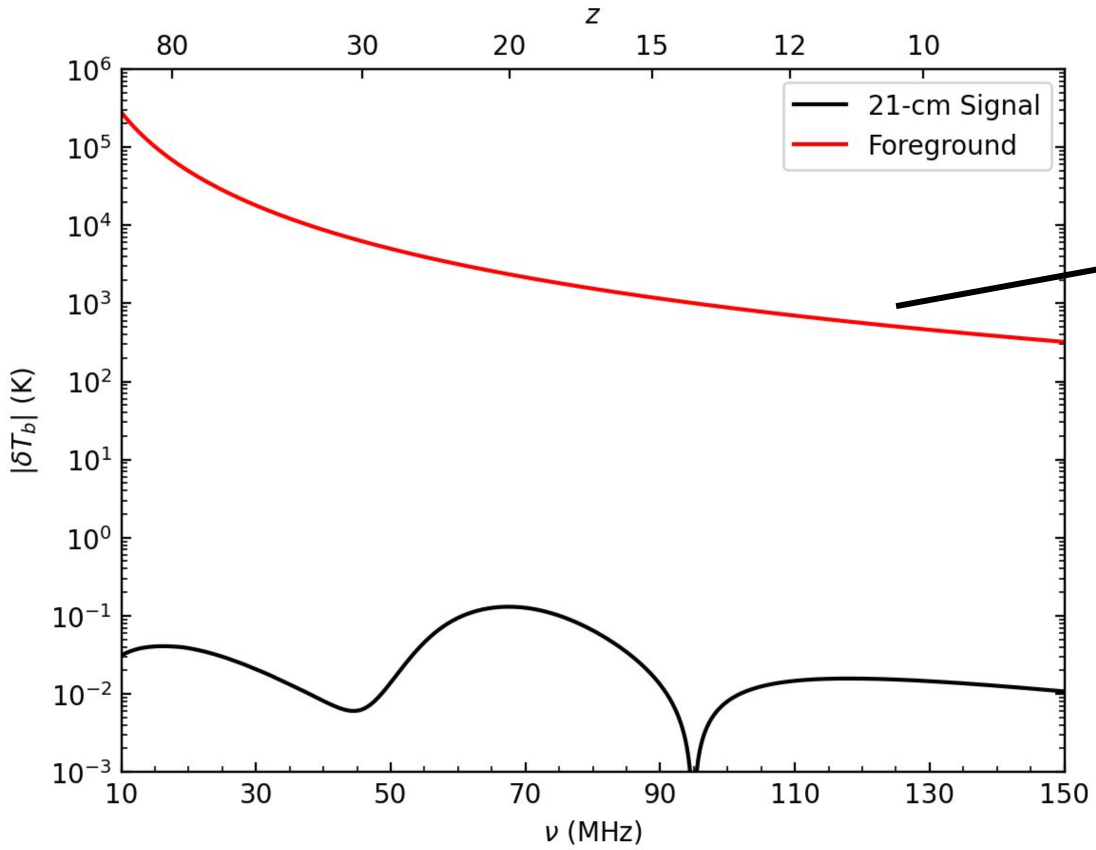
A Pattern Recognition Pipeline for DAPPER Spectra

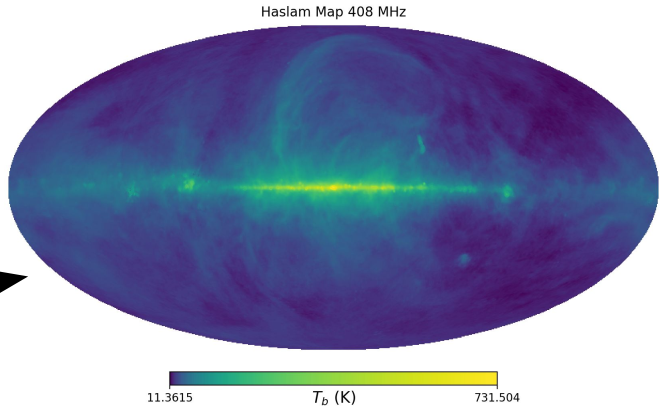
Neil Bassett
University of Colorado Boulder

NESS
Network for Exploration and Space Science

University of Colorado Boulder

Observational Difficulties





$$T_{\rm sky}(\nu,\theta,\phi) = T_{\nu_0}(\theta,\phi)\left(\frac{\nu}{\nu_0}\right)^{\beta(\theta,\phi)}$$

Synchrotron radiation will follow a power law, but large beams introduce complicated spectral structure.

Observational Difficulties

$$\text{Data} = \underbrace{(\text{Beam} \ast \text{Sky})}_{\text{“Foreground”}} + \text{21-cm Signal} + \text{noise}$$

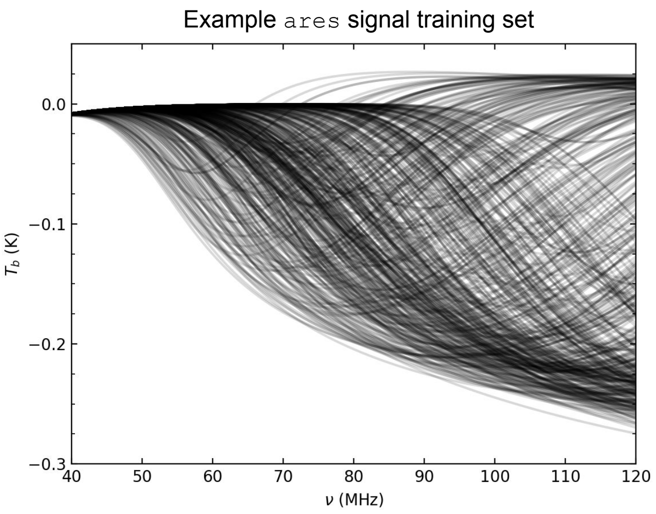
- Experiments such as EDGES attempt to calibrate out beam chromaticity and then fit a polynomial in log(T)-log(v) for the foreground component (Monsalve et al. 2017):

$$\ln\left(T_{\mathrm{fg}}\right) = \ln\left(T_0\right) + \sum_{i=1}^{n} a_i \left[\ln\left(\frac{\nu}{\nu_0}\right)\right]^i$$

- Our pipeline models each component through “training sets” which include many realizations of each component within realistic uncertainties.
  - This method allows us to create models specifically suited to a given dataset, rather than an a priori model.
  - We then utilize a specialized MCMC sampler to explore the (non-linear) parameters of the signal model.
  - Python code publicly available (https://bitbucket.org/ktausch/pylinex/)

Linear Pipeline

Step 1: Create training sets for each component of data

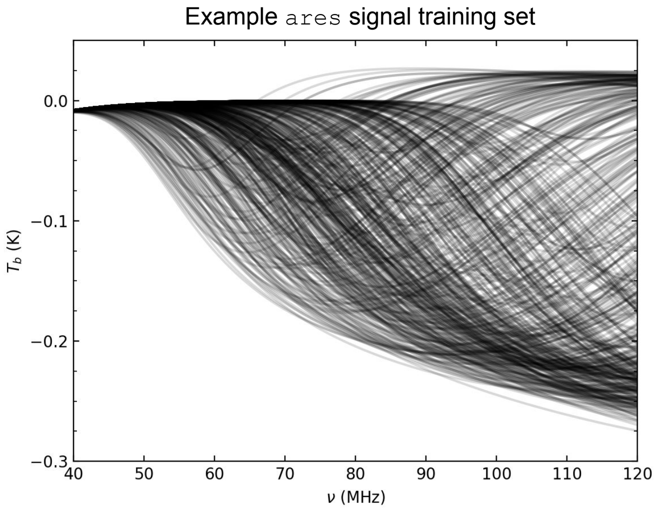


Training sets require knowledge of how components may vary rather than precise, absolute knowledge of their exact form

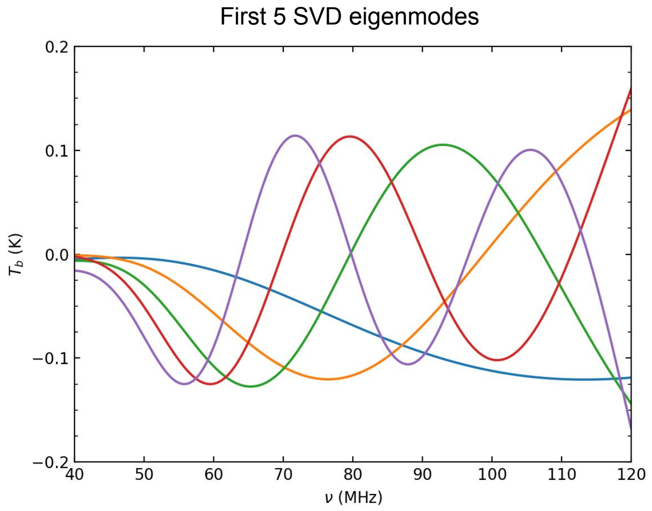
Linear portion of pipeline detailed in Tauscher et al. 2018a (Paper I; ApJ, 853, 187)

Linear Pipeline

Step 2: Perform singular value decomposition (SVD) on training sets



$$\boldsymbol{B} = \boldsymbol{U\Sigma V}^T$$



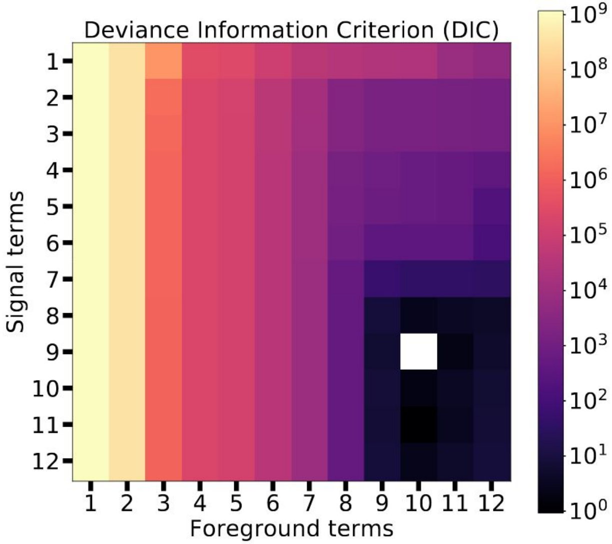
Linear portion of pipeline detailed in Tauscher et al. 2018a (Paper I; ApJ, 853, 187)

Linear Pipeline

Step 3: Minimize information criterion to determine number of modes to use

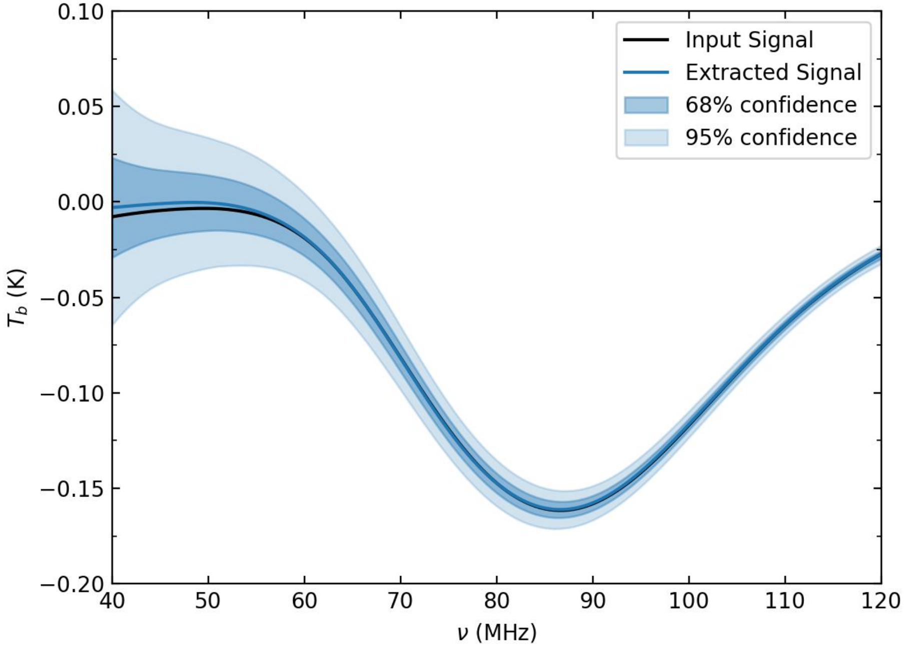
$$\mathrm{DIC} = \boldsymbol{\delta}^T \boldsymbol{C}^{-1} \boldsymbol{\delta} + 2N_p$$

Information criterion chooses number of parameters large enough to describe data, but small enough to provide reasonable uncertainties.



Linear portion of pipeline detailed in Tauscher et al. 2018a (Paper I; ApJ, 853, 187)

Linear Pipeline

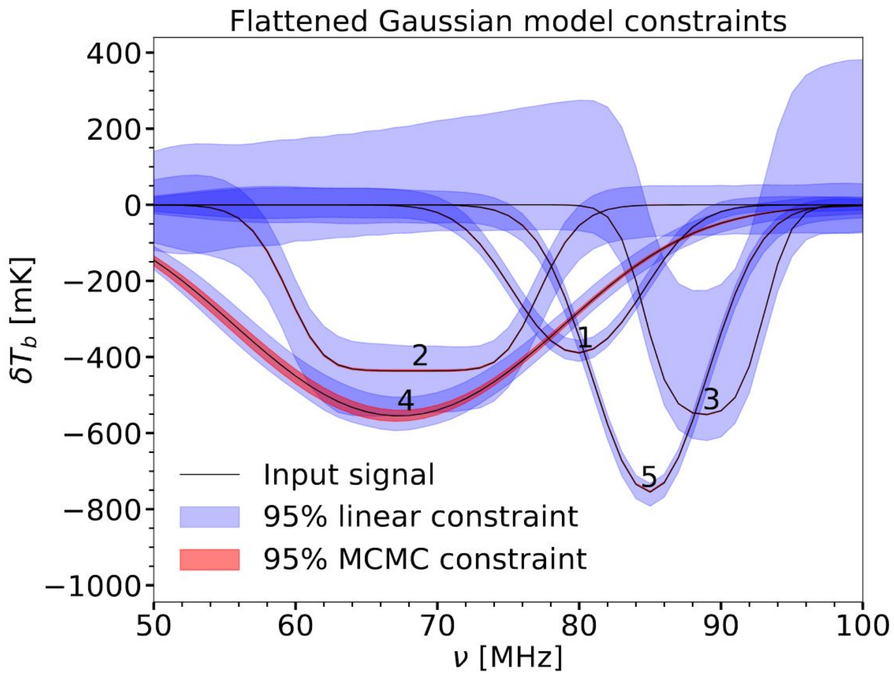


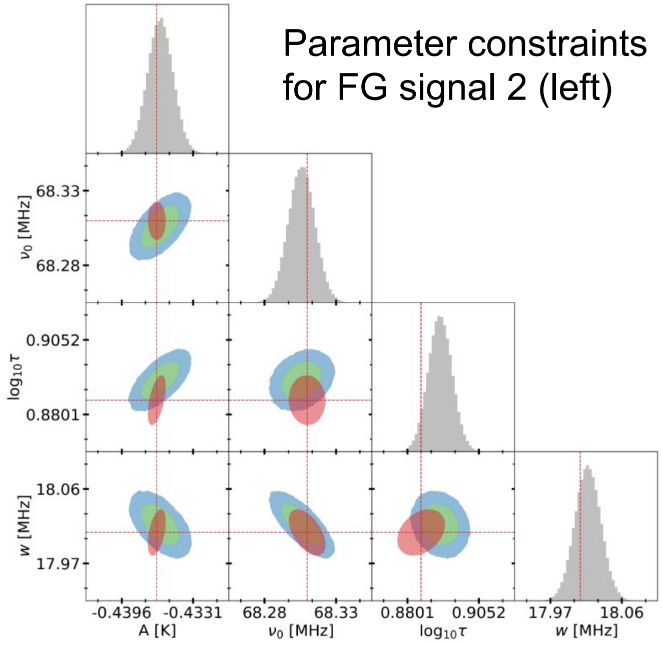
Final outputs are spectral constraints on the 21-cm signal.

Using the spectral constraints from the linear fit to obtain constraints on signal model parameters requires the nonlinear portion of the pipeline which uses an MCMC (Rapetti et al. 2020, Paper II).

Linear portion of pipeline detailed in Tauscher et al. 2018a (Paper I; ApJ, 853, 187)

Non-linear Pipeline





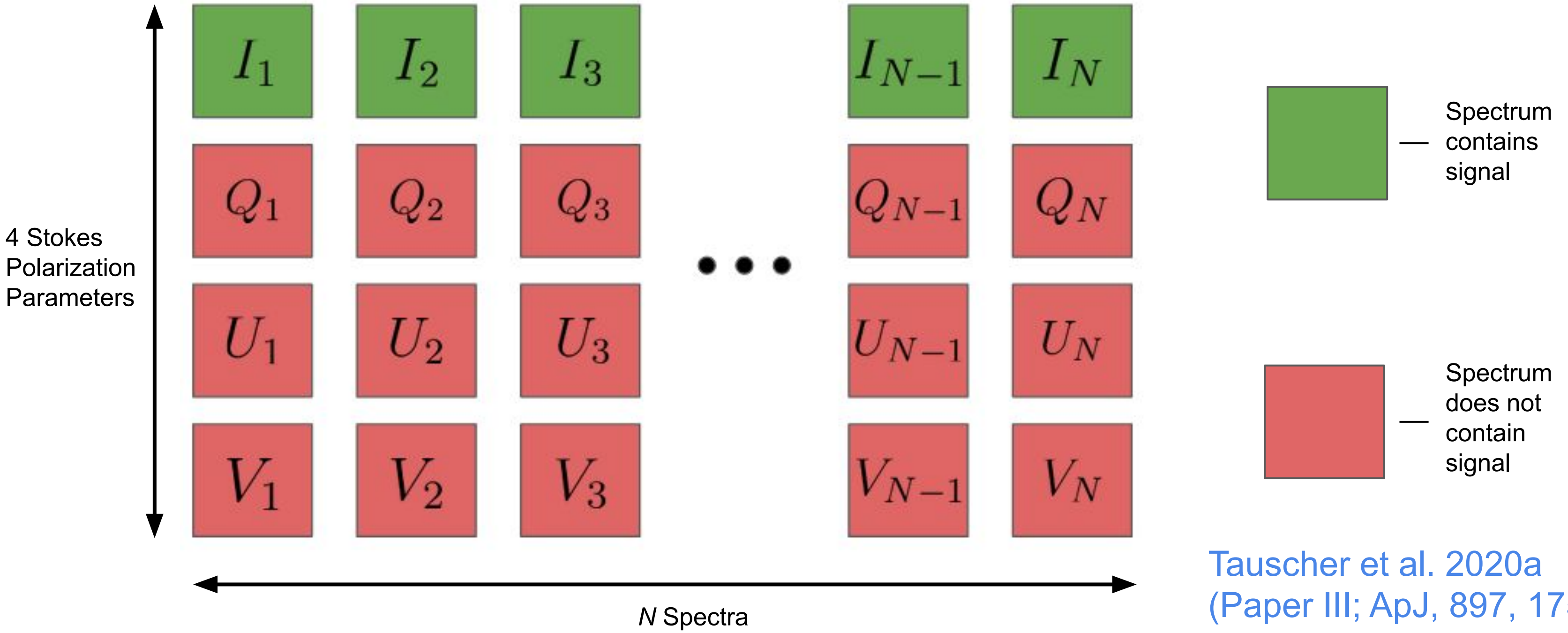
Parameter constraints for FG signal 2 (left)

- Conditional MCMC analytically marginalizes SVD foreground parameters to efficiently explore signal parameter space (Rapetti et al. 2020; Paper II; ApJ, 897, 174).
- Overlap between the signal and systematics is still properly accounted for in parameter constraints.

Dynamic Polarization

Uncertainty of signal extraction depends primarily on overlap between foreground and signal models. This overlap can be decreased by using:

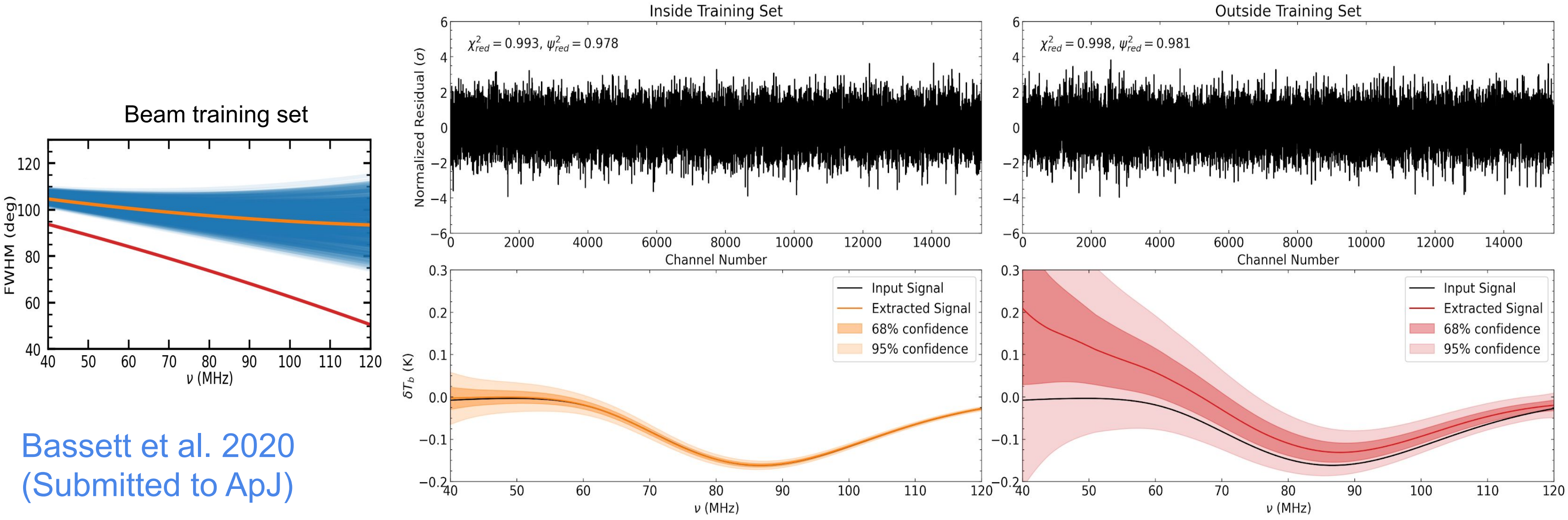
1. Polarization
2. Many Correlated Spectra



Tauscher et al. 2020a (Paper III; ApJ, 897, 175)

Assessing Training Sets

How do we have confidence that our training sets are sufficient to extract the 21-cm signal both accurately and precisely?



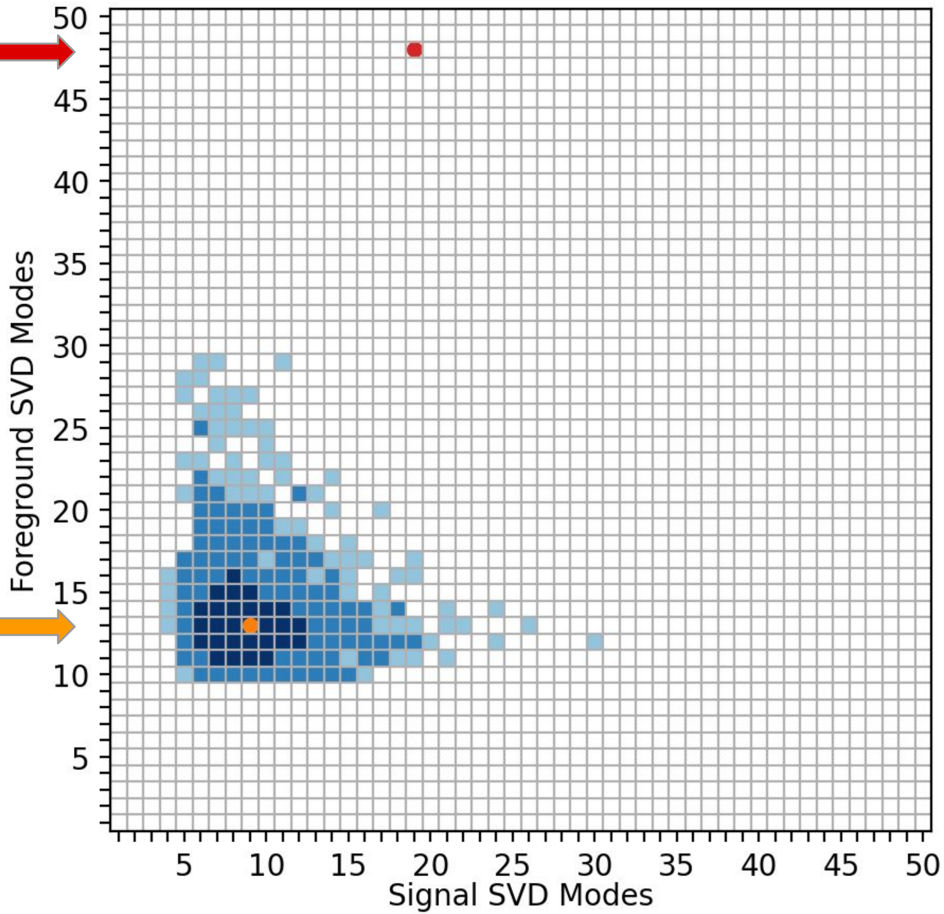
Bassett et al. 2020 (Submitted to ApJ)

In this example, traditional goodness-of-fit statistics are unable to detect when the signal extraction is suboptimal

Assessing Training Sets

Red fit used beam much different than training set

Orange fit used beam consistent with training set



The blue distributions contain 68, 95, and 99.7 percent of 5000 fits with data realizations made directly from the training set.

Number of SVD modes chosen by pipeline can tell us when training set need to be updated!

Bassett et al. 2020 (Submitted to ApJ)

Summary

- We have a developed a novel pipeline for global 21-cm signal extraction in the presence of systematics that is publicly available (https://bitbucket.org/ktausch/pylinex/)
- Uncertainty on 21-cm signal extraction can be decreased by including multiple spectra and polarization measurements in the analysis, which decreases overlap between foreground and signal models.
- We are able to identify when training sets are inadequate for analyzing a given dataset by comparing the number of SVD modes chosen for the linear fit to a distribution of simulated realizations from the training set.

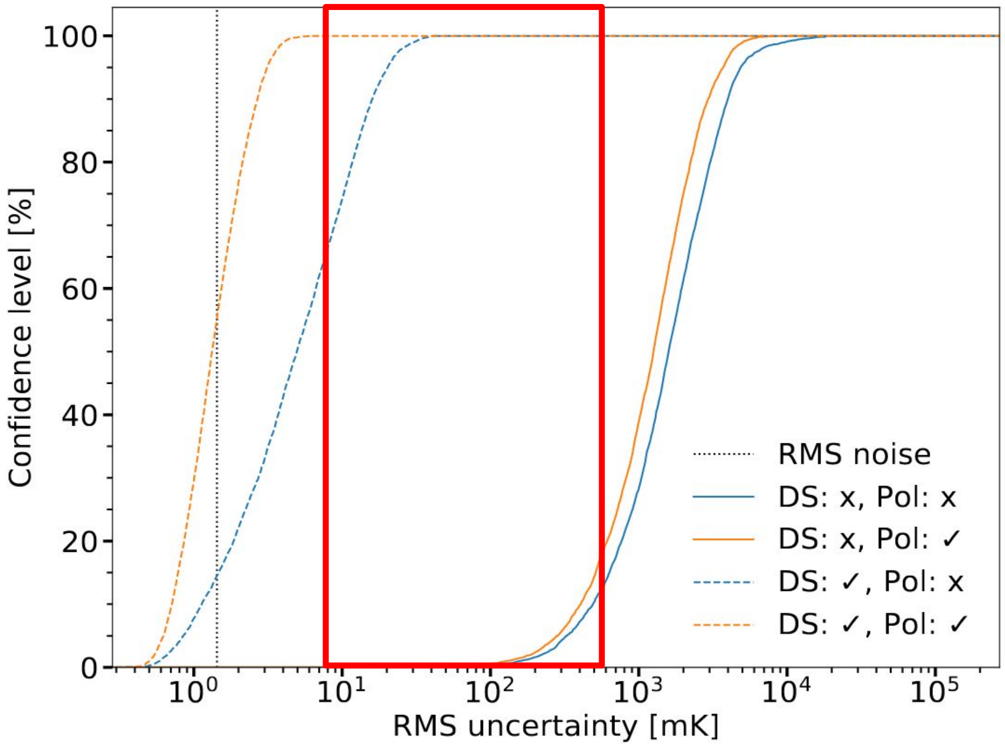
Ongoing and Future Work

- We are currently re-analyzing EDGES data with our pipeline to confirm published EDGES result (Bowman et al. 2018) using an independent analysis strategy.
- Work ongoing for Paper IV, which will include instrumental effects such as receiver gain within the pipeline formalism.

Extra Slides

Dynamic Polarization

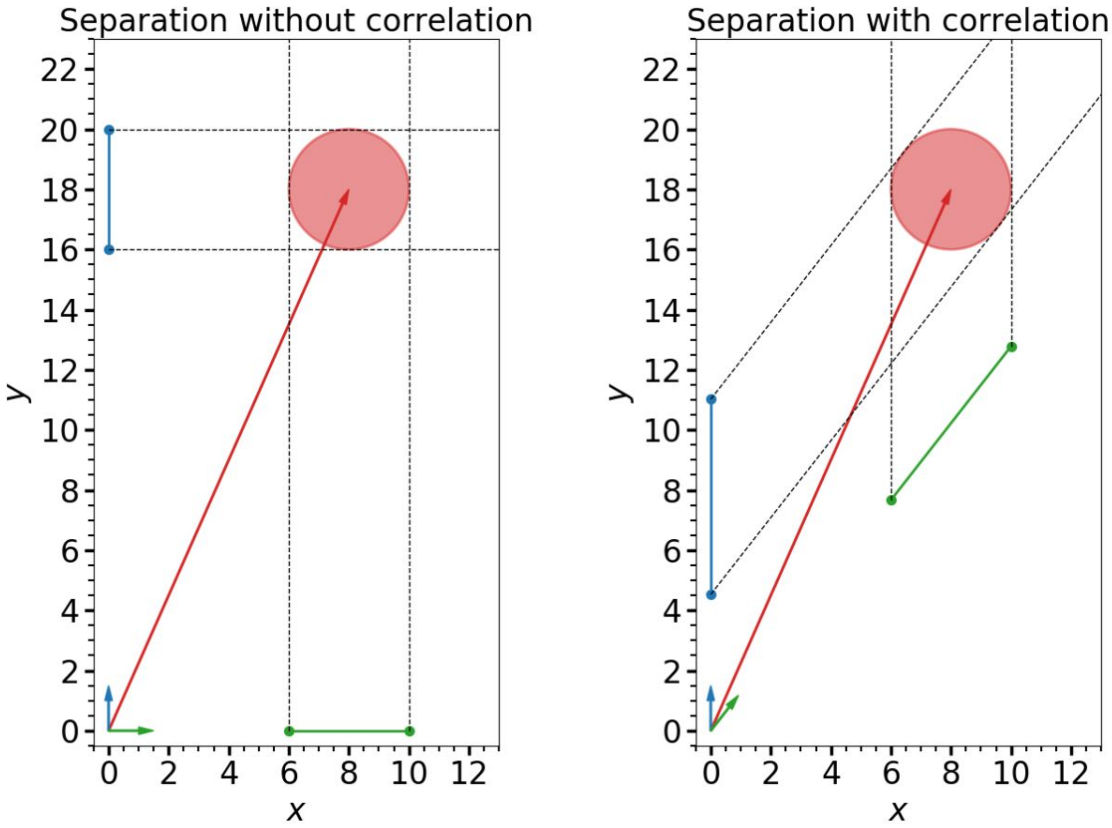
Magnitude of signal



- CDFs of RMS uncertainty levels for 5000 simulated fits using four different observation strategies.
  - Smallest uncertainties produced by including both multiple spectra and polarization

See Tauscher et al. 2020a (Paper III)

Simultaneously Fitting Multiple Components

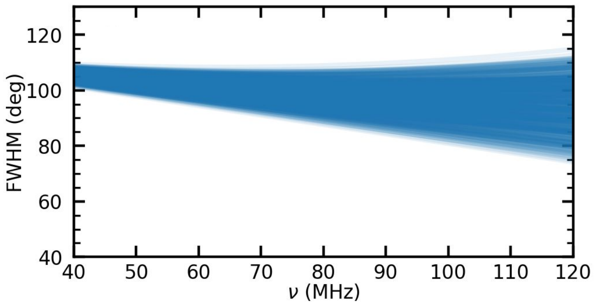


- Foreground and signal models are fit to data simultaneously.
- Uncertainty on each component depends on how similar the models are.
    - Similarity of models highly dependent on experimental design (single spectrum experiment will lead large overlap, wheras using polarization, e.g., can decrease this overlap)

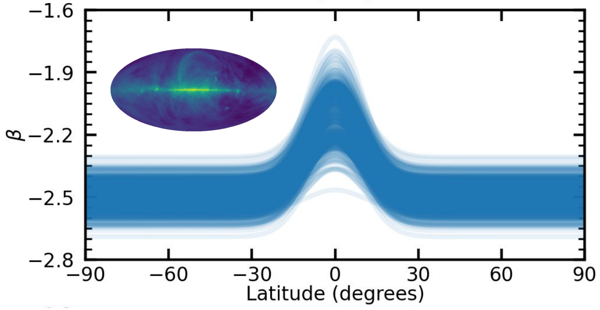
Simulating Observations

$$\text{Data} = \underbrace{(\text{Beam} \ast \text{Sky})}_{\text{“Foreground”}} + \text{21-cm Signal} + \underbrace{\text{noise}}_{\sigma(\nu) = \frac{T_{\mathrm{sys}}}{\sqrt{\Delta\nu\Delta t}}}$$

Beam



Sky



Signal