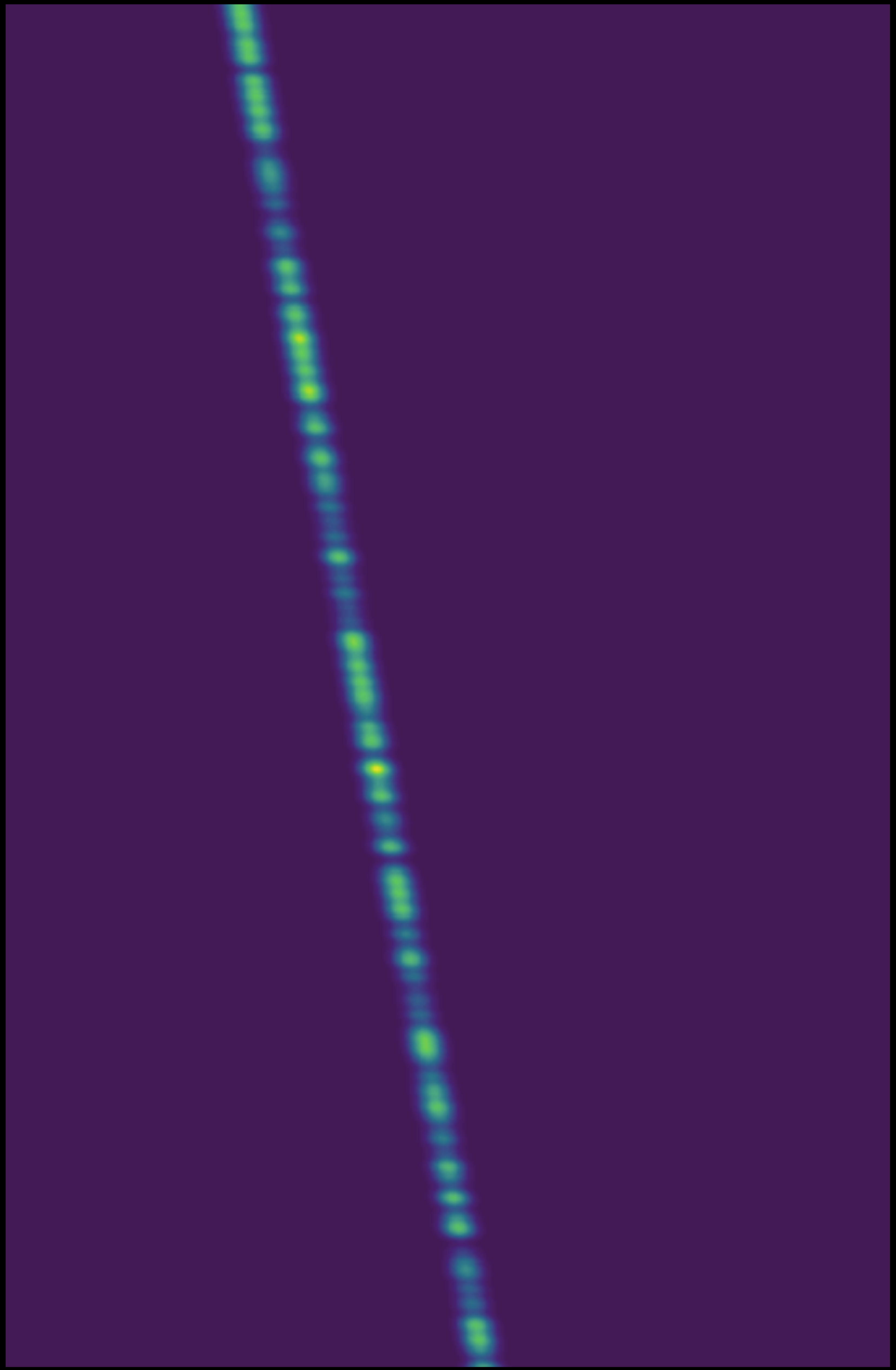
BREAKTHROUGH
LISTEN



Searching for narrow-band signals with ML

Bryan Brzycki
University of California Berkeley

Breakthrough Advisory, June 26, 2020

BERKELEY SETI
RESEARCH CENTER

Scintillation: A Potential ETI Discriminant

- Radio waves interact with the inhomogenous plasma of the ISM, resulting in scintillation and broadening
- We readily observe these effects in dynamic spectra of pulsars
- A series of papers by Jim Cordes and Joe Lazio characterized scintillation effects on narrow-band radio signals (Cordes & Lazio 1991, 2002; Cordes, Lazio, Sagan 1997)
- **We claim that ISM scintillation could be used as a novel discriminant for detecting technosignatures!**

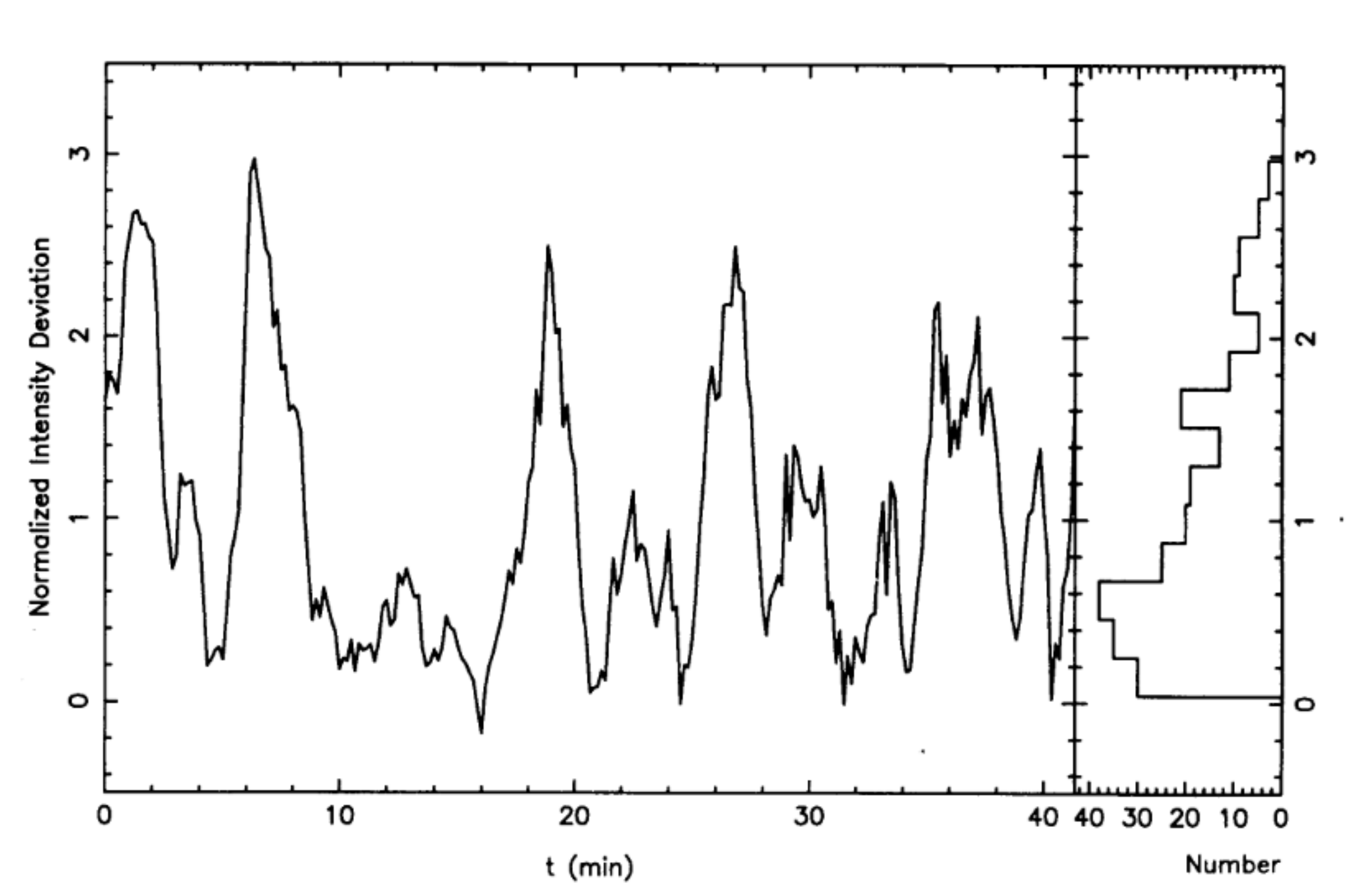


Figure 1. Left panel: Intensity vs. time for pulsar PSR 1933+16. Right panel: histogram of intensity values, showing an exponential-like distribution (Cordes & Lazio 1991).

Search Technique: Machine Learning

- We can visualize BL data as waterfall plots (spectrograms), of intensity as a function of frequency and time
- Major advances in machine learning with respect to image classification
- Computer vision techniques are good at classifying images based on morphological features
- There is a lot of potential in machine learning for identifying scintillation features!

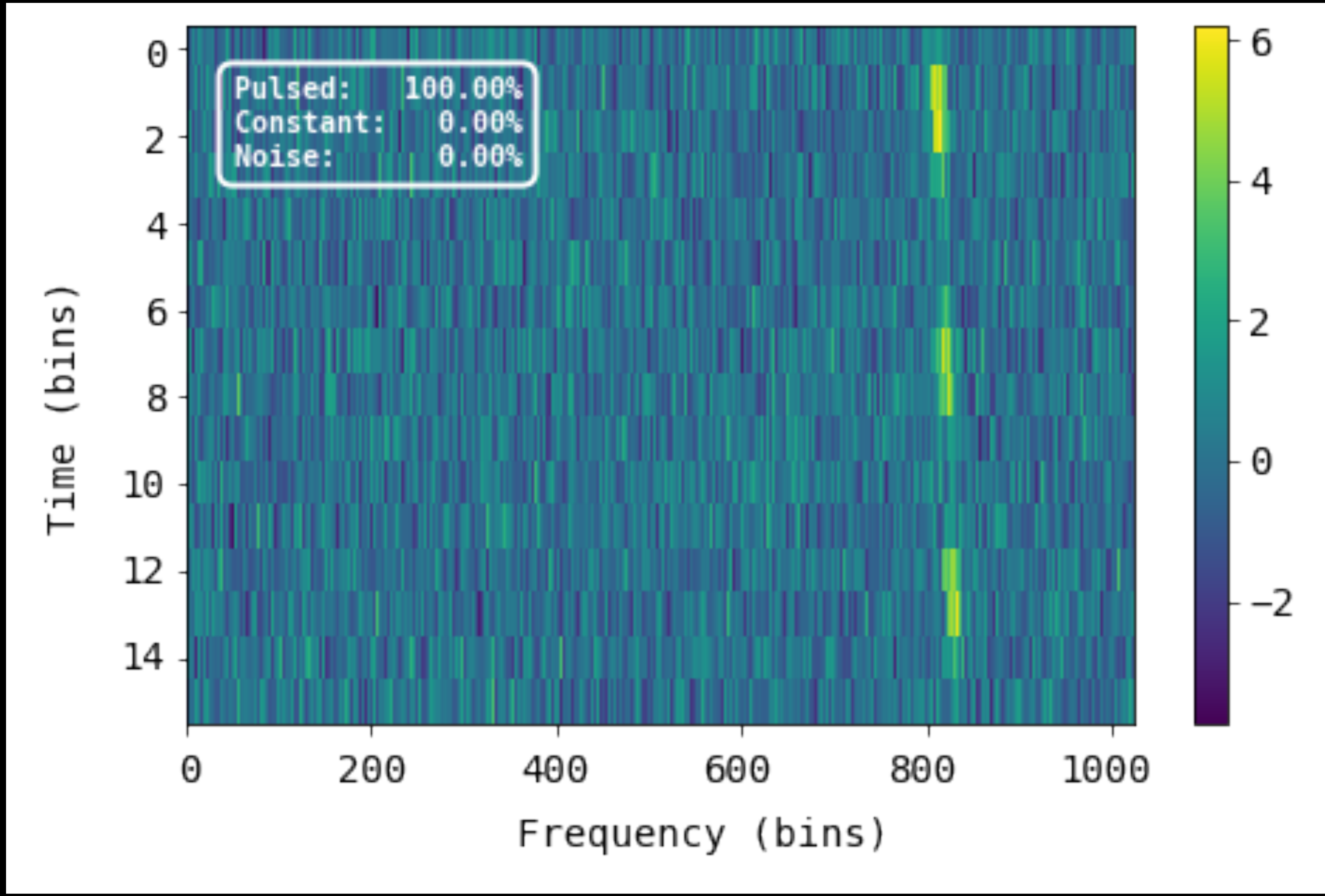


Figure 2. Simple example of ML classification (between noise, constant intensity, or pulsed) with a synthetic signal

Approaching ML for Narrow-Band Signals

- We don't have any examples of ISM-scintillated narrow-band signals, so we need to generate our own!
- Created **setigen**, a Python module that facilitates the creation of synthetic data frames of varying complexity
- Others have already used setigen for ML experiments and injection recovery for signal search pipelines!

github.com/bbrzycki/setigen

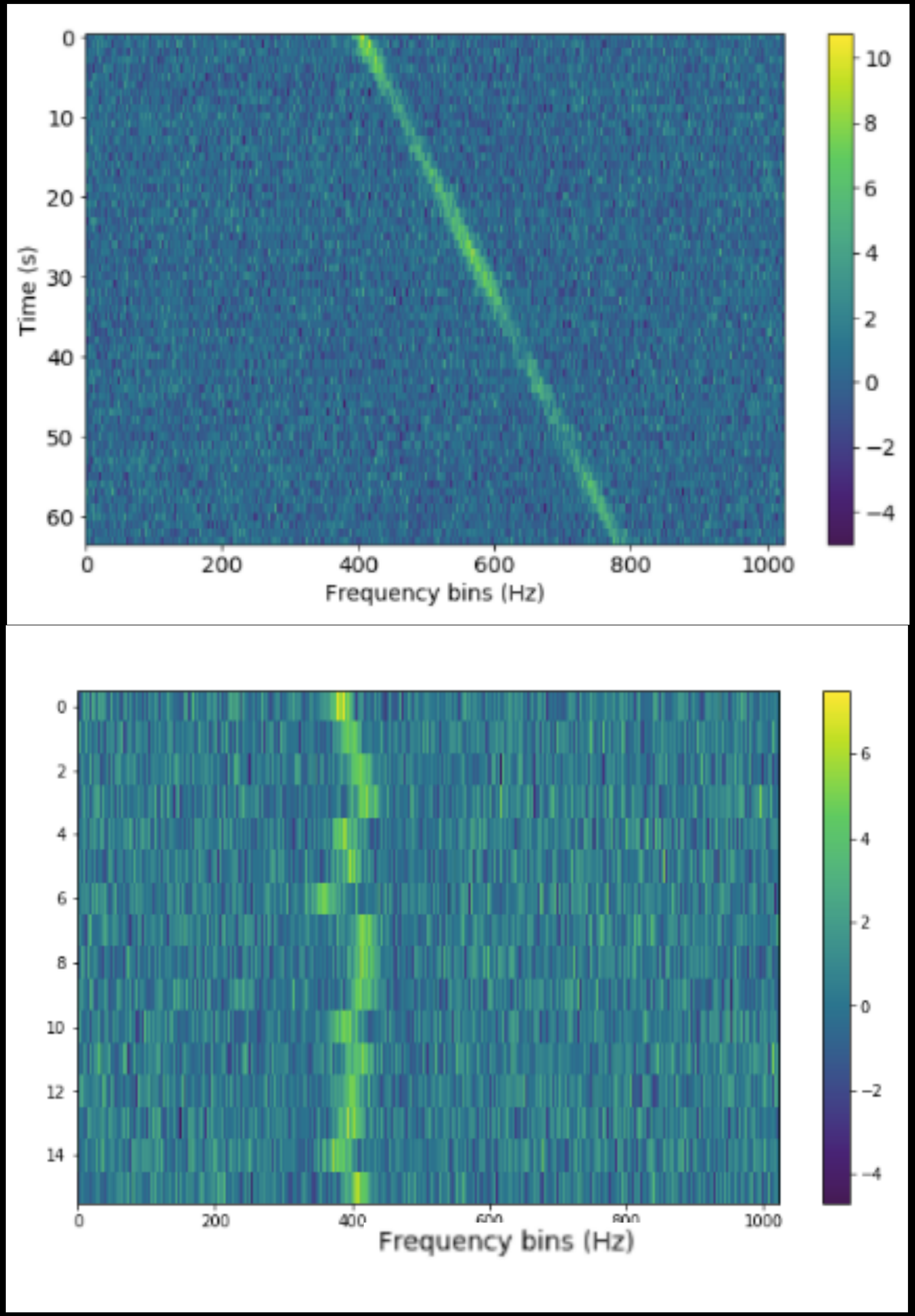


Figure 3. Top: synthetic scintillation signal. Bottom: synthetic RFI signal.

Narrow-Band Signal Localization (Brzycki et al. 2020, submitted)

- Finding signals in general data frames is important in itself
- Dedoppler search methods (such as TurboSETI) struggle to find dim signals in the presence of bright ones
- ML potentially offers a method for improving this
- Localization of narrow-band signals is a good starting problem because it's a relatively simple task; predict 2 numbers per signal

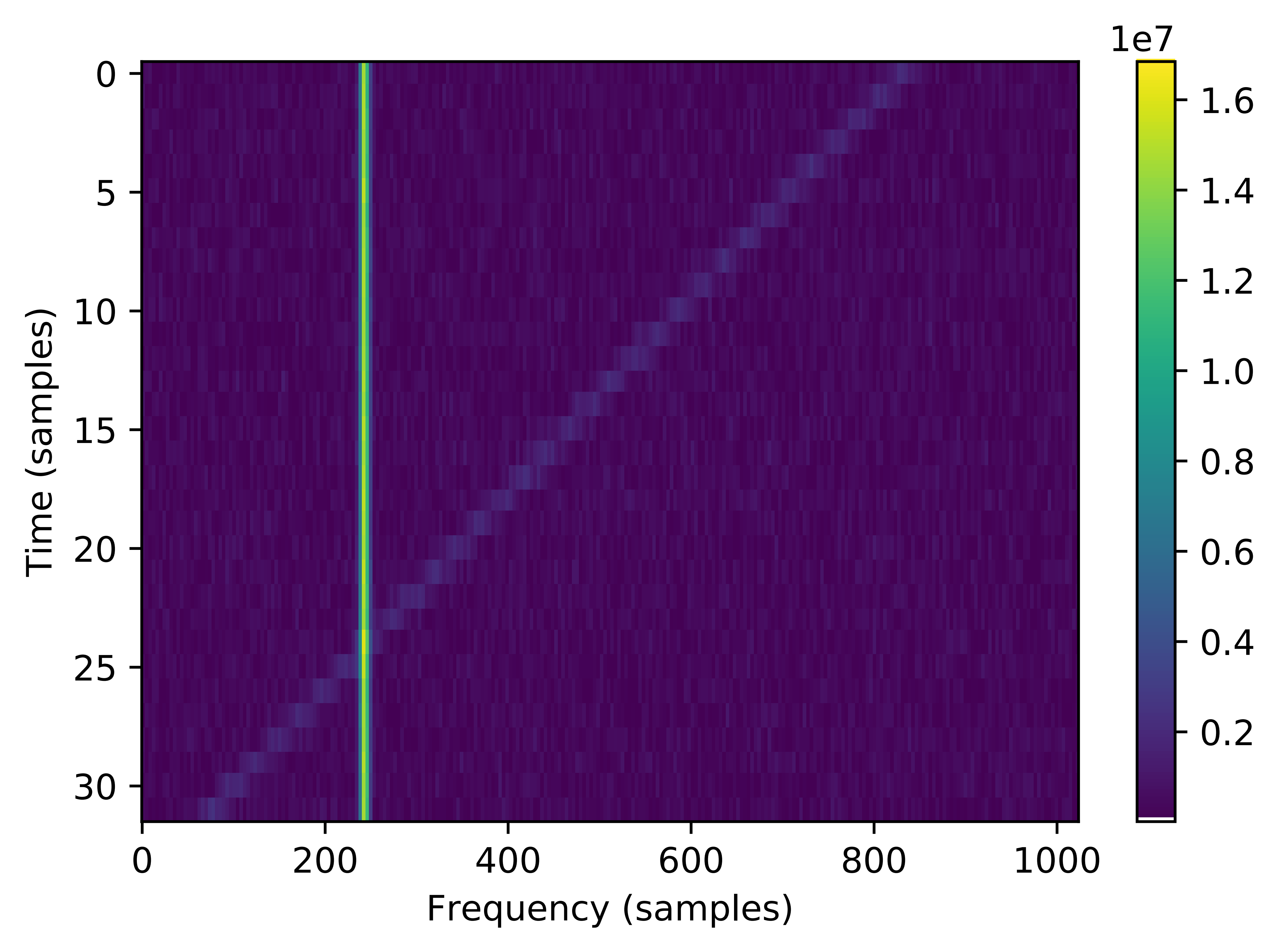


Figure 4. Example of a frame with 2 synthetic signals.

Narrow-Band Signal Localization (Brzycki et al. 2020, submitted)

- Created data frames with a primary bright “RFI” signal and a dimmer drifted signal with random slope
- Explored multiple neural network architectures
- Accuracy goes up with SNR, as expected
- Even for SNR > 100, best models were close but not pixel-perfect

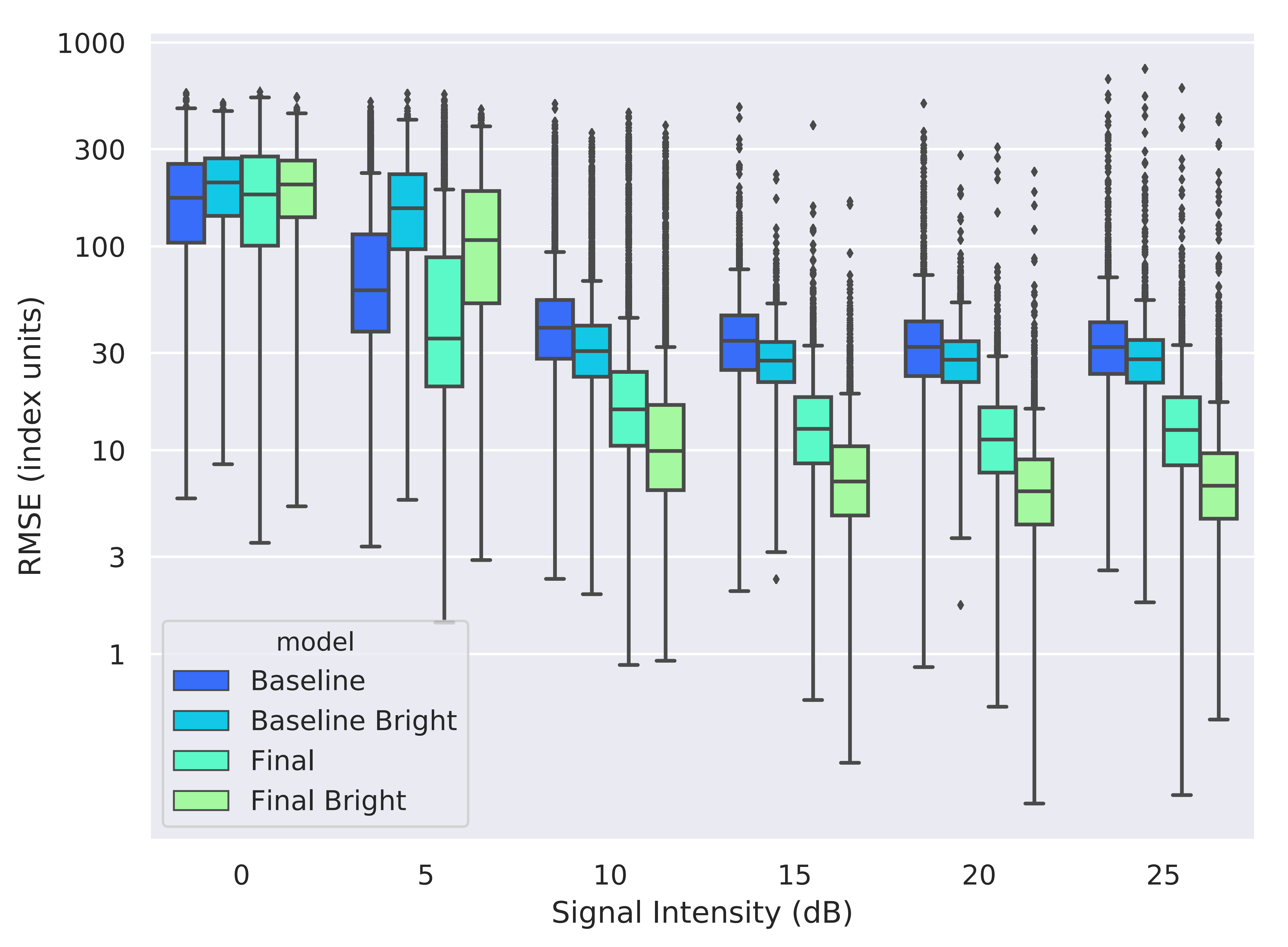


Figure 5. Mean squared error across different signal intensities, in pixels, for 2 signal case.

Back to Scintillation & Future Steps

- We have an observation plan for ML searches using scintillation as a discriminant, focused around the galactic center
- We have a procedure for generating synthetic scintillated signals, based on theory
- To do:
  - Create ML dataset with injected synthetic scintillated signals
  - Evaluate search procedure on galactic center pointings
- Compare ML localization procedure to the dedoppler search algorithm for non-ideal RFI signals

Thank you!

Thank you!

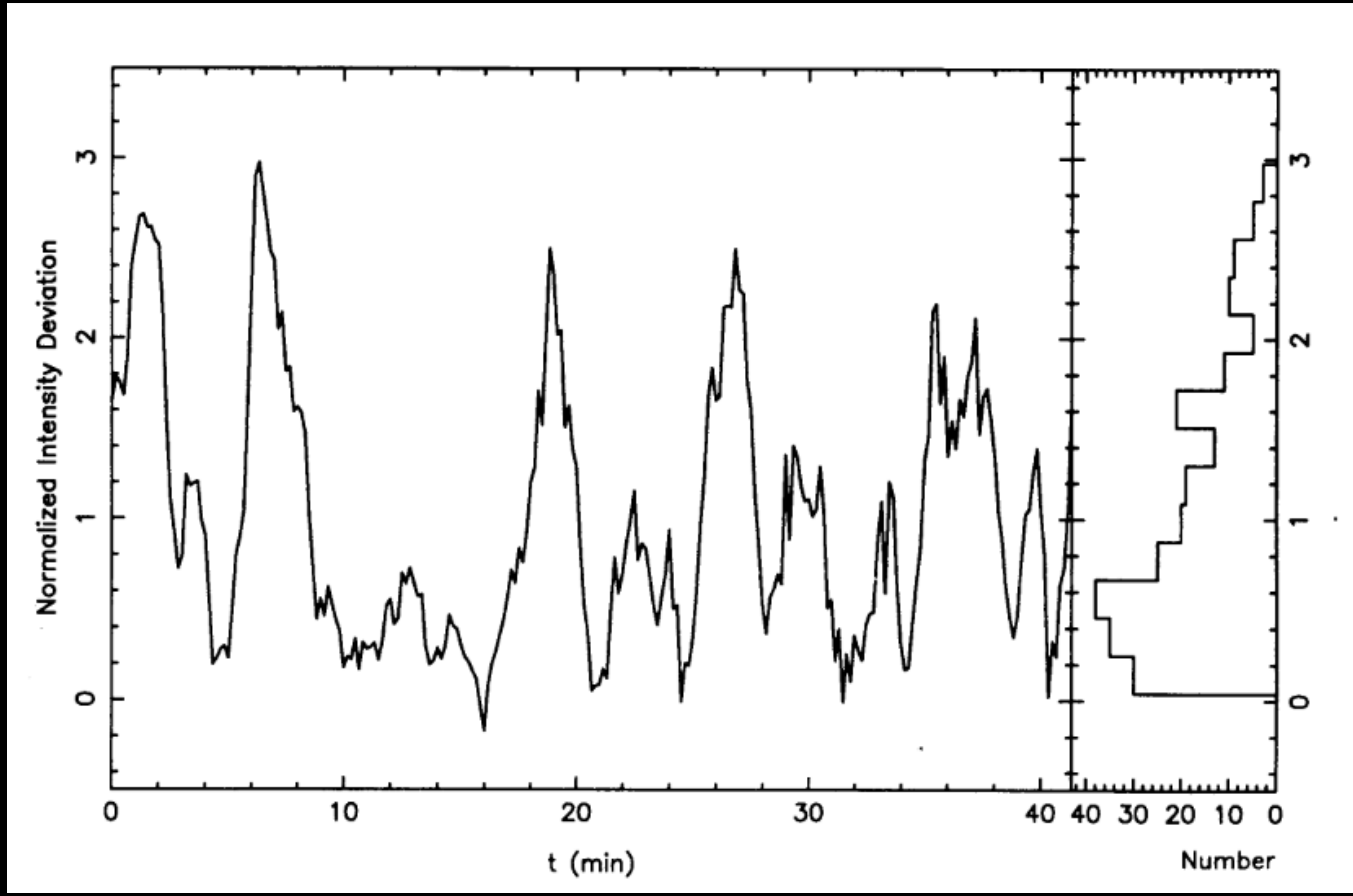
Acknowledgements

- My advisor, Andrew Siemion
- Conversations with Jim Cordes & Carl Heiles
- The entire BSRC team
- Breakthrough Listen

- Cordes, J. M. & Lazio, T. J. 1991, ApJ
- Cordes, J. M. & Lazio, T. J., Sagan, C. 1997, ApJ
- Cordes, J. M. & Lazio, T. J. W. 2002, arXiv, astro-ph
- Zhang et al. 2018, ApJ, submitted

Autoregressive to Anything (ARTA)

- We need a method for generating synthetic data that satisfy these constraints:
  - Exponential intensity distribution
  - Gaussian autocorrelation, with FWHM = scintillation timescale
- Autoregressive to Anything (ARTA) is a method for generating "stationary time series [data] with arbitrary marginal distributions and autocorrelation structures" (Cario & Nelson 1996).



Cordes & Lazio 1991

Autoregressive to Anything (ARTA)

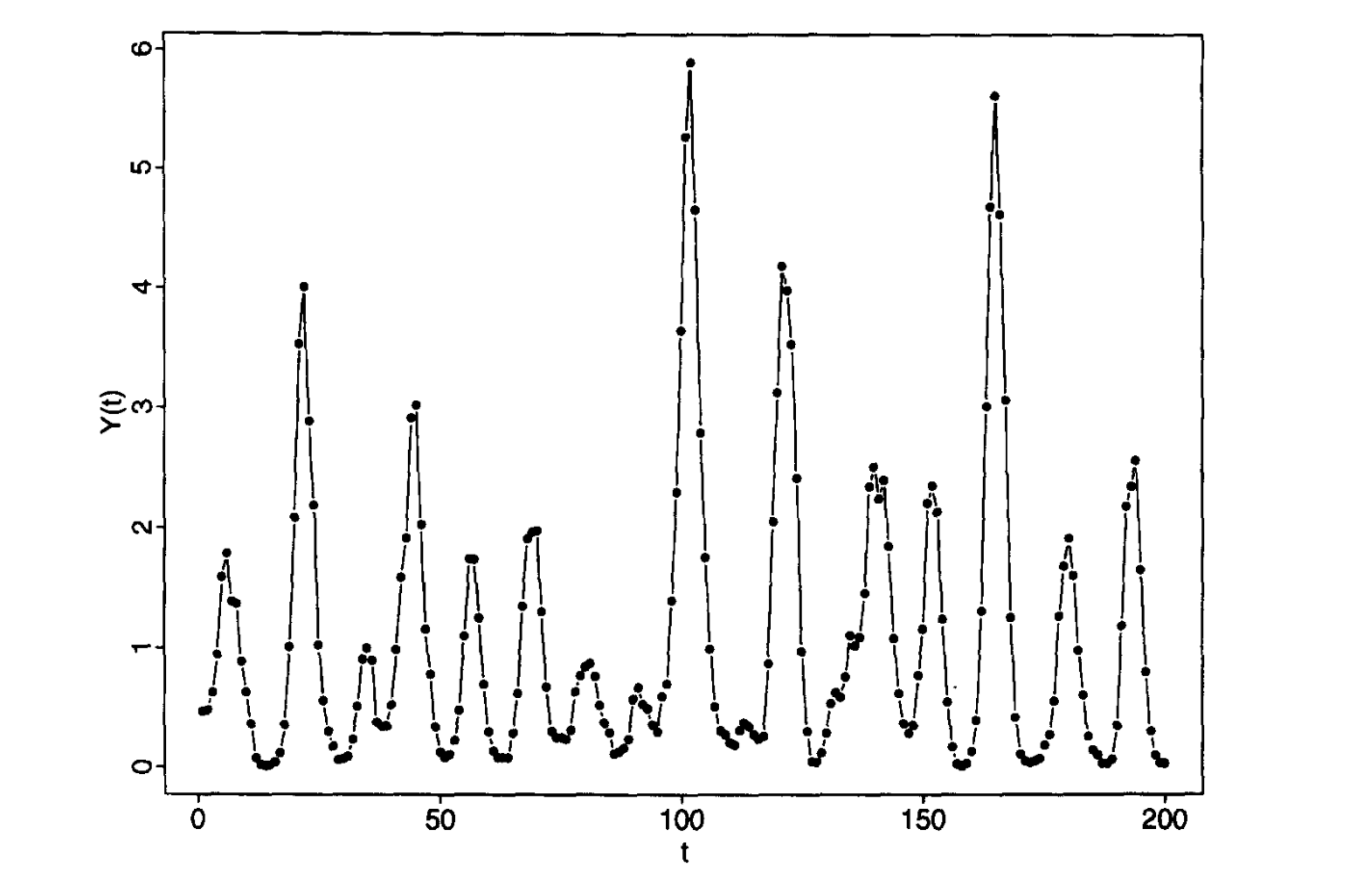
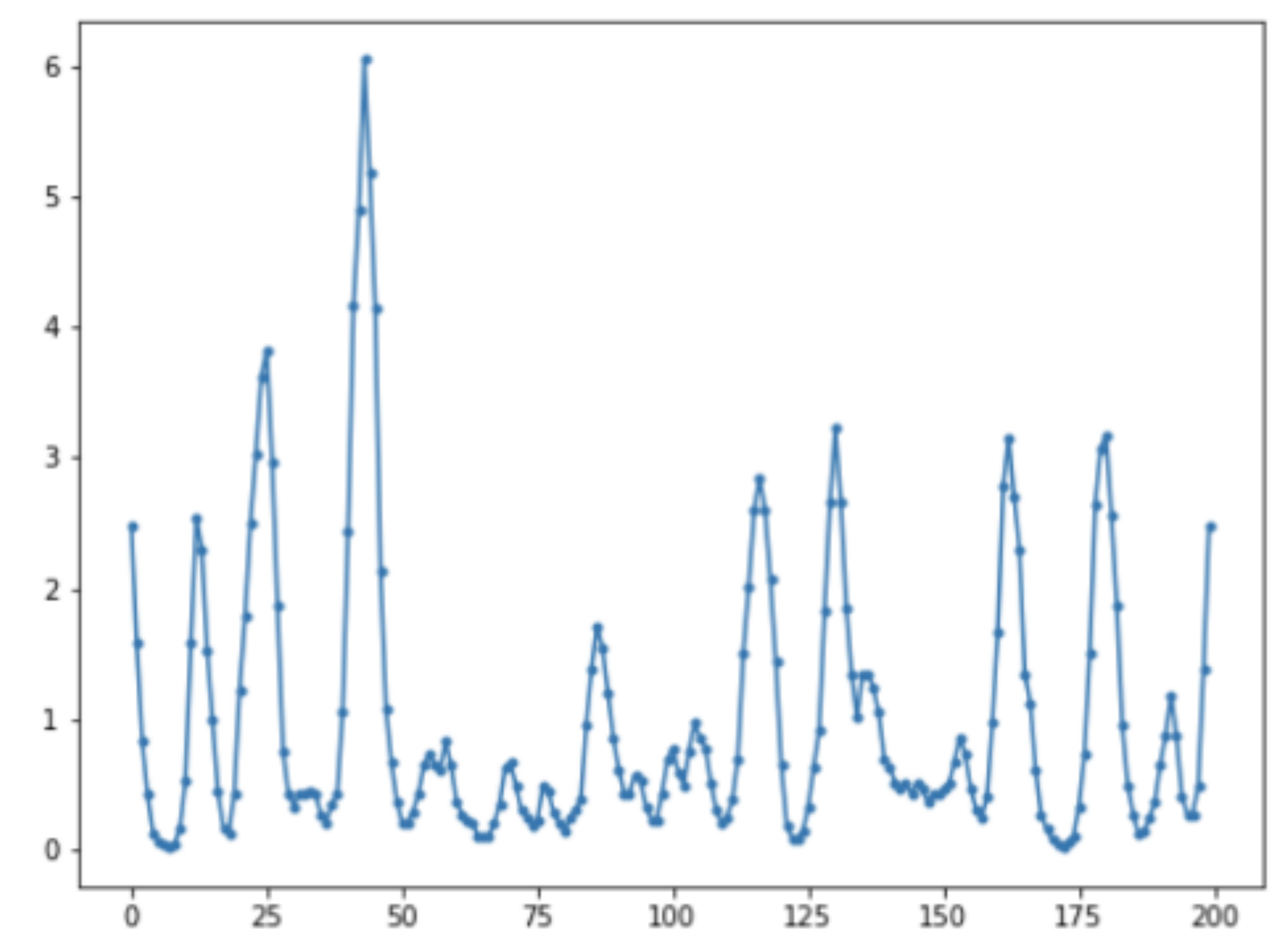


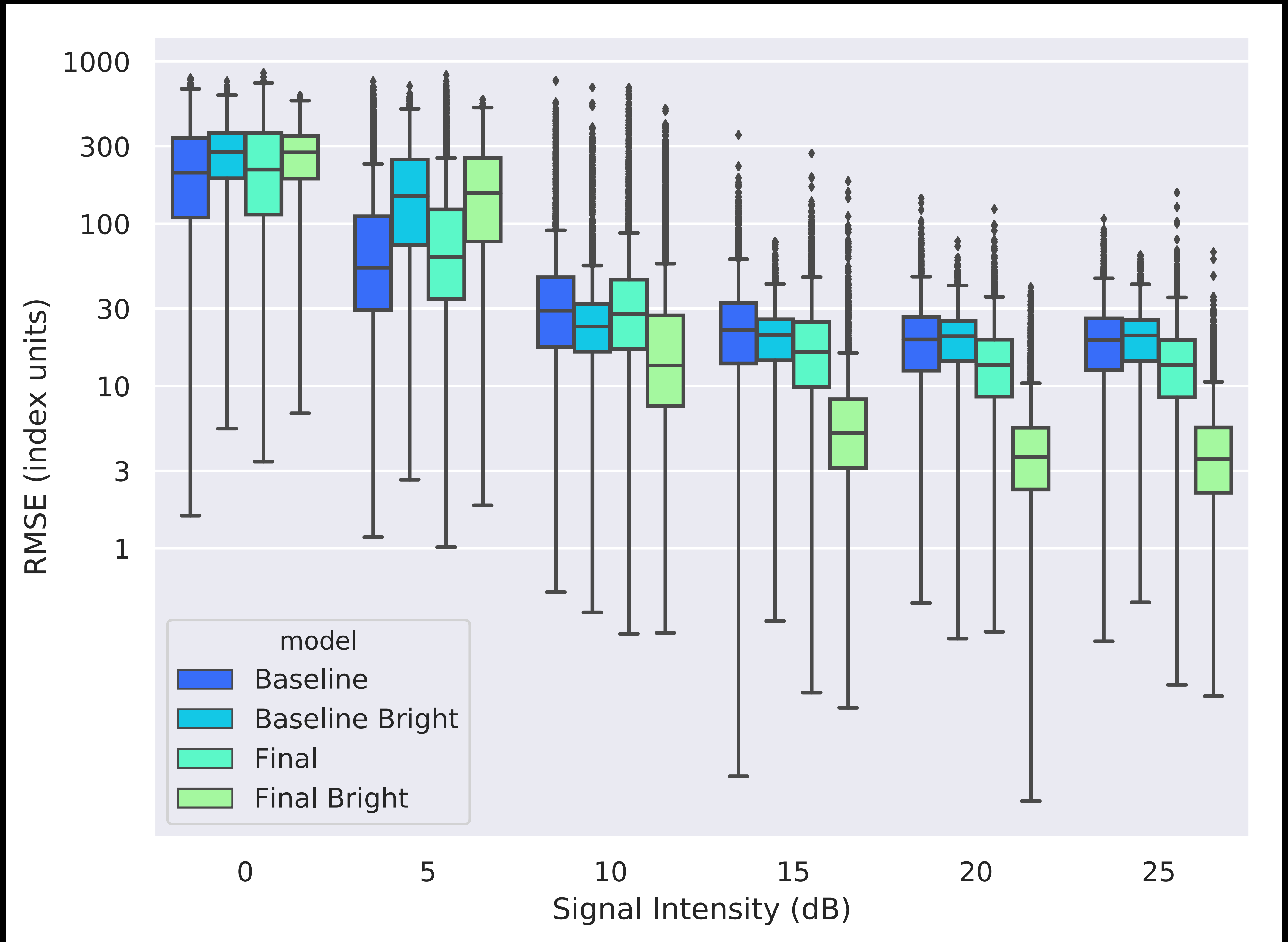
Fig. 1. Sample path of an ARTA process with exponential marginals and autocorrelations $\rho_1 = 0.9$ and $\rho_2 = 0.6$.



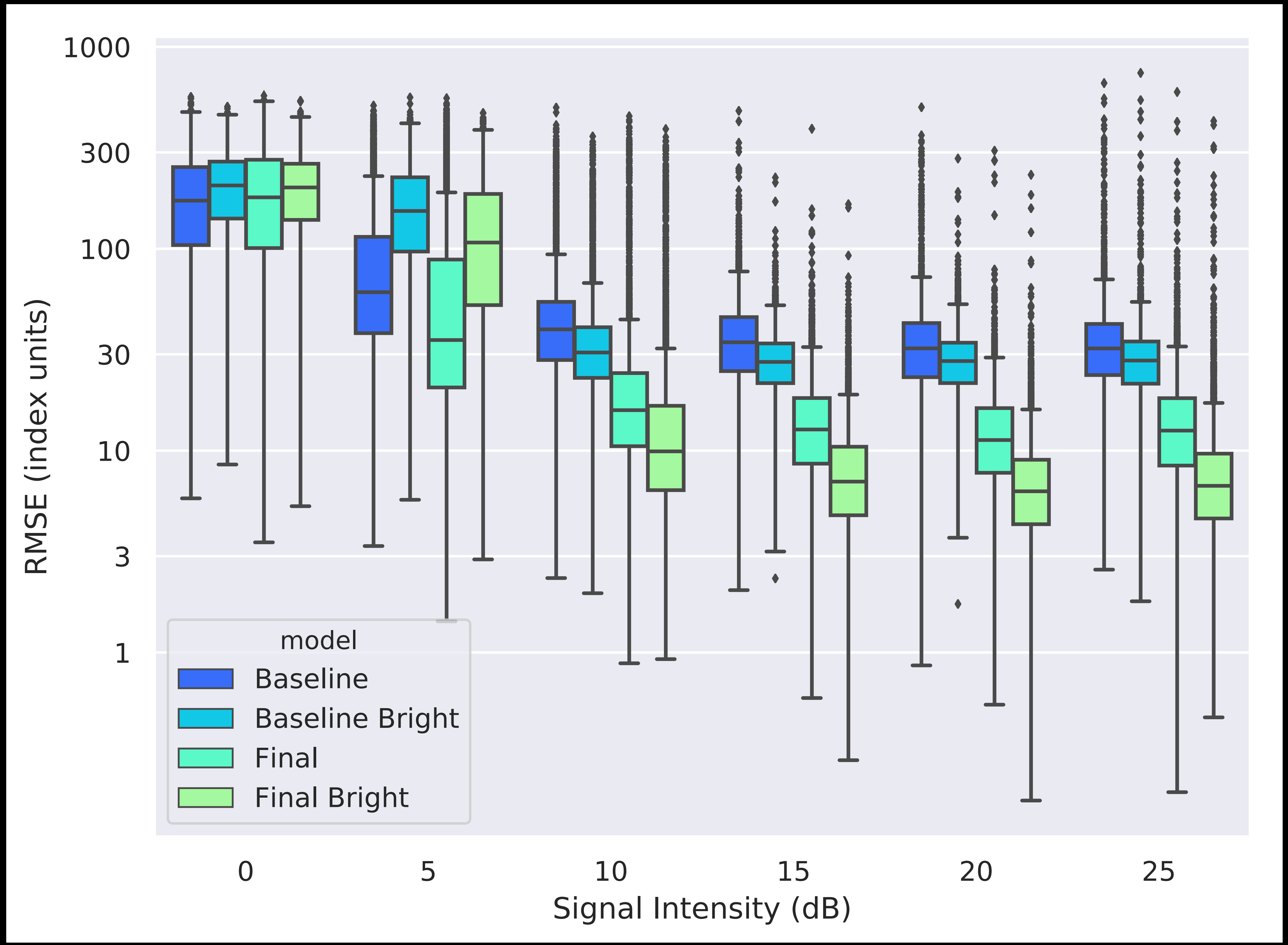
Cario & Nelson 1996

Our implementation

Signal Localization in Spectrograms



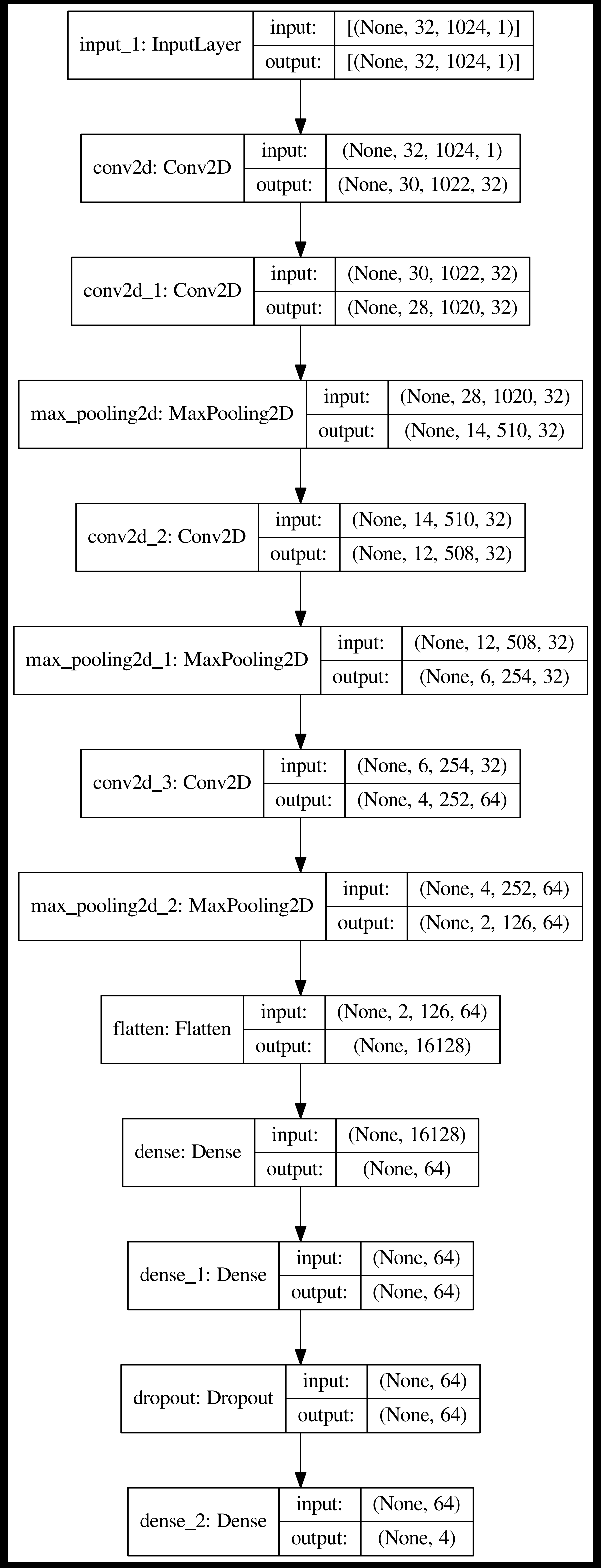
Mean squared error across different signal intensities, in pixels, for 1 signal case (Brzycki et al. 2020, submitted)

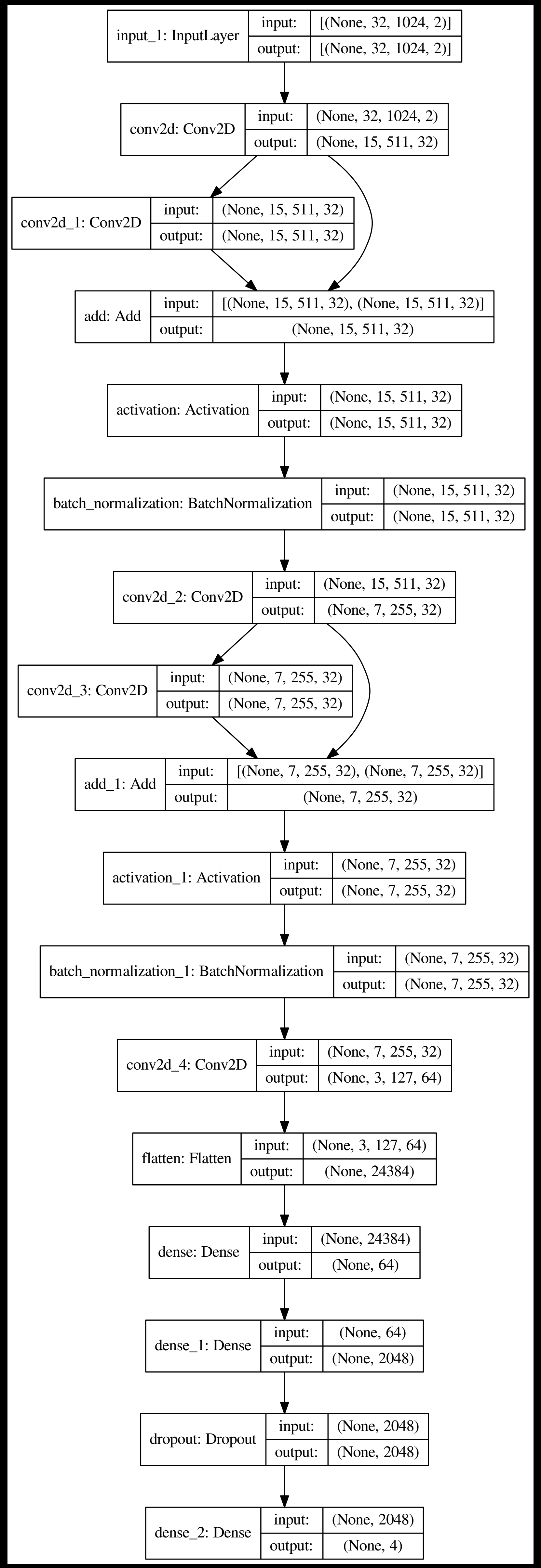


Mean squared error across different signal intensities, in pixels, for 2 signal case (Brzycki et al. 2020, submitted)

Localization Models

Baseline





Final