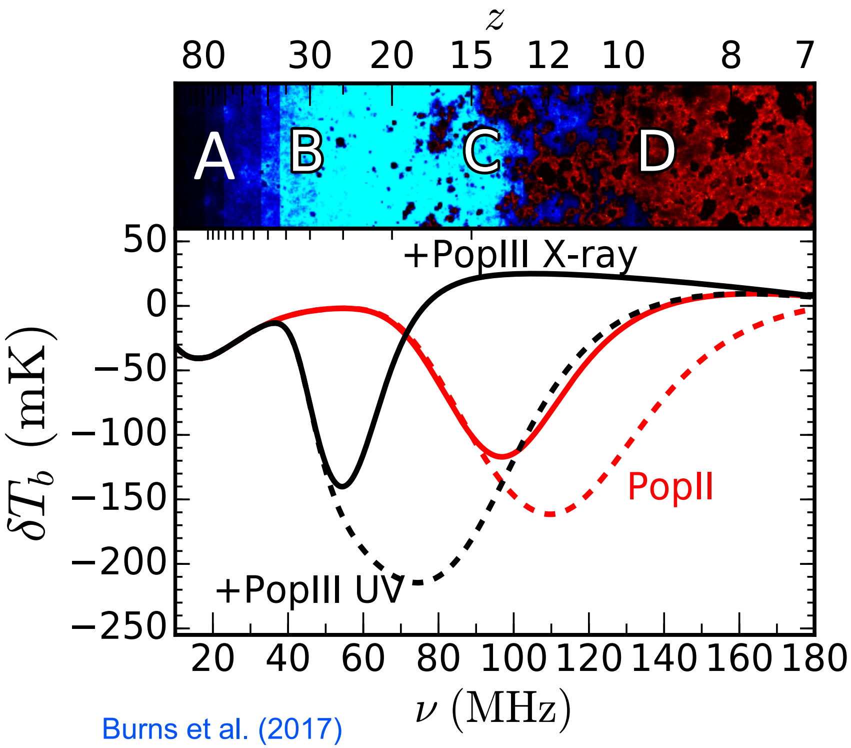
Full Data Analysis Pipeline for Low Radio Frequency Measurements of the Dark Ages and Cosmic Dawn

David Rapetti
University of Colorado Boulder / NASA Ames Research Center

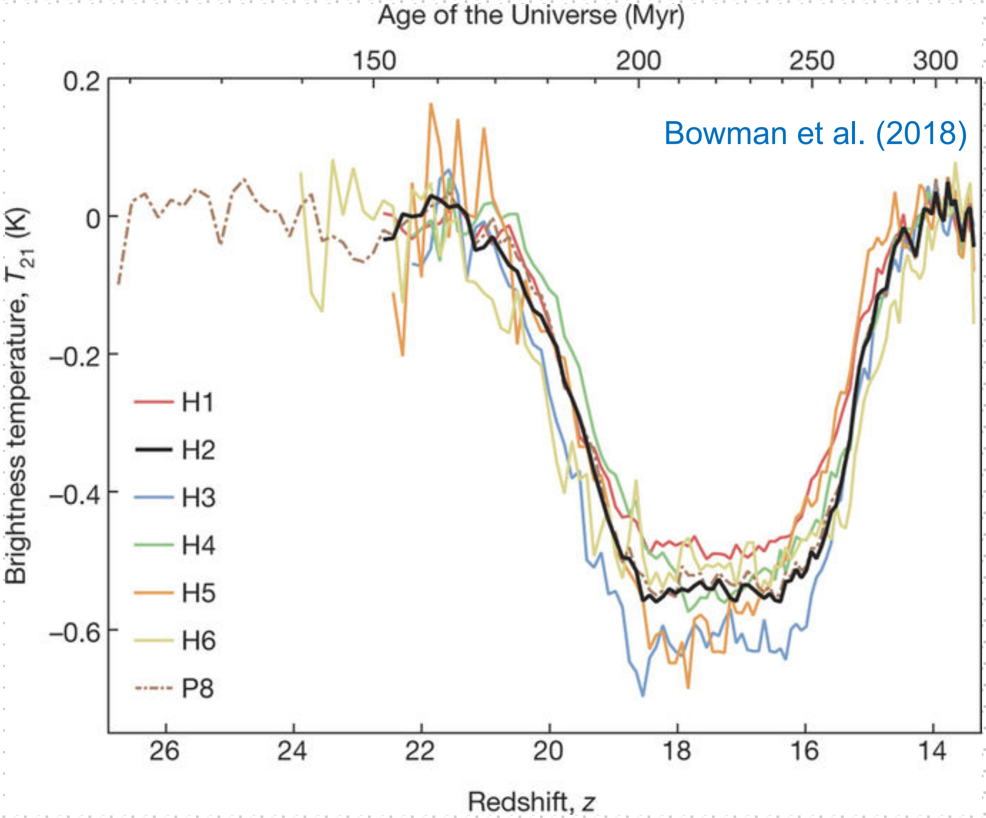
The work presented is in collaboration with:
Keith Tauscher (CU Boulder), Jack O. Burns (CU Boulder), Jordan Mirocha (McGill U.), and Eric Switzer (NASA Goddard), Neil Bassett (CU Boulder), Julian Merten (INAF-Bologna/Oxford)

NESS
Network for Exploration
and Space Science

HYDROGEN COSMOLOGY AND FIRST OBSERVATIONAL MEASUREMENTS

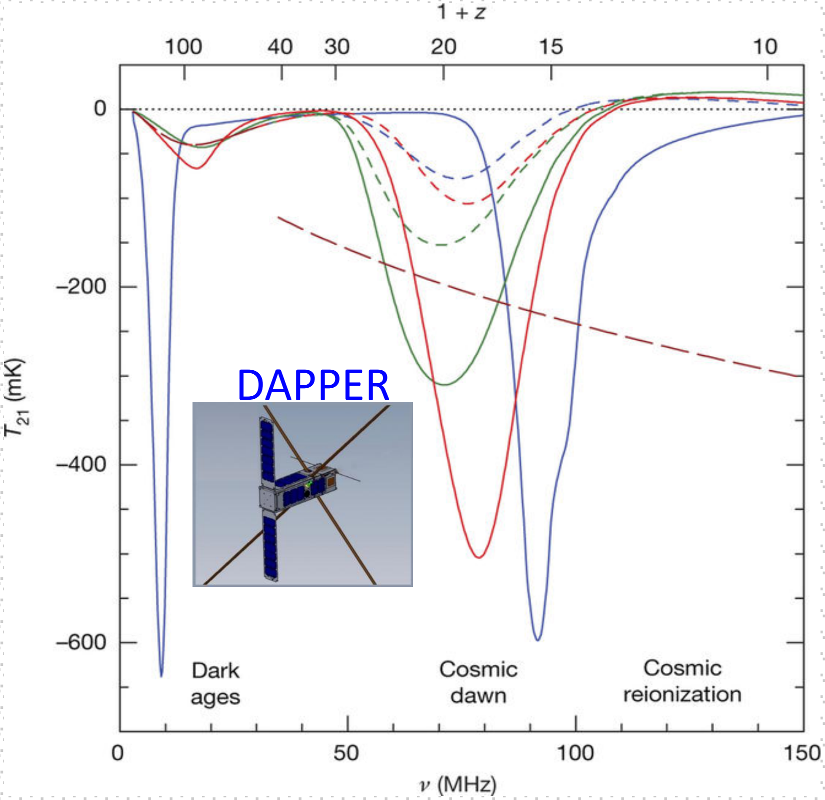


Burns et al. (2017)



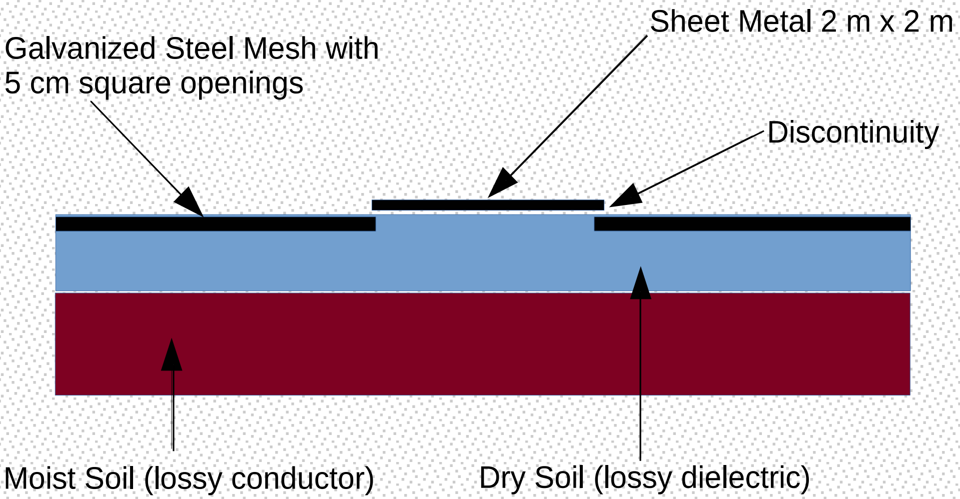
- Left: δT_b is a combination of temperatures: T_S spin, T_k kinetic, T_α Lyman-α, T_γ background (CMB).
- A: Expansion recouples $T_S \rightarrow T_\gamma$; B: First stars Ly-α emission couples back $T_S \rightarrow T_k$; C: Heating sources including initial black hole accretion drive $T_k \rightarrow T_\gamma$; D: Reionization onset removes signal ($x_{HI} \rightarrow 0$).
- Right: EDGES measured a 78 MHz absorption profile at a frequency consistent with those expected for a Cosmic Dawn signal in the global 21-cm spectrum.

FIRST THEORETICAL EXPLANATION AND POSSIBLE GROUND PLANE ARTIFACT



Barkana, Nature 555, 71 (2018)

Theory for larger than standard amplitude: scattering between baryons and dark matter (solid curves) and same models without (short-dashed curves); the brown long-dashed curve is the adiabatic limit.

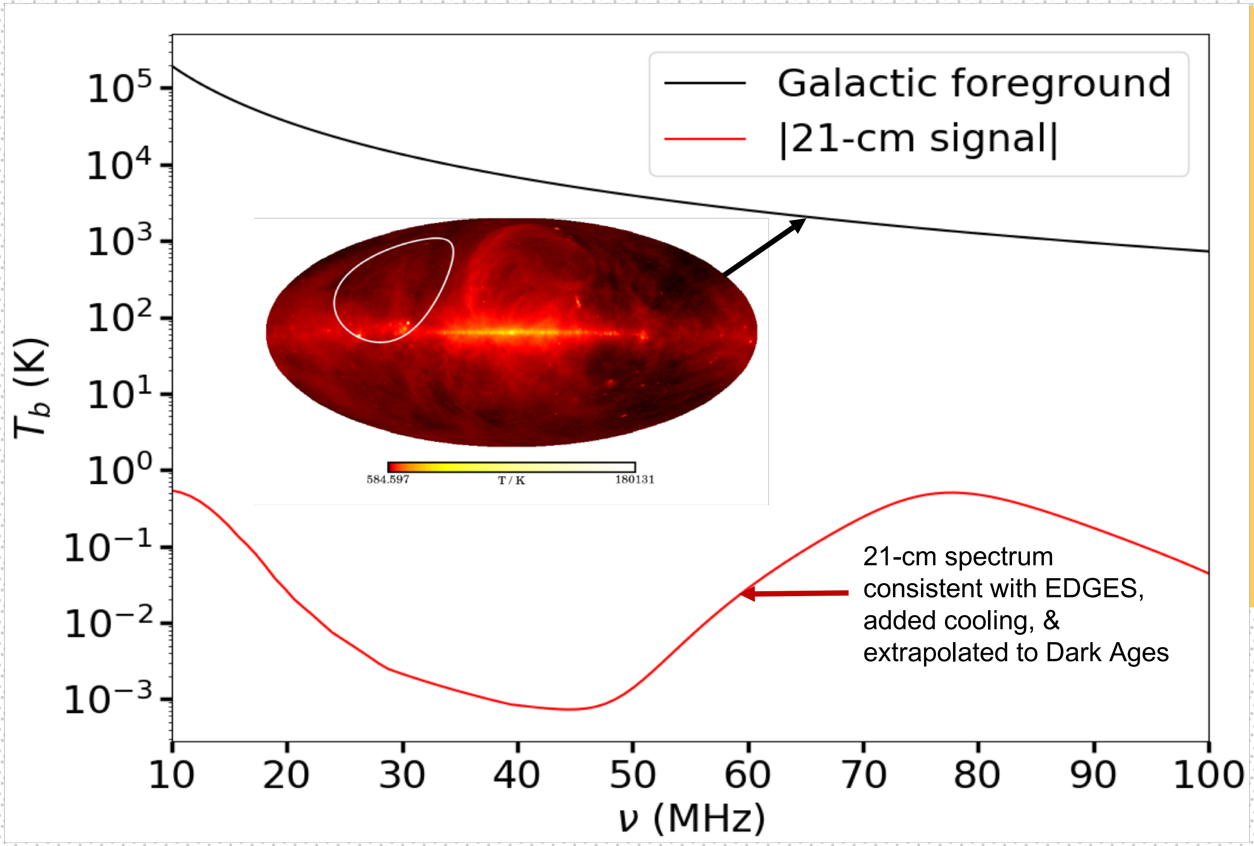


A sketch of the ground plane region of the EDGES instrument showing the soil layers and ground discontinuity between the sheet metal and steel mesh that can lead to leakage of power.

(See further details in K. Tauscher presentation this afternoon)

Bradley, Tauscher, Rapetti & Burns, arXiv:1810.09015

CHALLENGES OF GLOBAL 21-CM OBSERVATIONS



21-cm spectrum consistent with EDGES, added cooling, & extrapolated to Dark Ages

$$T_{ant}(\nu) = \frac{\int_0^{2\pi}\int_0^{\pi/2} T_{sky}(\nu,\theta,\phi)F(\theta,\phi,\nu)\sin\theta \,d\theta d\phi}{\int_0^{2\pi}\int_0^{\pi/2} F(\theta,\phi,\nu)\sin\theta \,d\theta d\phi}$$

Foreground Characteristics

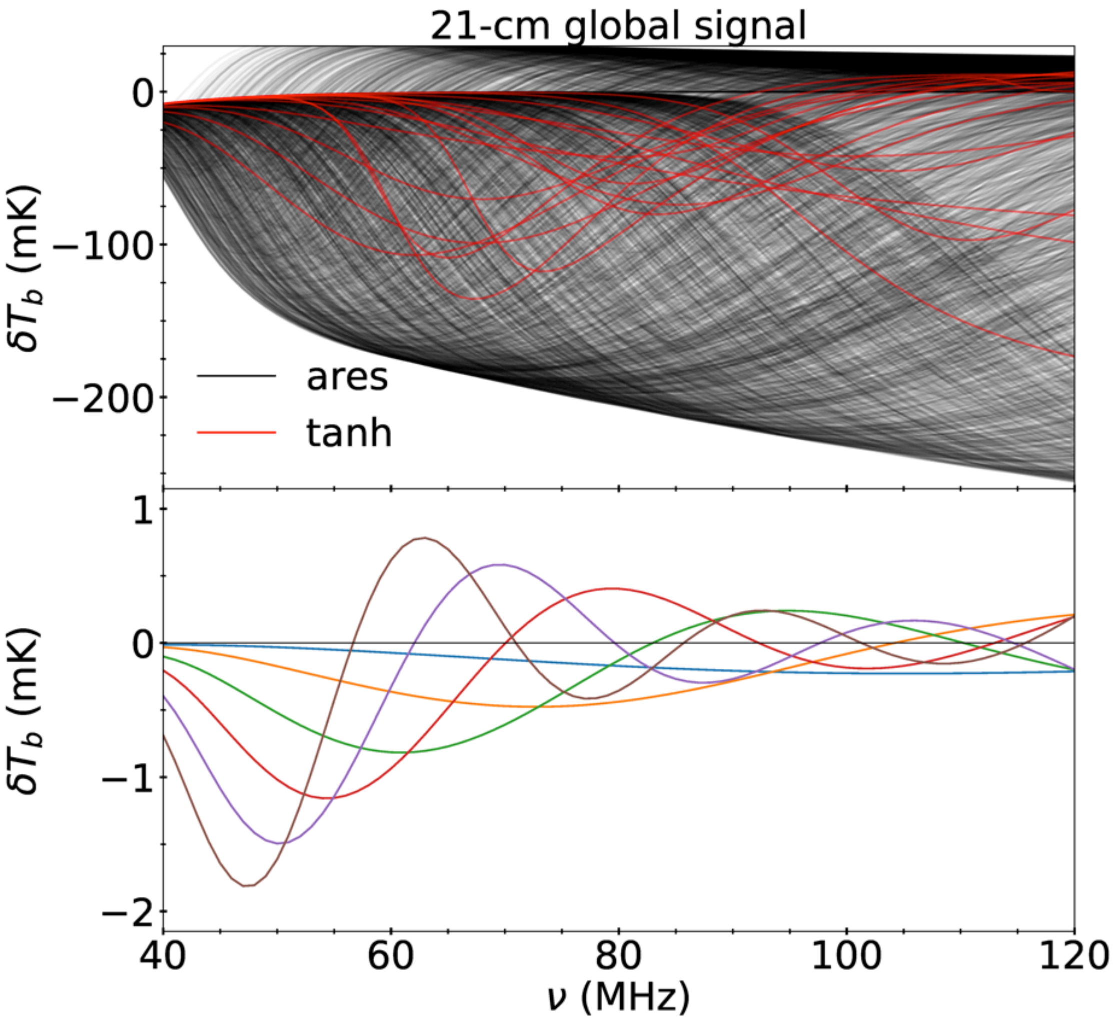
- Spectrally smooth
- Spatial structure
- Polarized

Signal Characteristics

- Spectral structure
- Spatially isotropic
- Unpolarized

Weighting by antenna beam introduces spectral structure in foreground
(e.g., Bernardi *et al.* 2015, Mozdzen *et al.* 2016)

GLOBAL 21-CM SIGNAL USING PATTERN RECOGNITION AND TRAINING SETS



Tauscher, Rapetti, Burns, Switzer (2018)

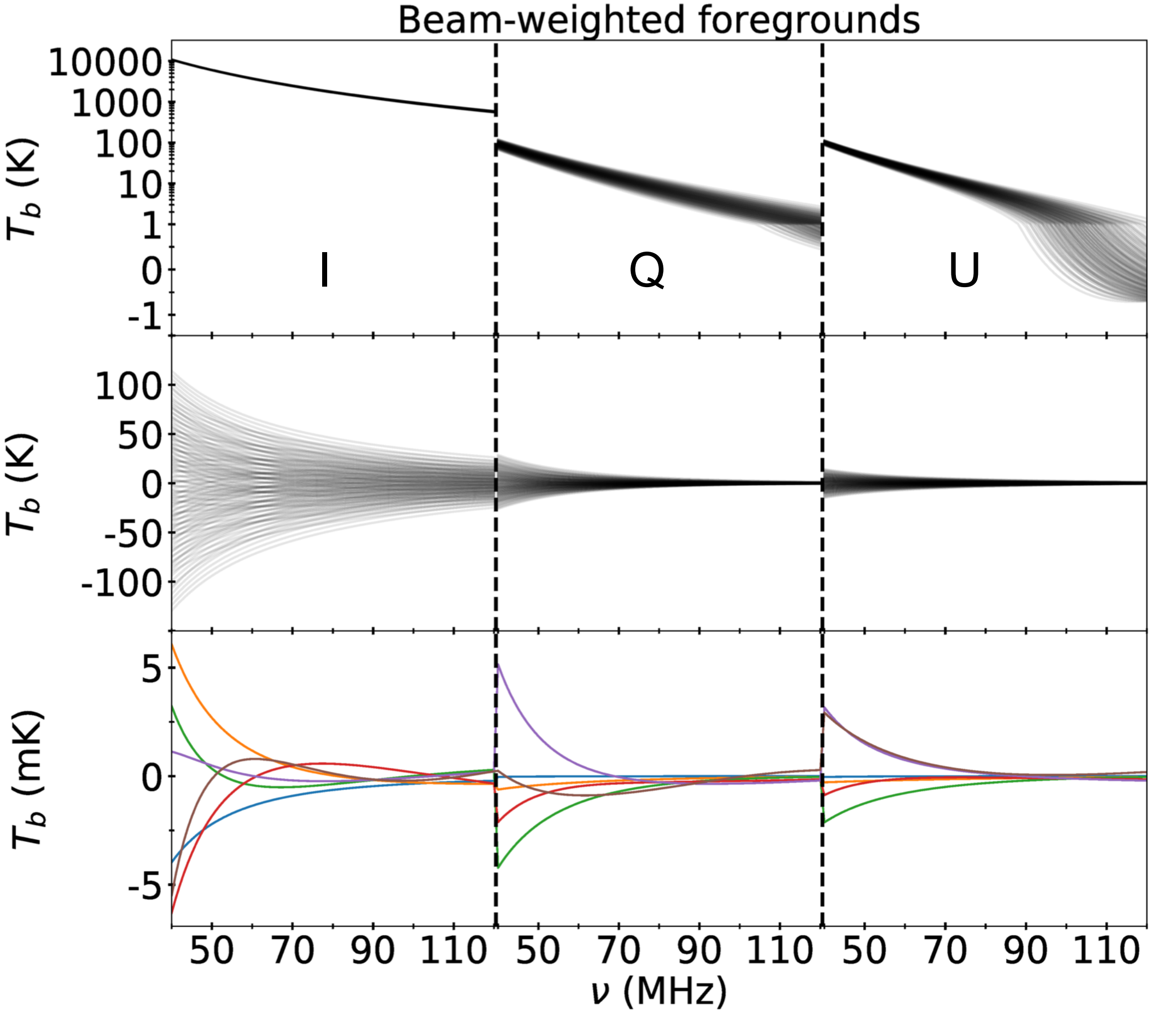
- Signal training set generated by running the *ares* code 7×10^5 times within reasonable parameter bounds.
- The top panel shows a thinned sample of that set (black curves).

$$\underbrace{M}_{\substack{\text{Training Set:}\\(N_{channel}\\ \times N_{curves})}} = \underbrace{U}_{\substack{\text{Ordered basis}\\ \text{functions:}\\(N_{channel}\\ \times N_{channel})}}\Sigma V^T$$

- The SVD modes are ordered from most to least important.

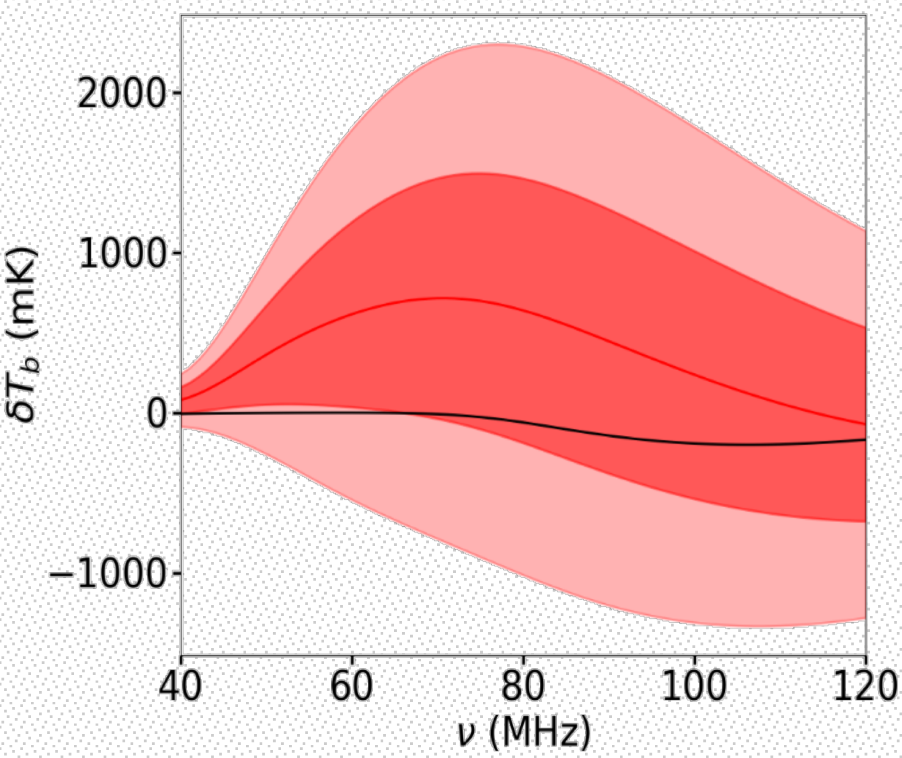
EXPERIMENTAL DESIGN: INCLUDING STOKES PARAMETERS INTO THE LIKELIHOOD FUNCTION

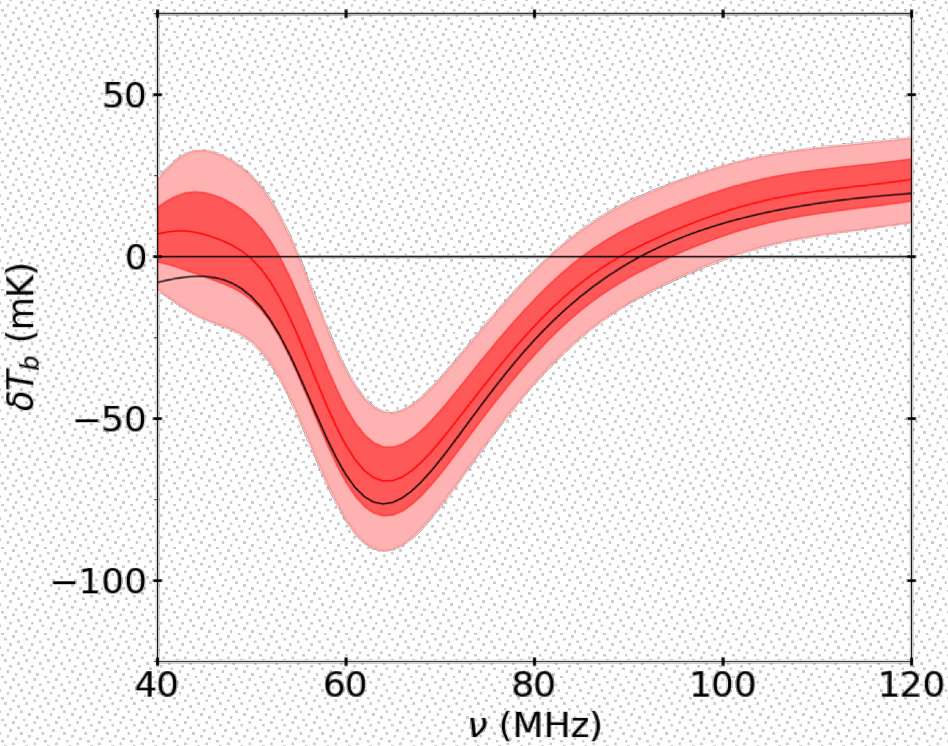
Tauscher, Rapetti, Burns, Switzer (2018)



- Top: Beam(Gaussian)-weighted foreground training set for a single rotation angle about one of the 4 antenna pointing directions.
- Middle: The same training set with its mean subtracted.
- Bottom: The first 6 SVD basis functions obtained from the training set.
- The different rotation angles about each antenna pointing direction are part of the same training set so that SVD can pick up on angle-dependent structure and imprint it onto the basis functions.

IMPORTANCE OF USING POLARIZATION DATA

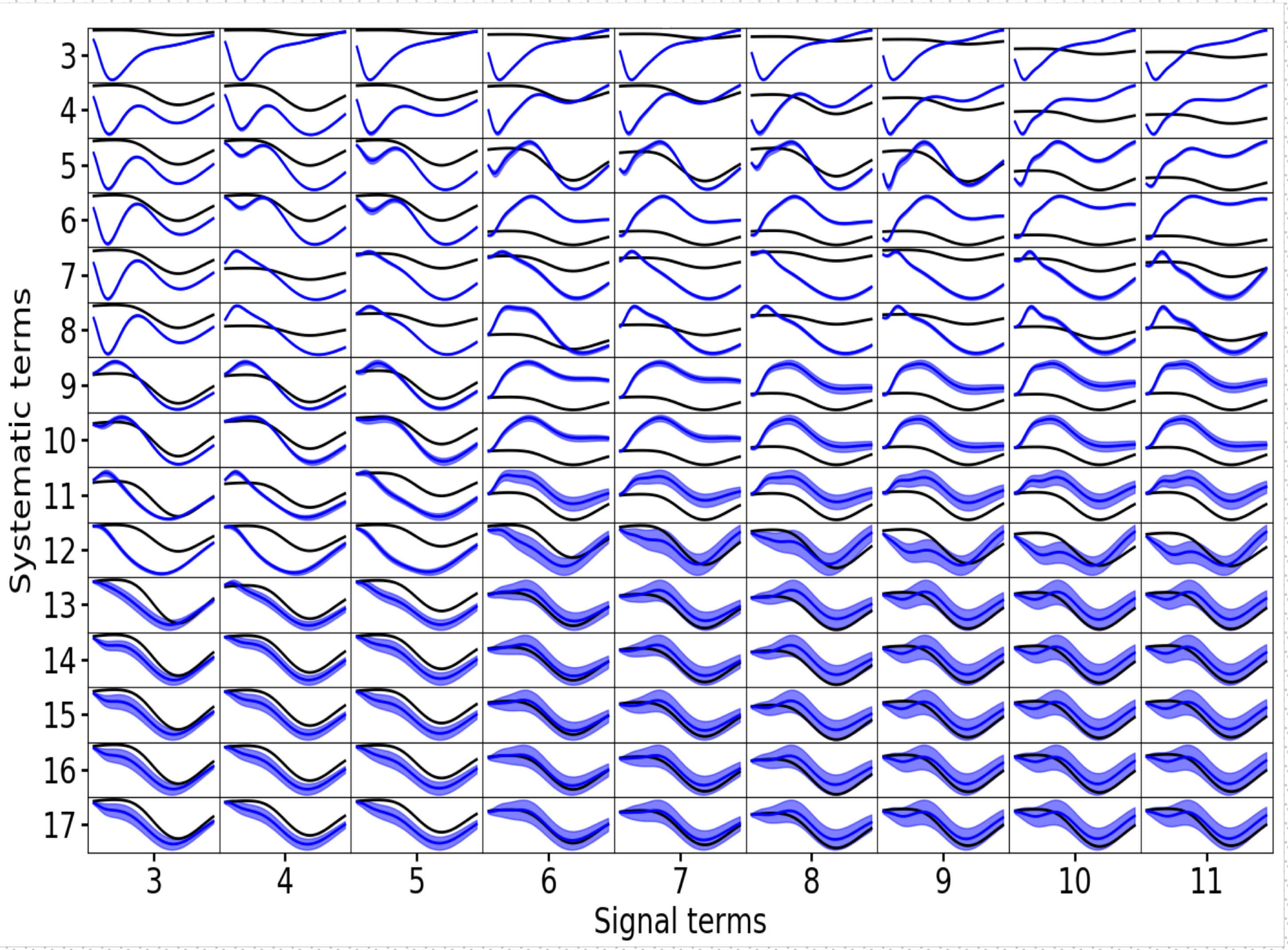

Stokes I Only


All 4 Stokes Parameters

(Nhan et al., submitted to ApJ)

Note in addition the large difference in scale between both panels

MODEL SELECTION: EXAMPLE USING BPIC

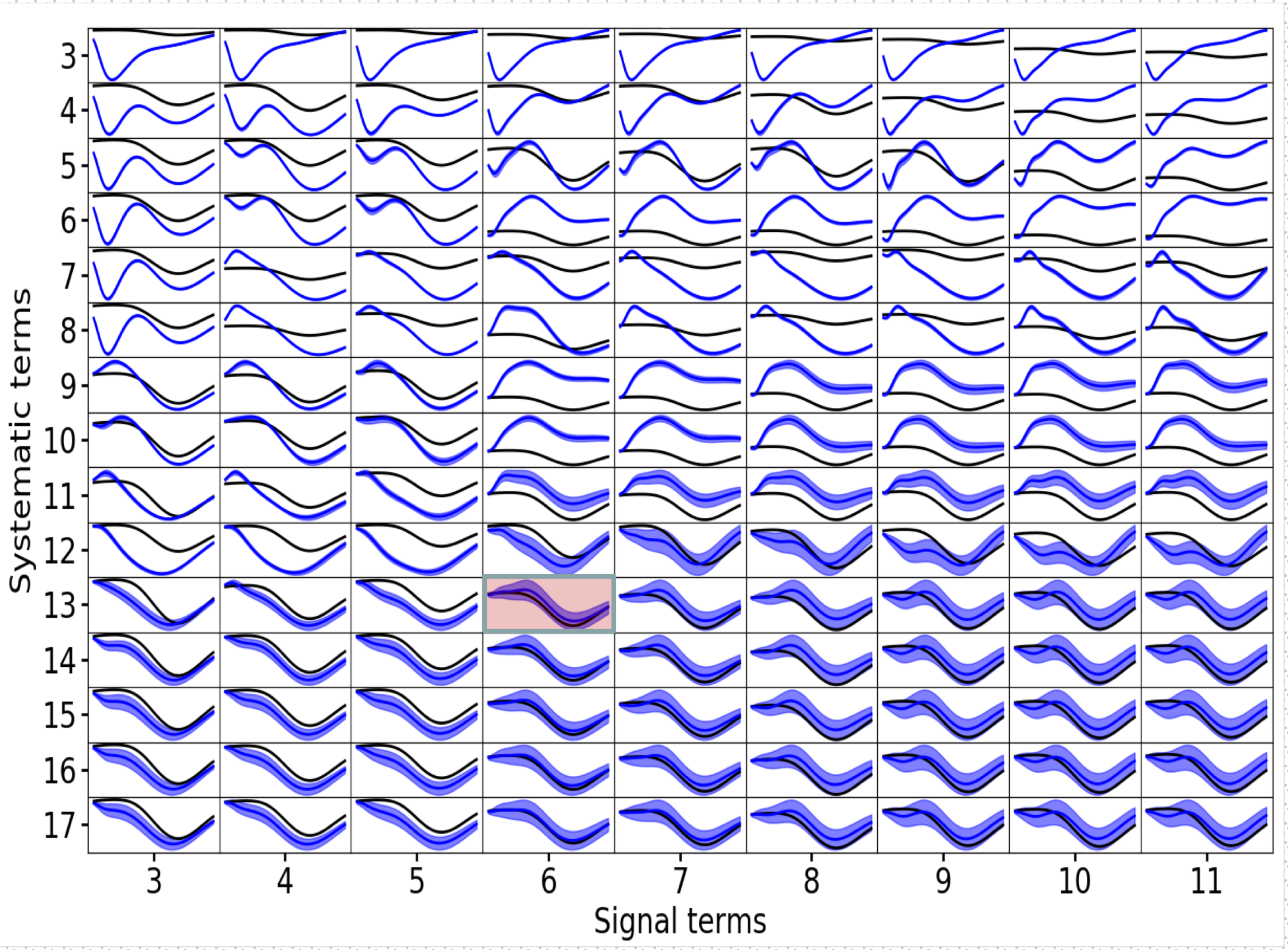


Signal Extraction optimization:

Black line for all panels: input 21-cm signal.

Blue bands: signal reconstructions for given numbers of SVD signal and systematic modes (parameters).

MODEL SELECTION: EXAMPLE USING BPIC



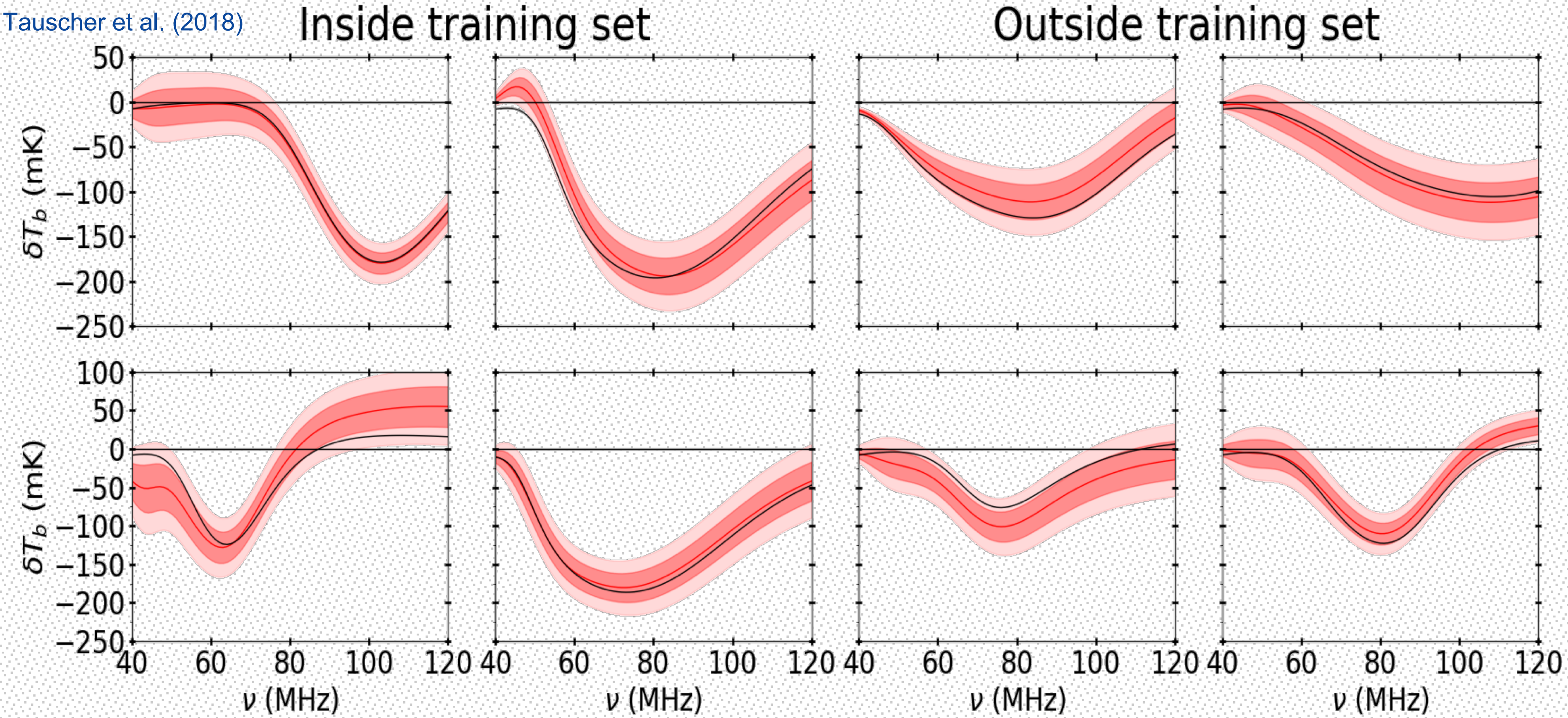
Signal Extraction optimization:

Black line for all panels: input 21-cm signal.

Blue bands: signal reconstructions for given numbers of SVD signal and systematic modes (parameters).

SIGNAL EXTRACTION WITH THE CODE PYLINEX

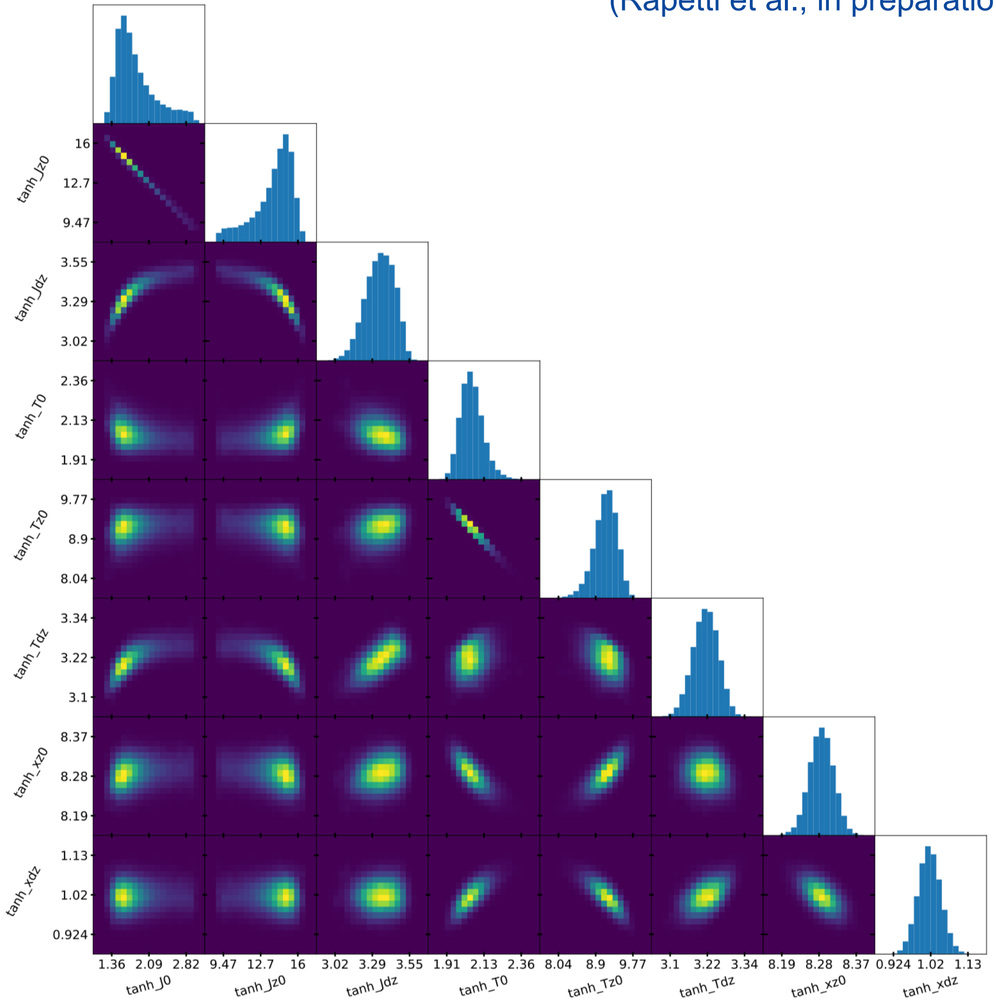
Tauscher et al. (2018)



Signal Linear Estimates from SVD eigenmodes. **Black** curves: Input signals. Red curves: Signal estimates. Dark/light red bands: posterior 68/95% confidence intervals. Left: 4 input signals from the *ares* set. Right: 4 from the tanh set (e.g. Harker et al. 2016).

See the pylinex in this link: https://bitbucket.org/ktausch/pylinex

MARKOV CHAIN MONTE CARLO (MCMC) RESULTS (PRELIMINARY)



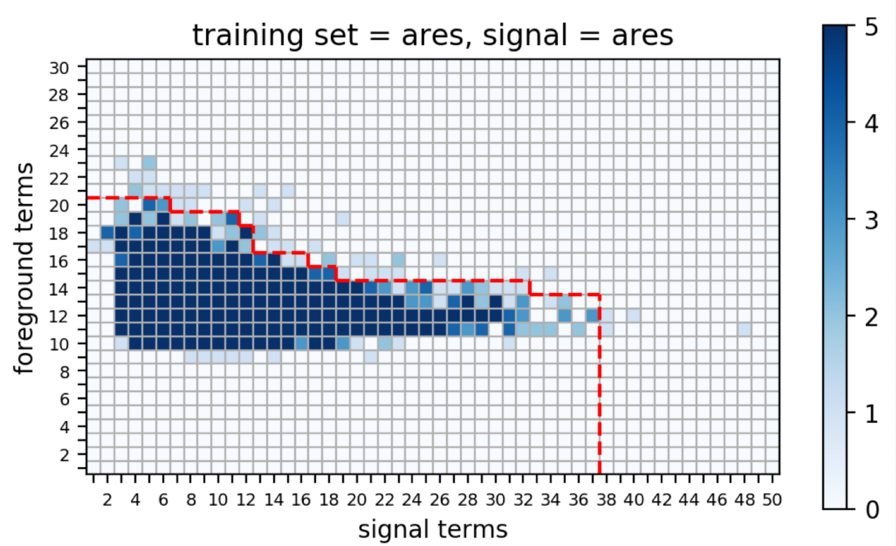
(Rapetti et al., in preparation)

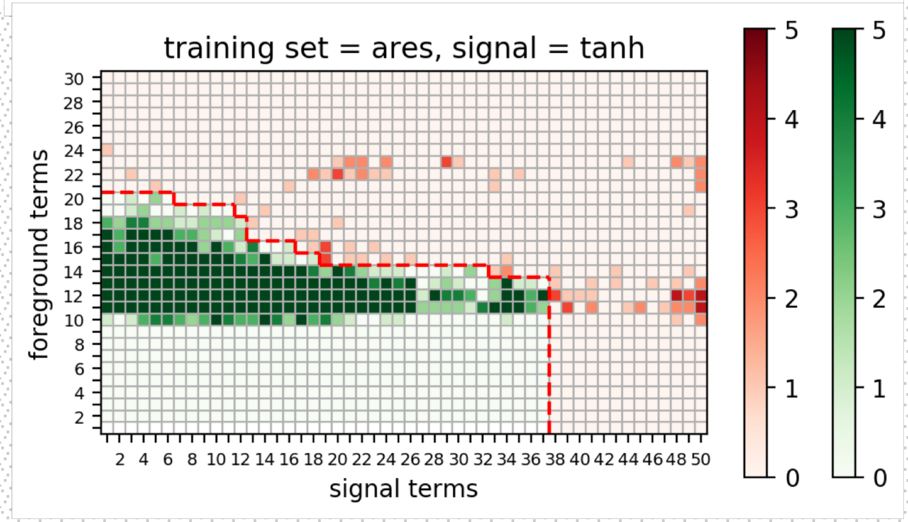
- Importantly, note that having an starting point within the estimated error is crucial for the convergence of the MCMC in a vast parameter space without otherwise any prior information.

- Preliminary recovered posterior probability distributions for tanh signal parameters.

- The full model (79 parameters) also includes SVD foreground parameters (not shown here).

FAILURE MODE STATISTICS: FINDING BASELINE MODEL SPACE REGION (PRELIMINARY)





- Each grid box represents an SVD model (given numbers of signal and systematic parameters).

- Finding the distribution of chosen models for 50,000 *ares* signals taken from an *ares* training set.

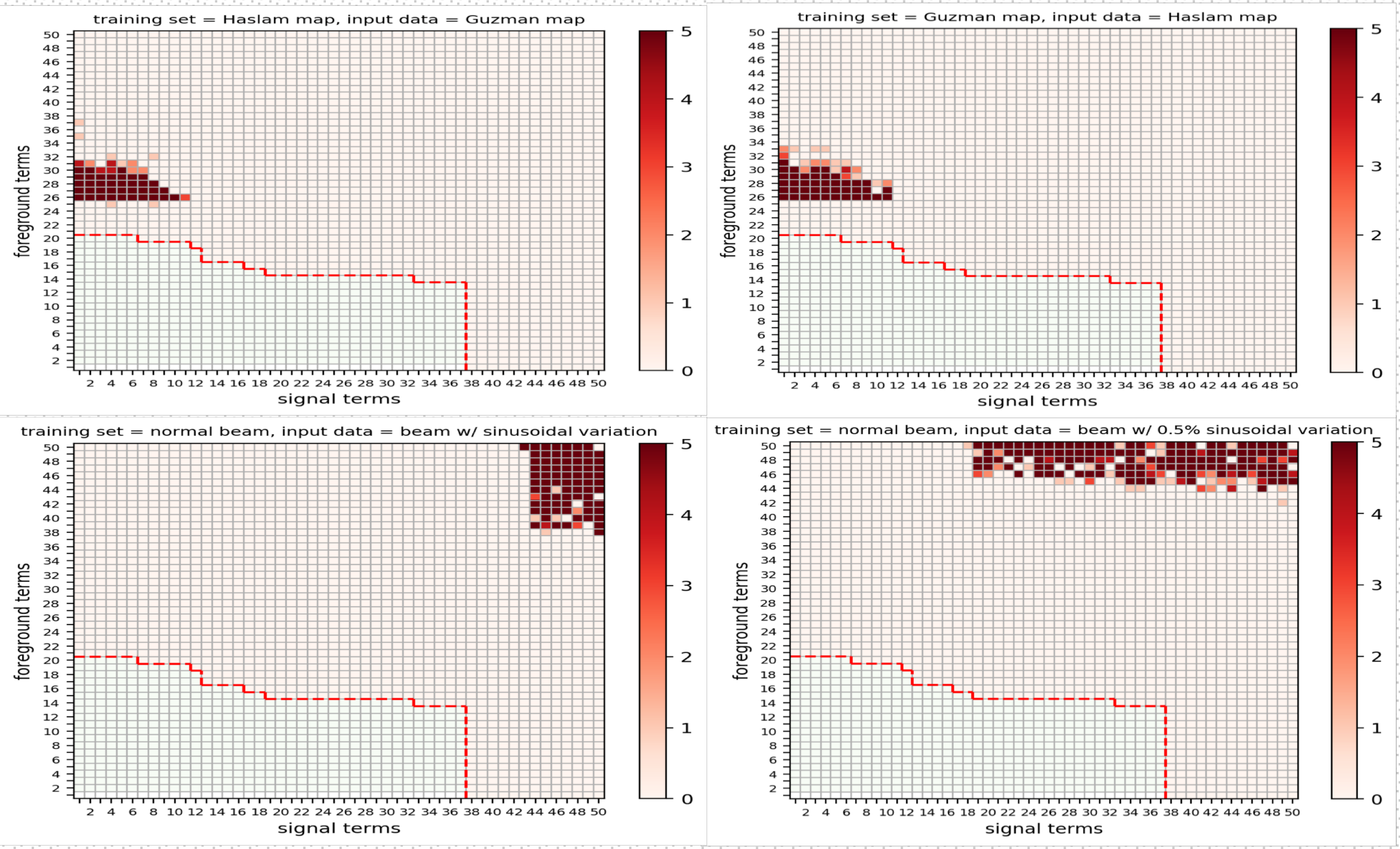
- Top: The color bar indicates the height of the distribution.

- Top: Red dashed lines enclose 99.9% of the distribution.

- Bottom: 5,000 tanh signals; green, models fitted well; red, those that not.

Bassett, Tauscher, Rapetti, Burns, in preparation

FAILURE MODE STATISTICS: FOREGROUND MAPS & BEAM CORRECTIONS (PRELIMINARY)



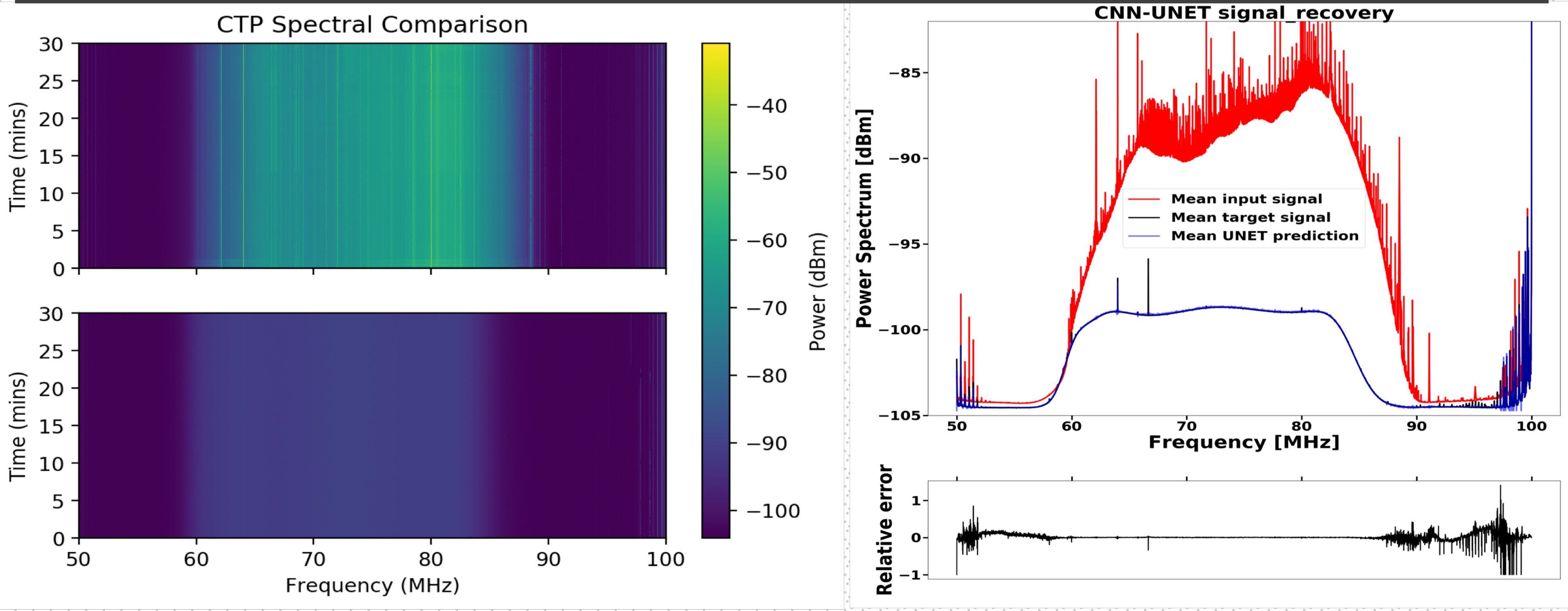
Green, models fitted well; red, those that not.

Guzman / Haslam maps (top left/right).

Beam chromaticity corrections of a sinusoidal form with smaller/larger amplitude & wavelength (bottom left/right).

Bassett, Tauscher, Rapetti, Burns, in preparation

RFI REMOVAL WITH NEUTRAL NETWORKS (PRELIMINARY)



Left: RFI-contaminated data (top) produced by an antenna while the clean data (bottom), recorded concurrently, was obtained by terminating the input with a 50 ohm load. The 30 minute interval contains 900 individual spectra, both dirty and clean. 810 pairs of spectra were used to train a UNET machine learning algorithm while 90 were left as a validation set.
Right: The top panel shows the RFI contaminated signal in red, the **clean target signal in black** and the UNET prediction of the target signal in blue. All curves are a sample average over the 90 spectra in the validation set. In a next step we are injecting a signal to be detected and will then do so experimentally.

CONCLUSIONS

- One of the main challenges of extracting the global 21-cm signal is the large foreground.
- However, unlike the foreground, the signal is spatially uniform, has well-characterized spectral features, and is unpolarized.
- We benefit from these differences using our novel approach for signal extraction and physical parameter constraints, using an SVD/IC/MCMC pipeline.
- We obtain a highly significant improvement by using a pioneering experimental design of induced polarization and we can do the same with a time series drift scan. Note that these are not mutually exclusive.
- DAPPER will be able to constrain exotic physics at lower-frequencies during the Dark Ages and probe Cosmic Dawn at higher frequencies, where EDGES data could be affected by a ground plane artifact.
- We are also working on running our pipeline on current/ongoing ground based data from EDGES and CTP using our Pattern Recognition/Information Criteria/MCMC methodology to measure absorption features.