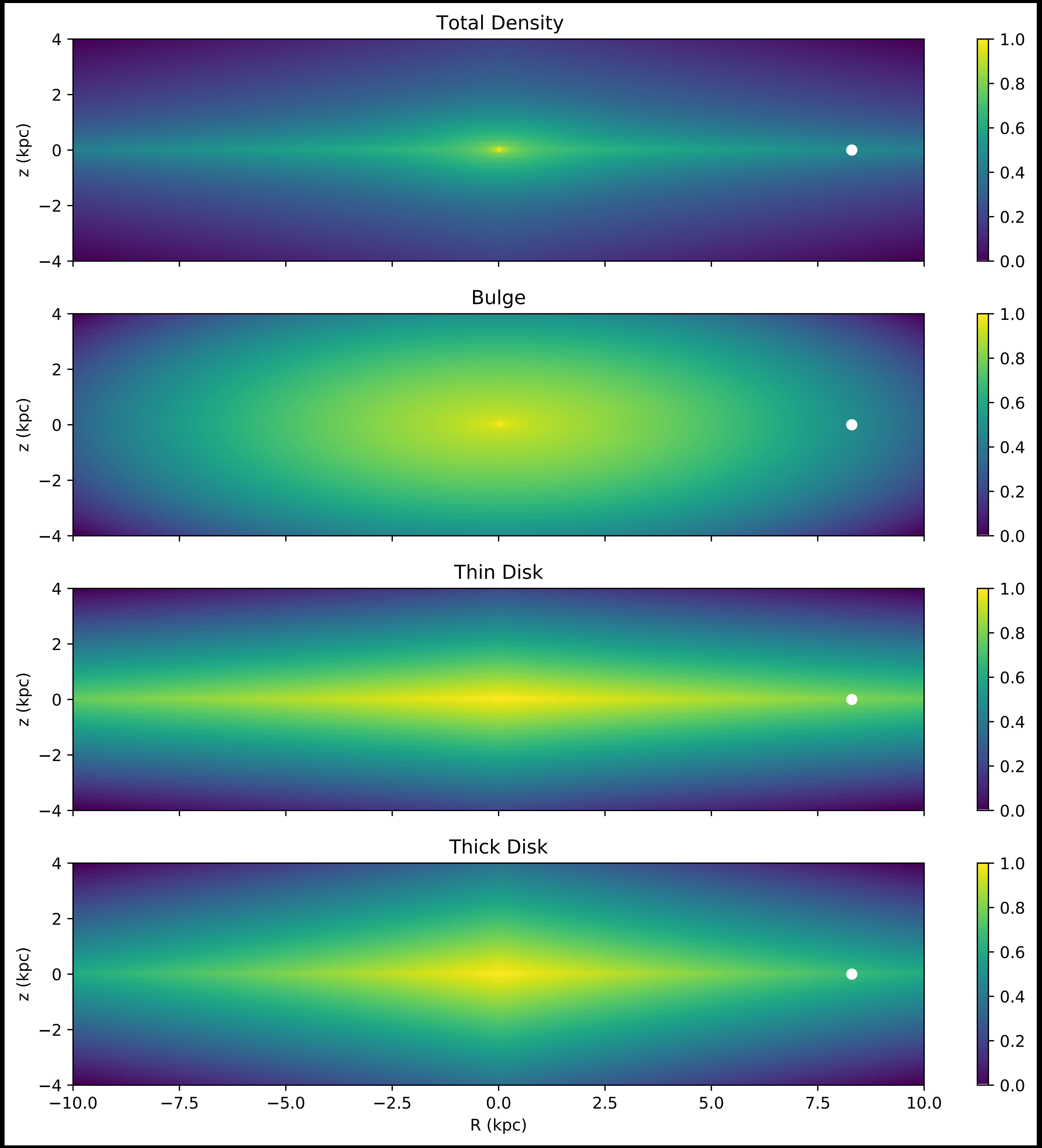
$$\Delta t_d \propto \nu^{6/5} v_T^{-1}$$

• Sky direction	→	Chosen parameter
• Frequency	→	Uniform sampling within chosen band
• Distance	→	Uniform or density based sampling
• Transverse velocity	→	Uniform sampling

Density-based sampling

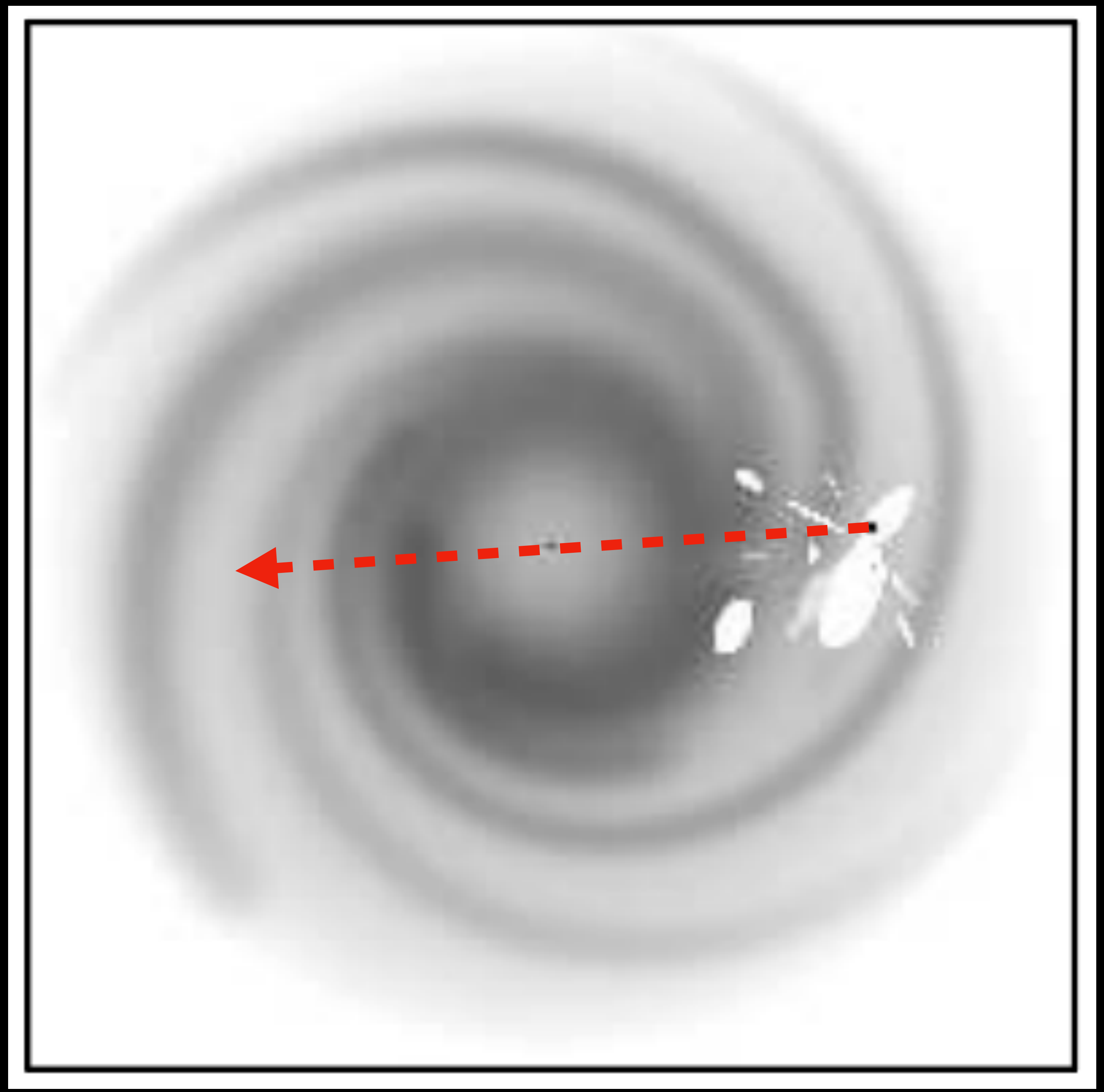


Cordes & Lazio 2002

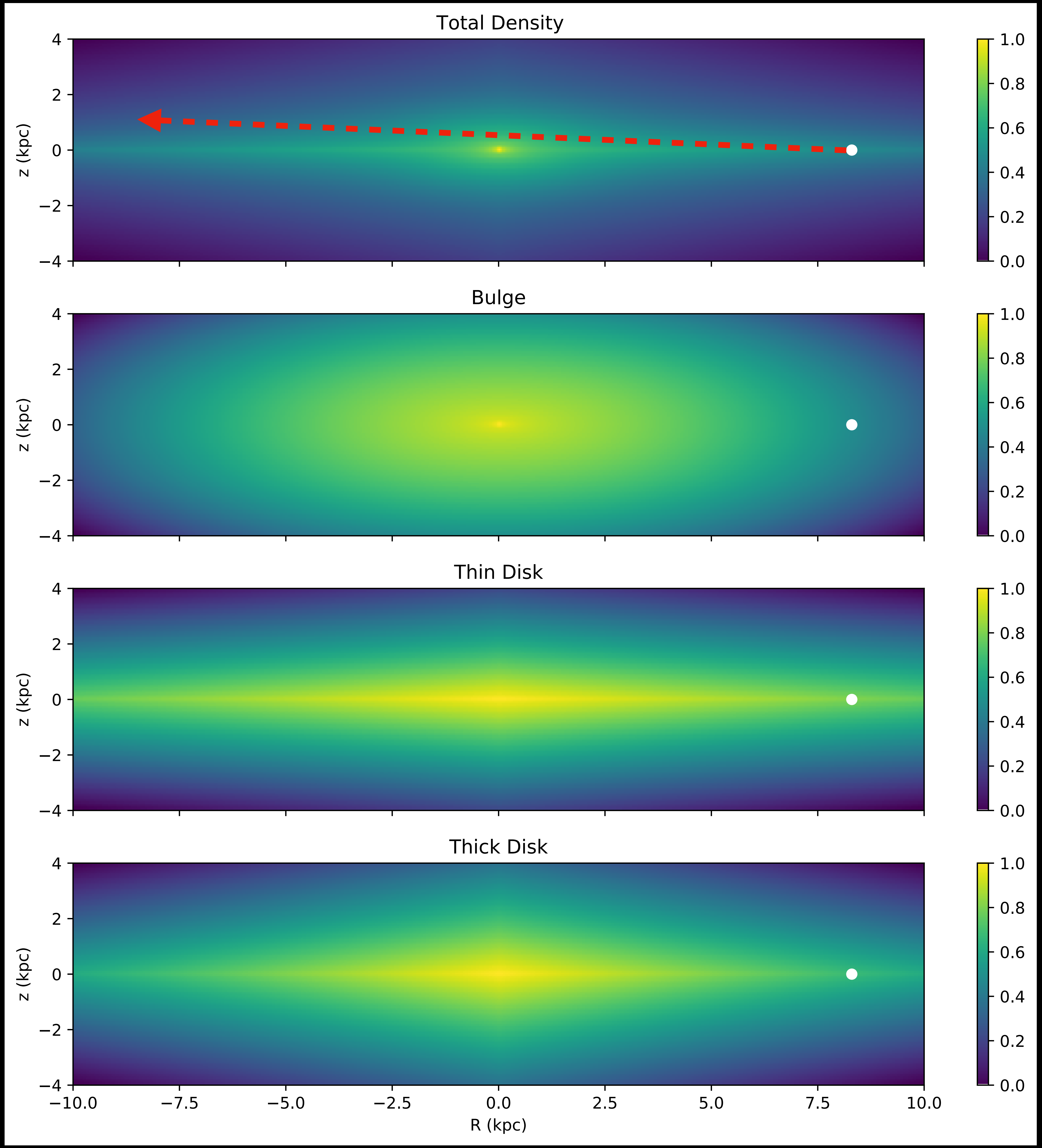


McMillan 2017, Gowanlock et al. 2011, Carroll & Ostlie 2007

Density-based sampling

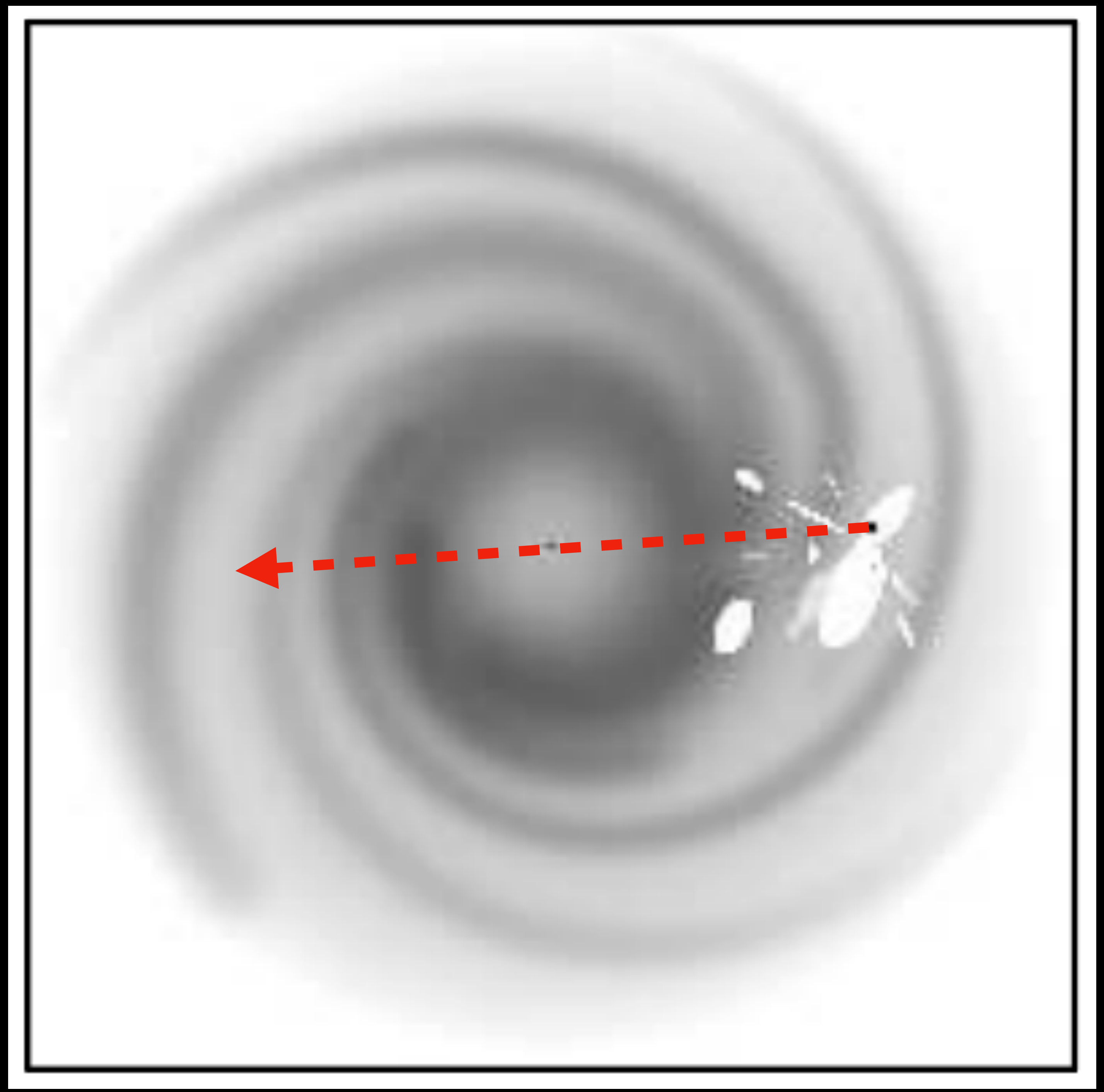


Cordes & Lazio 2002



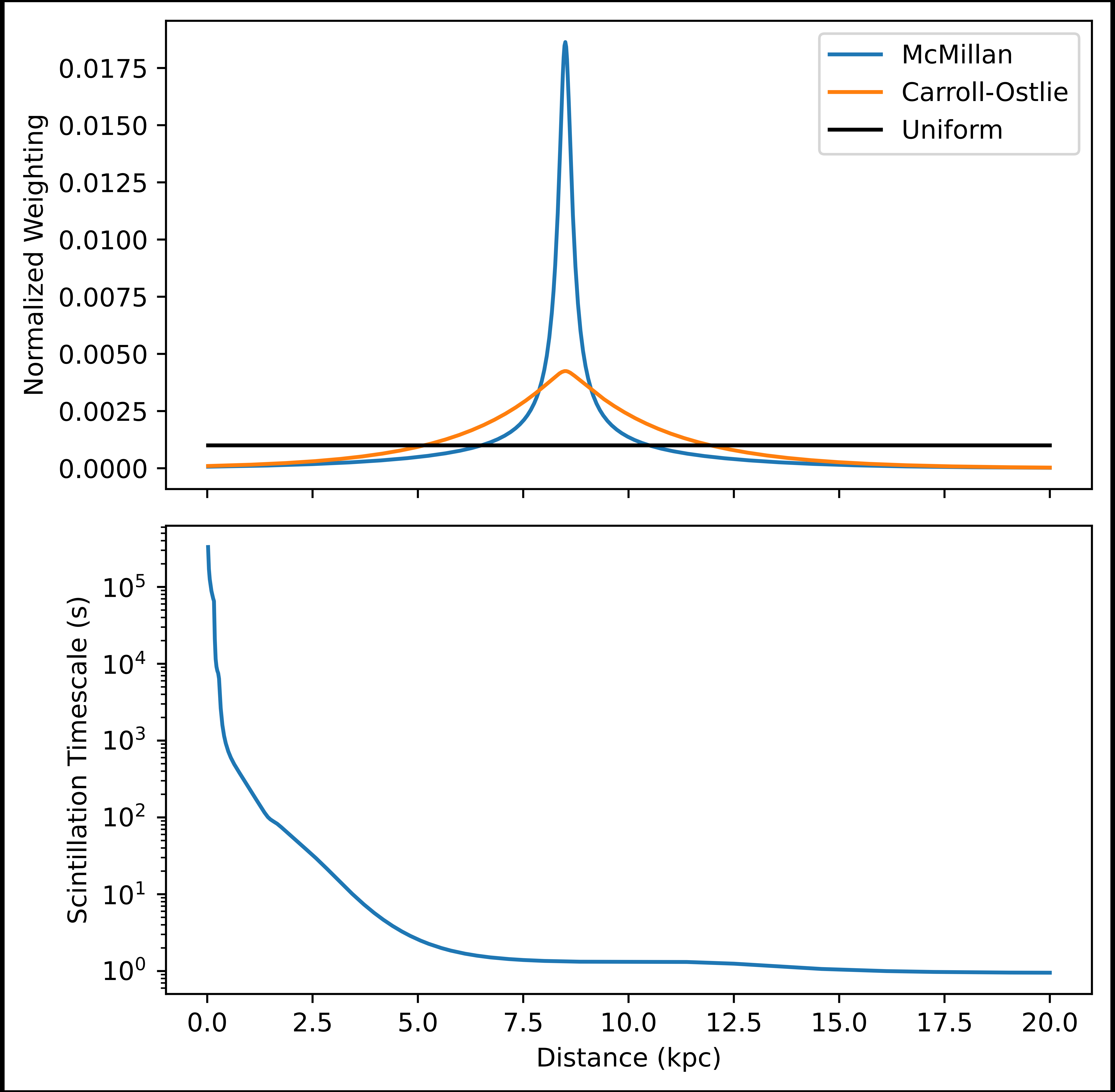
McMillan 2017, Gowanlock et al. 2011, Carroll & Ostlie 2007

Density-based sampling



Cordes & Lazio 2002

(l, b) = (1, 0)



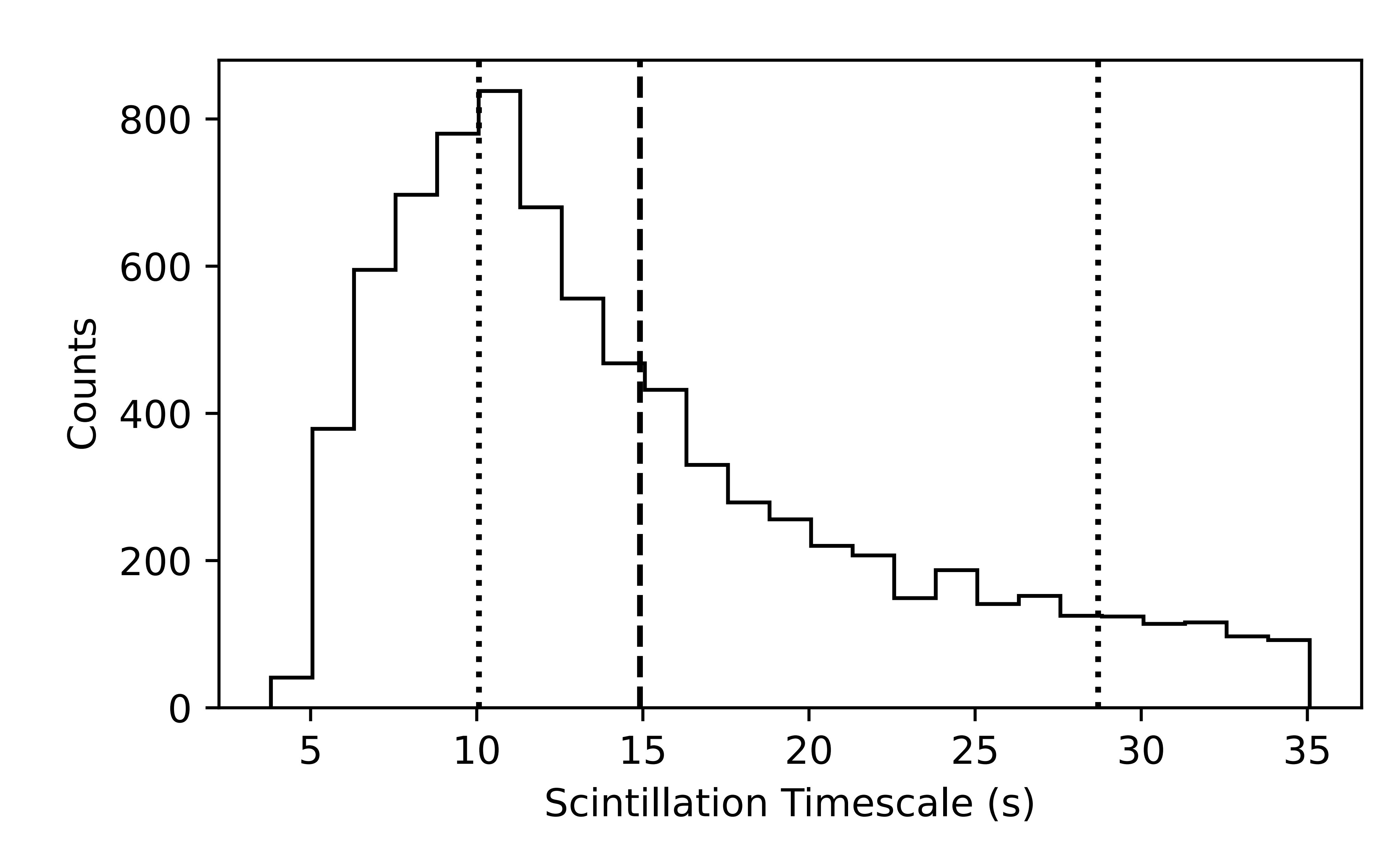
McMillan 2017, Gowanlock et al. 2011, Carroll & Ostlie 2007

Inter-quartile range

Median

C-band

(l, b) = (1, 0)

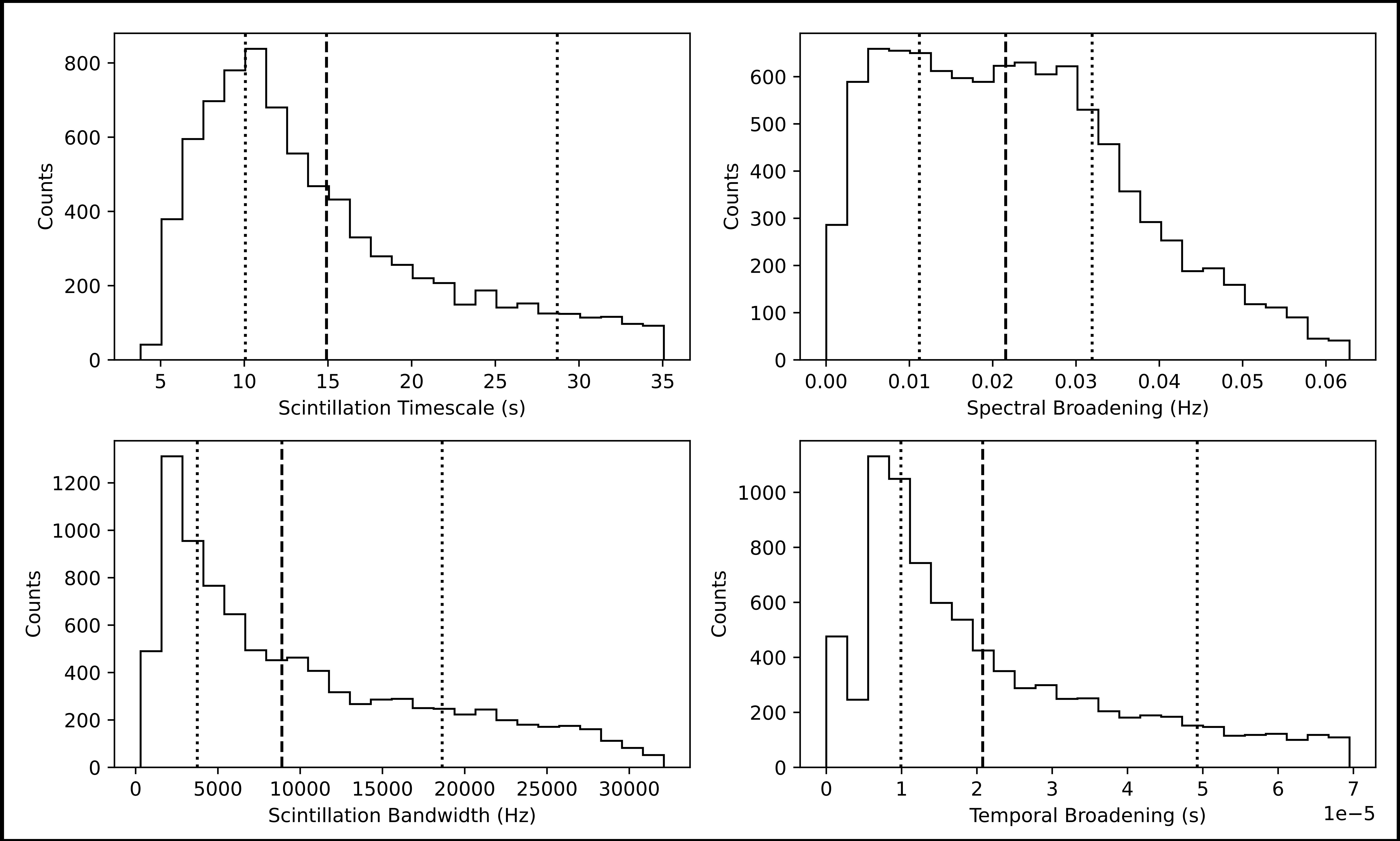


Inter-quartile range

Median

C-band

(l, b) = (1, 0)



Monte Carlo-sampled timescales

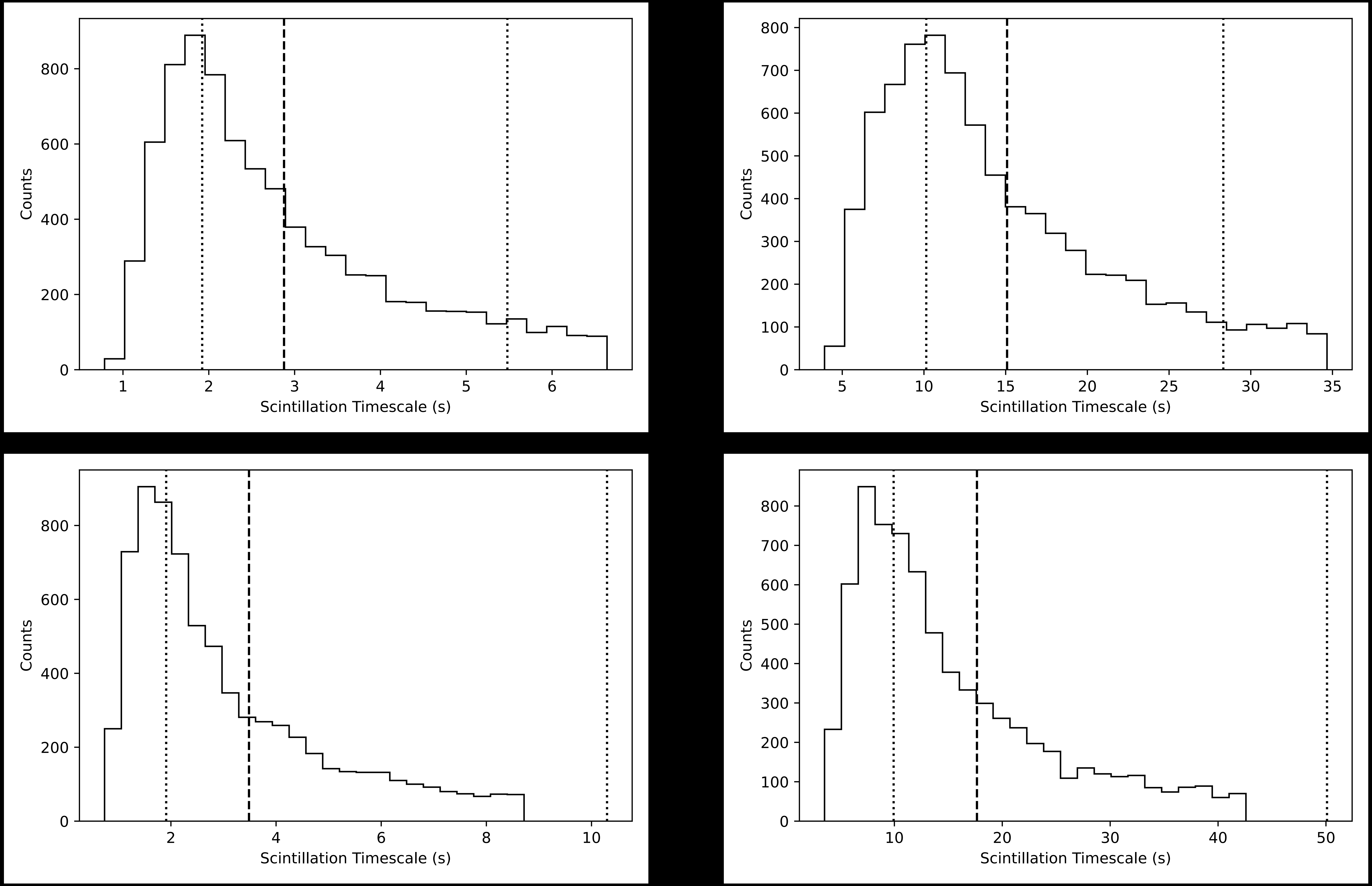
L band

C band

Distance sampling

Weighted

Uniform



Example: Statistics at different bands

(l, b) = (1, 0)

Band	Frequency (GHz)	Median (s)	IQR (s)	Mode (s)
LOFAR	0.110 – 0.240	0.22	0.14 – 0.41	0.14
L	1.1 – 1.9	2.9	1.9 – 5.6	1.9
S	1.8 – 2.8	4.8	3.3 – 9.0	3.1
C	3.95 – 8	15	10 – 28	11
X	8 – 11.6	28	19 – 52	16

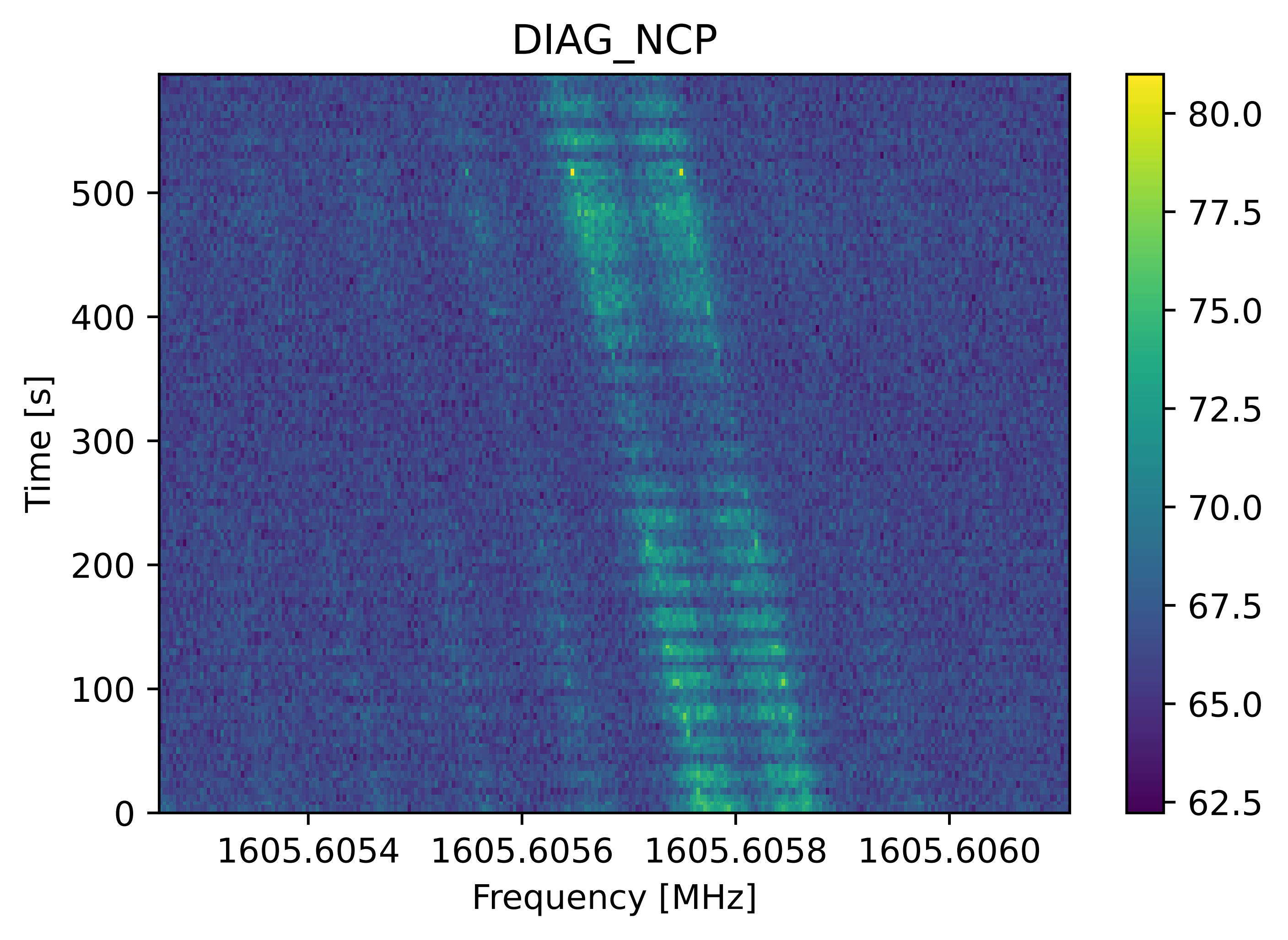
$$\Delta t_d \propto \nu^{6/5} v_T^{-1}$$

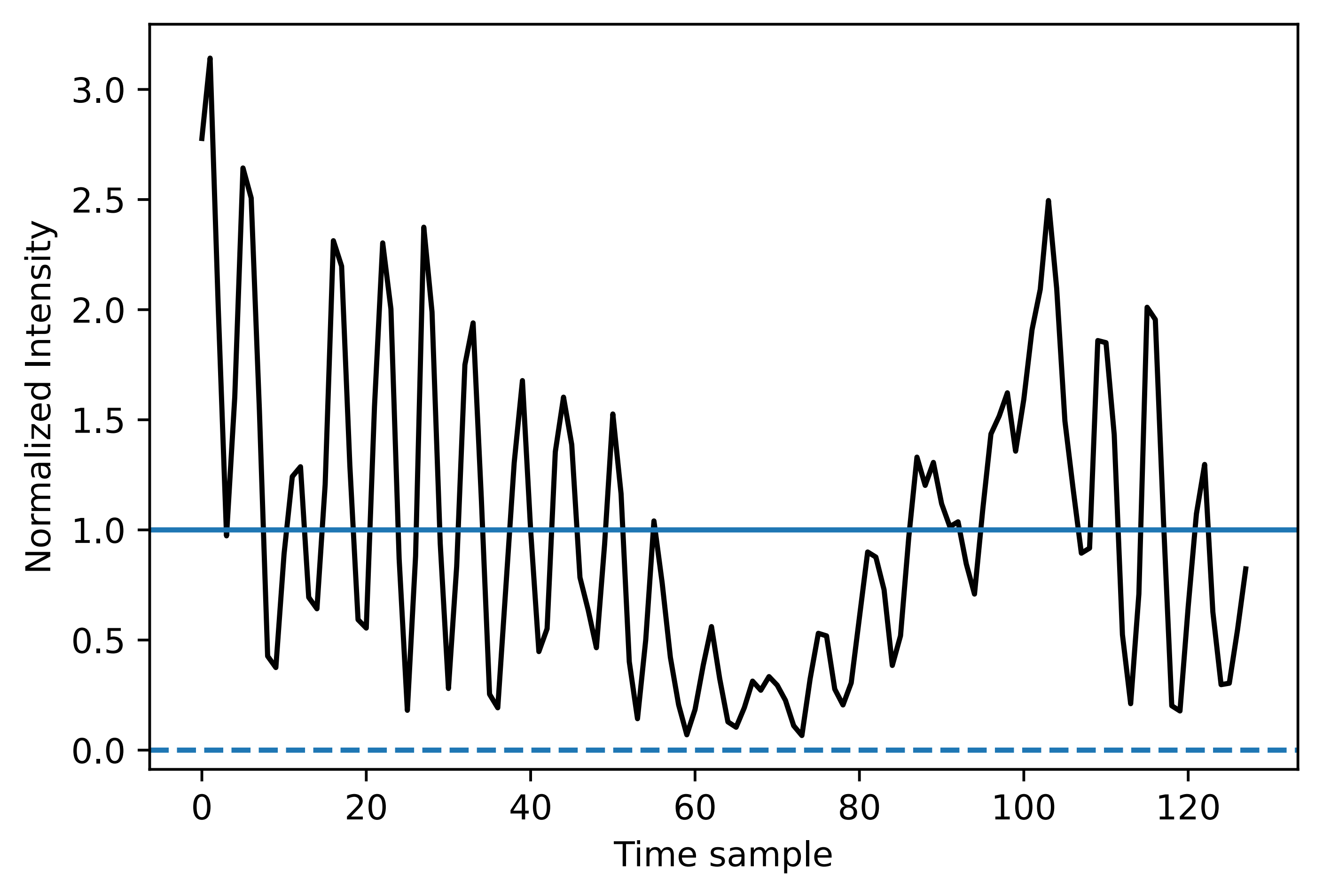
Can we detect scintillated narrowband technosignatures?

1. What scintillation timescales should we expect?
2. **How can we probe asymptotic statistics?**
3. Can we differentiate scintillated signals from existing RFI?
4. How can we design a survey to search for scintillated technosignatures?

How might we detect scintillation?

- Estimate intensity time series from signals detected with deDoppler methods
- Since scintillation is stochastic, identify **measurable statistics** for asymptotic behavior
- Would existing RFI modulation confound real scintillation?
    - Methods for creating synthetic scintillated intensities
    - Compare statistics of detected signals with those of synthetic scintillated signals





Set of diagnostic statistics

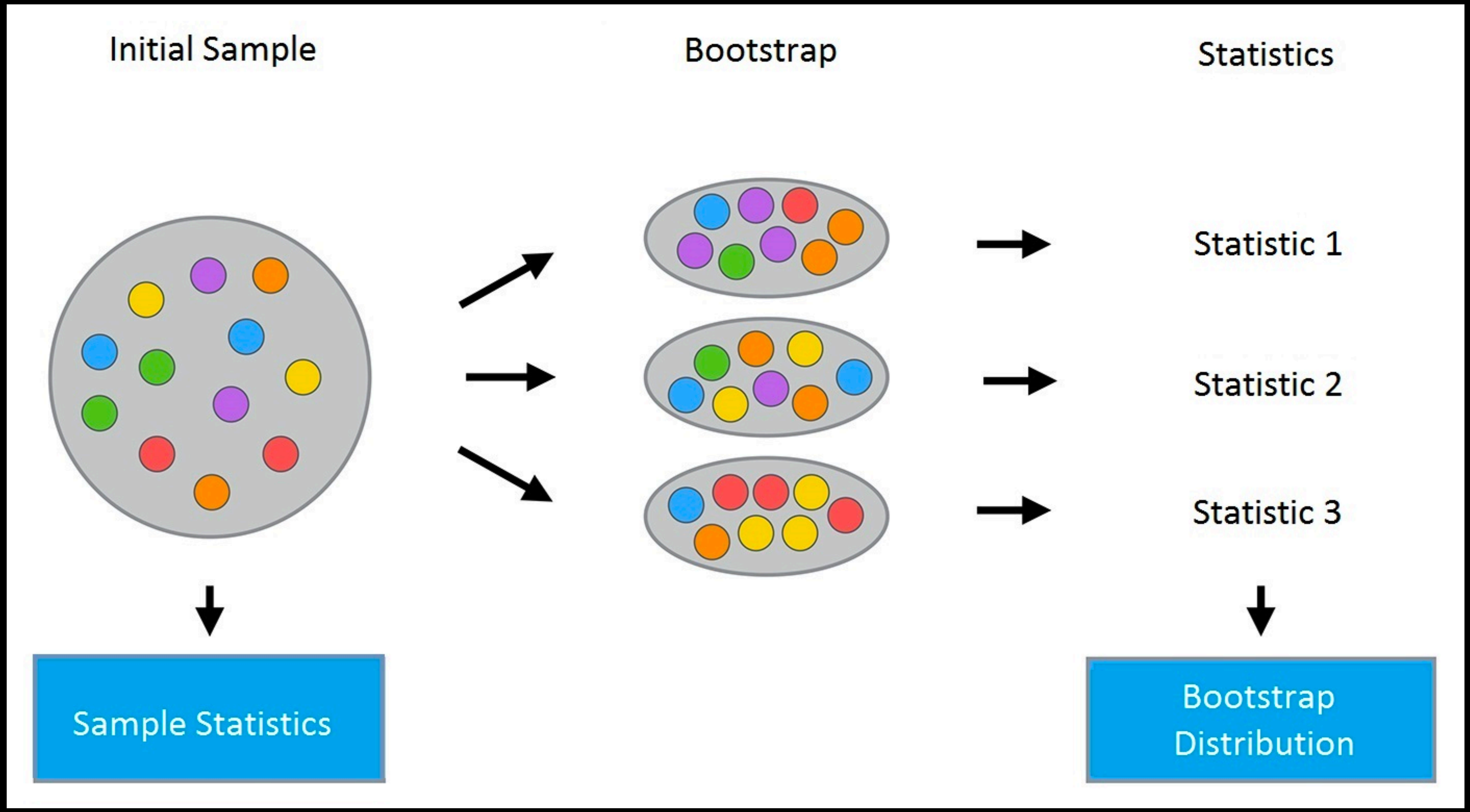
Statistic	Asymptotic Value (with no noise)	Data Type	Theoretical Behavior
Standard Deviation (RMS)	1	Intensity	Exponential
Minimum	0	Intensity	Exponential
Kolmogorov-Smirnoff statistic	0	Intensity	Exponential
Autocorrelation lag	Variable	Autocorrelation	Near-Gaussian
Least squares fit to autocorrelation	Variable	Autocorrelation	Near-Gaussian

There are a number of constraints...

- Time resolution - Sufficiently resolve scintles
- Observation time - Collect enough scintles, gain stability
- Signal brightness - Compute accurate statistics embedded in noise
- RFI environment - Bad normalization, false narrowband detections, confounding modulation

Low sample regime

- Spread of values around the asymptotic "truth"
- Both correlated and uncorrelated samples within the same observation
- How can we evaluate this?



Kaggle

Synthetic random processes: Autoregressive-to-anything (ARTA)

- (Cario & Nelson 1996) The ARTA process:
    - Matches a target intensity distribution
    - Matches a target autocorrelation structure (with custom asymptotic precision)

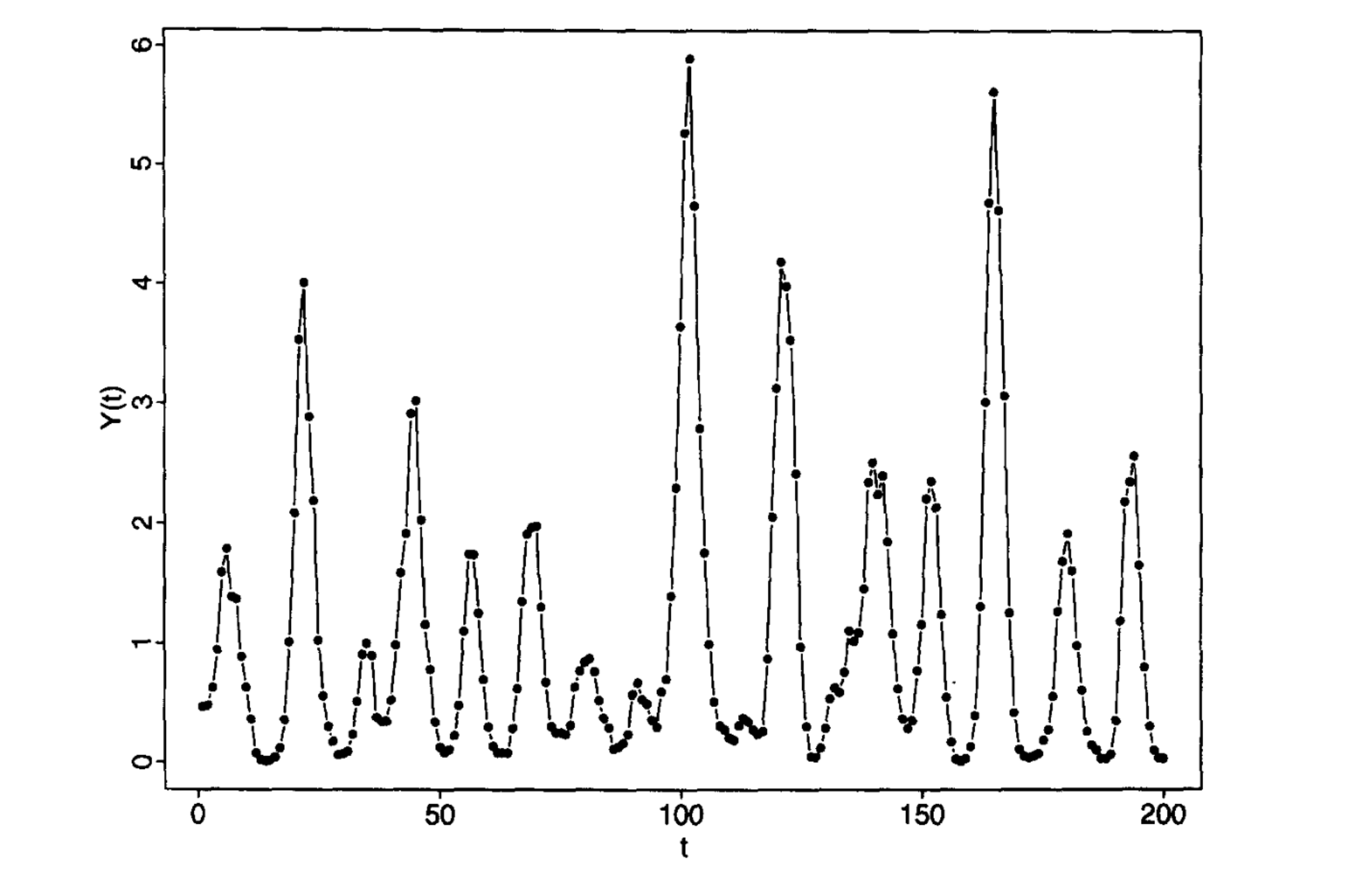
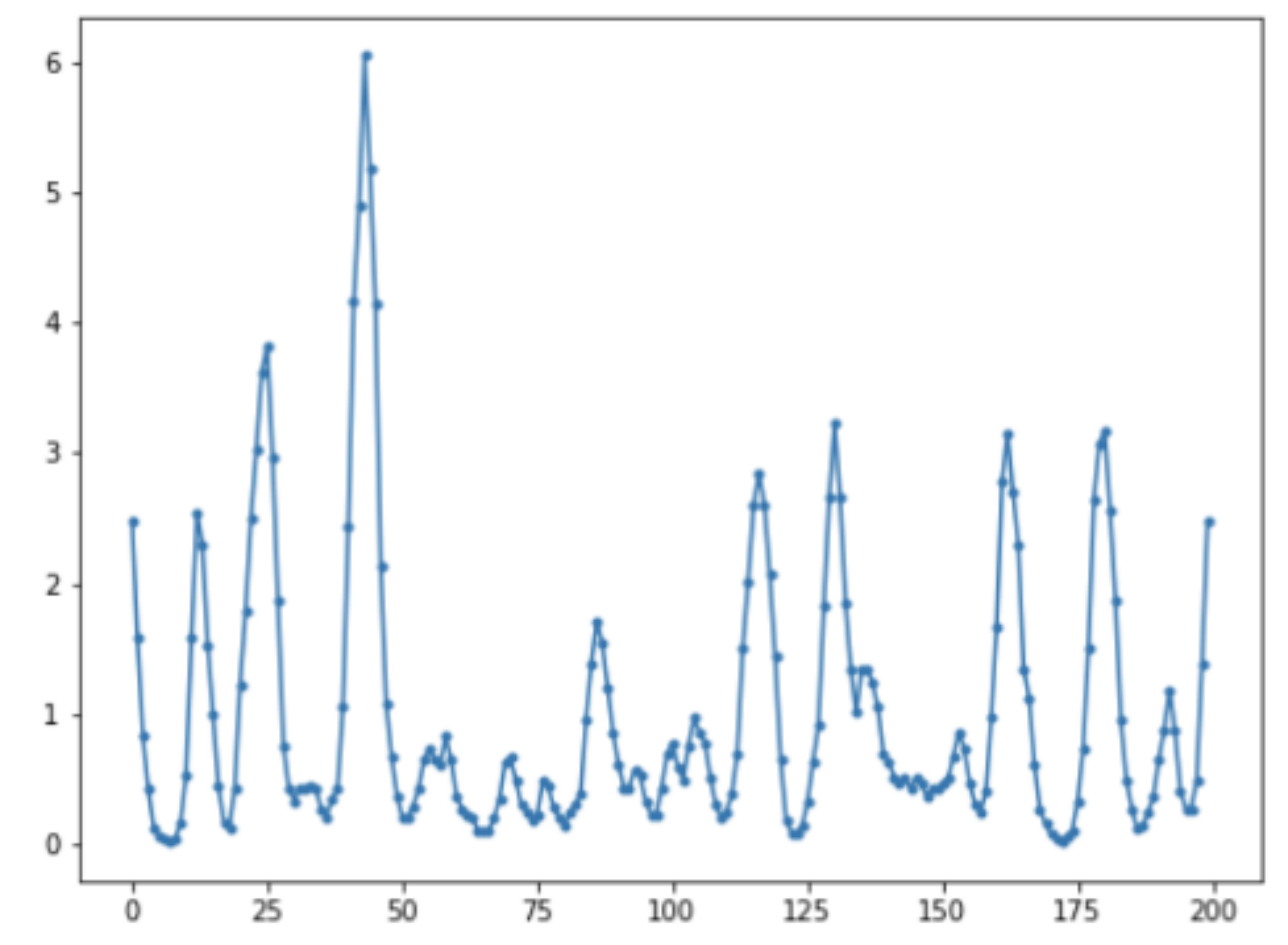
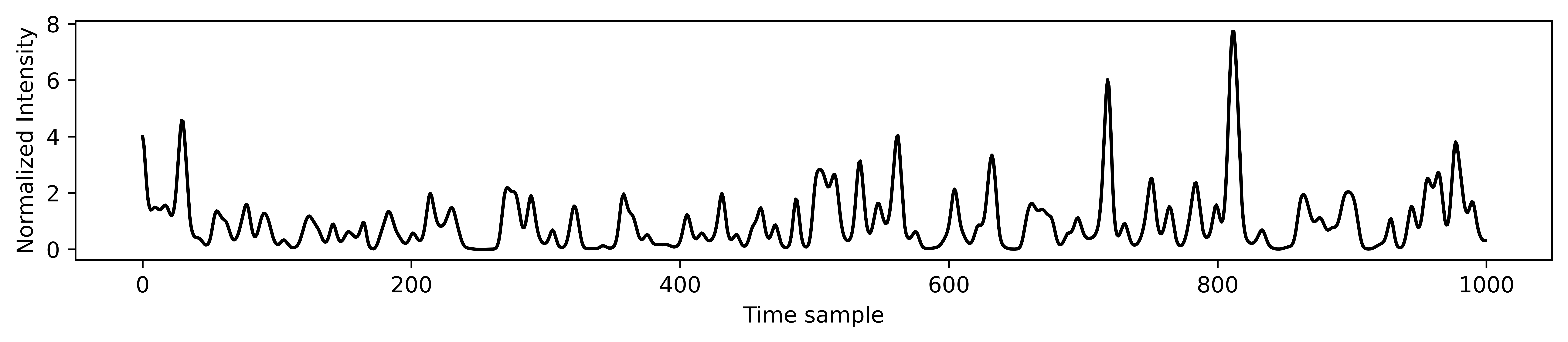


Fig. 1. Sample path of an ARTA process with exponential marginals and autocorrelations $\rho_1 = 0.9$ and $\rho_2 = 0.6$.

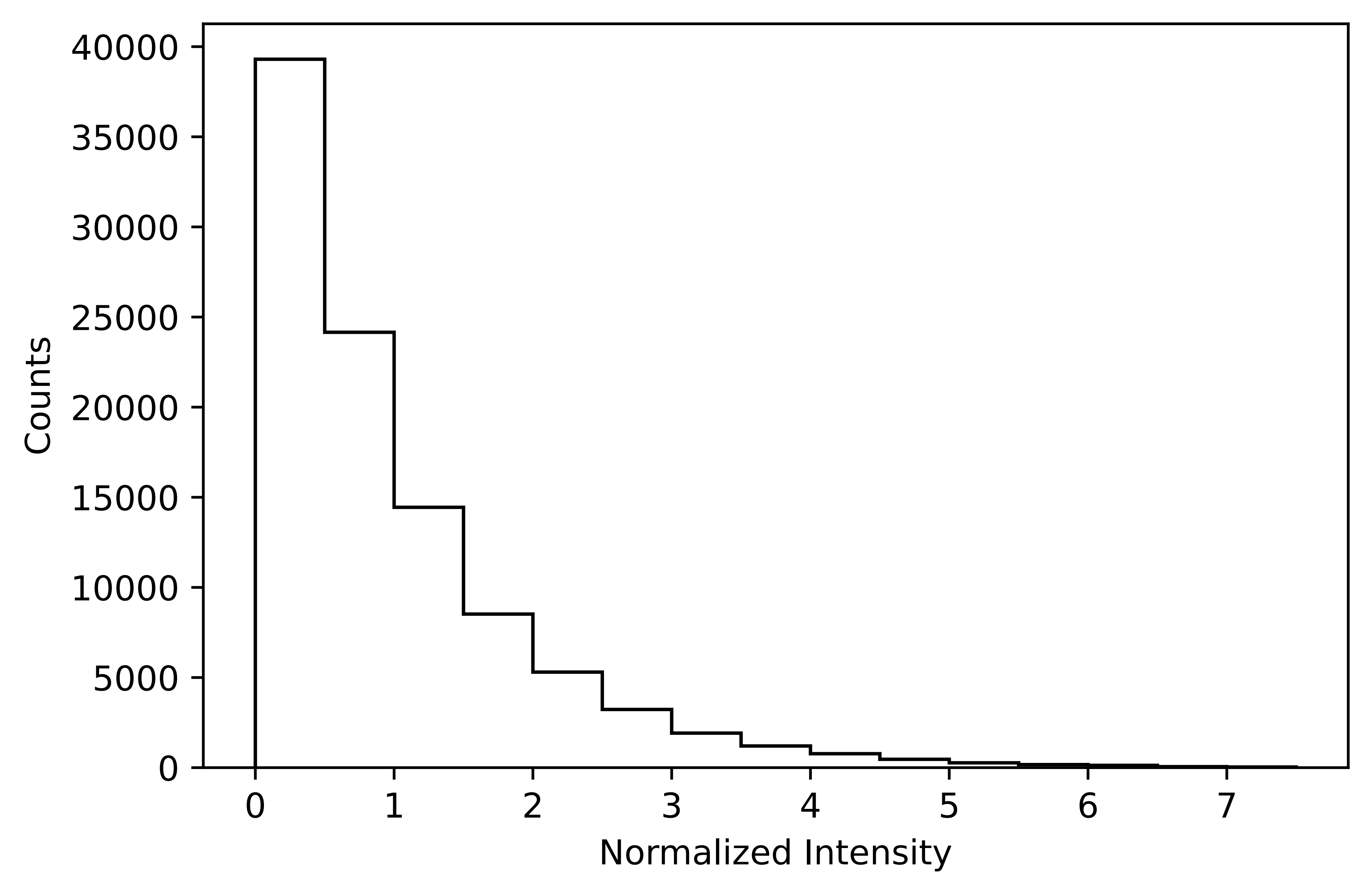


Cario & Nelson 1996 **Our implementation**

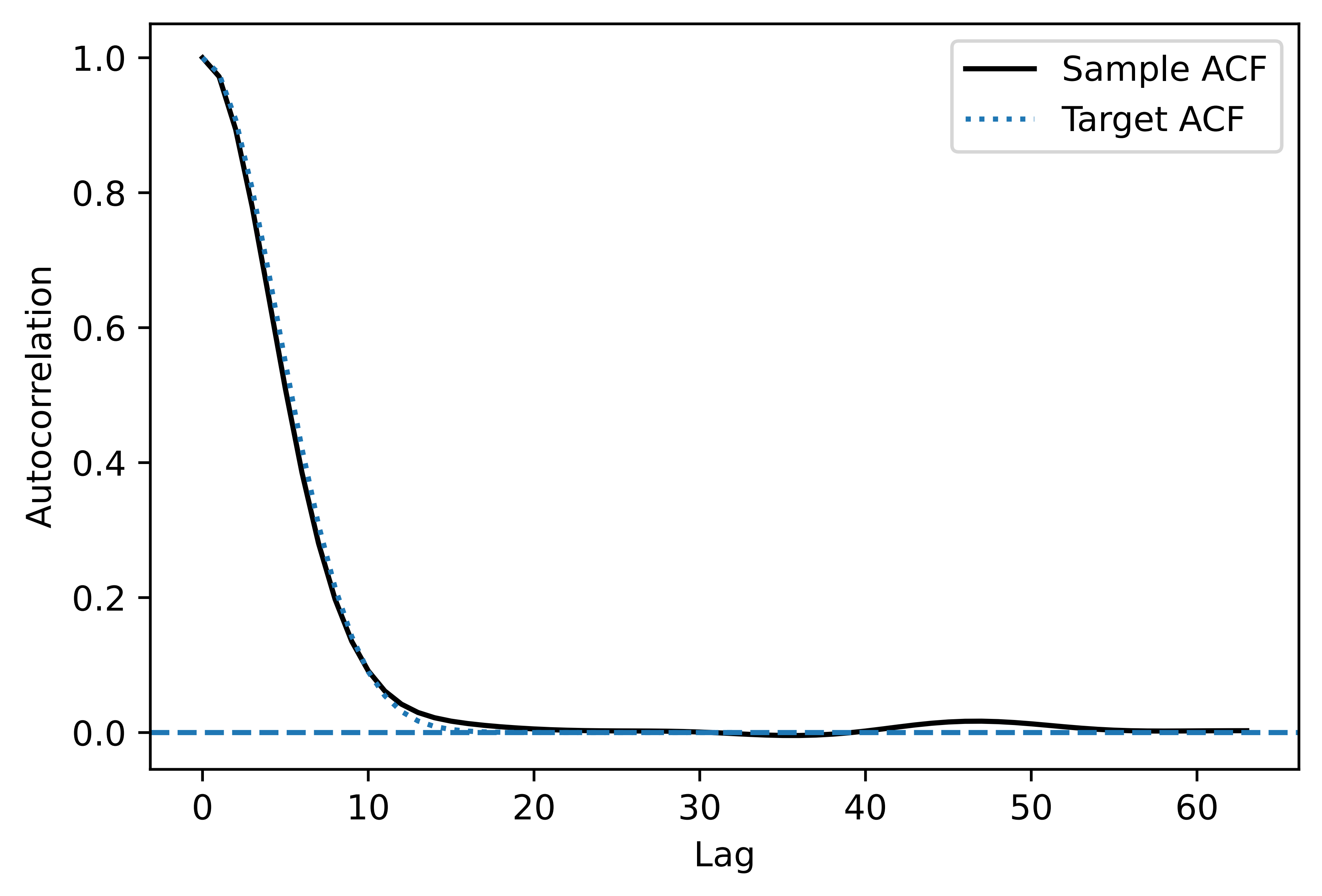
Large sample limit (n = 1e5) for $\Delta t_d = 30$ s



Intensity histogram

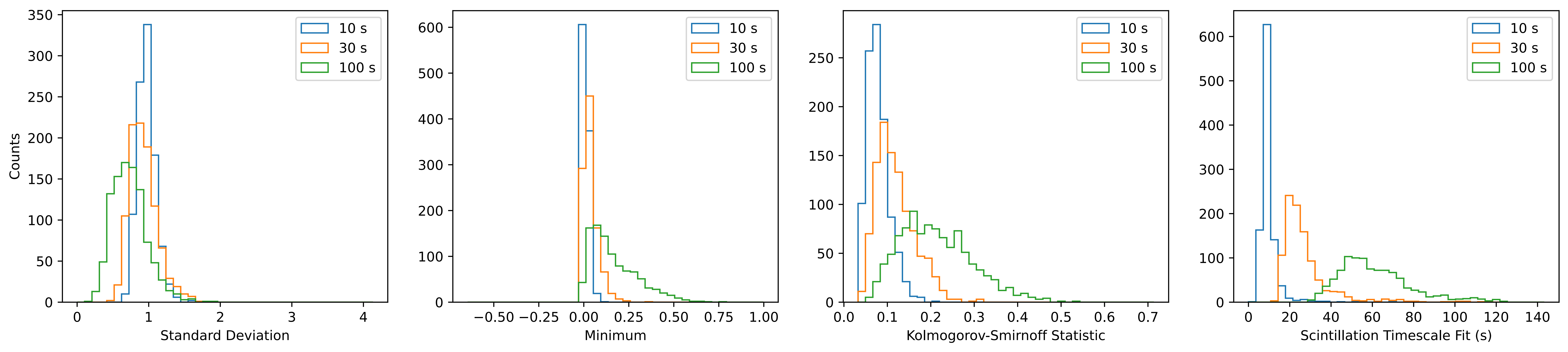


Autocorrelation



Statistics using low number of synthetic samples

10 min “observation”, 4.65 s resolution



Standard Deviation

Minimum

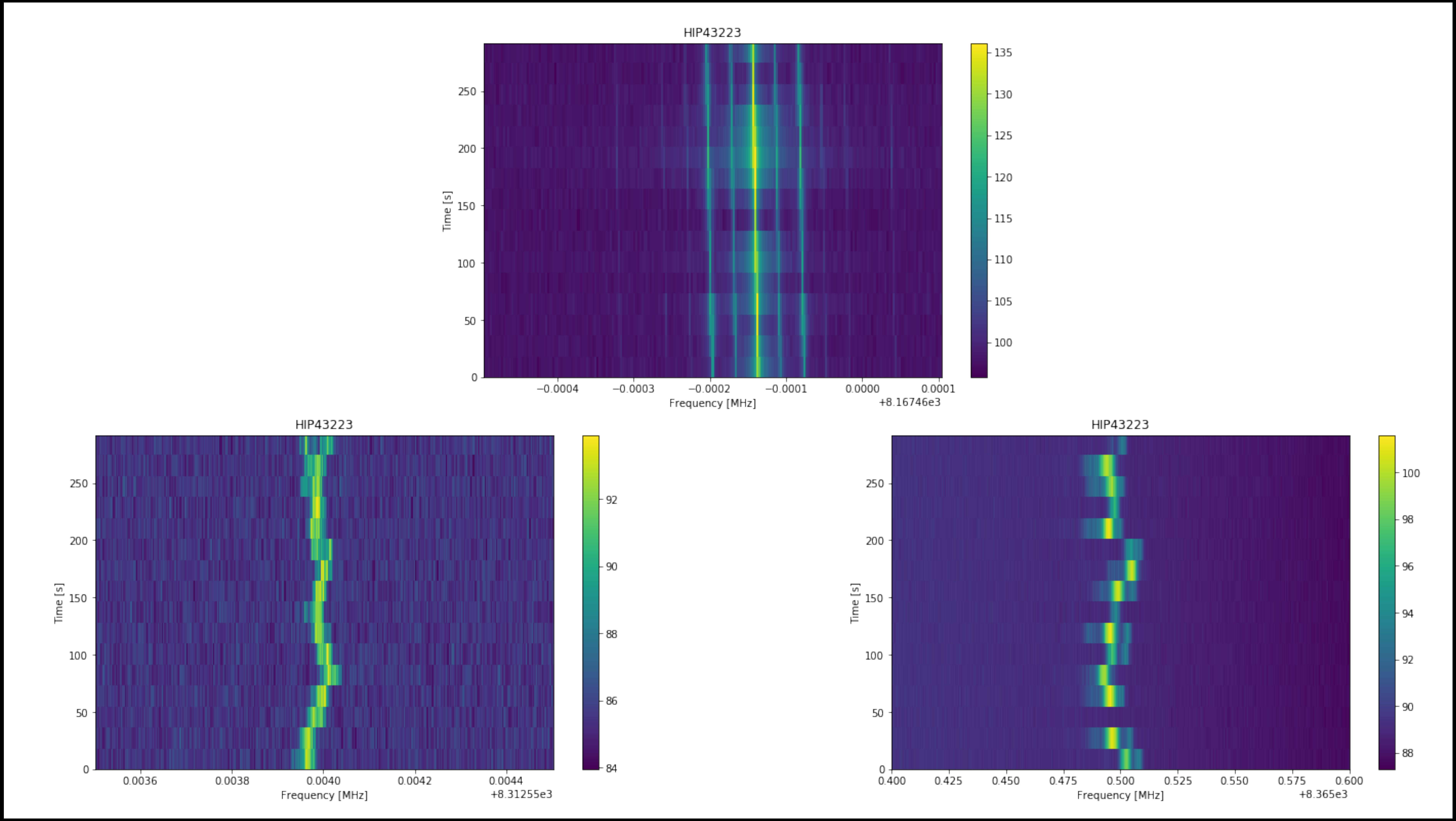
Kolmogorov-Smirnoff Statistic

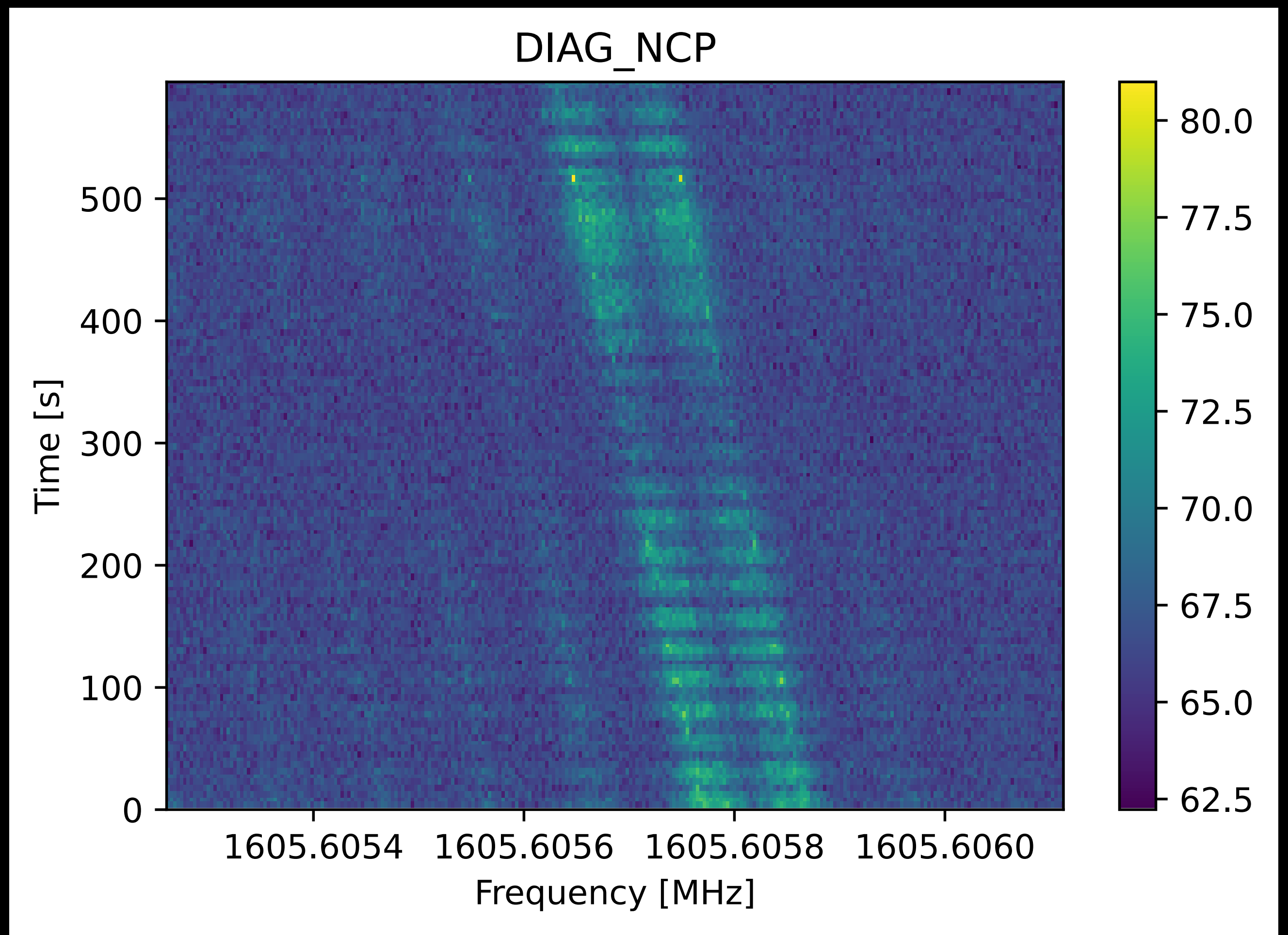
Scintillation Timescale Fit

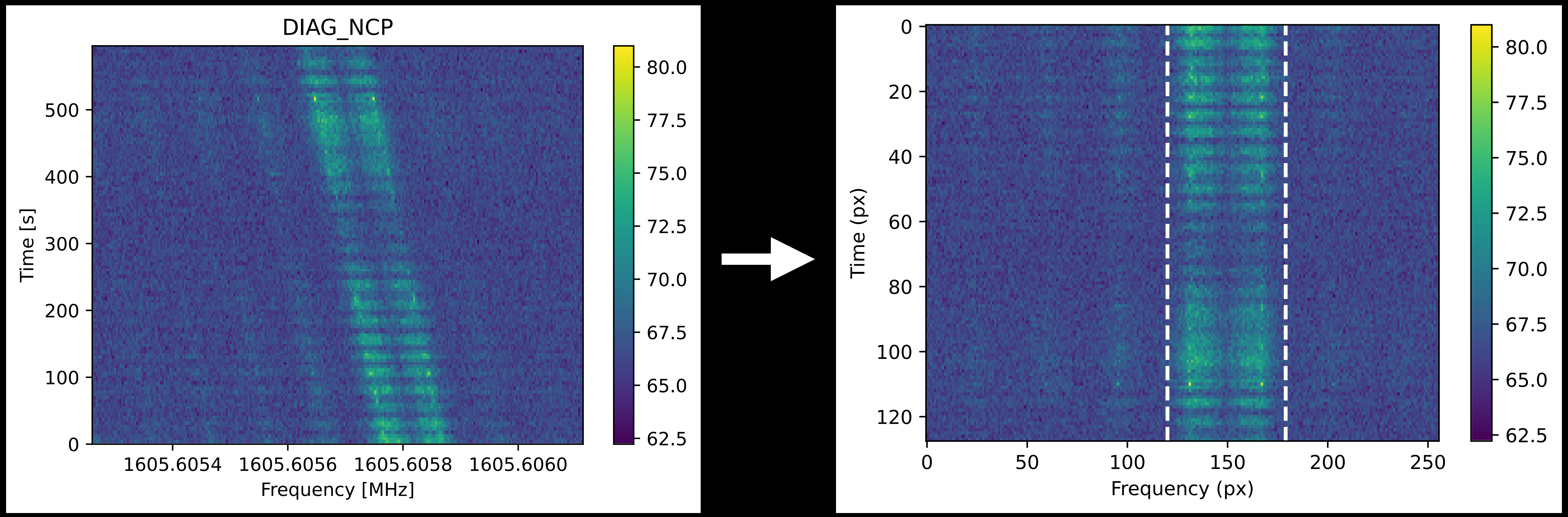
Can we detect scintillated narrowband technosignatures?

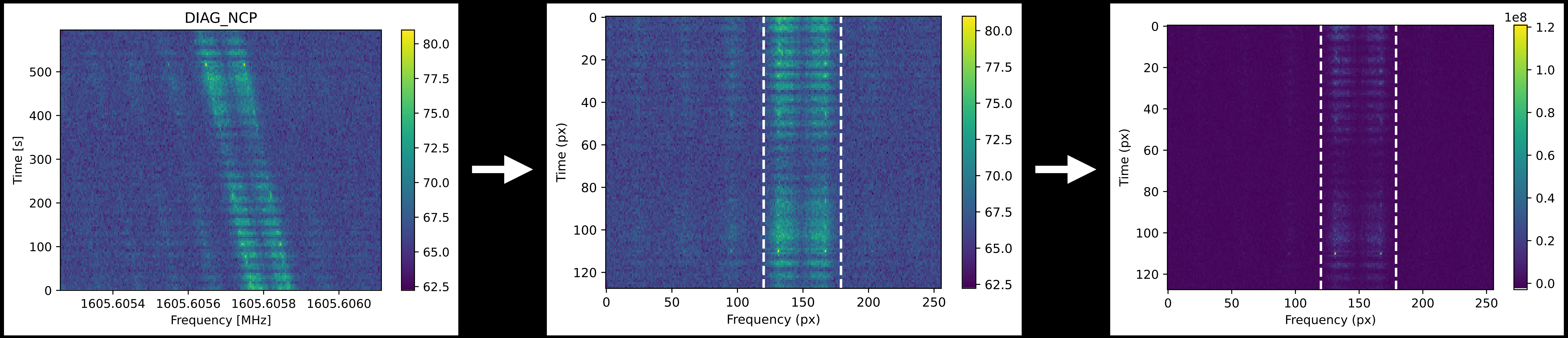
1. What scintillation timescales should we expect?
2. How can we probe asymptotic statistics?
3. **Can we differentiate scintillated signals from existing RFI?**
4. How can we design a survey to search for scintillated technosignatures?

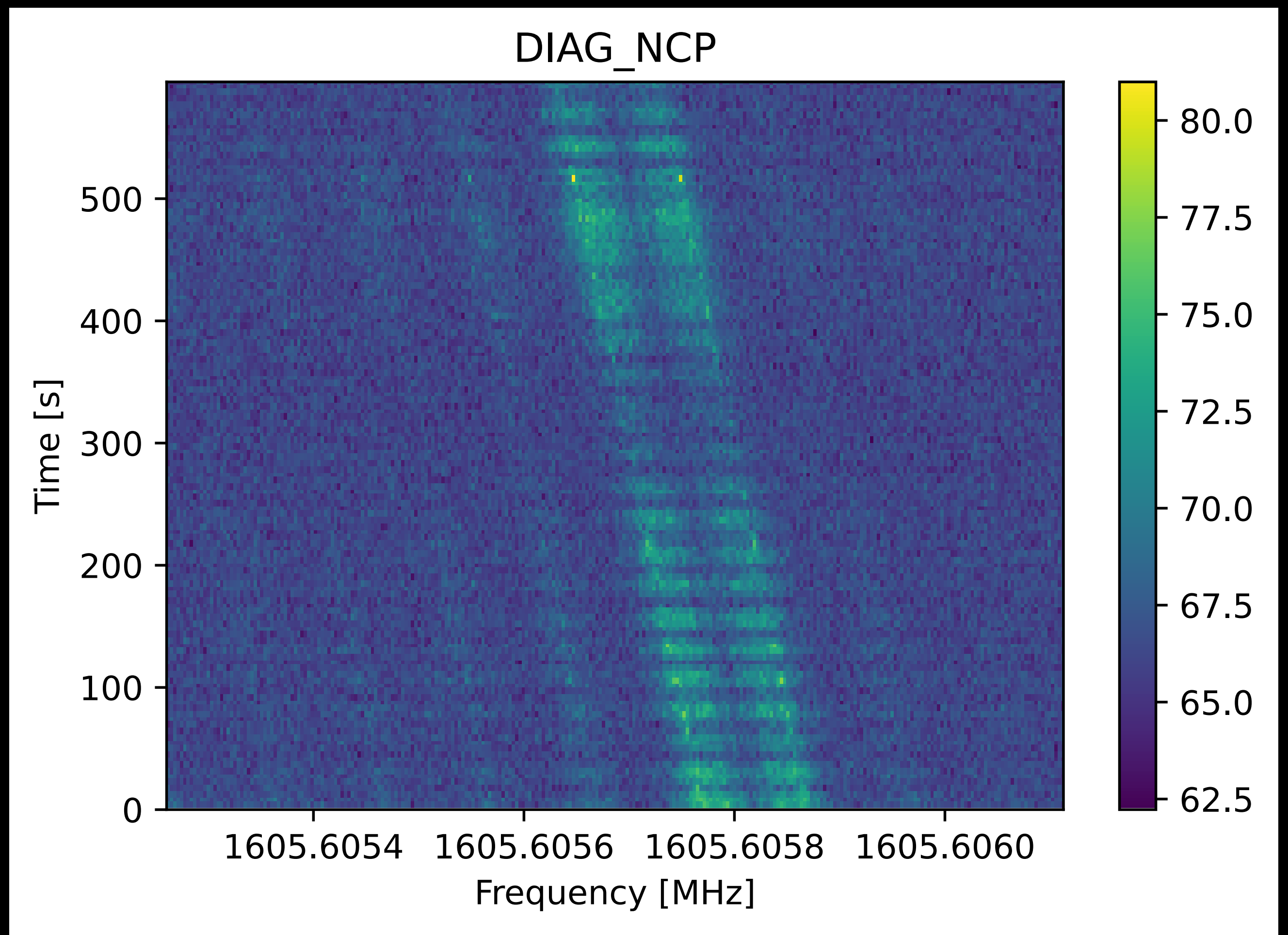
What does the RFI environment look like?

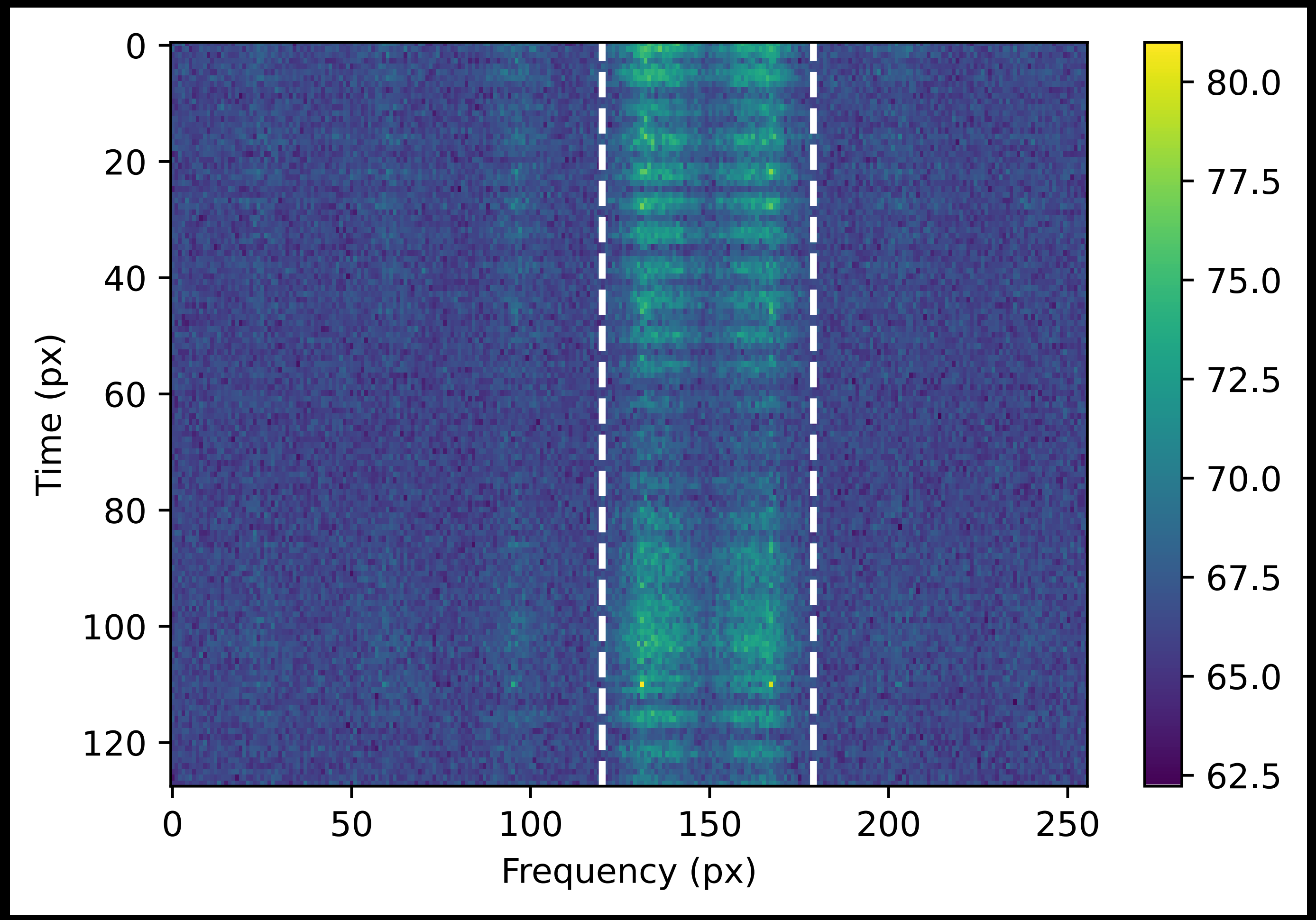


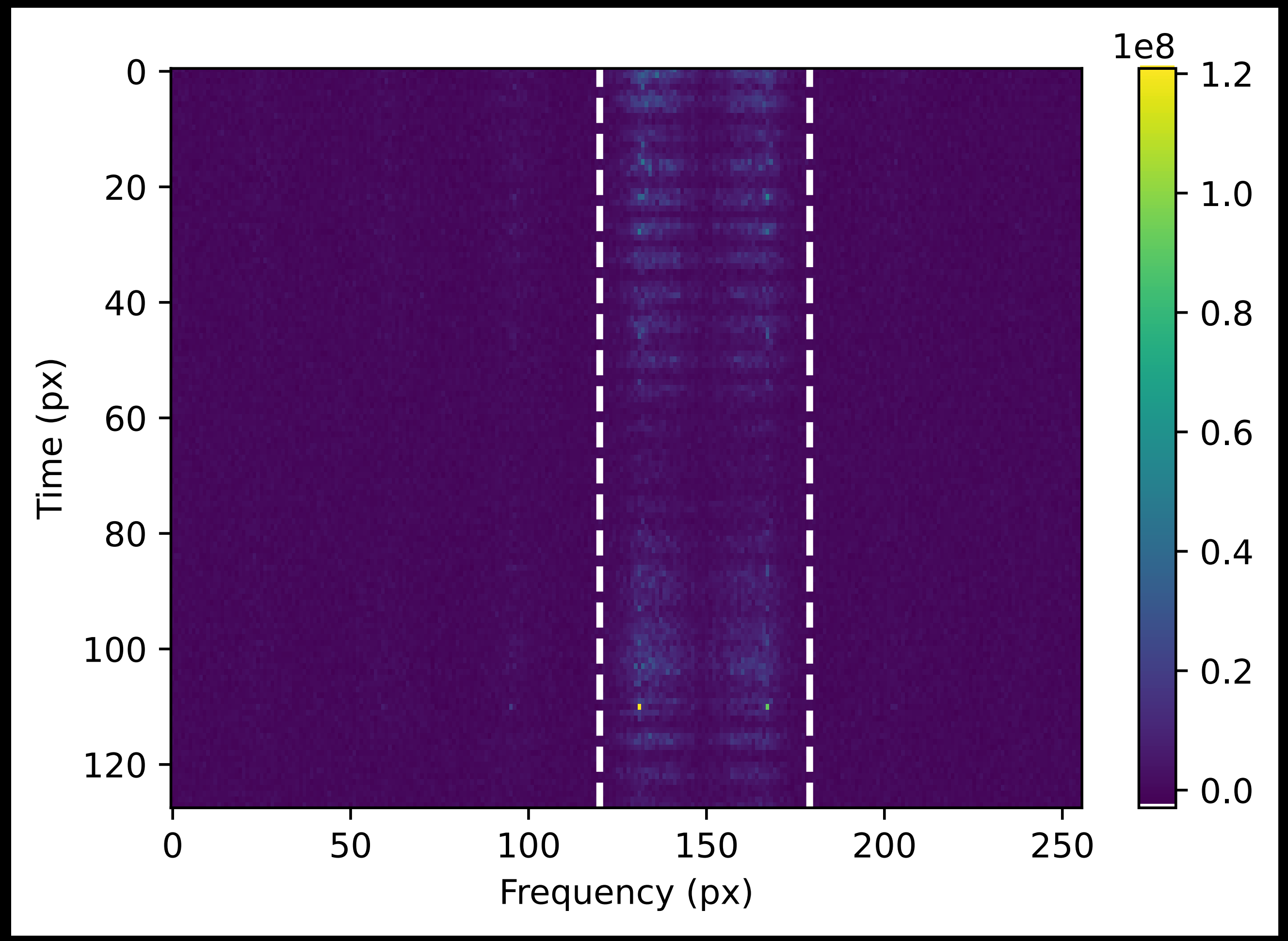

DIAG_NCP
Time [s]
500
400
300
200
100
0
1605.6054
1605.6056
1605.6058
1605.6060
Frequency [MHz]
80.0
77.5
75.0
72.5
70.0
67.5
65.0
62.5

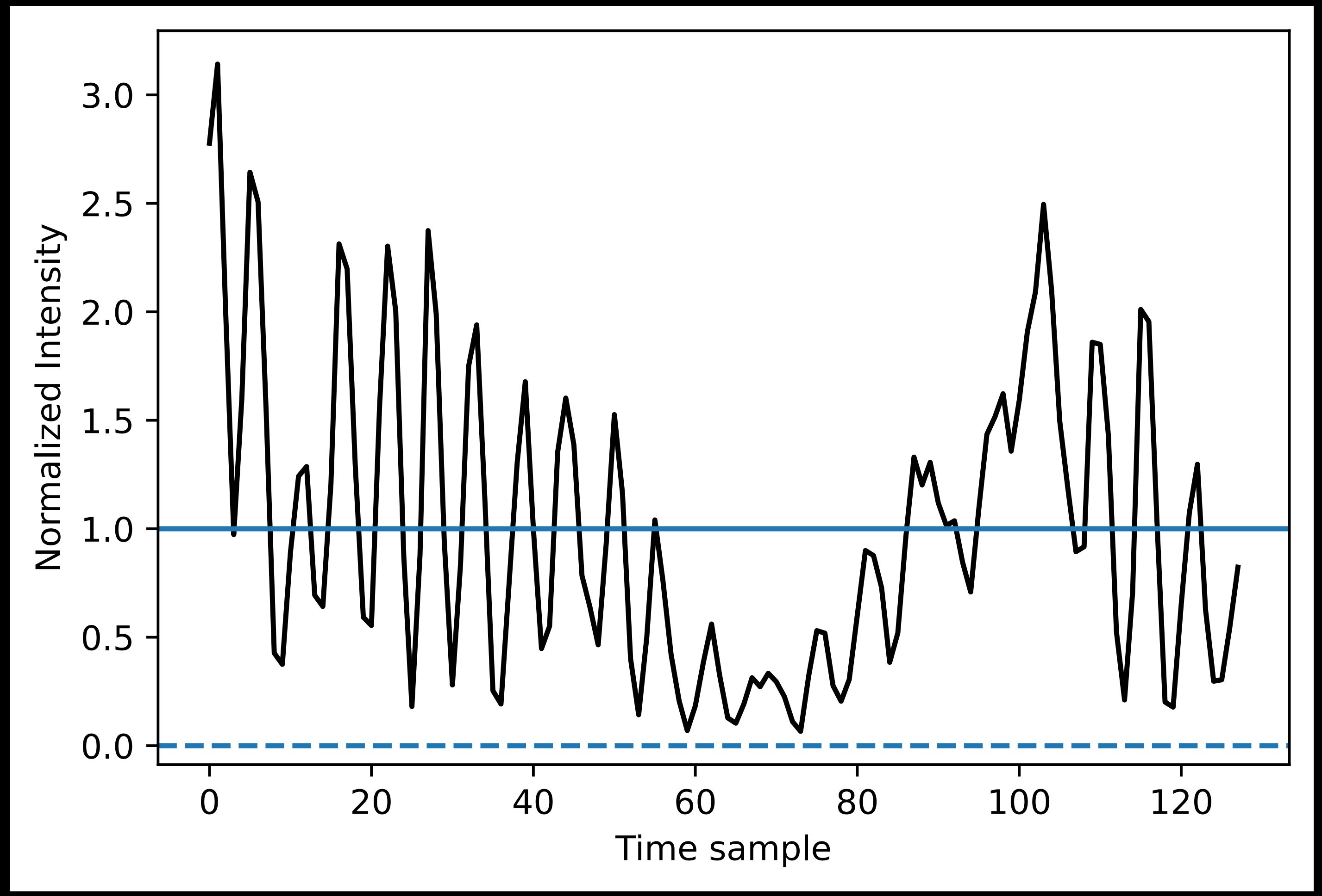

DIAG_NCP
Time [s]
Frequency [MHz]
1605.6054
1605.6056
1605.6058
1605.6060
Time (px)
Frequency (px)

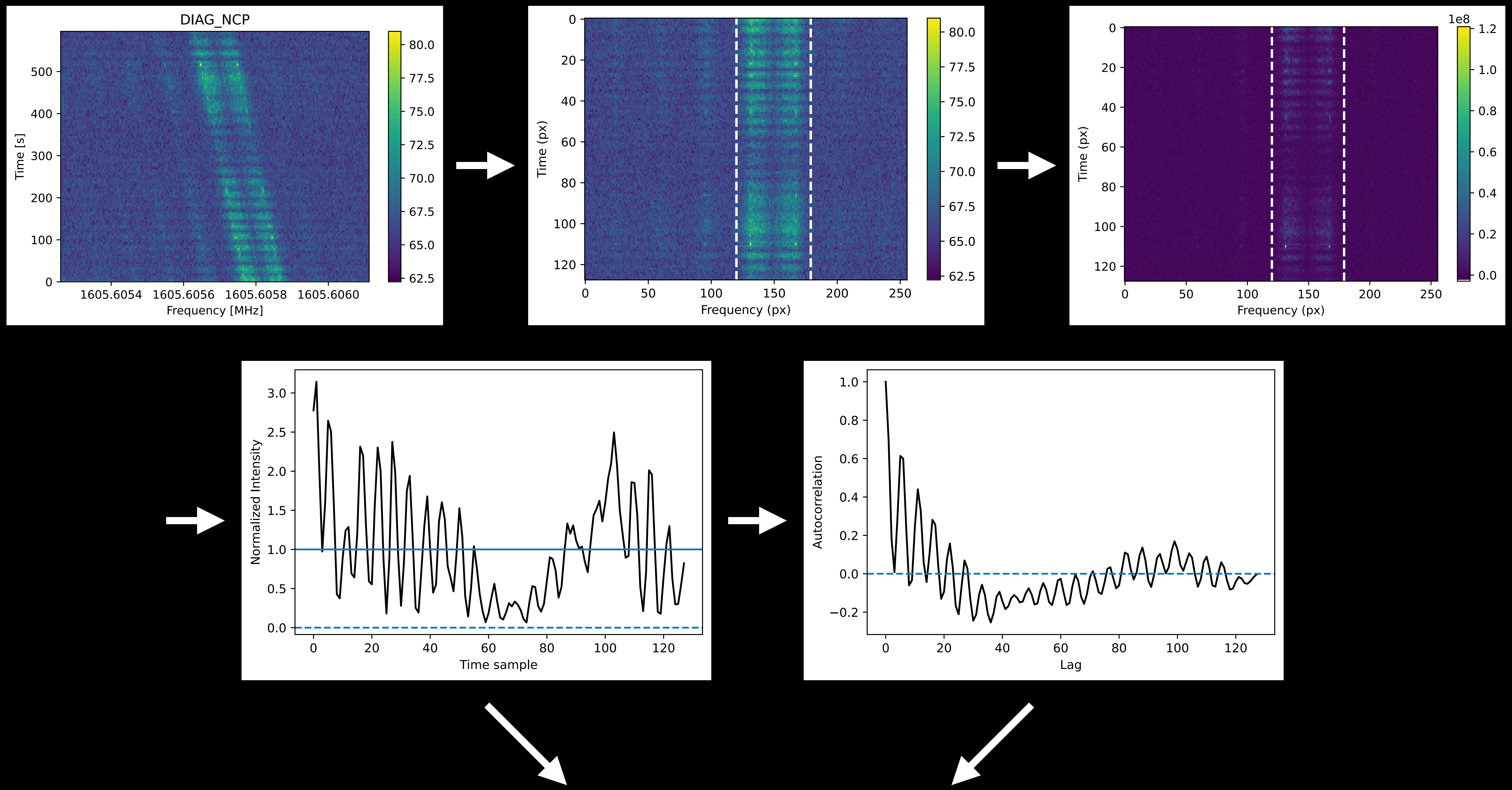

DIAG_NCP
Time [s]
Frequency [MHz]
1605.6054
1605.6056
1605.6058
1605.6060
Time (px)
Frequency (px)
1e8


DIAG_NCP
Time [s]
Frequency [MHz]
1605.6054
1605.6056
1605.6058
1605.6060


Time (px)
Frequency (px)


1e8
Time (px)
Frequency (px)

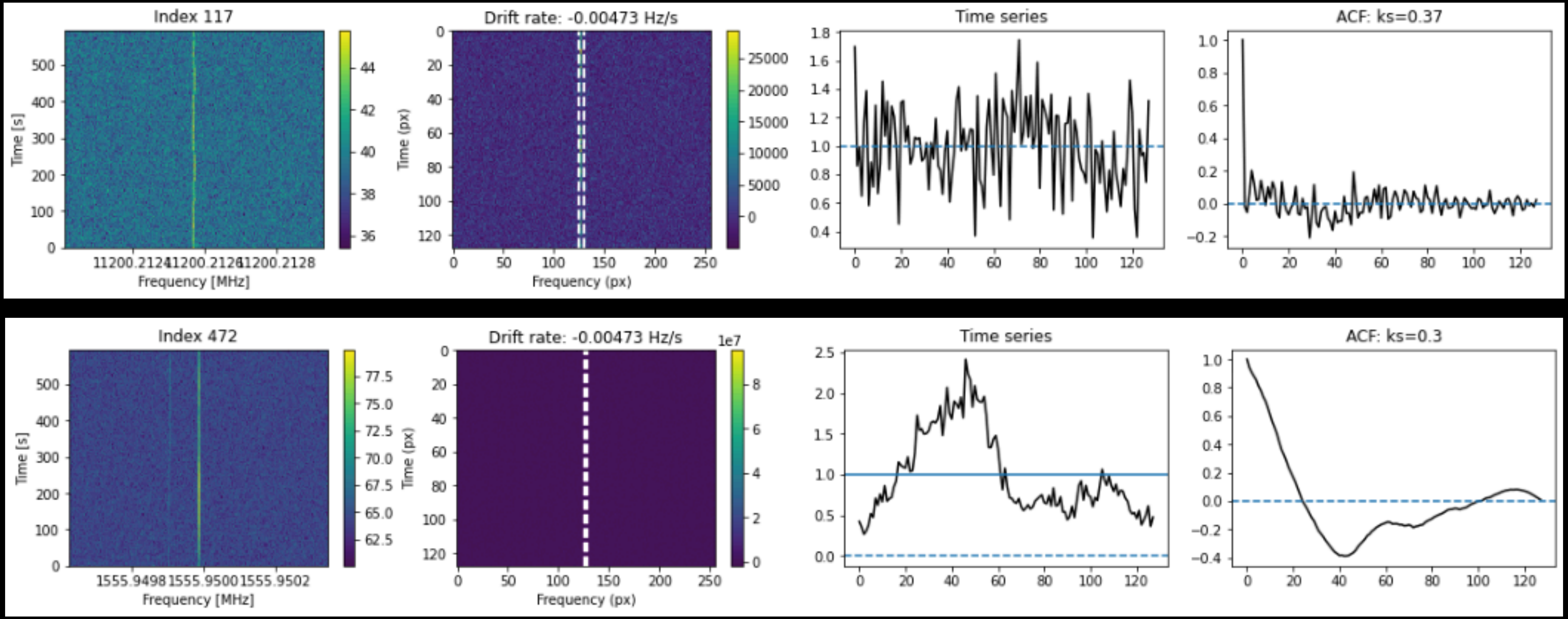

Normalized Intensity
Time sample



Diagnostic statistics

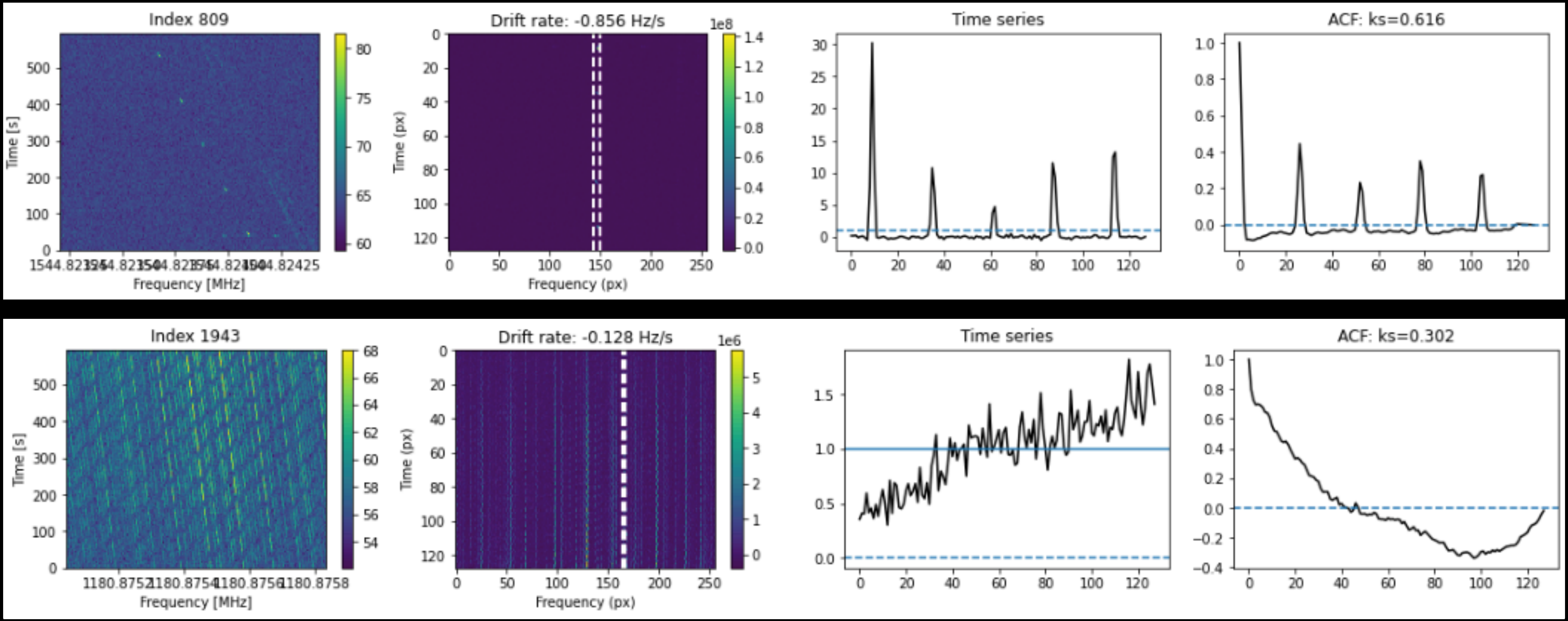
Some examples

Original | De-drifted | Intensity time series | Autocorrelation



Some more examples

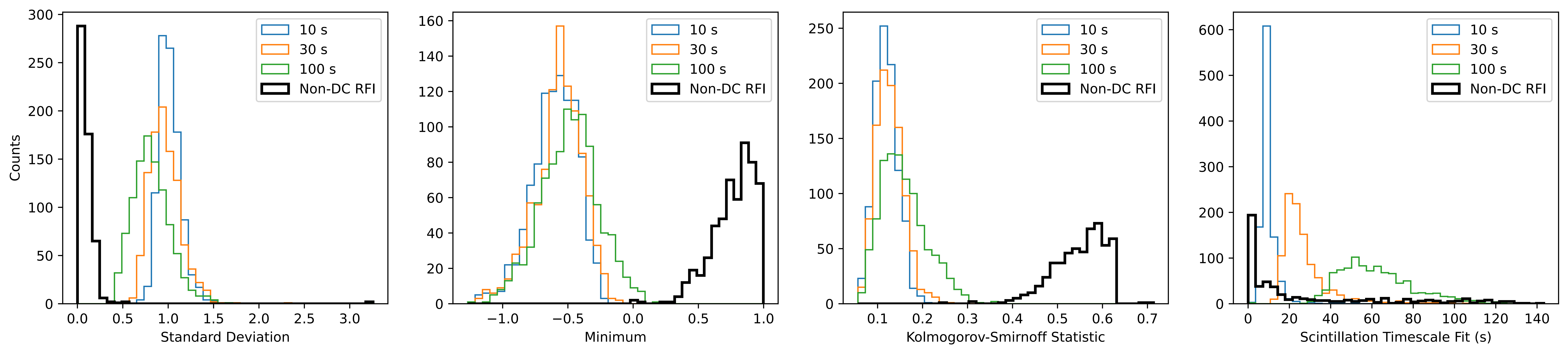
Original | De-drifted | Intensity time series | Autocorrelation



GBT RFI vs. synthetic scintillated signals

C band

S/N > 25



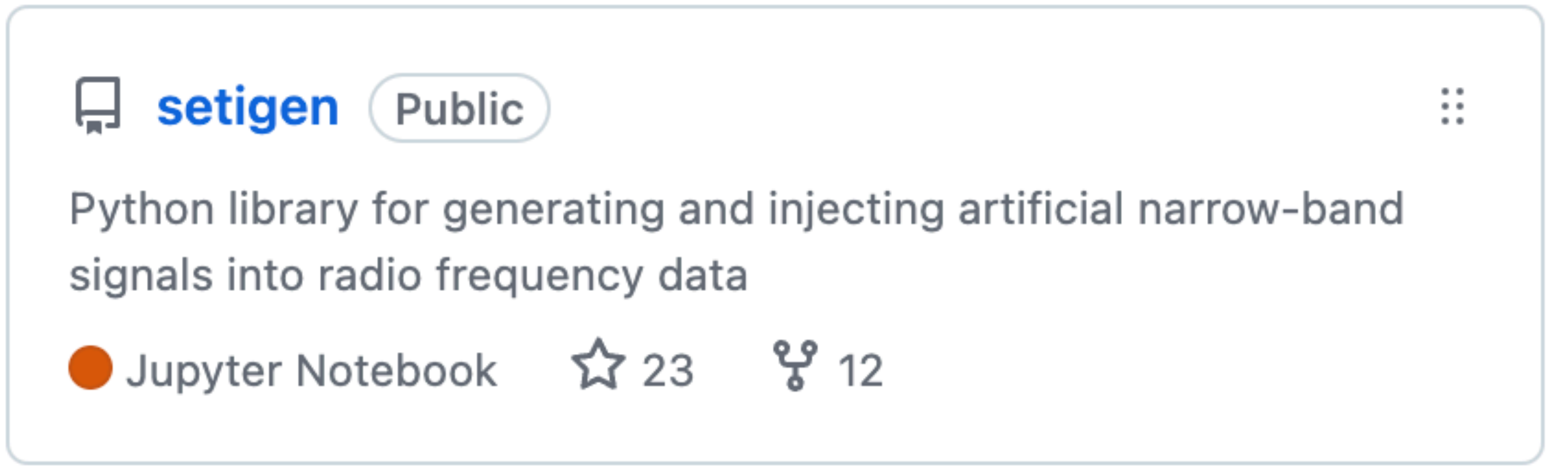
Standard Deviation

Minimum

Kolmogorov-Smirnoff Statistic

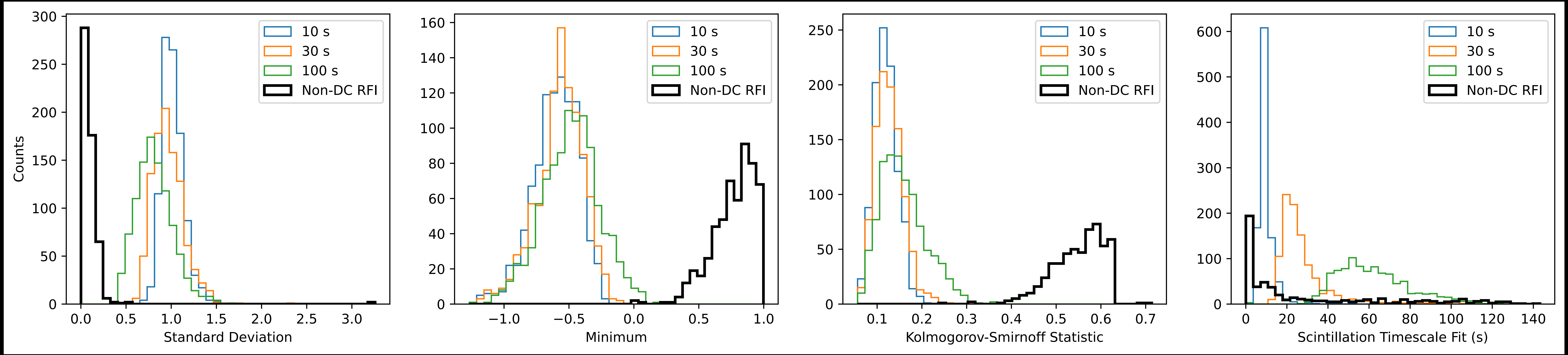
Scintillation Timescale Fit

GBT RFI vs. synthetic scintillated signals



C band

S/N > 25



Standard Deviation

Minimum

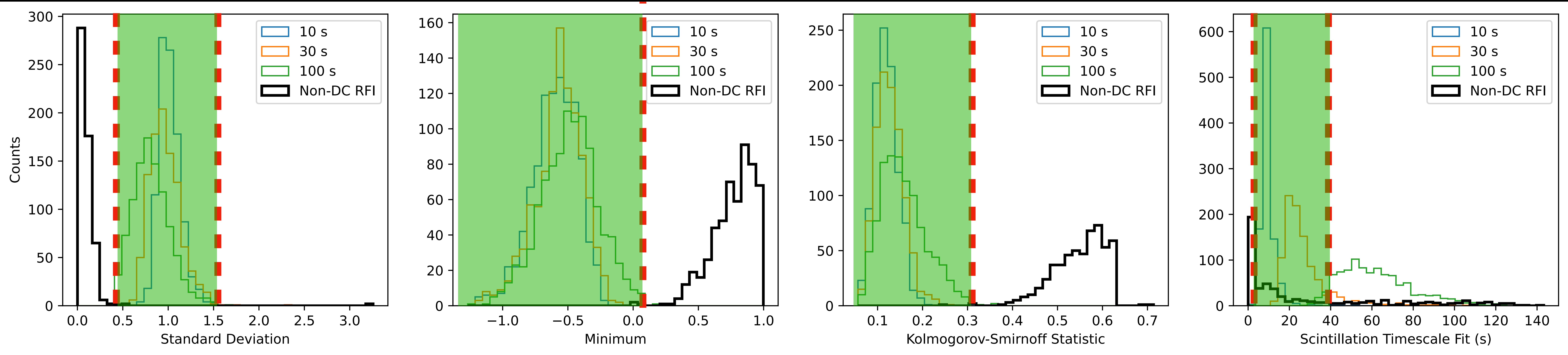
Kolmogorov-Smirnoff Statistic

Scintillation Timescale Fit

GBT RFI vs. synthetic scintillated signals

C band

S/N > 25



Standard Deviation

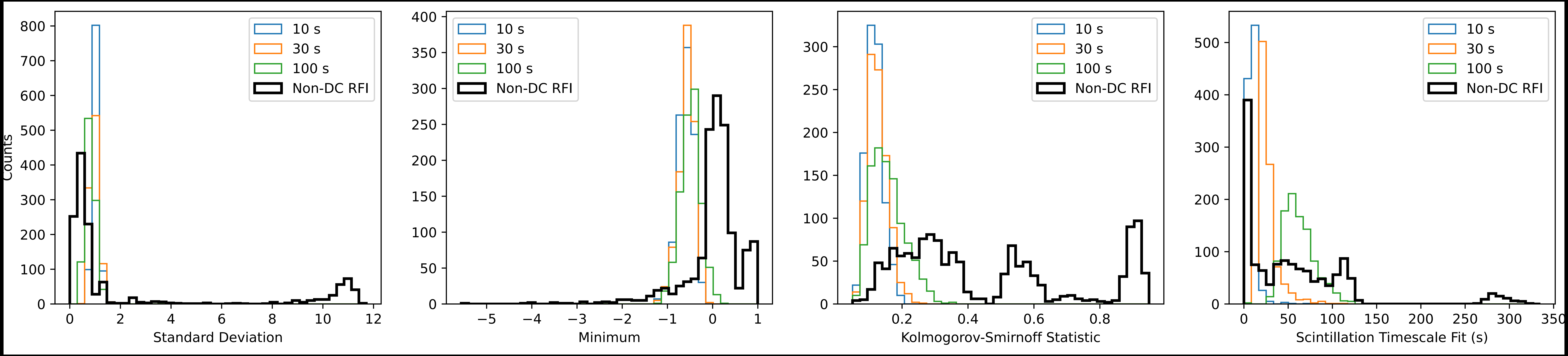
Minimum

Kolmogorov-Smirnoff Statistic

Scintillation Timescale Fit

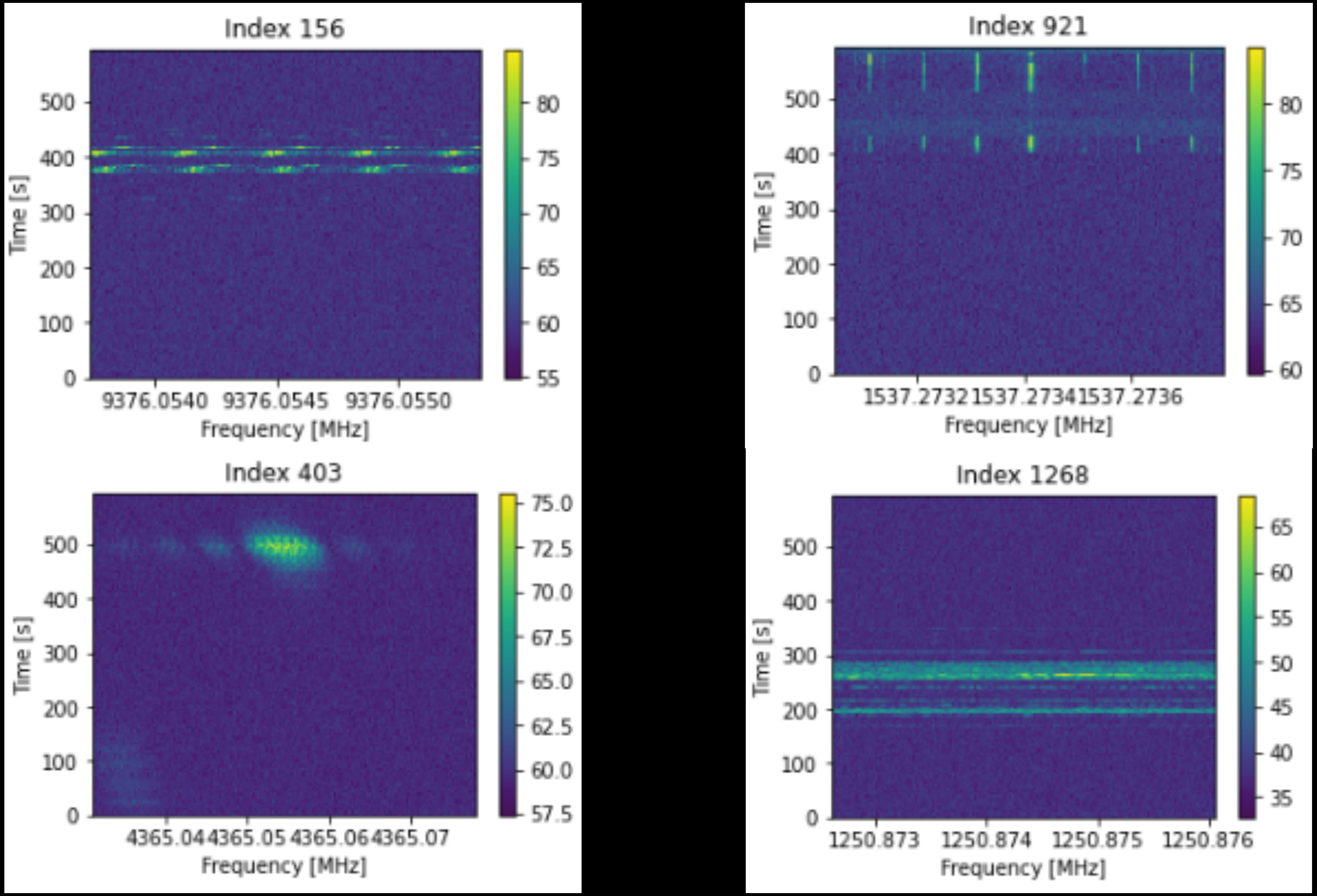
L band

S/N = 25



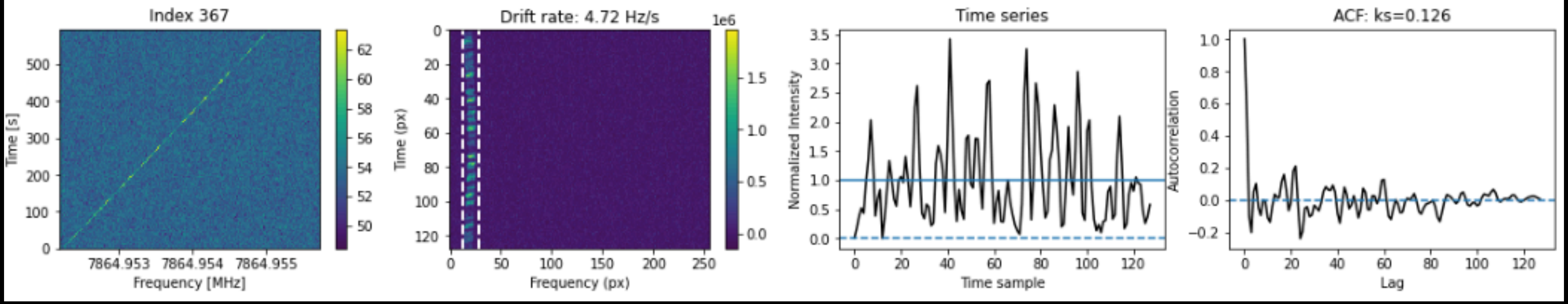
Counts
Standard Deviation
Minimum
Kolmogorov-Smirnoff Statistic
Scintillation Timescale Fit (s)
10 s
30 s
100 s
Non-DC RFI

High standard deviation (RMS) signals are pulsed and/or broadband

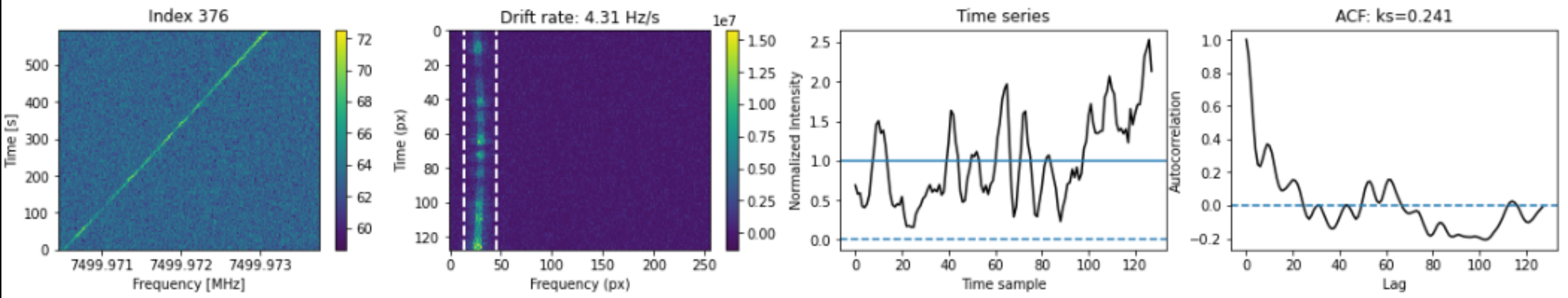


What signals pass the threshold?

- At C-band, S/N > 25, 3 out of 1102



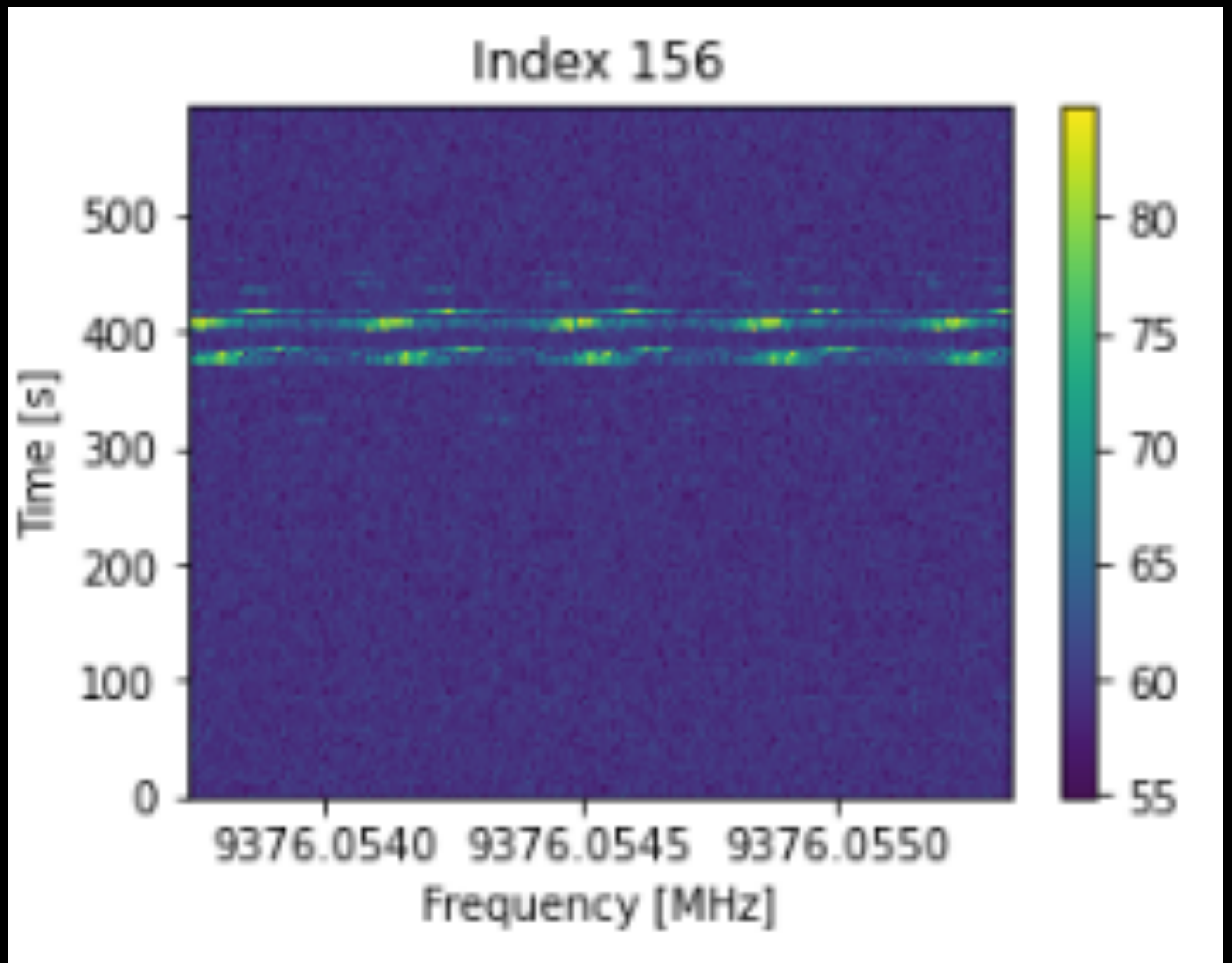
Timescale fit ~ 2 s

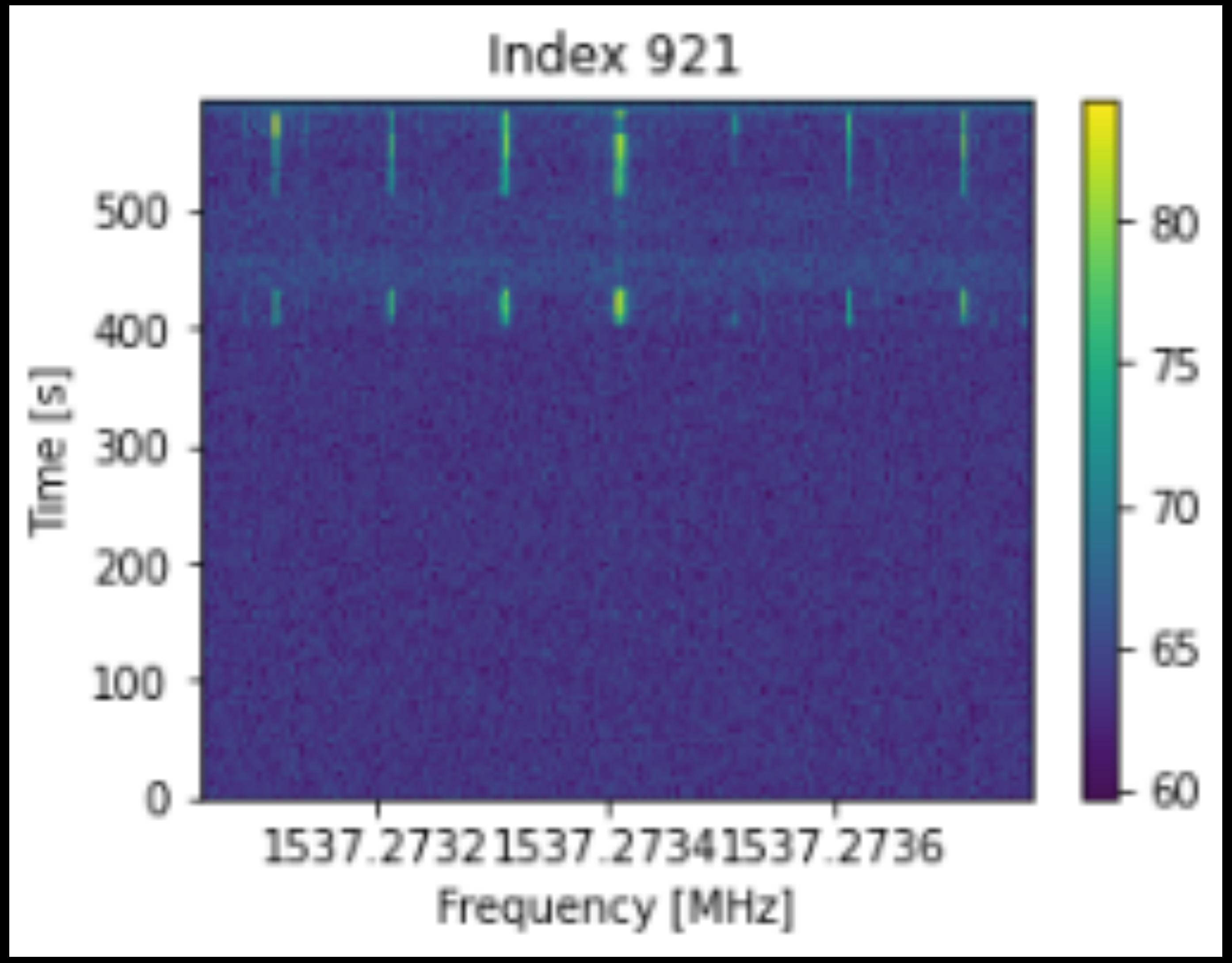


Timescale fit ~ 60 s

Limitations from RFI analysis?

- L and S bands in particular are very noisy
- Non-narrowband signals detected just because they are above the SNR threshold
- Difficult to apply a one-size-fits-all bounding box method
- Perhaps ML can help!





Can we detect scintillated narrowband technosignatures?

1. What scintillation timescales should we expect?
2. How can we probe asymptotic statistics?
3. Can we differentiate scintillated signals from existing RFI?
4. **How can we design a survey to search for scintillated technosignatures?**

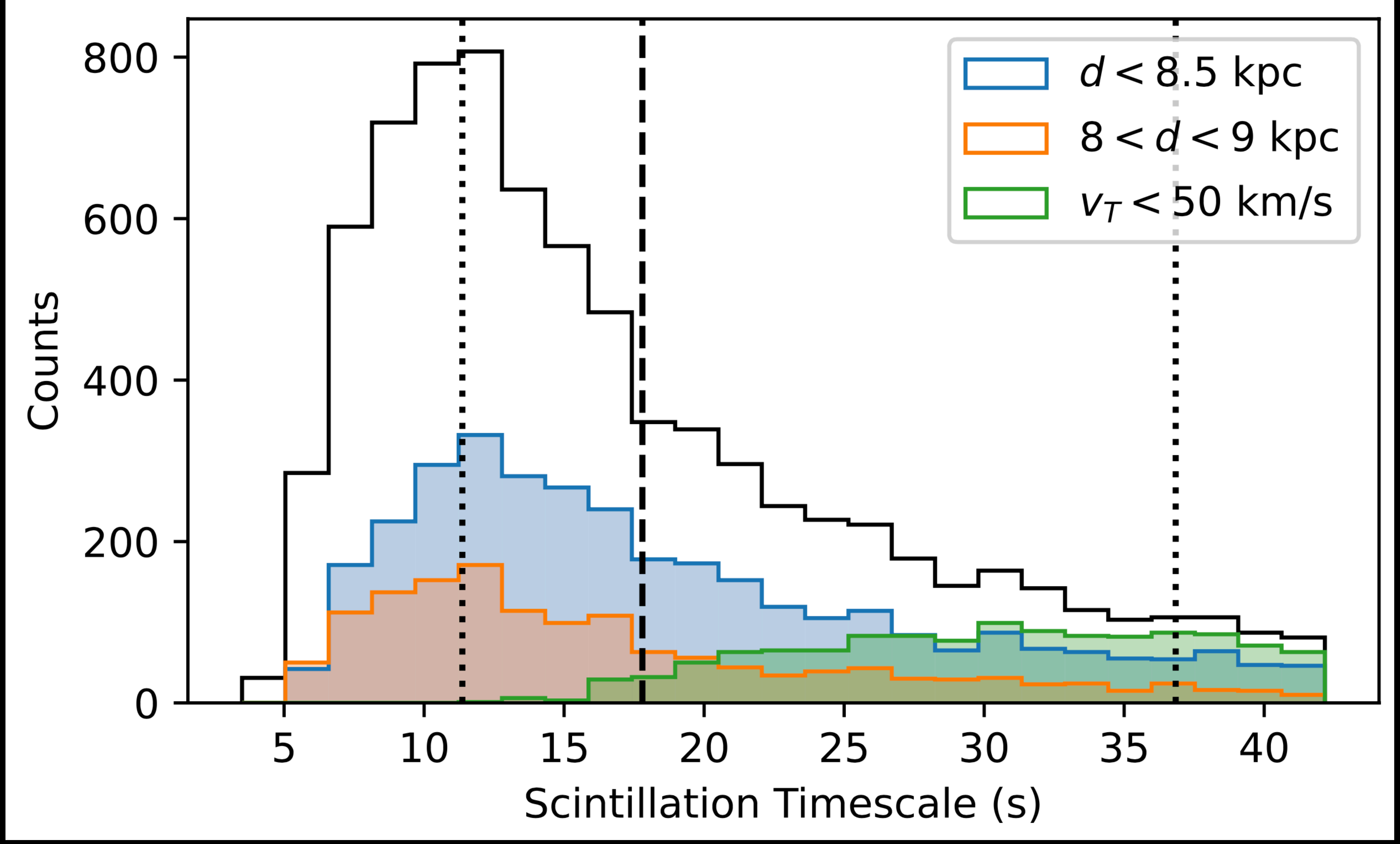
Planning Galactic Center observations — Monte Carlo sims with NE2001

- Estimate scintillation timescales with NE2001 (Cordes & Lazio 2002) and scale with different sets of parameters
  - Galactic coordinates
  - Distance
  - Frequency
  - Transverse velocities

- Monte Carlo sample to estimate most probable scintillation timescales

Inter-quartile range

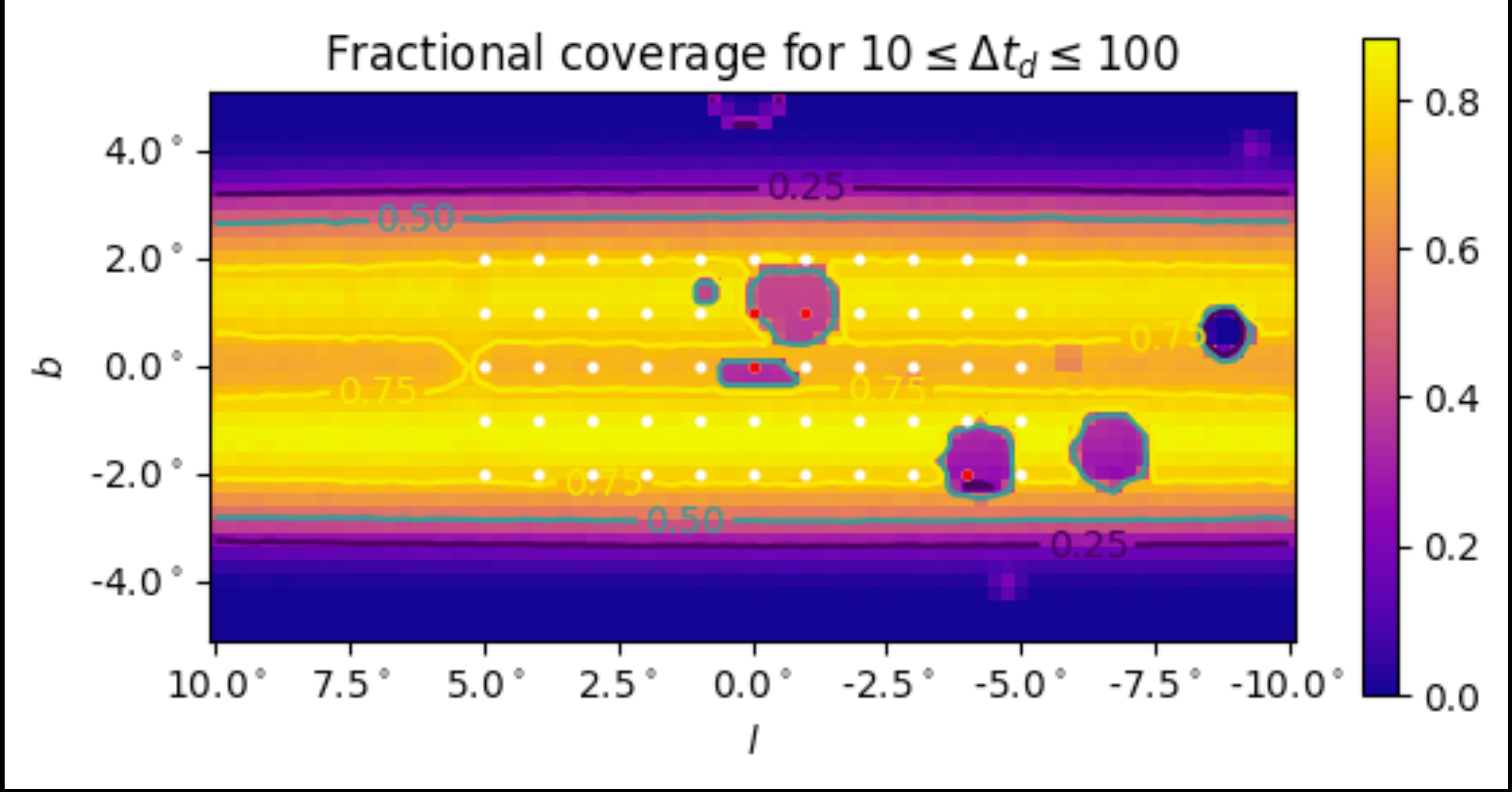
Median



(l, b) = (5, 0) at C-band

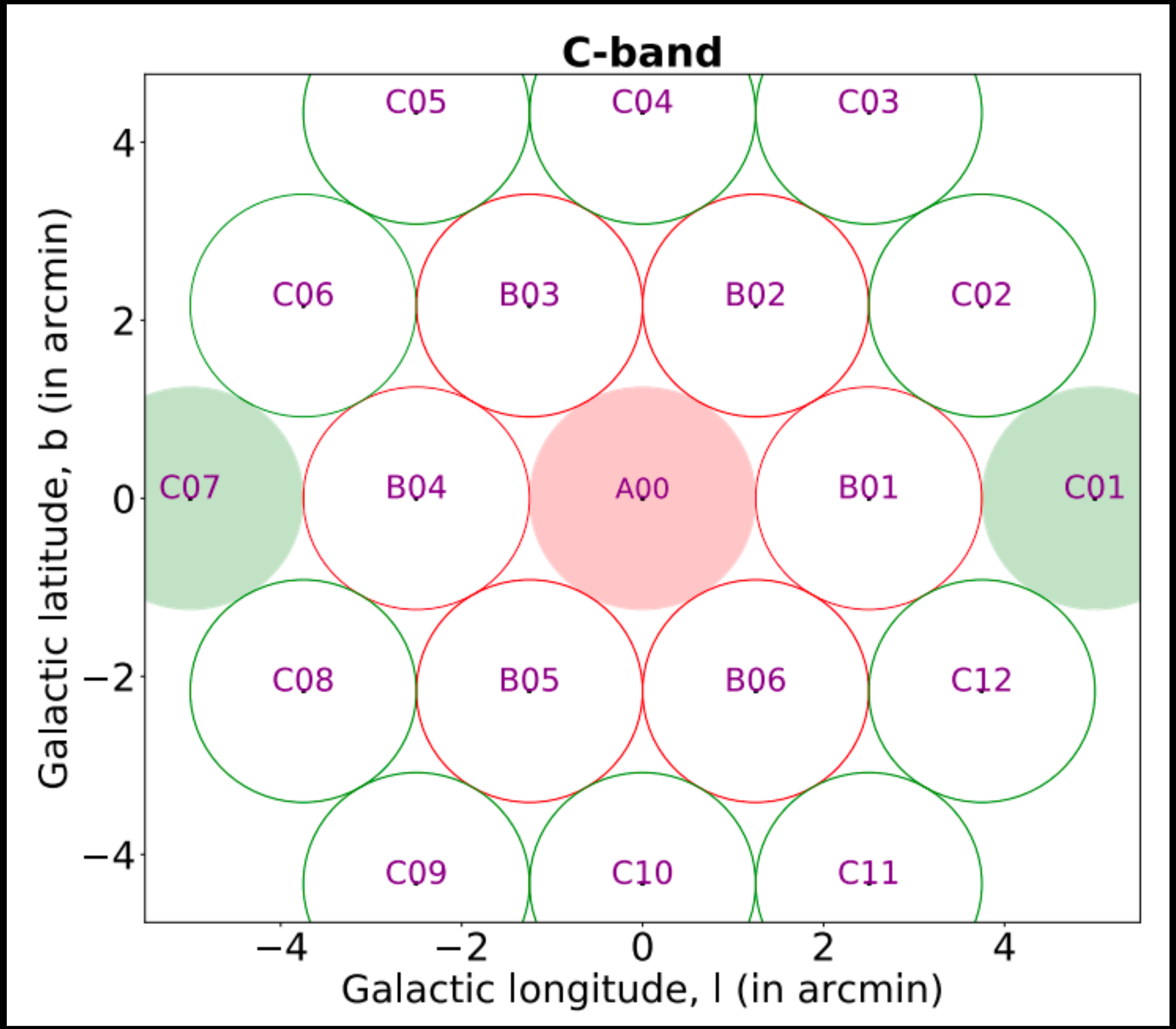
Current observing plan for scintillation survey of the Galactic center

- Galactic plane survey: 54 pointings, with |l| < 5 deg, |b| < 2 deg



Current observing plan for scintillation survey of the Galactic center

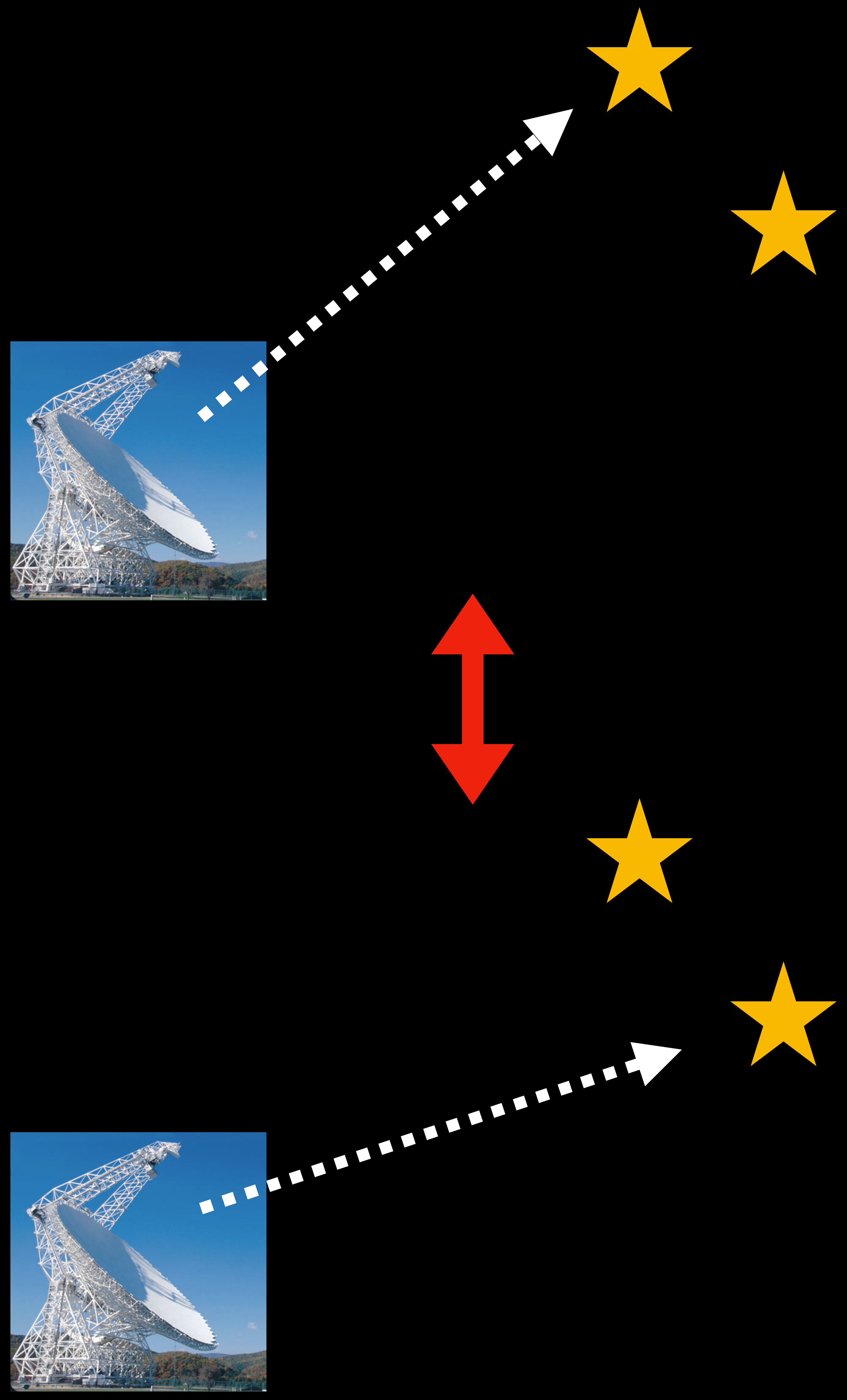
- Galactic center survey: 19 pointings (following Gajjar et al. 2021)



Gajjar et al. 2021

Current observing plan for scintillation survey of the Galactic center

- ABAB cadences
- 10 minutes per observation, so each pointing gets 20 minutes total
- 2.5 s, 2.8 Hz resolution
- Start each observing session with single pointing of North Galactic Pole as probe of local RFI environment



NRAO

Next Steps

- Currently, we have data for 16 out of 27 cadences of the Galactic plane survey, about 12 hours of data
    - 11 GP cadences and 9 GC cadences remain
- Filter collected data using established ON-OFF search methods and perform scintillation analysis
- Ultimate goal is to comment on the prevalance of scintillated technosignatures, as well as the prevalence of RFI that might pass the scintillation thresholds

Summary

- We developed a scintillation analysis framework, with accompanying codebase
- We can set statistical filter thresholds based on synthetic signals and the local RFI environment
- We've planned a survey to search for scintillated signals towards the Galactic center / plane, which is well under way

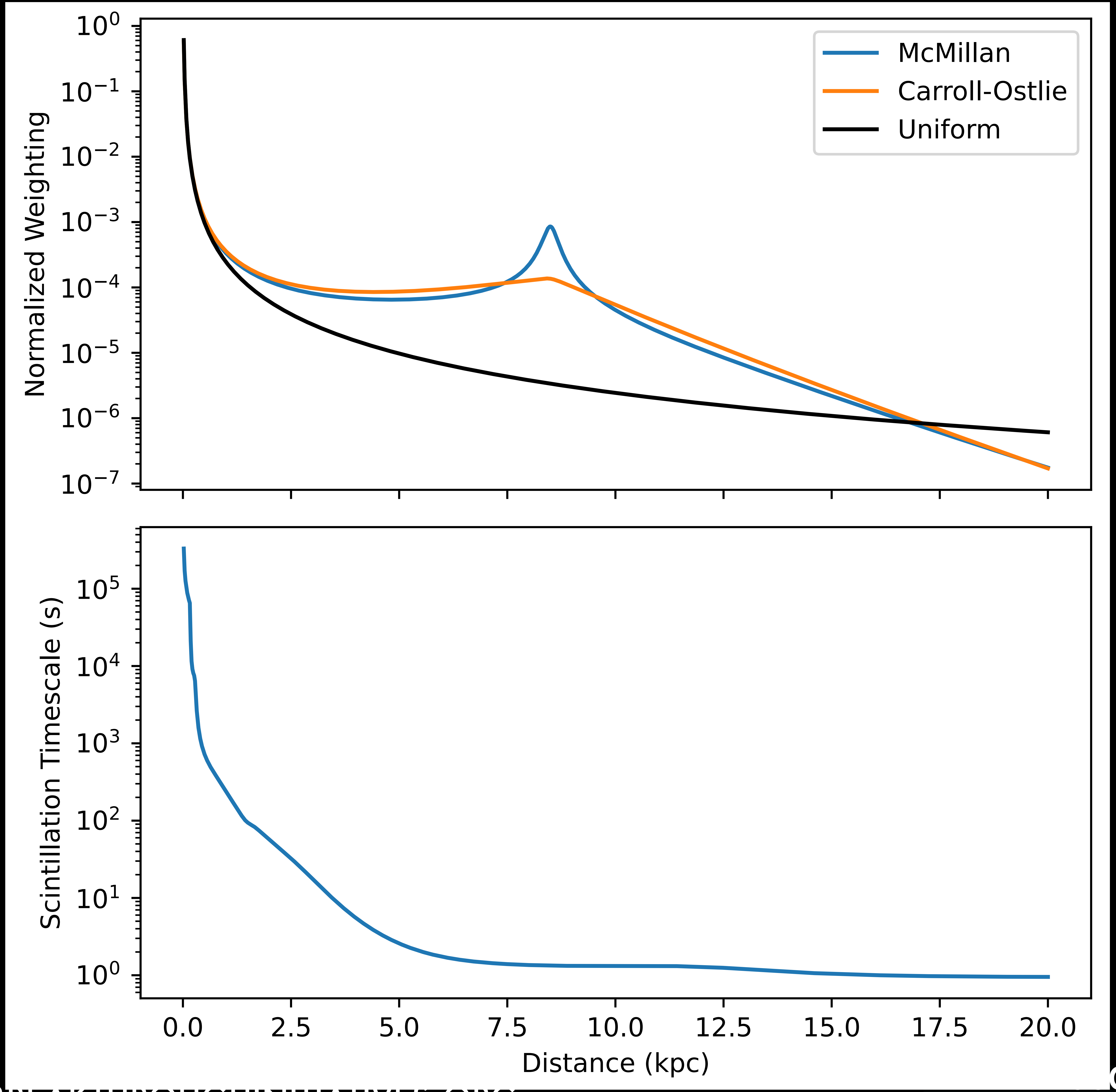
Extra Slides

Density-based sampling

Modulating by the inverse square-law for detectability:

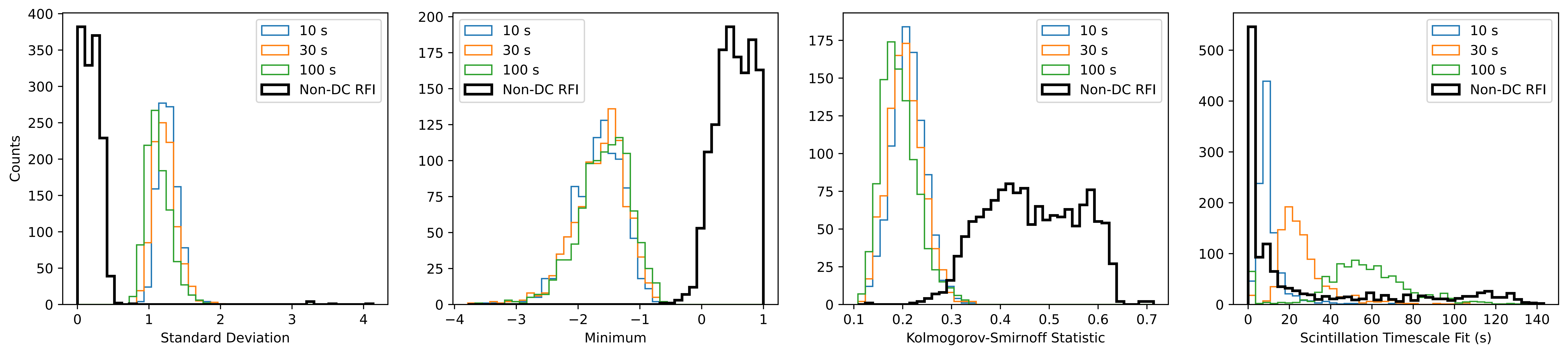
Depends on the assumptions made about transmission power and resources.

(l, b) = (1, 0)



C band

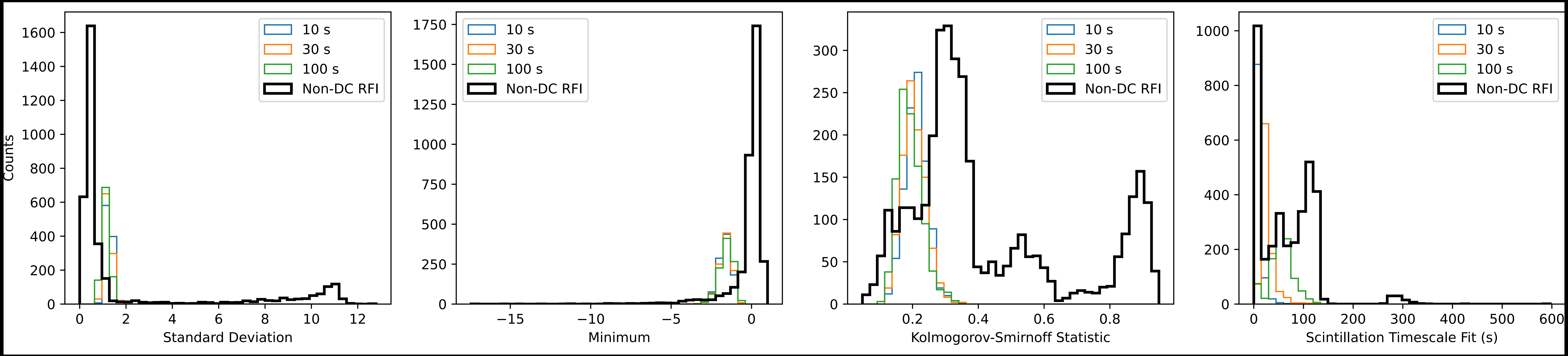
S/N = 10



Counts
Standard Deviation
Minimum
Kolmogorov-Smirnoff Statistic
Scintillation Timescale Fit (s)
10 s
30 s
100 s
Non-DC RFI

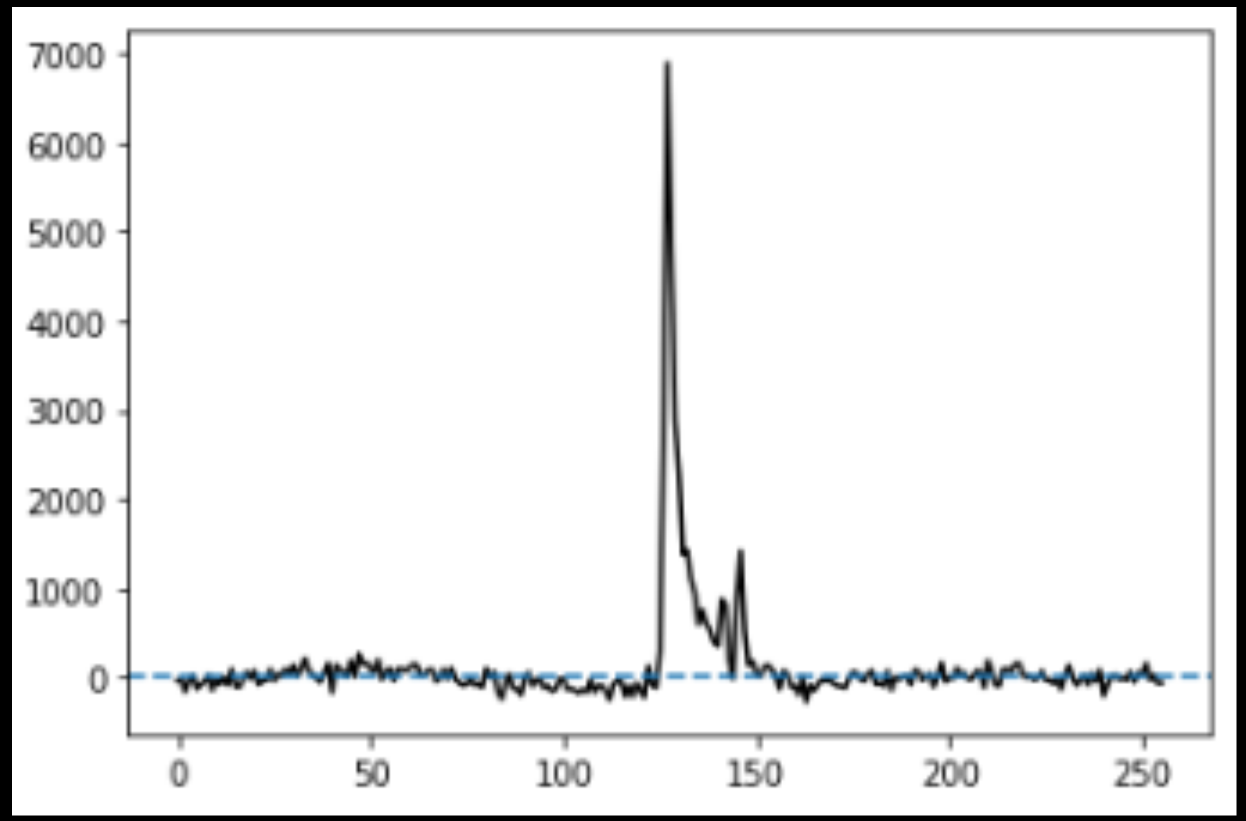
L band

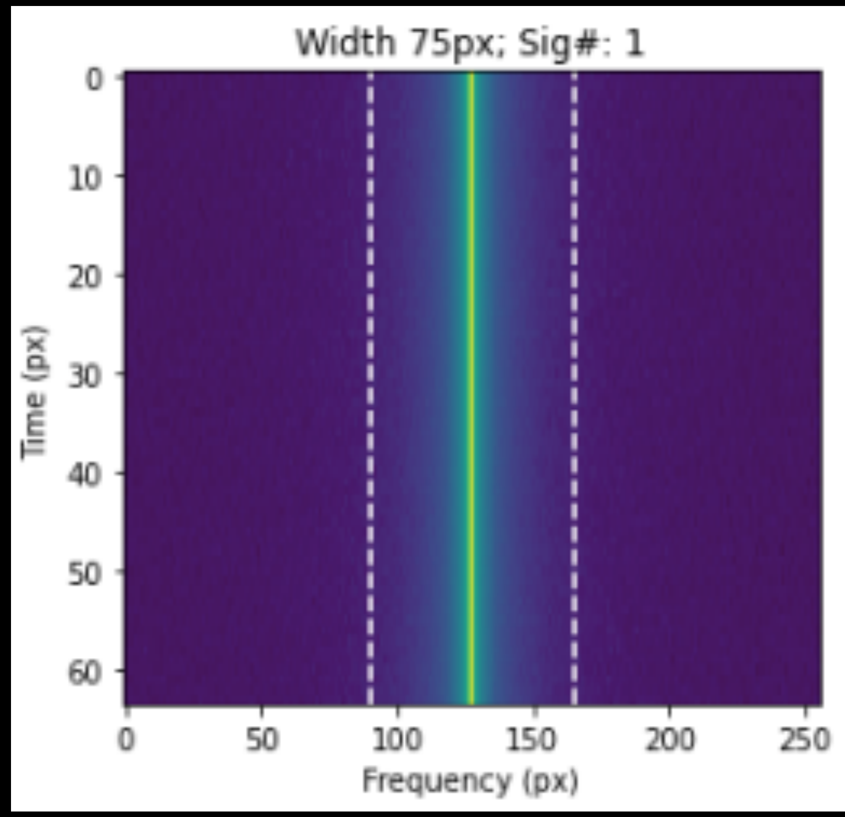
S/N = 10


Counts
Standard Deviation
Minimum
Kolmogorov-Smirnoff Statistic
Scintillation Timescale Fit (s)
10 s
30 s
100 s
Non-DC RFI

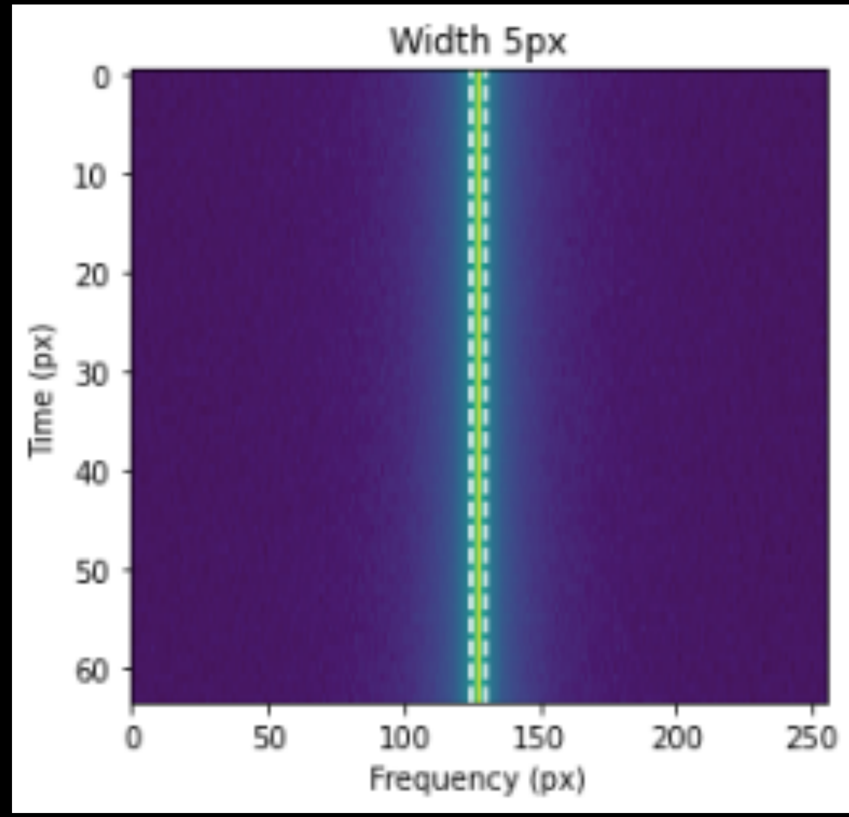
Selecting bounding boxes

- After experimentation with various methods, the final pipeline uses a combination of baseline fitting and peak detection to calculate the right size of frame to use
- The final bounds are created using a thresholding method, similar to PSRCHIVE
- Take the final bounded signal and integrate in the frequency direction to derive our raw time series – then we normalize to mean of 1 before calculating our scattering statistics





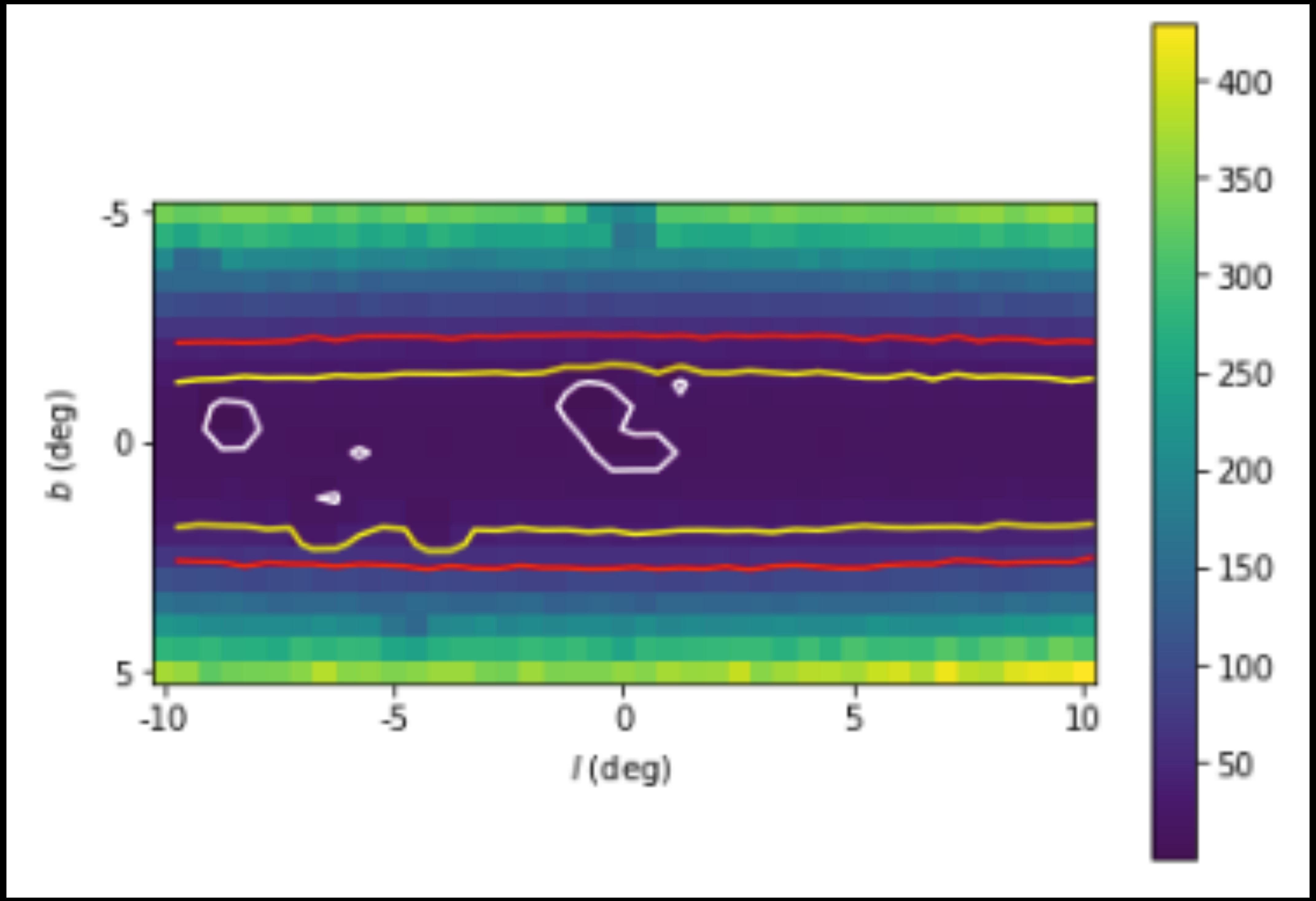
Polynomial fit



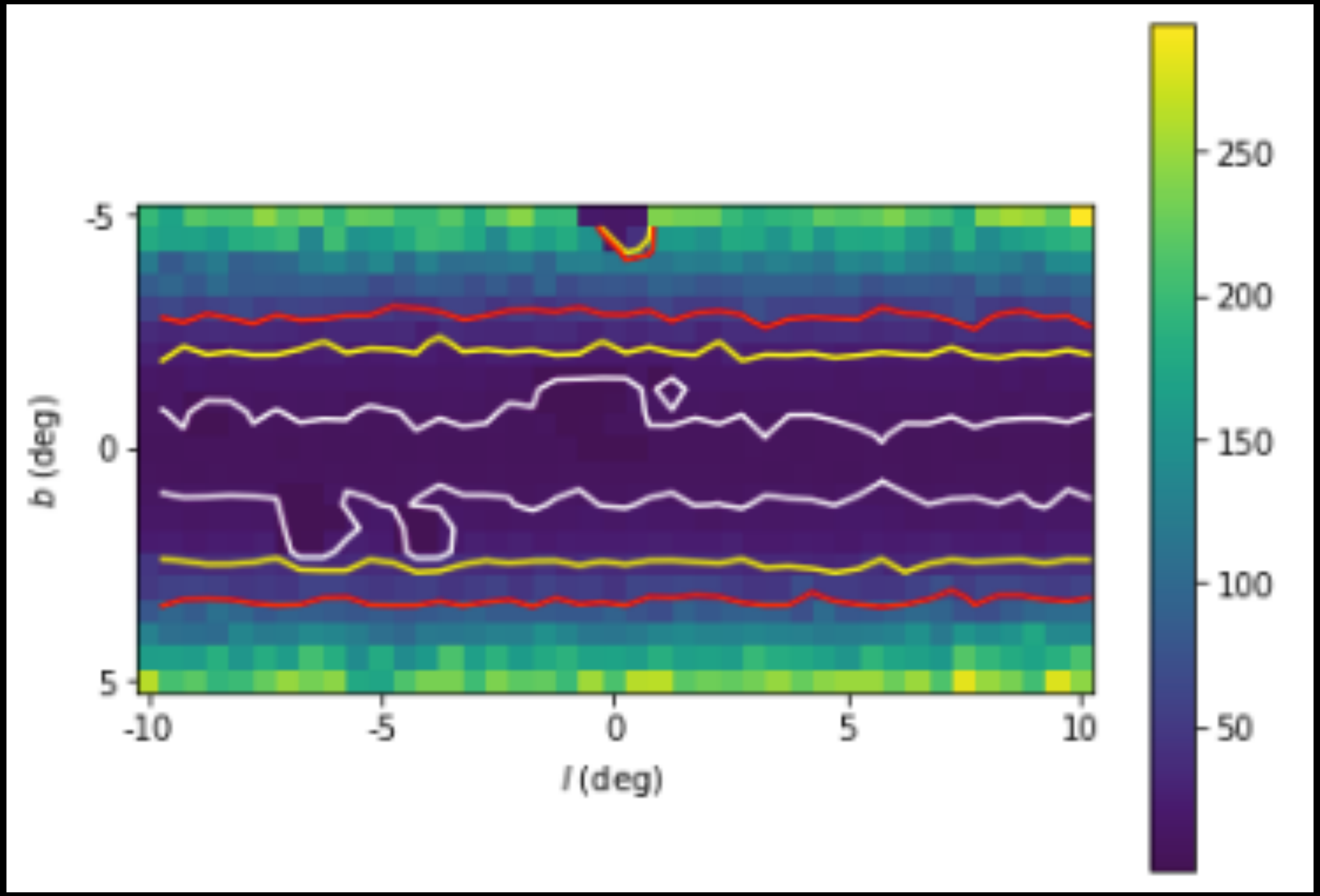
Threshold fit

Scintillation maps around the GC at C-band

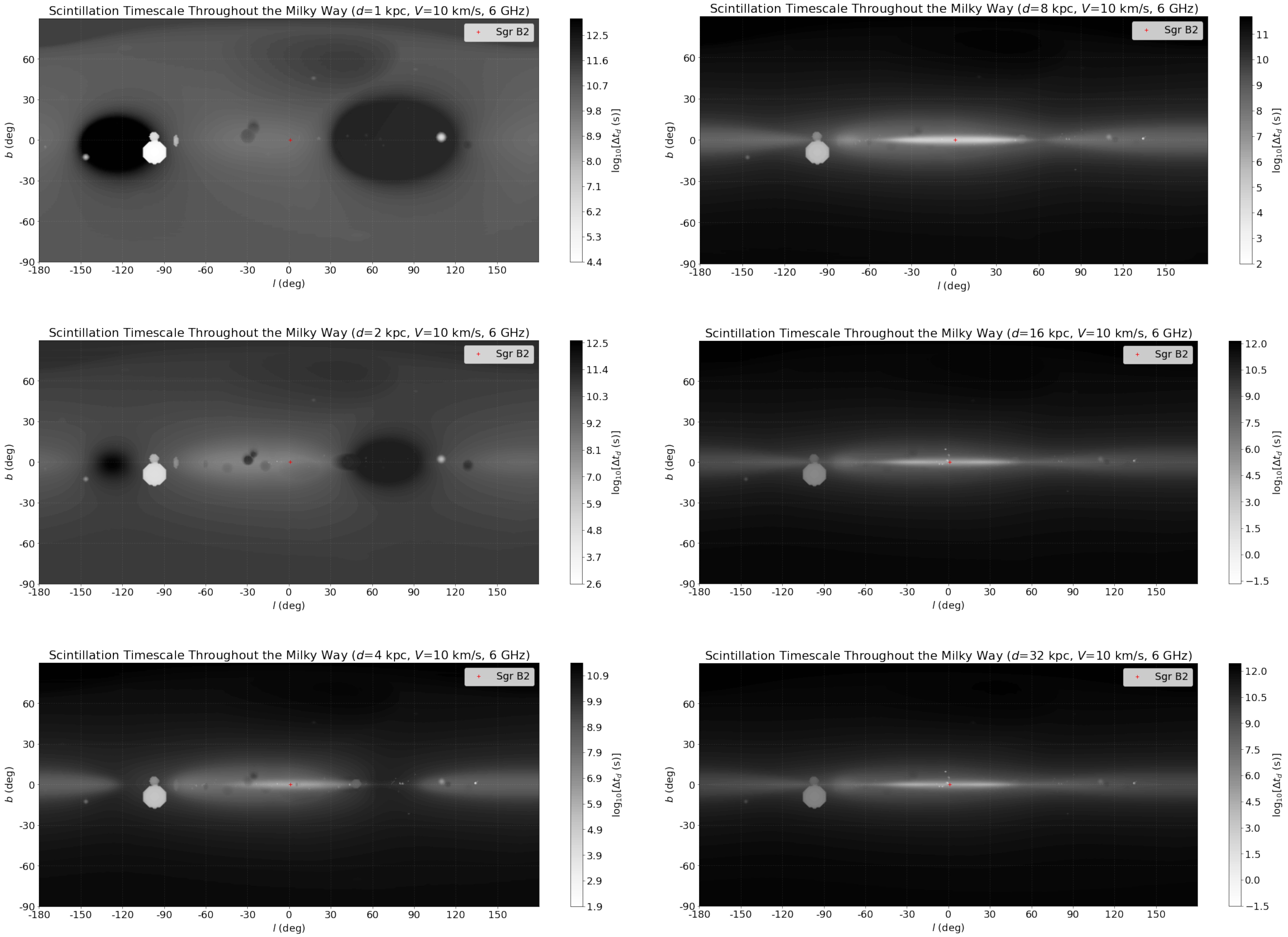
Median



Mode



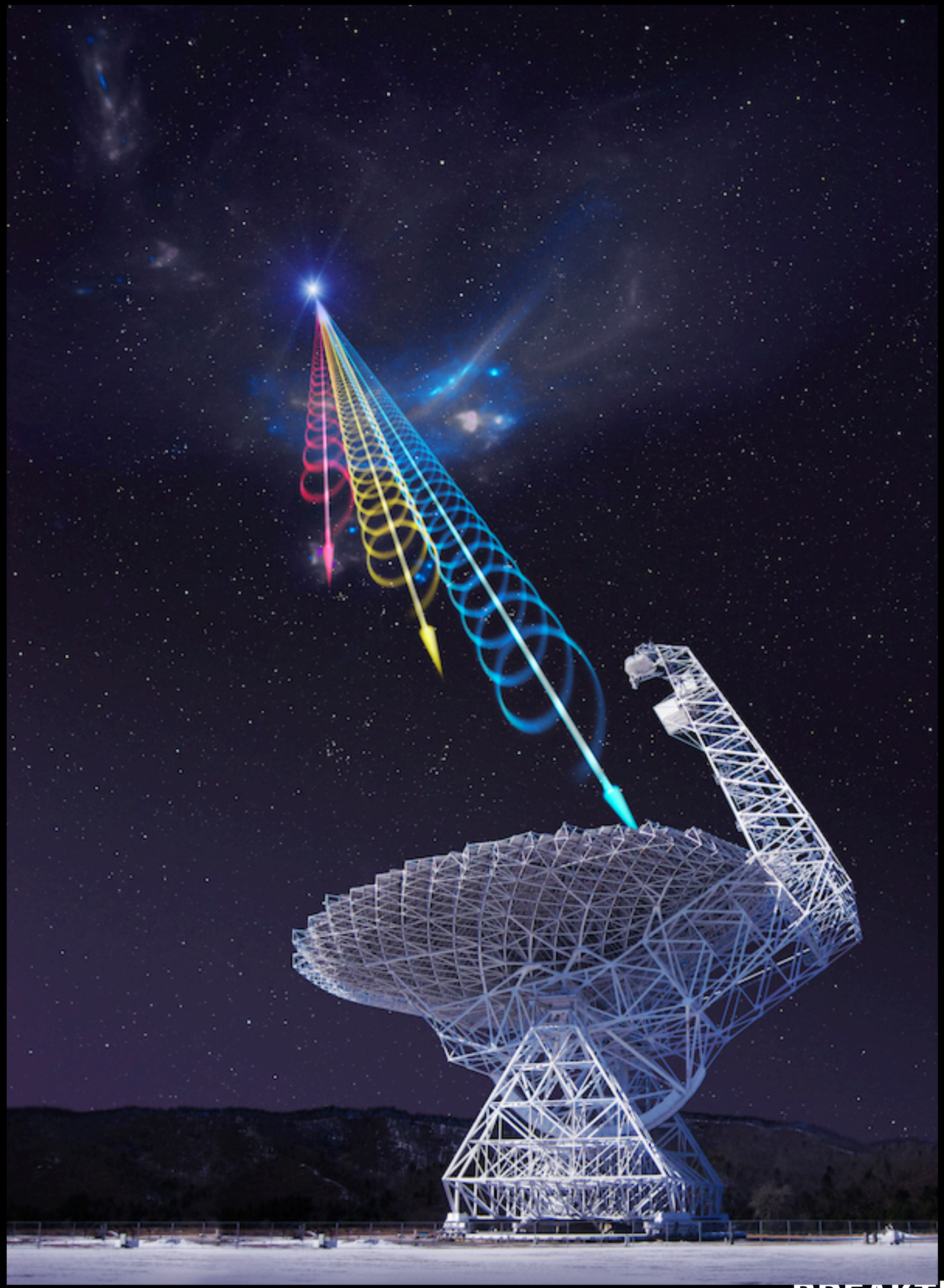
10 s, 30 s, 60 s


Scintillation Timescale Throughout the Milky Way (d=1 kpc, V=10 km/s, 6 GHz)
Sgr B2
b (deg)
l (deg)
log10[Δt_d (s)]
12.5
11.6
10.7
9.8
8.9
8.0
7.1
6.2
5.3
4.4
Scintillation Timescale Throughout the Milky Way (d=8 kpc, V=10 km/s, 6 GHz)
Sgr B2
b (deg)
l (deg)
log10[Δt_d (s)]
11
10
9
8
7
6
5
4
3
2
Scintillation Timescale Throughout the Milky Way (d=2 kpc, V=10 km/s, 6 GHz)
Sgr B2
b (deg)
l (deg)
log10[Δt_d (s)]
12.5
11.4
10.3
9.2
8.1
7.0
5.9
4.8
3.7
2.6
Scintillation Timescale Throughout the Milky Way (d=16 kpc, V=10 km/s, 6 GHz)
Sgr B2
b (deg)
l (deg)
log10[Δt_d (s)]
12.0
10.5
9.0
7.5
6.0
4.5
3.0
1.5
0.0
−1.5
Scintillation Timescale Throughout the Milky Way (d=4 kpc, V=10 km/s, 6 GHz)
Sgr B2
b (deg)
l (deg)
log10[Δt_d (s)]
10.9
9.9
8.9
7.9
6.9
5.9
4.9
3.9
2.9
1.9
Scintillation Timescale Throughout the Milky Way (d=32 kpc, V=10 km/s, 6 GHz)
Sgr B2
b (deg)
l (deg)
log10[Δt_d (s)]
12.0
10.5
9.0
7.5
6.0
4.5
3.0
1.5
0.0
−1.5

My goal: develop search methods for SETI from both angles

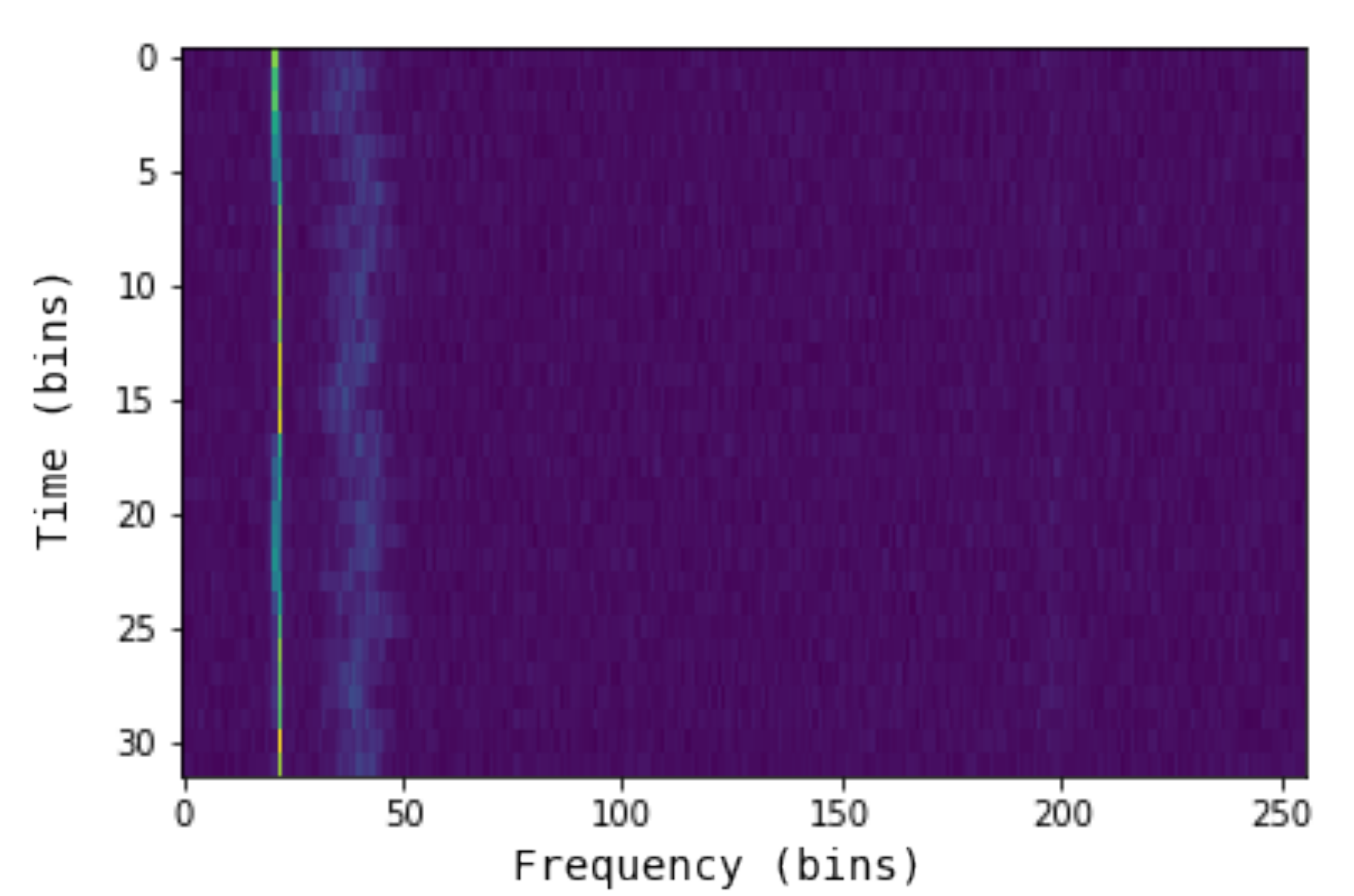
- Machine learning and software tools to support more complex detections

- Investigate astrophysical effects imprinted on technosignatures themselves



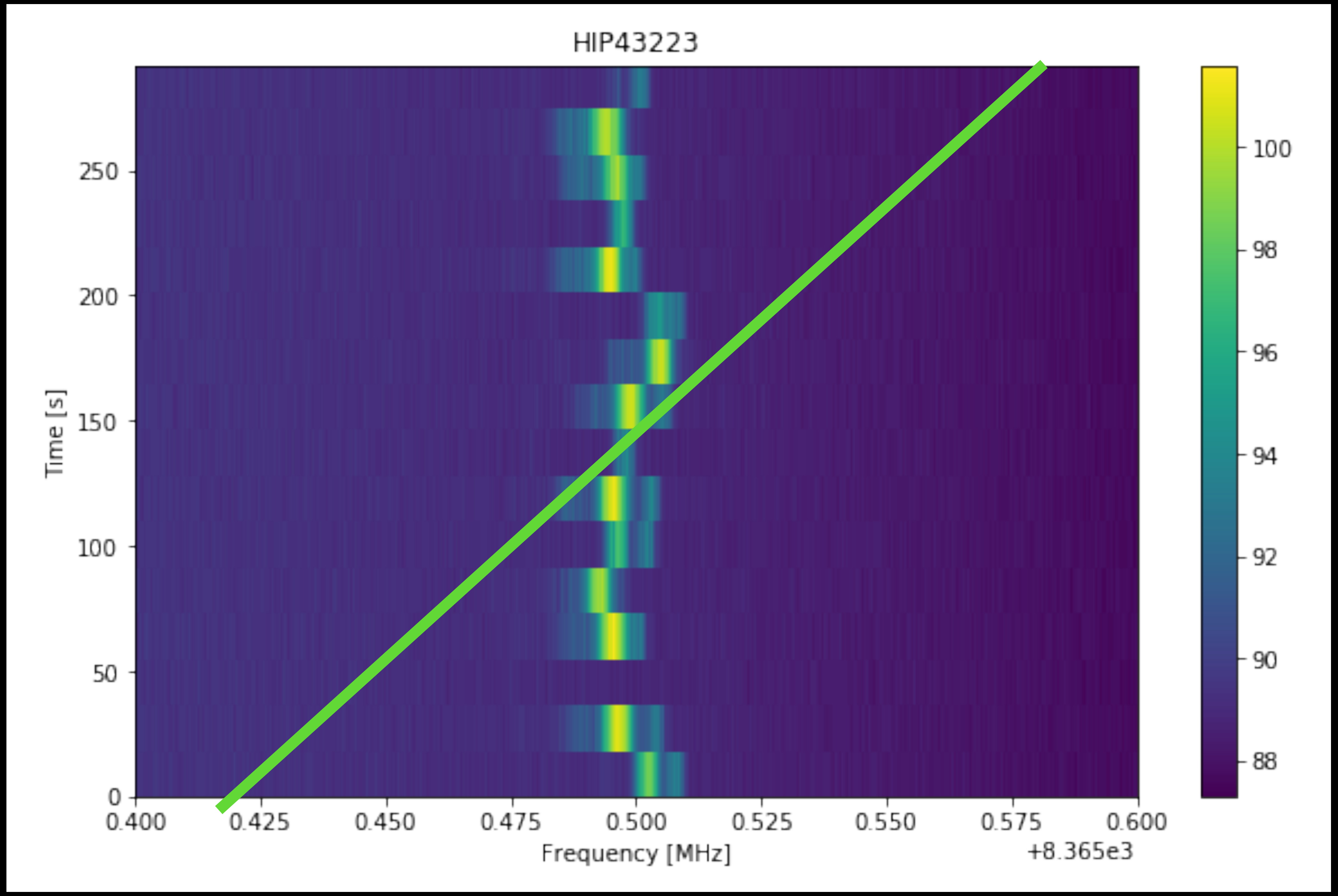
Narrowband signal localization with machine learning

- Standard deDoppler pipeline:
  - Dim signals concealed by nearby bright signals
  - Computationally expensive to search high drift rates



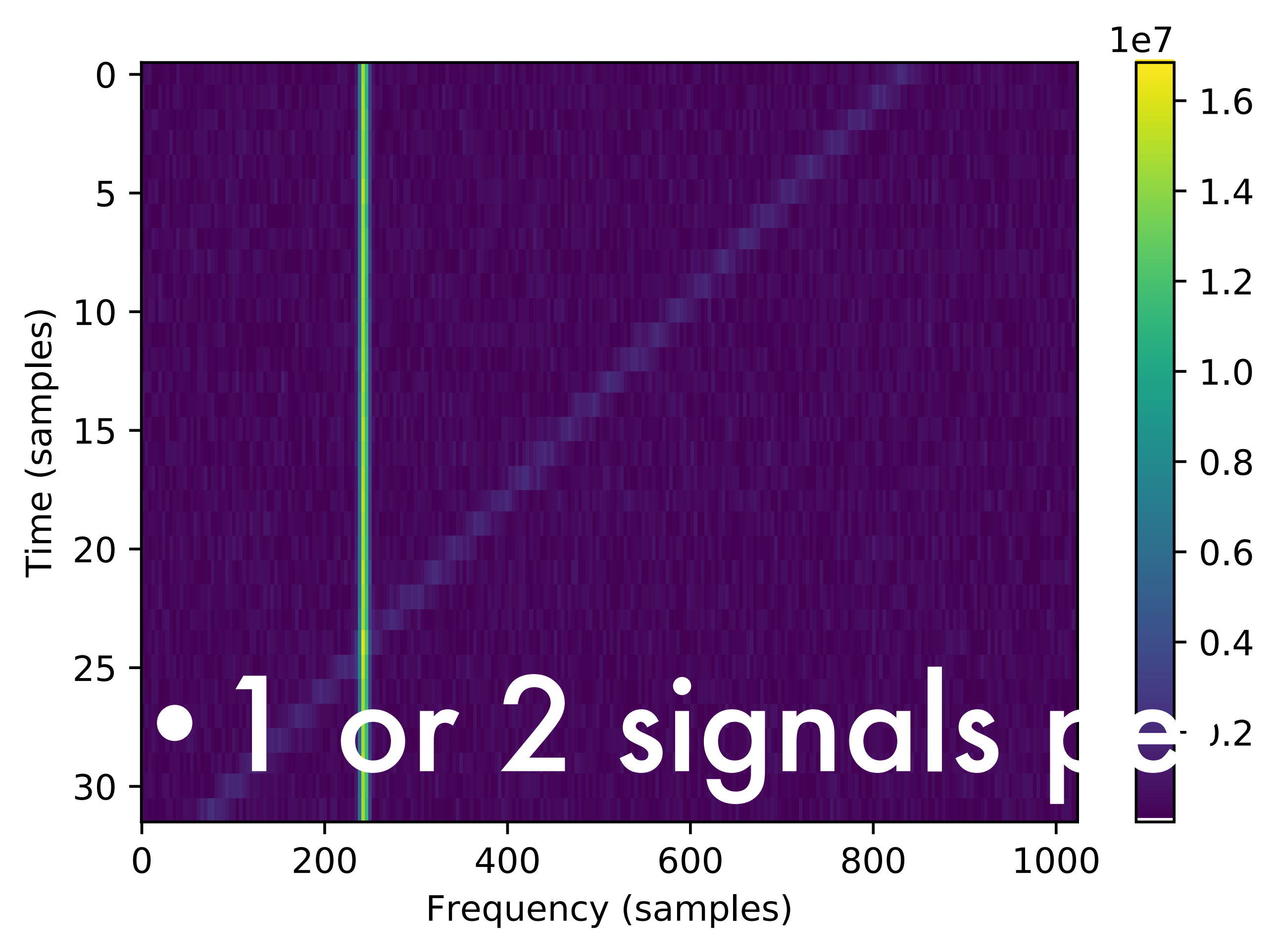
Small snippet of GBT data at C-band

• Masking?



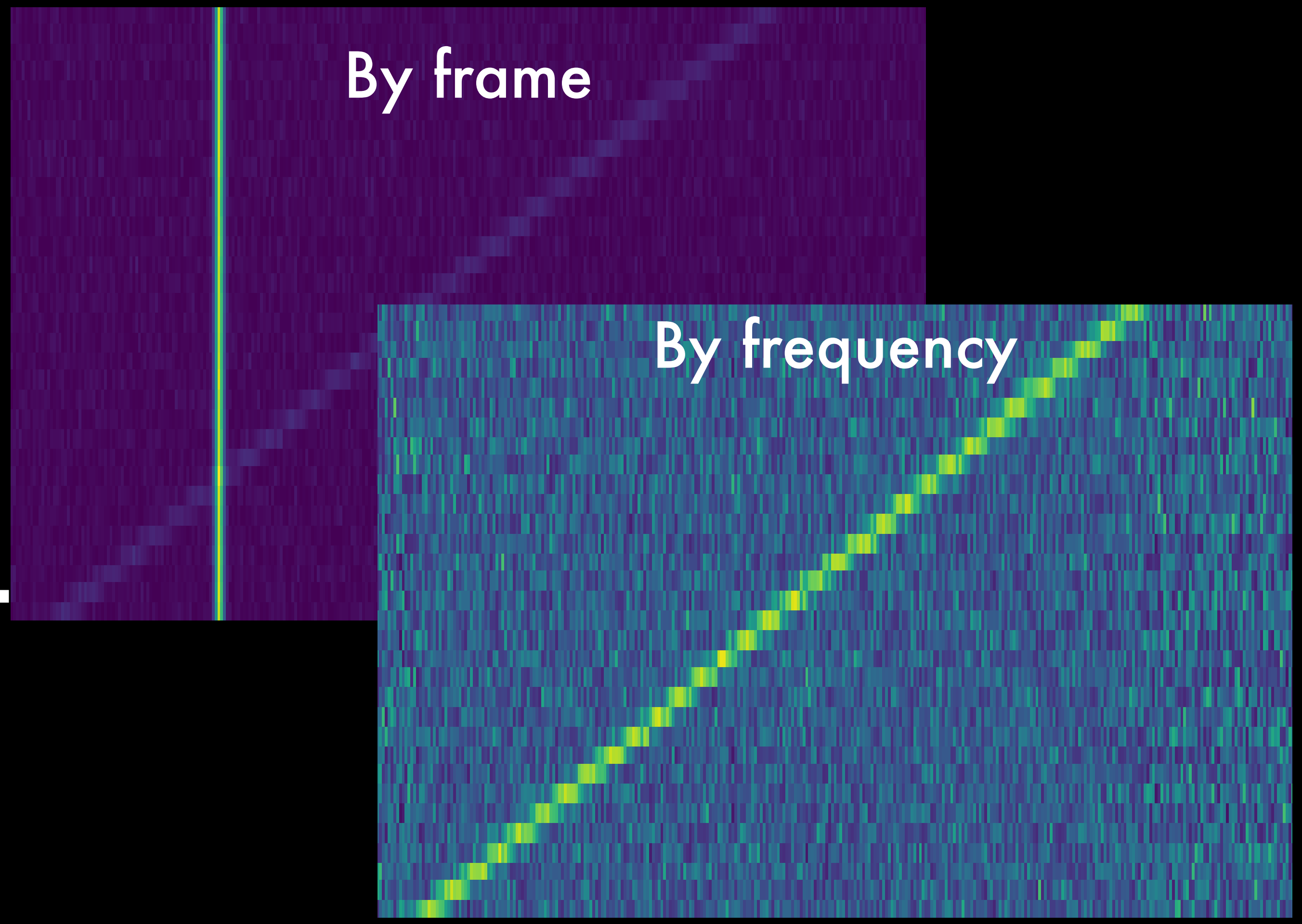
Brzycki et al. 2020, PASP

Synthetic training data



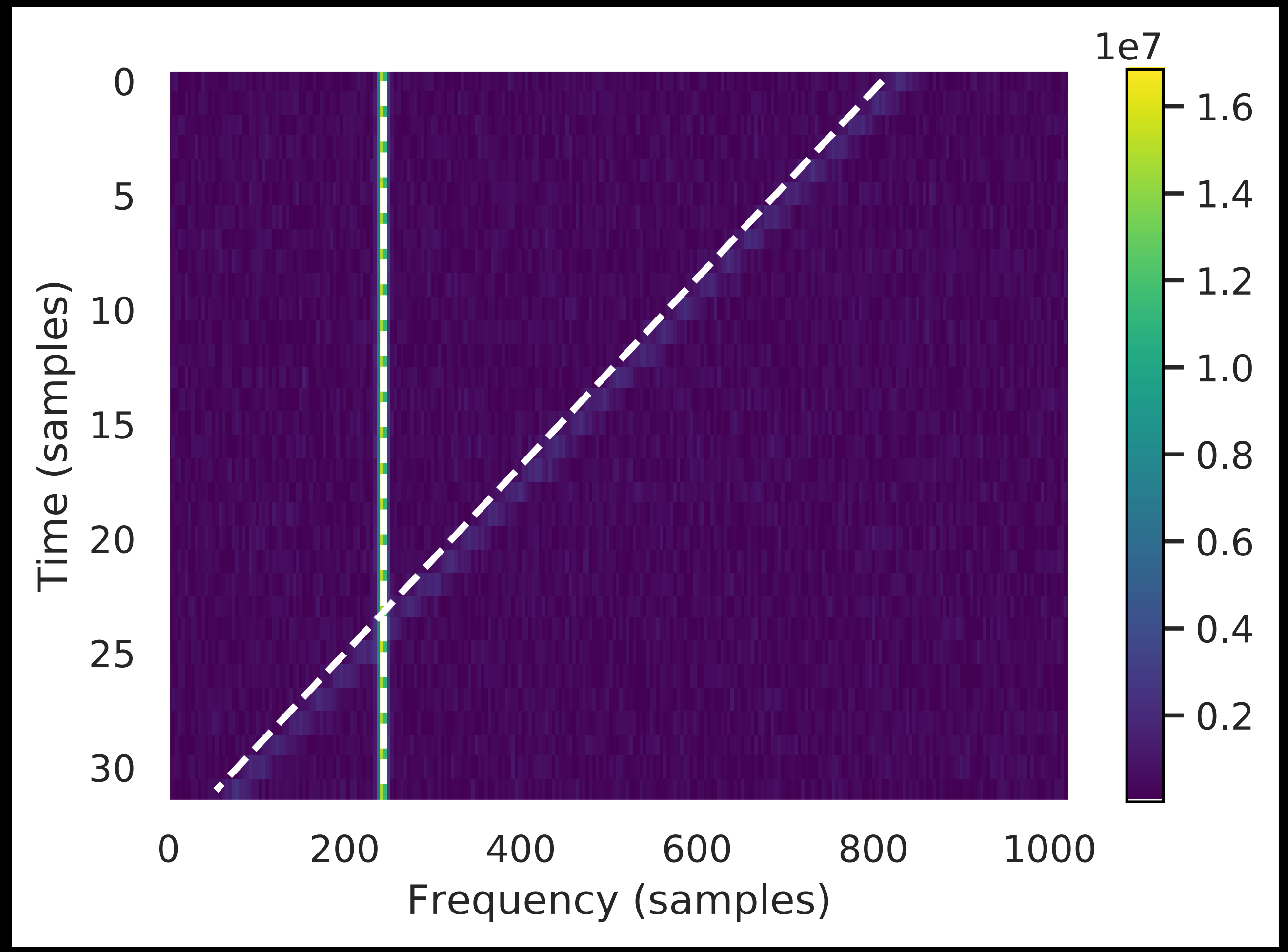
- 1 or 2 signals per input

Normalization



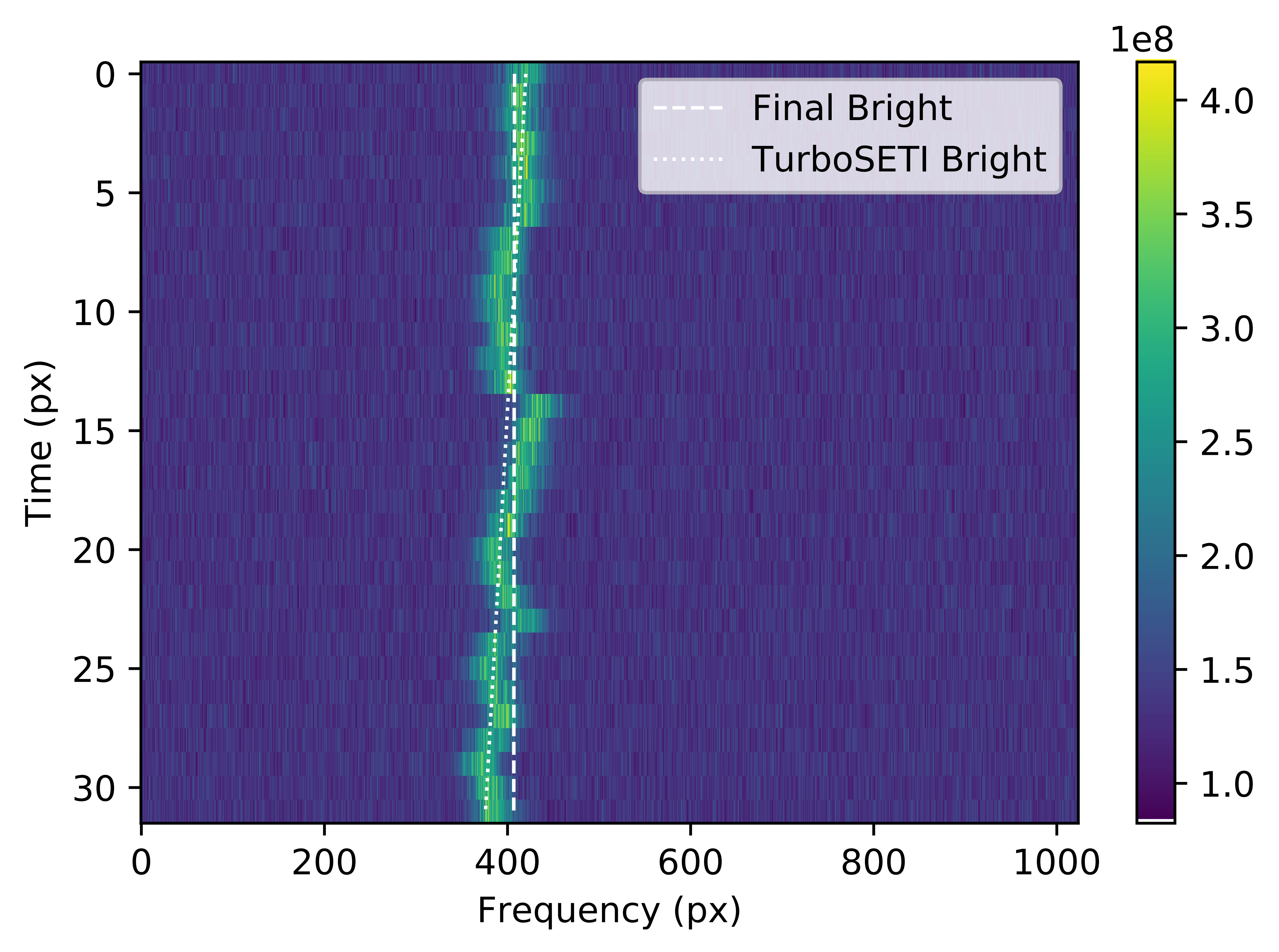
Neural Network

Predicted locations



Takeaways

- Less accurate than deDoppler methods, but generally 20-40x faster
- Trained on ideal signals but still relatively robust
- For production use, would need to extend to variable number of signals

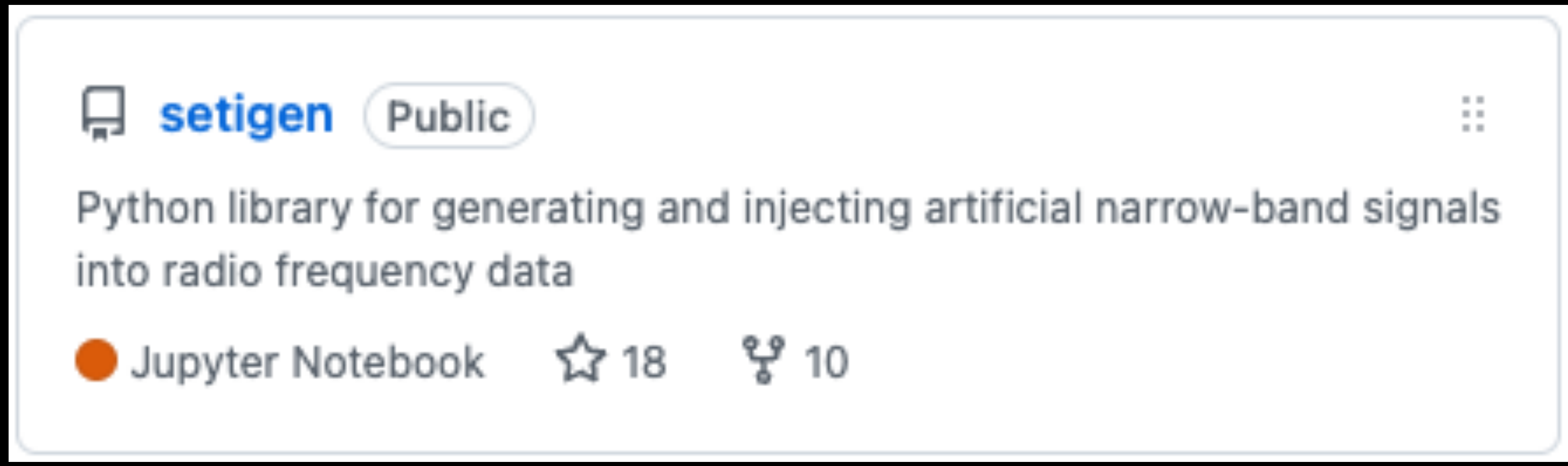


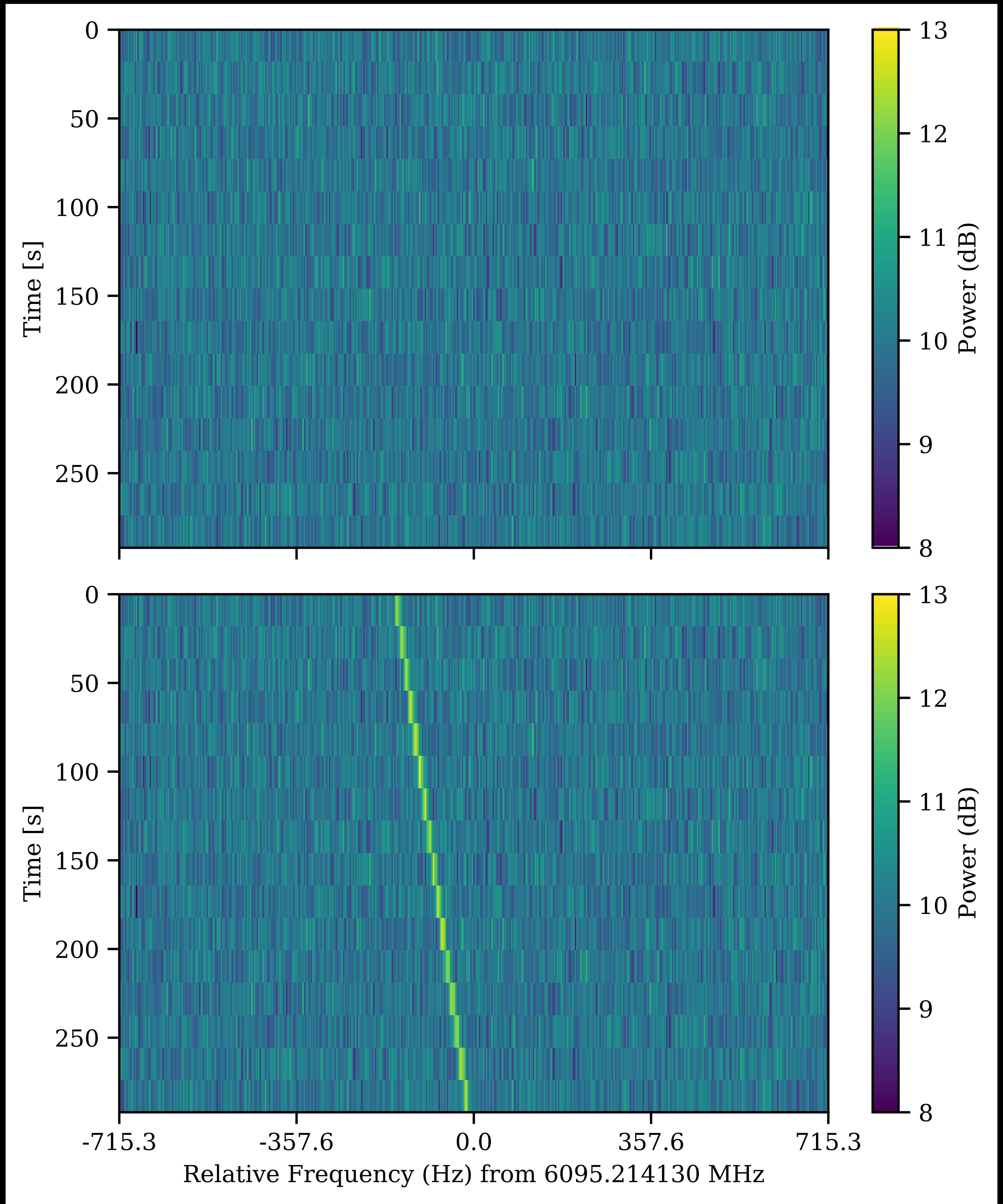
C-band RFI signal, with ML prediction dashed and TurboSETI localization dotted.

Brzycki et al. 2020, PASP

Setigen

- Python library for synthetic spectrogram and voltage data
- Specific focus on narrowband signal generation and injection





Brzycki et al. 2022, AJ

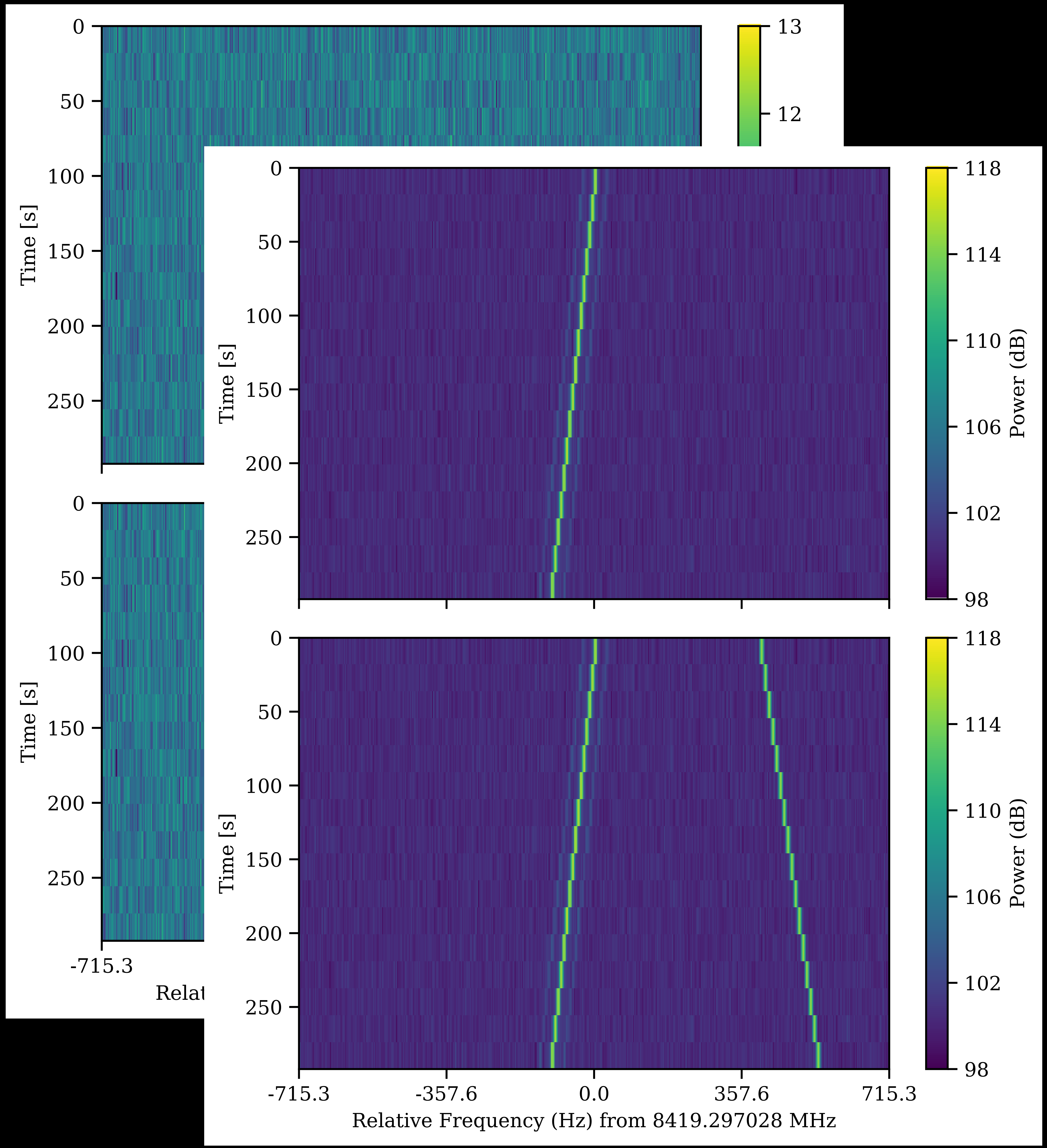
Setigen

- Python library for synthetic spectrogram and voltage data
- Specific focus on narrowband signal generation and injection

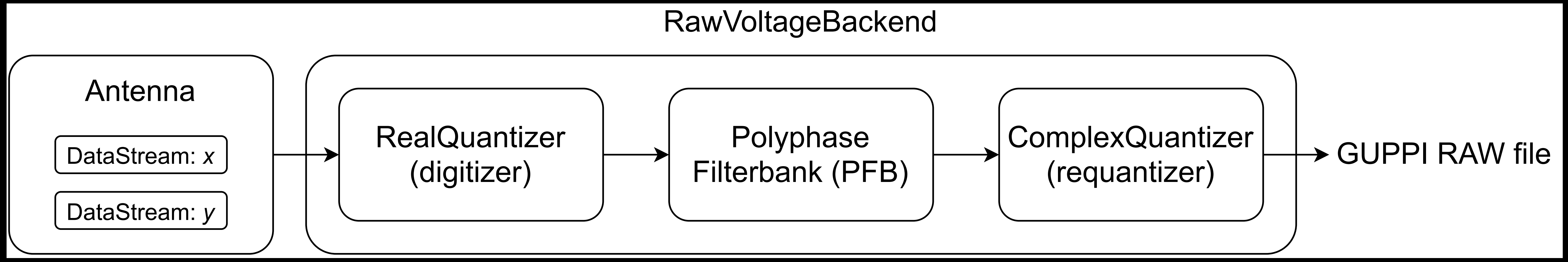
setigen Public

Python library for generating and injecting artificial narrow-band signals into radio frequency data

Jupyter Notebook ☆ 18 10

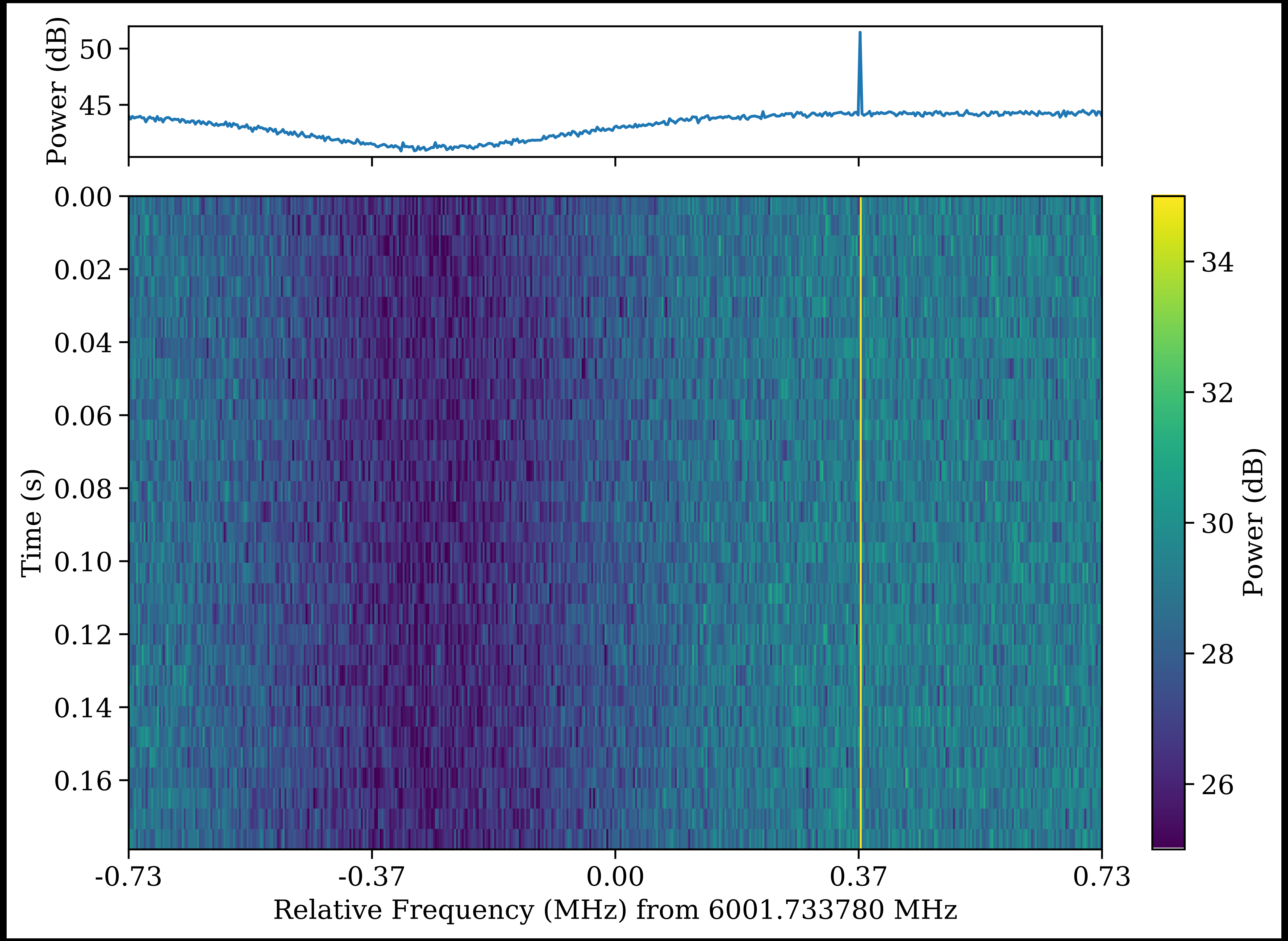


Brzycki et al. 2022, AJ



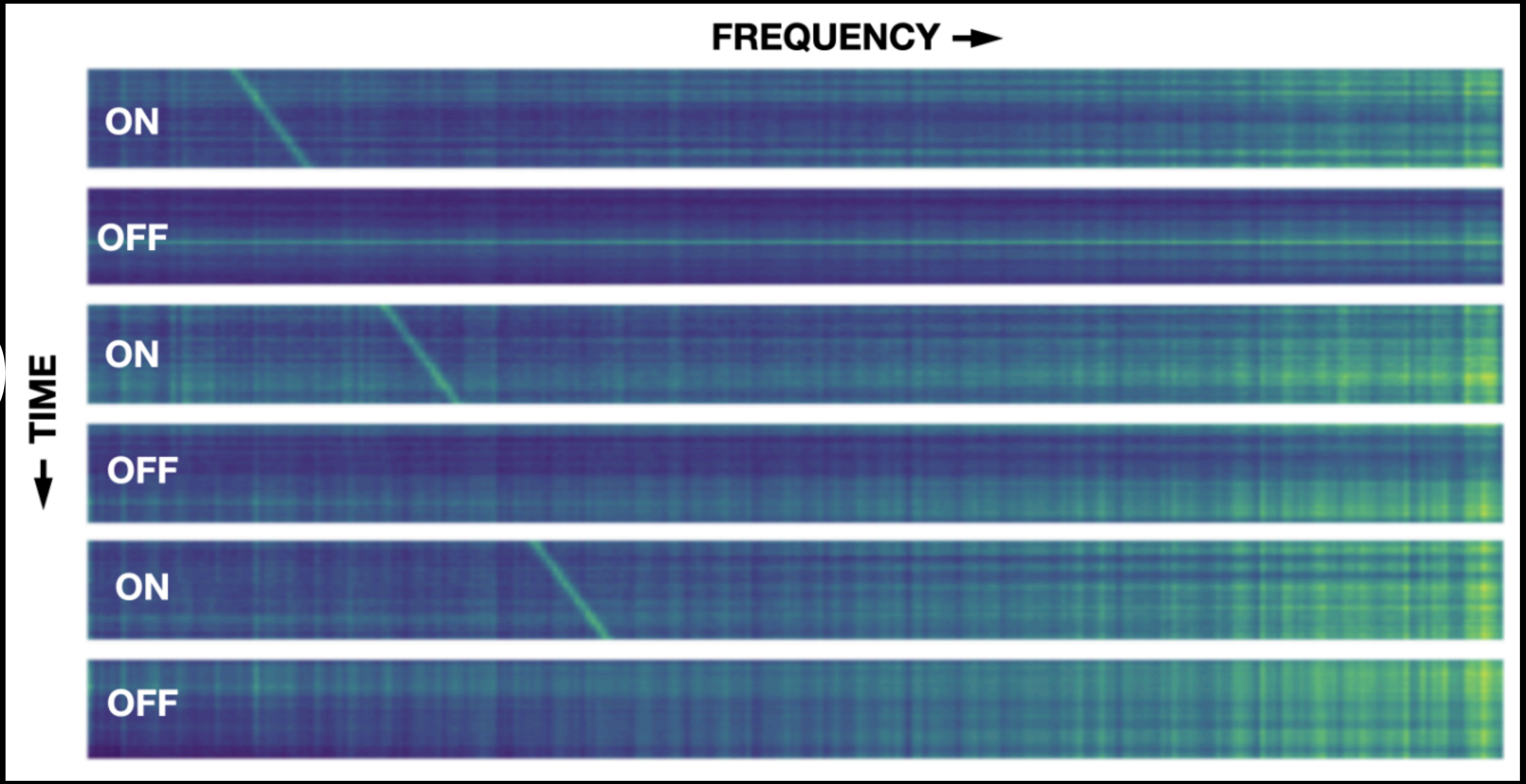
- Synthetic complex voltage data

- Simple models of backend components, such as a polyphase filterbank



Applications of Setigen beyond my research

- Injection – recovery testing
- ML dataset production (e.g. Kaggle)
- Multibeam search surveys
- Development of software for the Allen Telescope Array



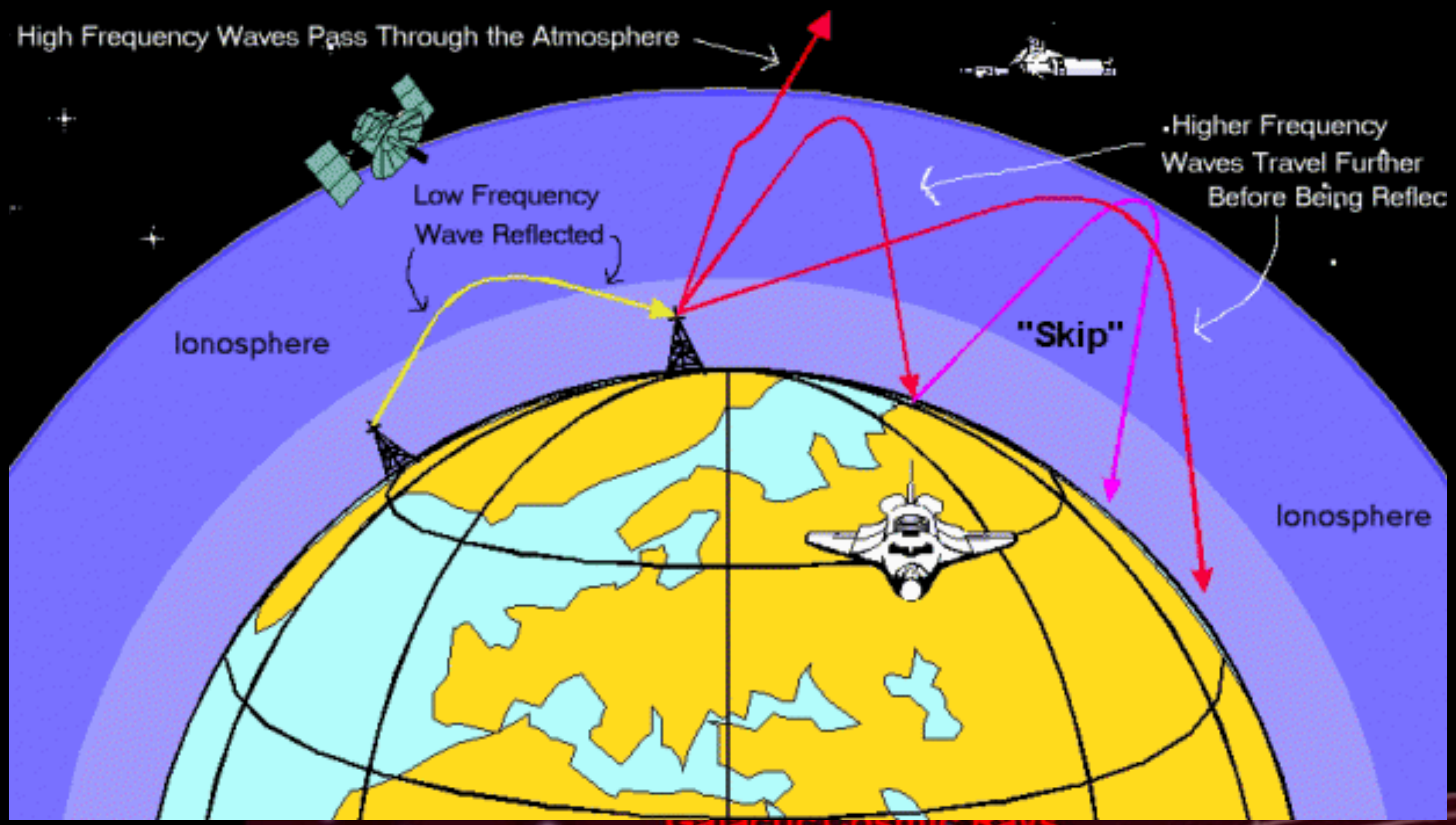
Breakthrough Listen x Kaggle 2021

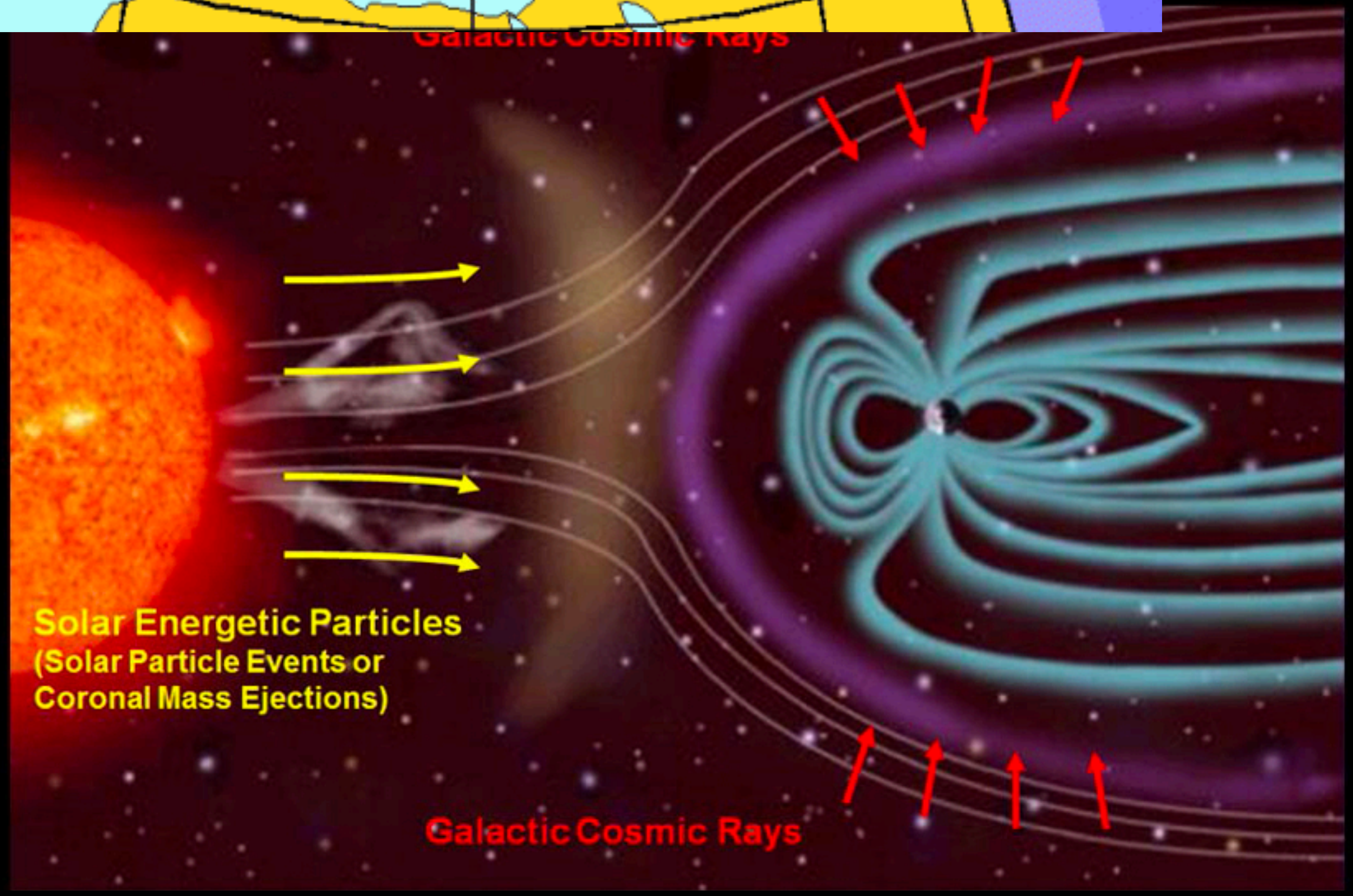
Scattering intensity

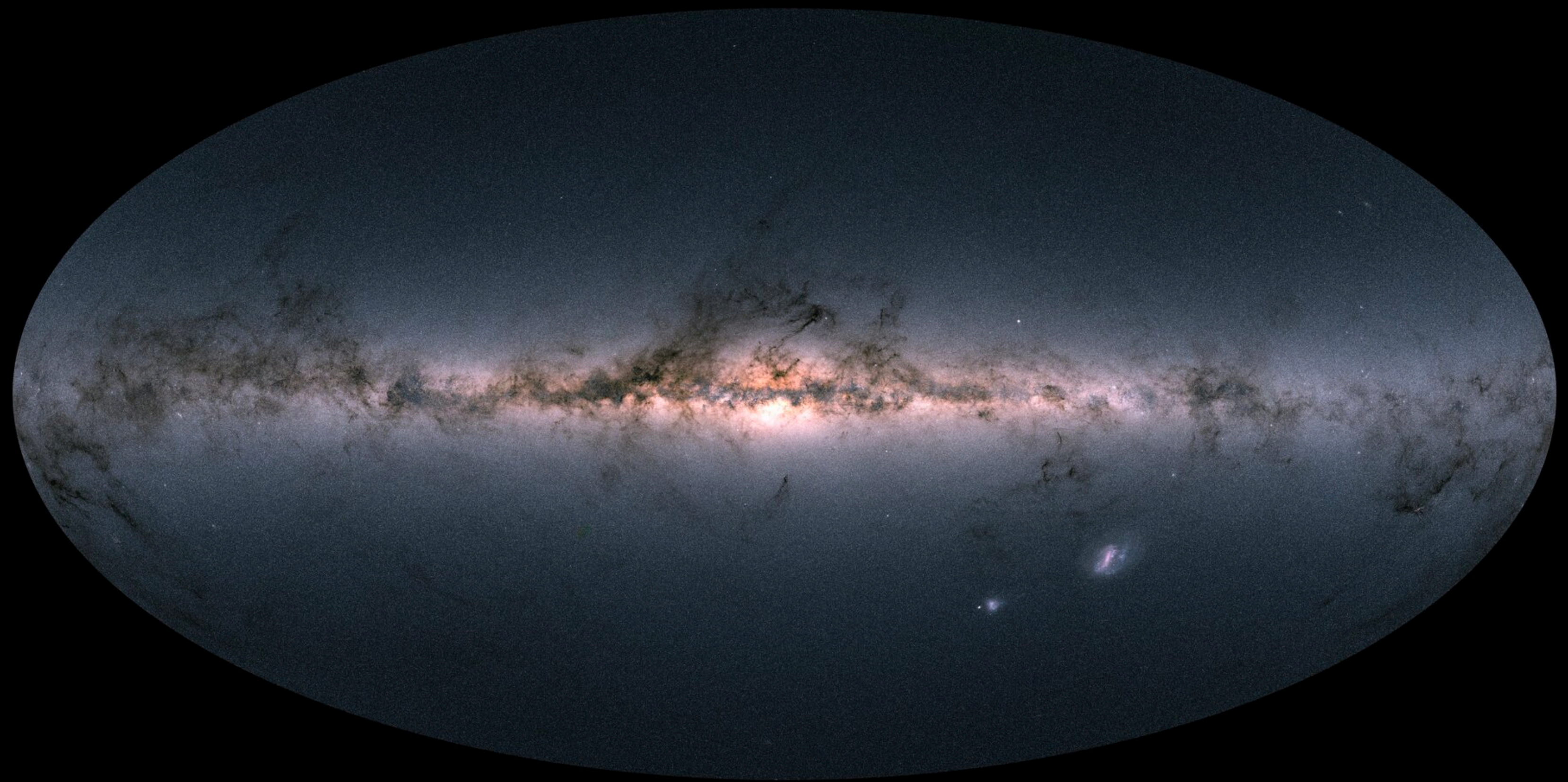
- Ionosphere – weak $m_d \ll 1$

- IPM – mostly weak

- ISM – can be strong! $m_d \approx 1$



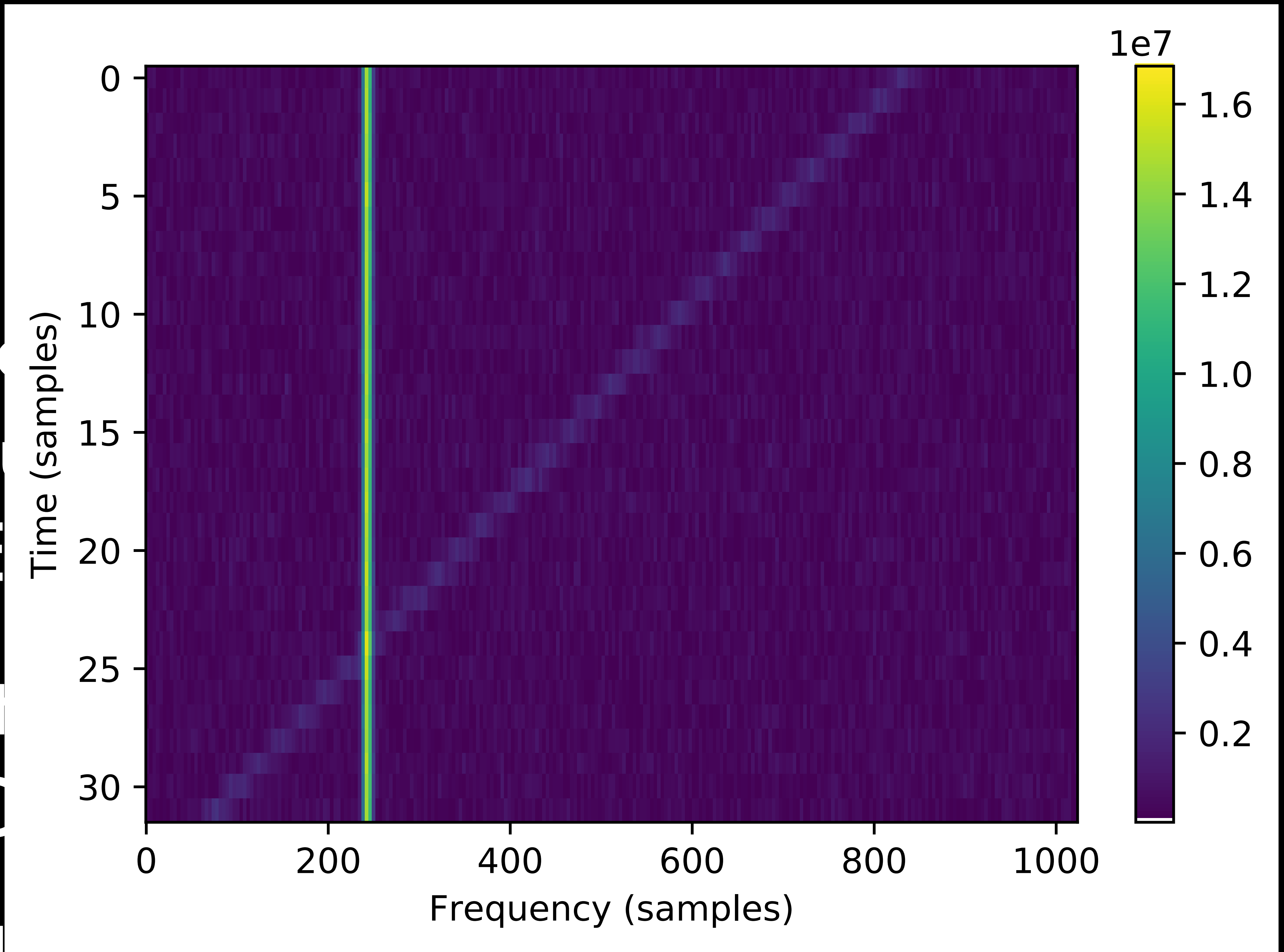




ROUGH

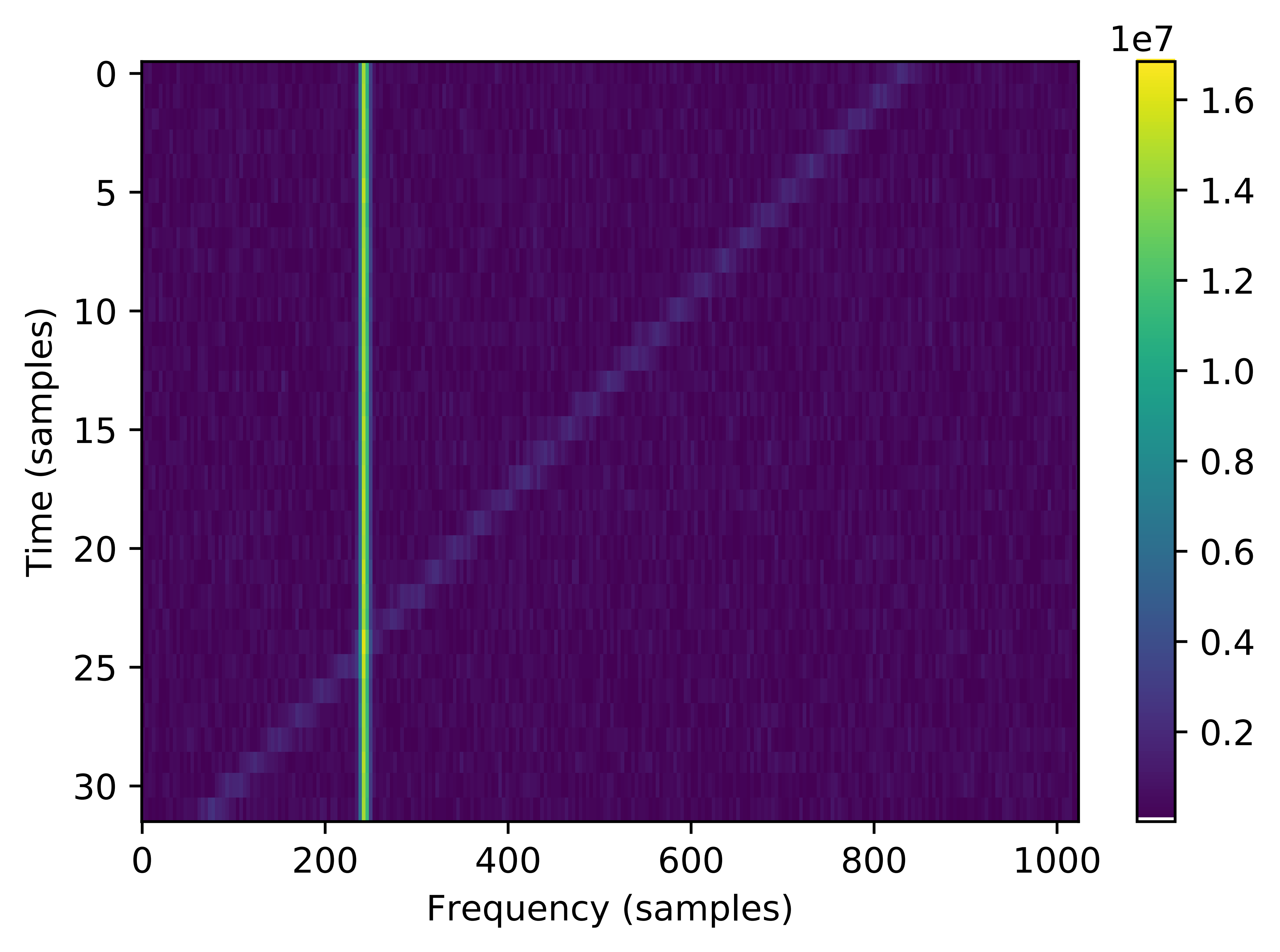
Narrow-band signal localization (Brzycki et al. 2020)

- With a means of simulating
  can produce frames that wo
  "organically" using TurboSE
- Importantly, we can generat
  synthetic data, and train a C
- Localization of narrow-band signals is a good initial ML problem because it's a relatively simple task; predict 2 numbers per signal



Example of a frame with 2 synthetic signals, at 25 and 15 dB

Narrow-band signal localization (Brzycki et al. 2020)

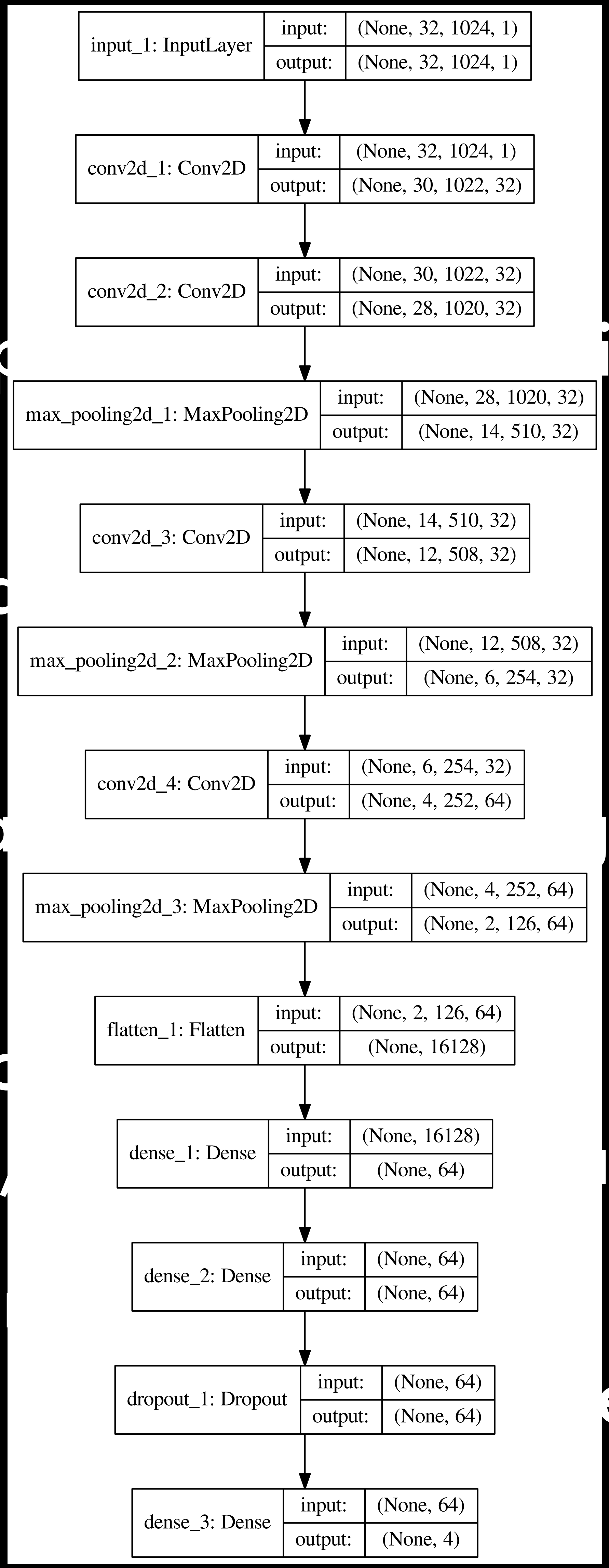


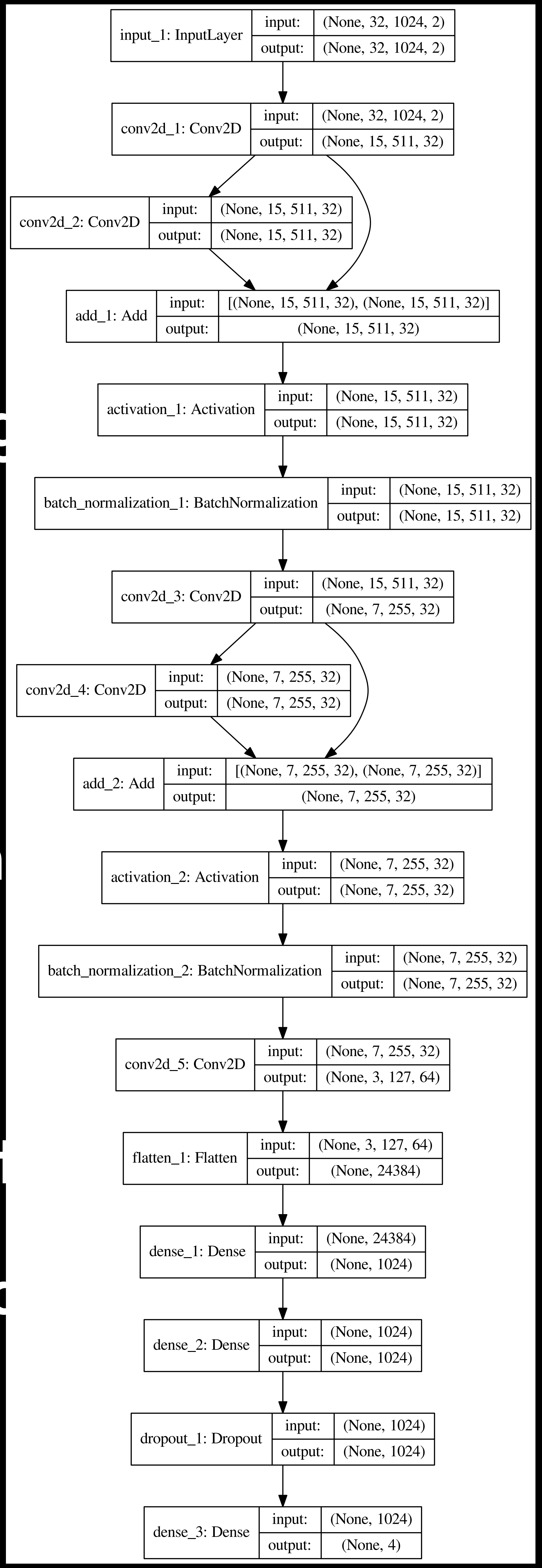
Example of a frame with 2 synthetic signals.

- Created two main datasets, b
  samples and 24,000 test sam
  - One signal, at 0, 5, ..., 25
  - Two signals, one at 0, 5, .
    rate, and the other at 25 c
    to simulate "bright" RFI)
- The one signal dataset allows for direct comparison with TurboSETI; the two signal dataset tests the effectiveness of localizing multiple signals simultaneously

Model architectures

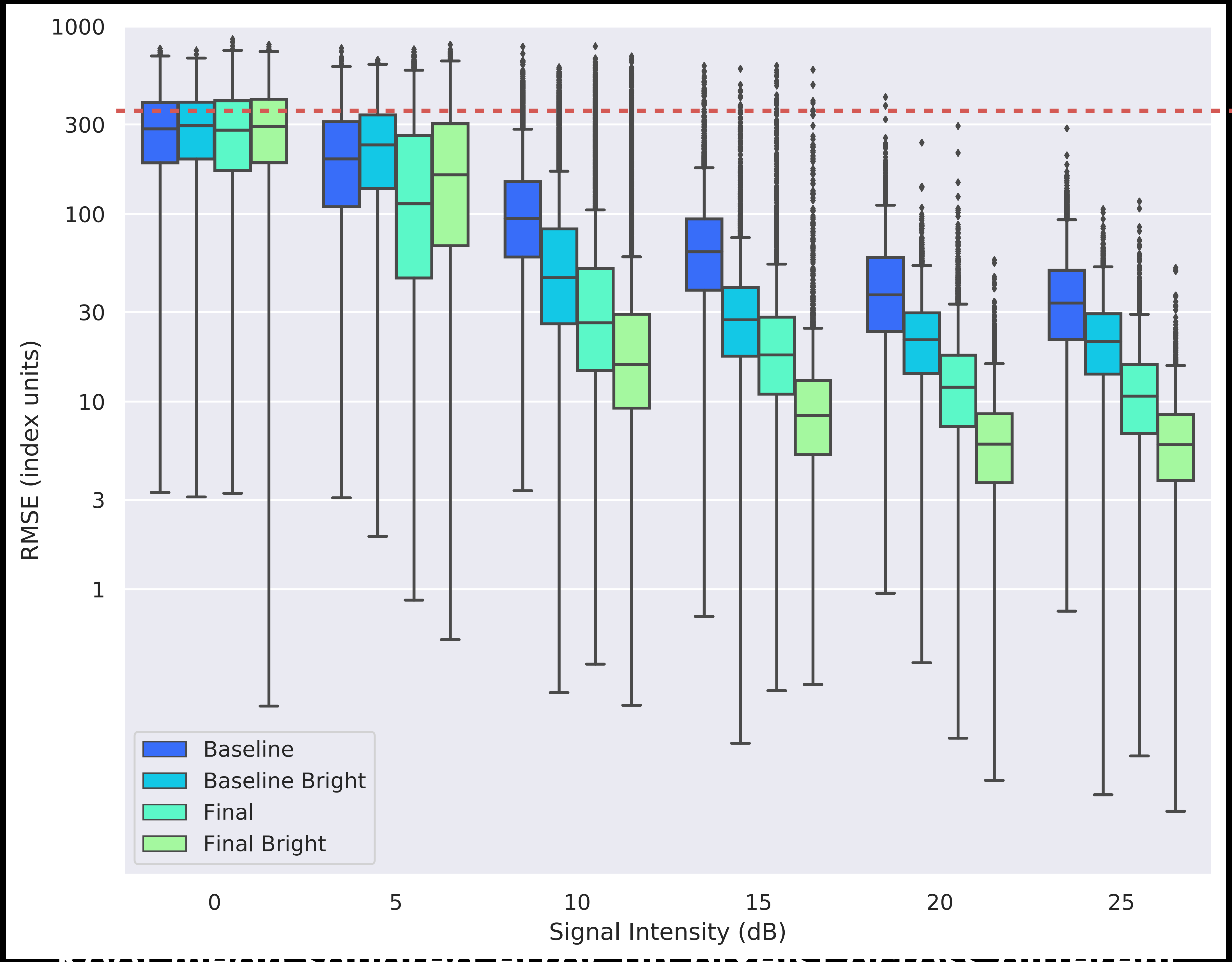
- Used convolutional neural networks, esp imag input data
- Created a "baseline" and a "final" mod performance:
  - Baseline model uses convolutional la , an fully connected layers
  - Final model includes residual connec convolutions instead of max pooling, lizat
- In addition to training these models ove data we did alternate training over only 10 - es, labeling these as "bright" models



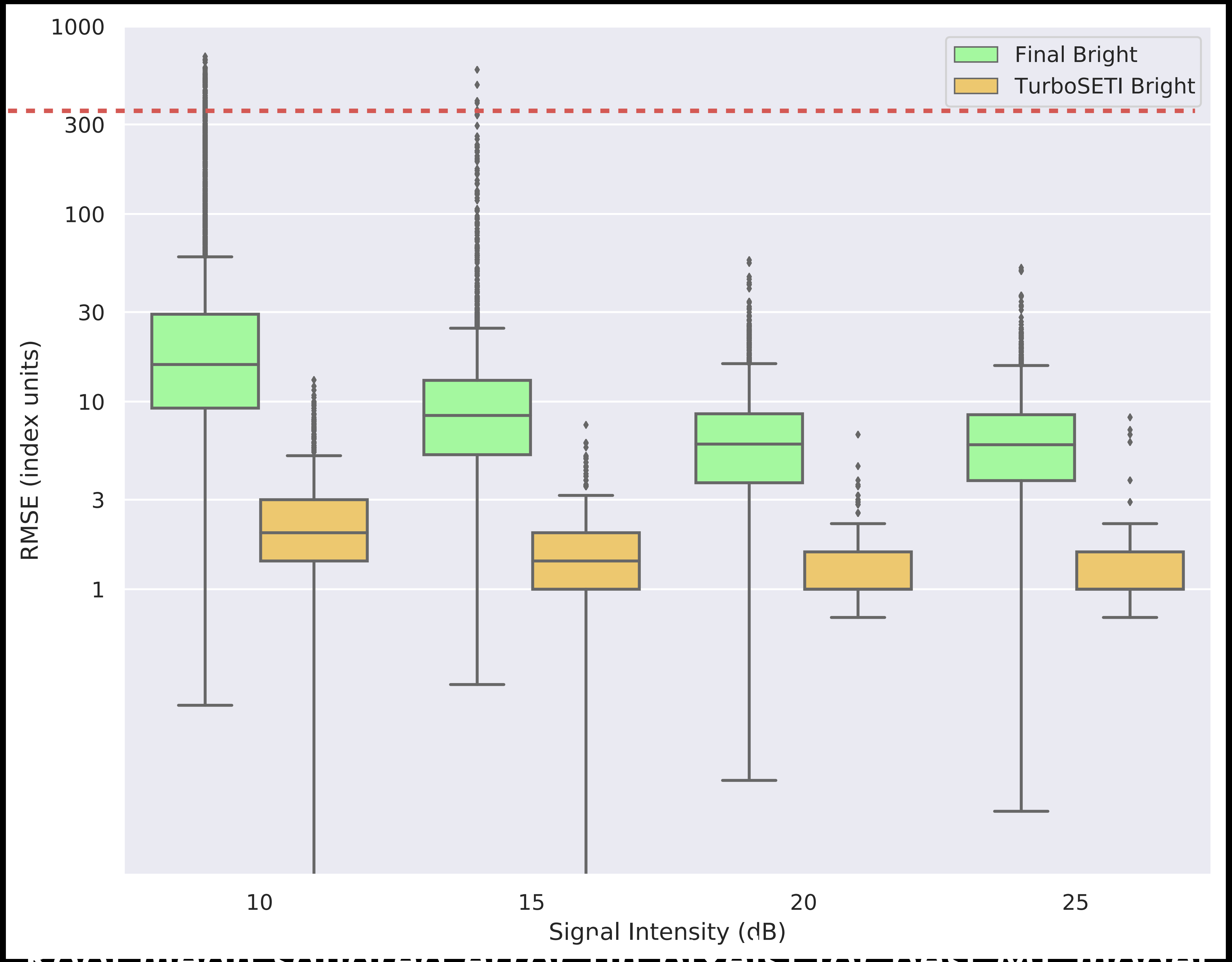


$$\text{RMSE (index units)} = 1024 \times \sqrt{\frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2}$$

ONE SIGNAL RESULTS ON TEST DATA



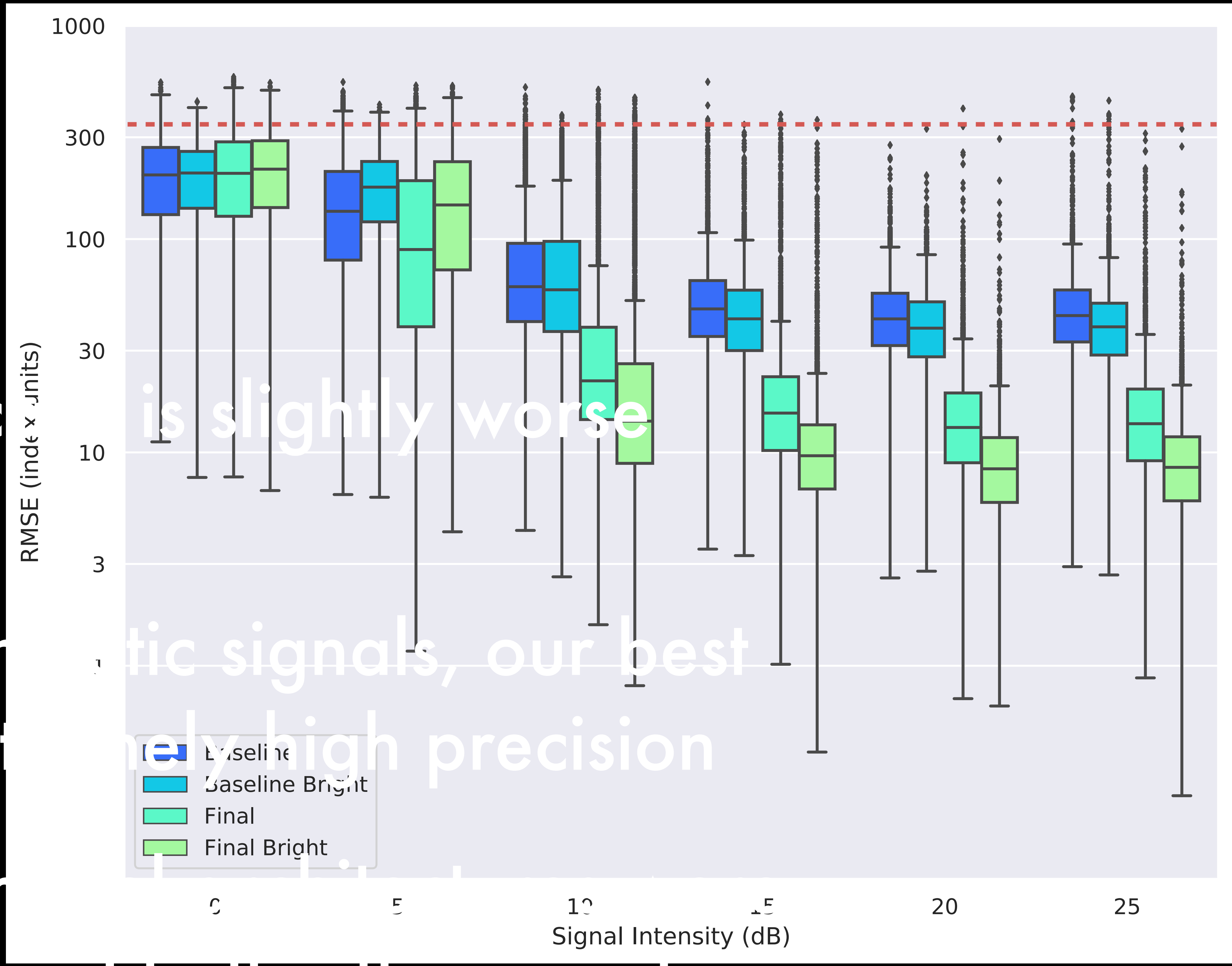
Root mean squared error (in pixels) across different models



Root mean squared error (in pixels) for best ML model vs. TurboSETI. Only calculated for SNR > 10.

$$\text{RMSE (index units)} = 1024 \times \sqrt{\frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2}$$

WO SIGNAL RESULTS ON TEST DATA

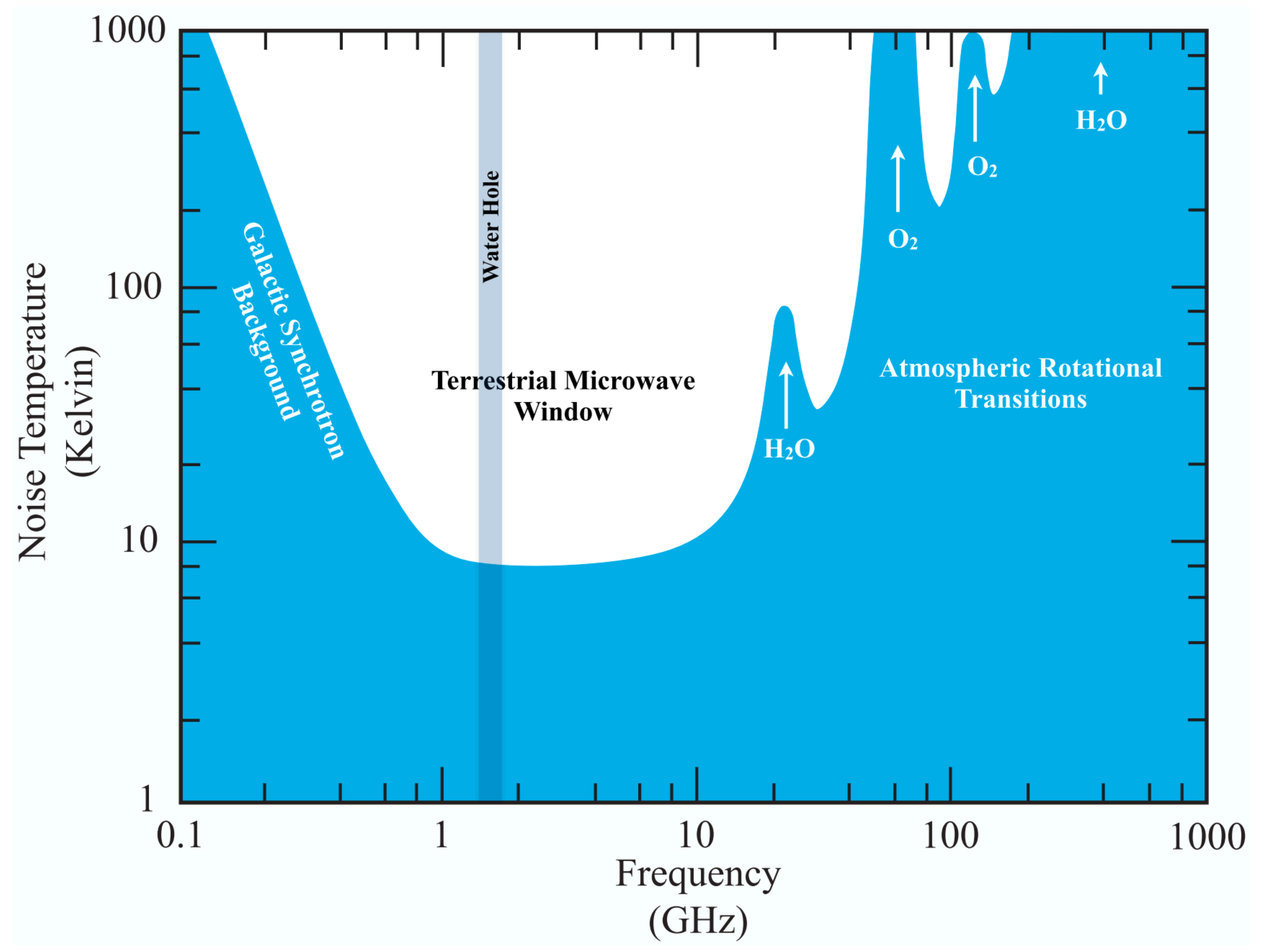


- Performance over two signal c is slightly worse than in the one signal case
- Even though we used ideal syn tic signals, our best models failed to localize to ext ely high precision
- Nevertheless, our two signal m able to localize the dimmer signal better than random, so these results are still encouraging

intensities, in pixels, for various neural network architectures in the 2 signal case.

Why radio?

- Low energy
- Low attenuation
- Produced by technology!



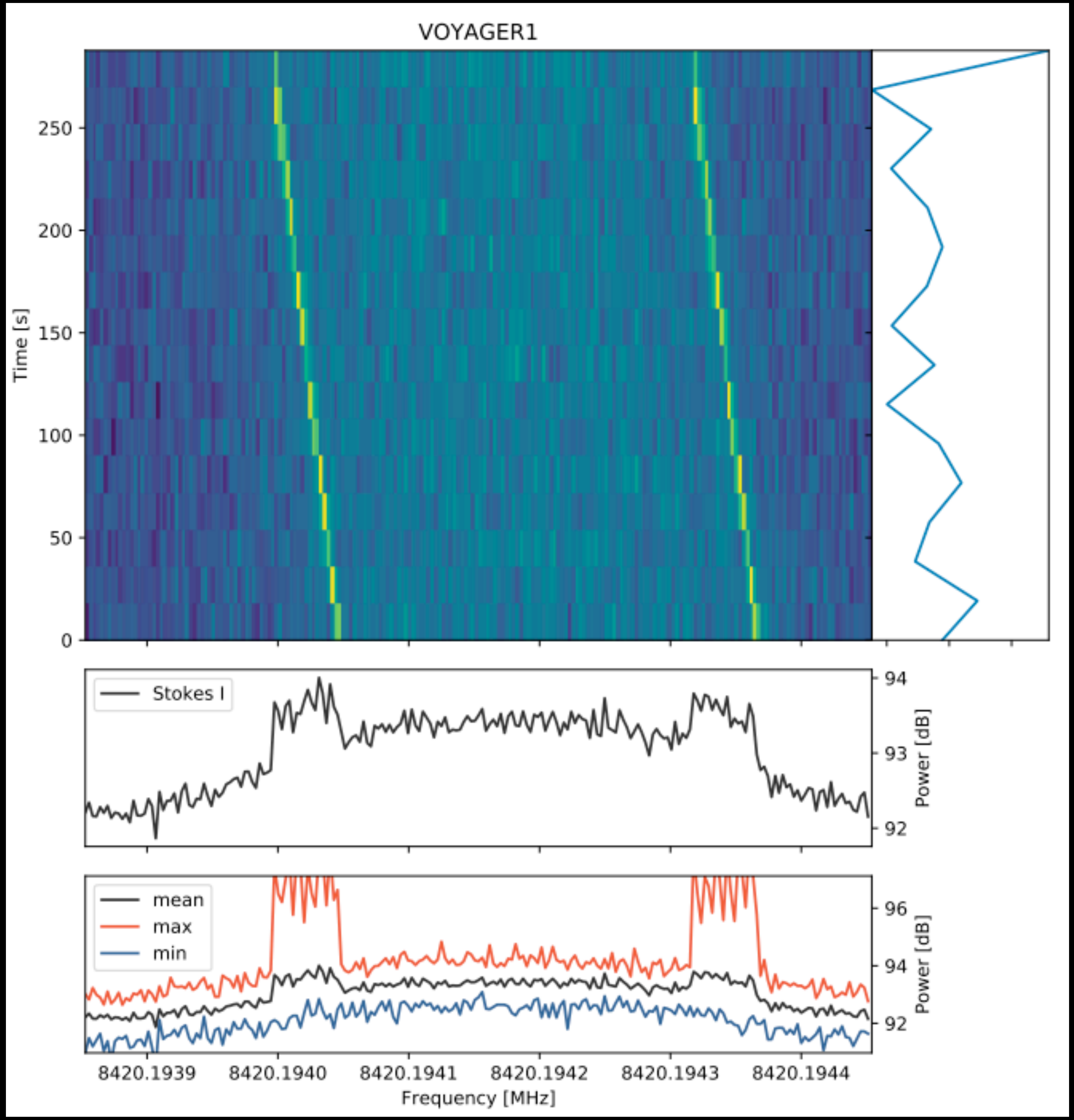
Siemion et al. 2014

Detection basis for SETI searches

- Raw signal detection
- Candidate identification and differentiation (filtering)

Raw signal detection

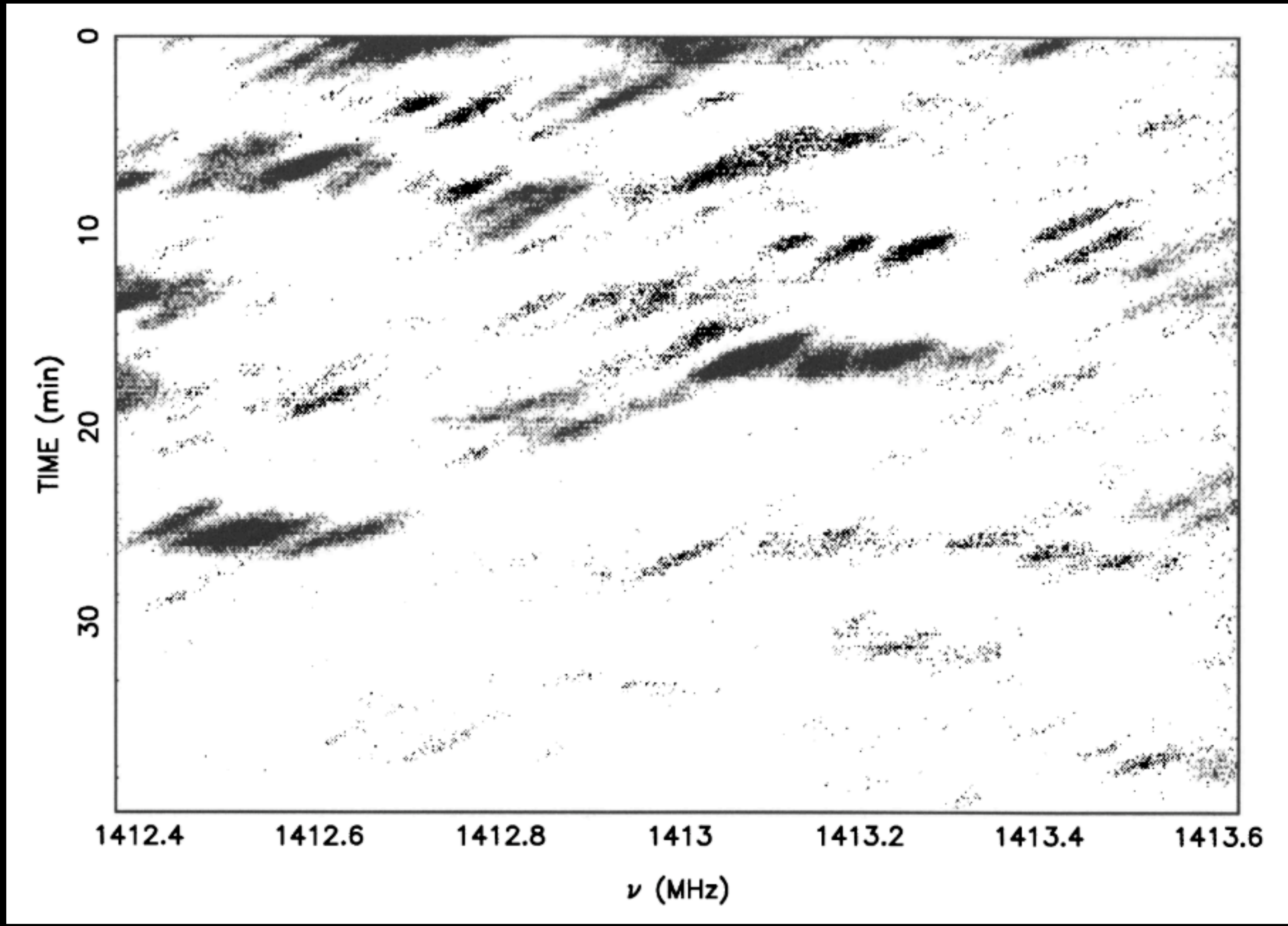
- Incoherent deDoppler (TurboSETI)
- Energy detection
- Machine learning (ML)



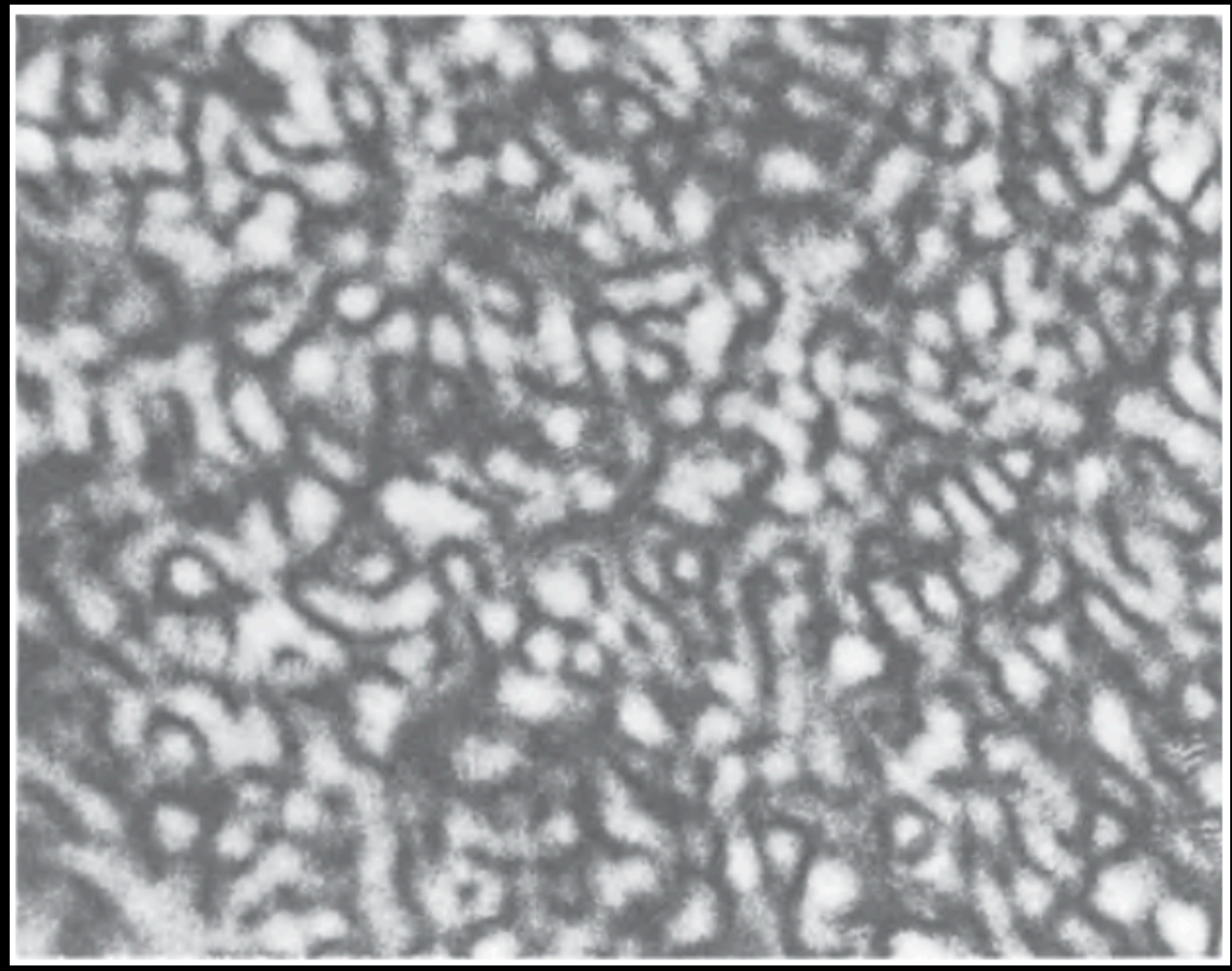
Lebofsky et al. 2019

Pulsar observations probe radio plasma effects

- Dispersion
- Scattering: scintillation and broadening
- Parallels with optical laser speckle



Cordes & Lazio 1991



Goodman 1984

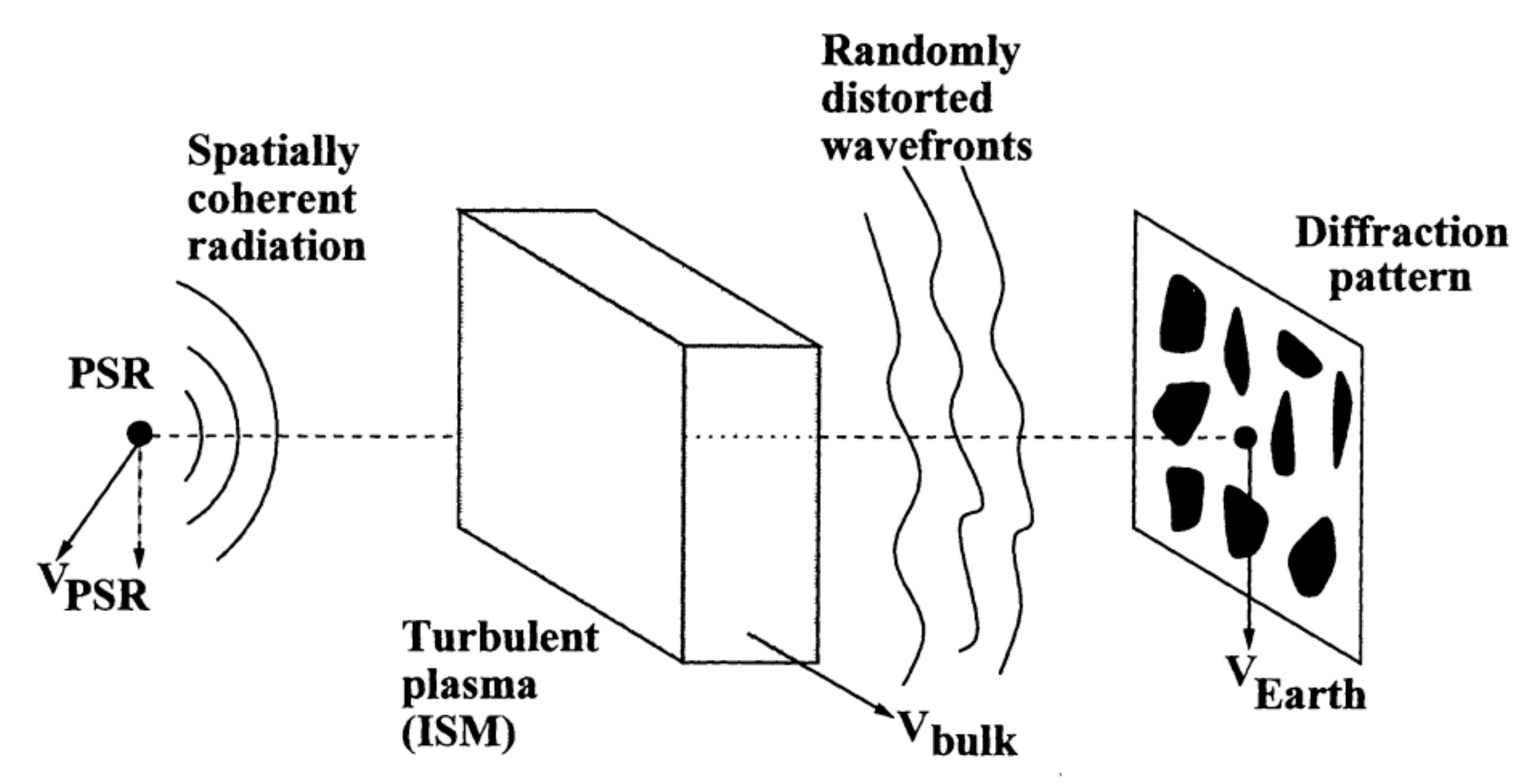
Bigger picture: research goals

- Where and how should we look to target scintillated narrowband sources? Is this feasible and worth trying?

- Develop a methodology and analysis framework for evaluating interesting signals and studies on a case-by-case basis

Diffractive scintillation in the ISM

- Electron density fluctuations in ionized plasma give rise to phase fluctuations
- Interference pattern at observer plane with characteristic spatial and spectral scales
- Can lead to 100% intensity modulation on characteristic temporal scales Δt_d, especially towards the Galactic center



Cordes 2002

Next steps: a Galactic Center / Galactic Plane survey

- Target most promising sections of parameter space

- Survey of Galactic plane with interesting targets

- Gaia DR3?