# Simultaneous Constraints on Foreground and Global 21-cm Models via a Novel Pattern Recognition Technique

David Rapetti
NASA Ames Research Center/Universities Space Research Association
University of Colorado Boulder

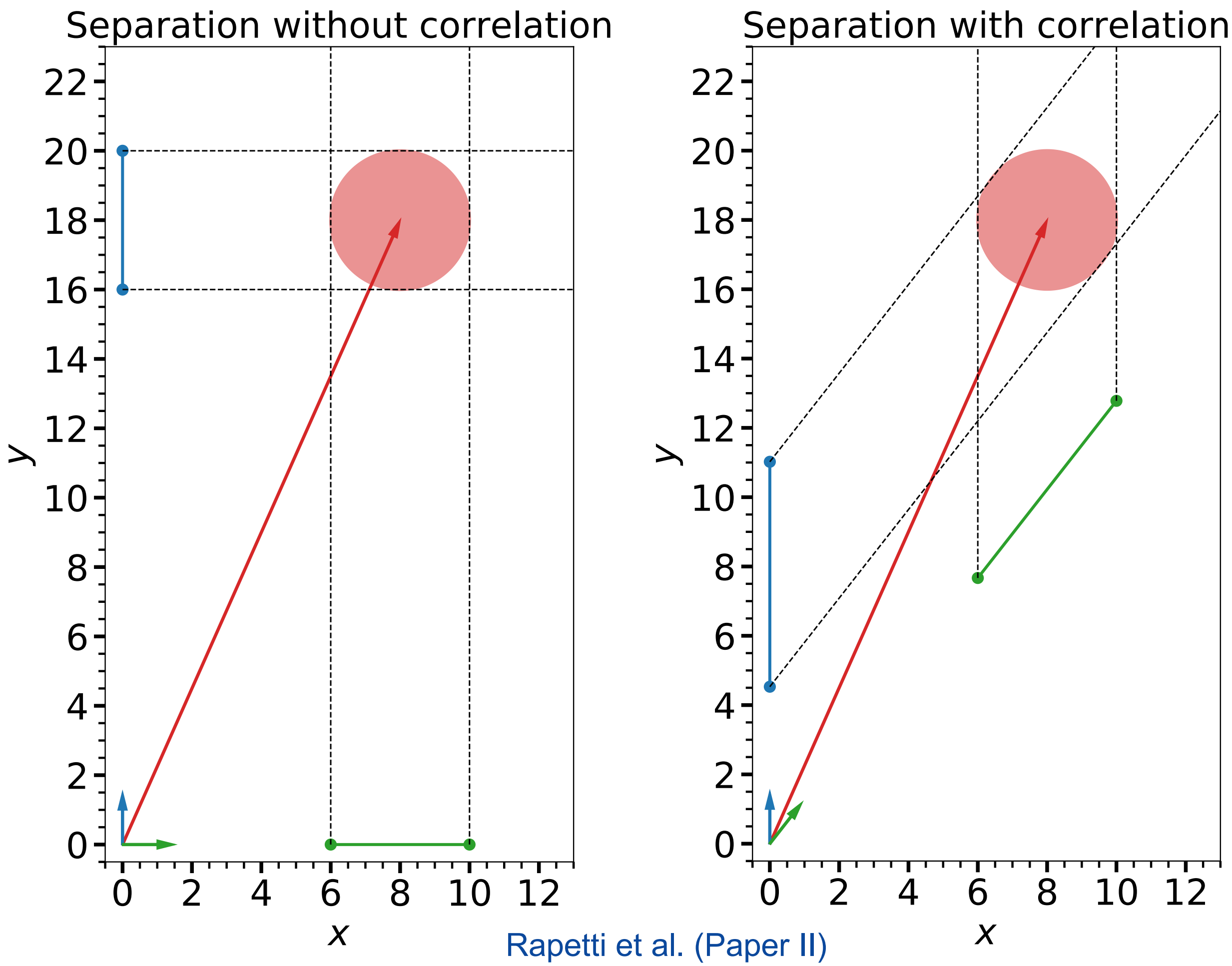
In collaboration with:
Keith Tauscher (CU Boulder), Jordan Mirocha (McGill U.), Jack O. Burns (CU Boulder)

NESS
Network for Exploration and Space Science

# JOINT CONSTRAINTS ON PHYSICAL PARAMETERS

- In Paper I (Tauscher, Rapetti, Burns & Switzer 2018), we presented a new method based on Singular Value Decomposition (SVD) and Information Criteria to obtain spectral constraints on the global 21-cm signal.
- Converting spectral constraints into constraints on any physical parameter space of choice is presented in Paper II (Rapetti, Tauscher, Mirocha & Burns, to be submitted).
- This allows us to analytically find a joint linear fit of the signal and systematics (currently, foreground) to be readily used as starting point (mean and covariance) for our simultaneous, nonlinear Markov Chain Monte Carlo (MCMC) fit.

# OVERLAP BETWEEN DATA COMPONENTS




Rapetti et al. (Paper II)

Left: Minimum uncertainties for each component (noise level) by using orthogonal modes. Right: Larger uncertainties due to overlap.

Schematic representation: overlap between signal and systematic modes increases the uncertainties of both components with respect to the statistical noise (red circle).

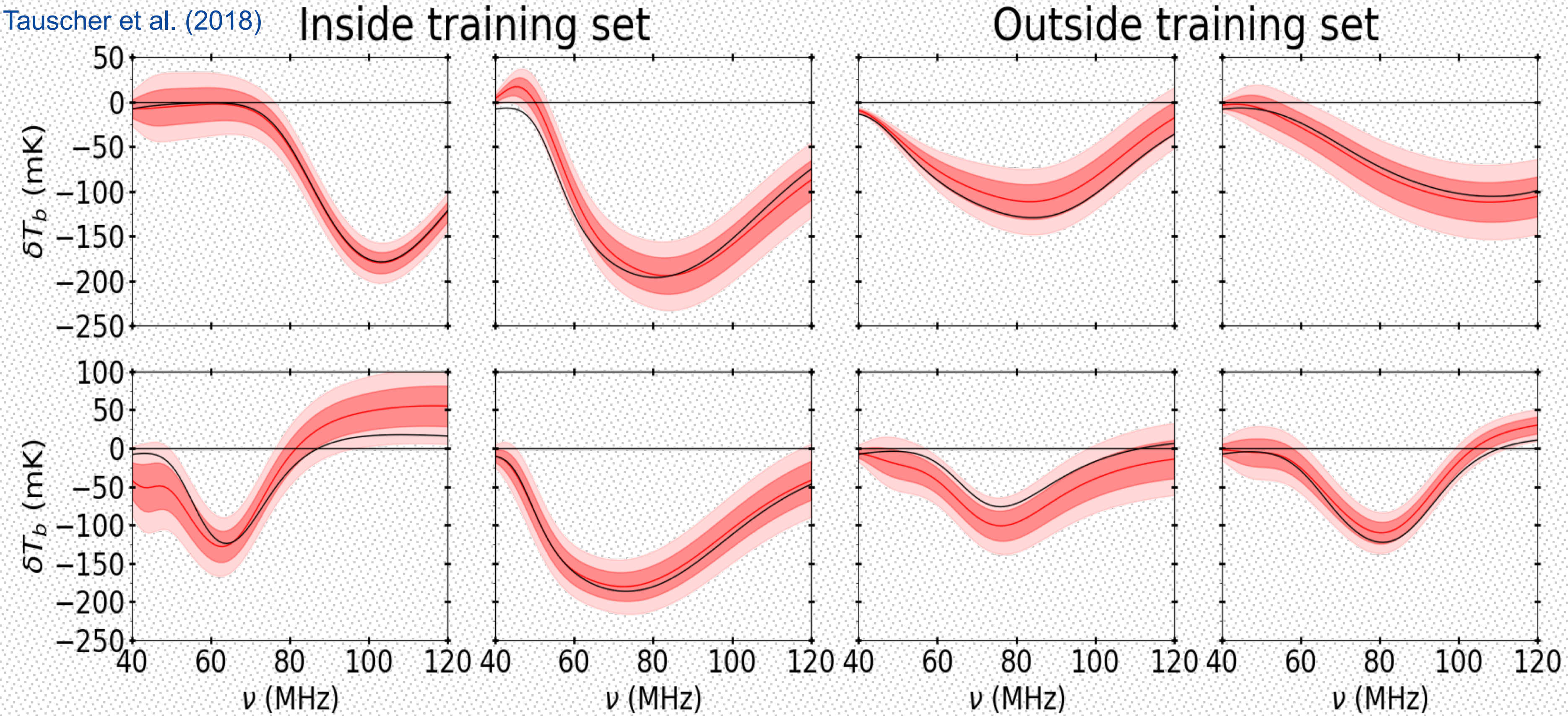
Data (red vector), signal (blue) and systematics (green) basis vectors vectors on the origin.

The blue and green intervals are the 1σ uncertainties on the signal and foreground.

The signal uncertainty is computed by projecting the noise ellipse parallel to the foreground basis vector onto the line defined by the signal basis vector and vice versa.

# LINEAR SIGNAL EXTRACTION WITH PYLINEX

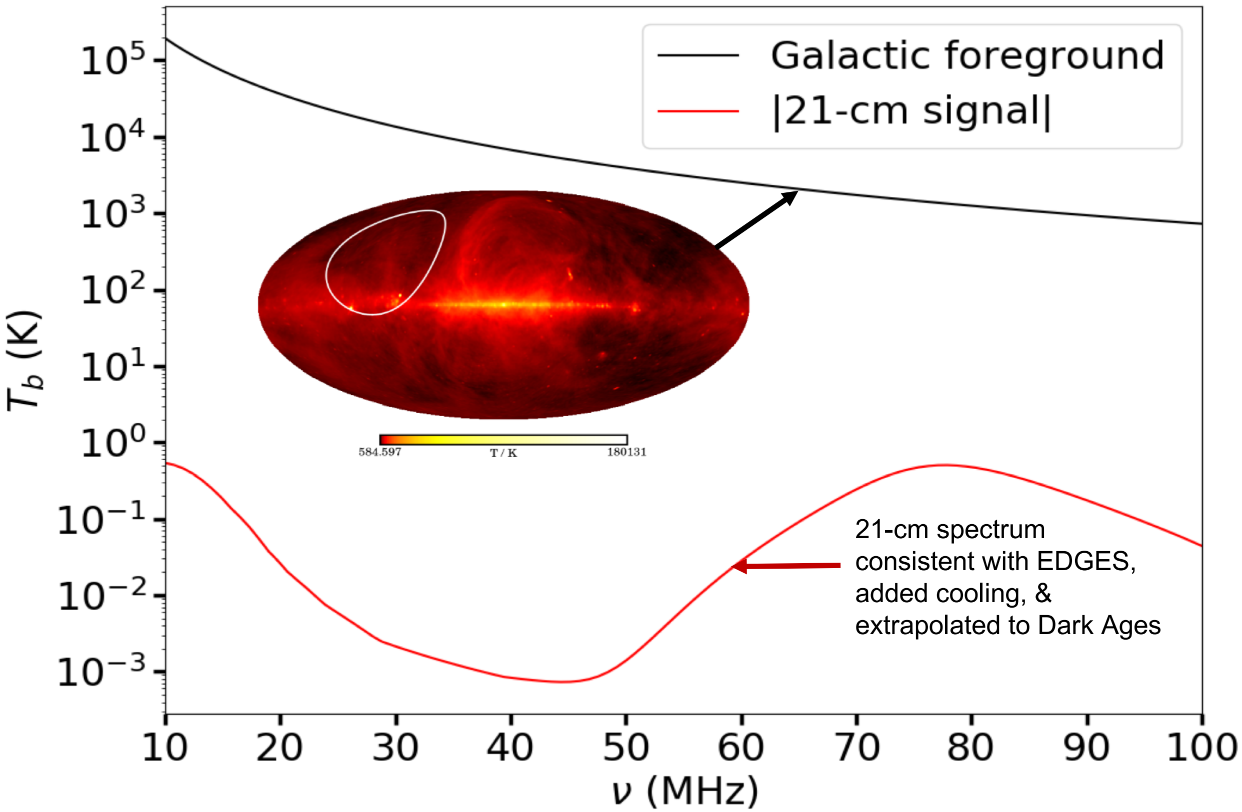
Tauscher et al. (2018)



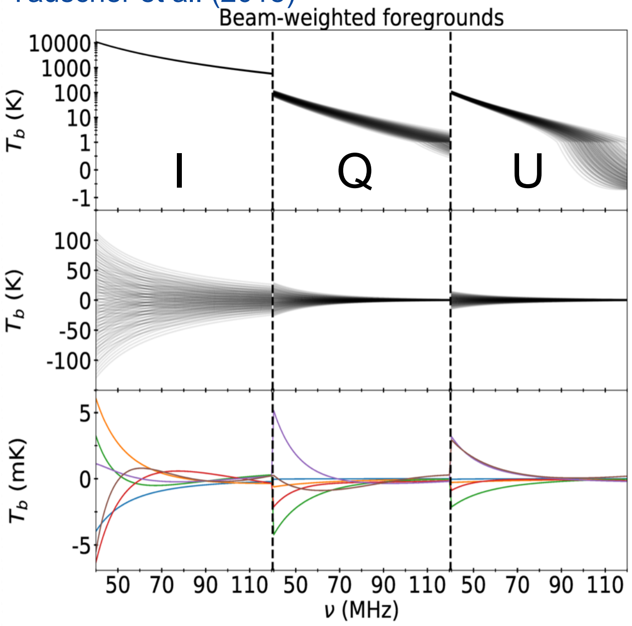

Signal Linear Estimates from SVD eigenmodes. **Black** curves: Input signals. Red curves: Signal estimates. Red bands: posterior 68/95% confidence intervals. Left: 4 input signals from the *ares* set. Right: 4 from the tanh set (e.g. Harker et al. 2016).

Find the code pylinex in this link: [https://bitbucket.org/ktausch/pylinex](https://bitbucket.org/ktausch/pylinex)

# CHALLENGES OF GLOBAL 21-CM OBSERVATIONS



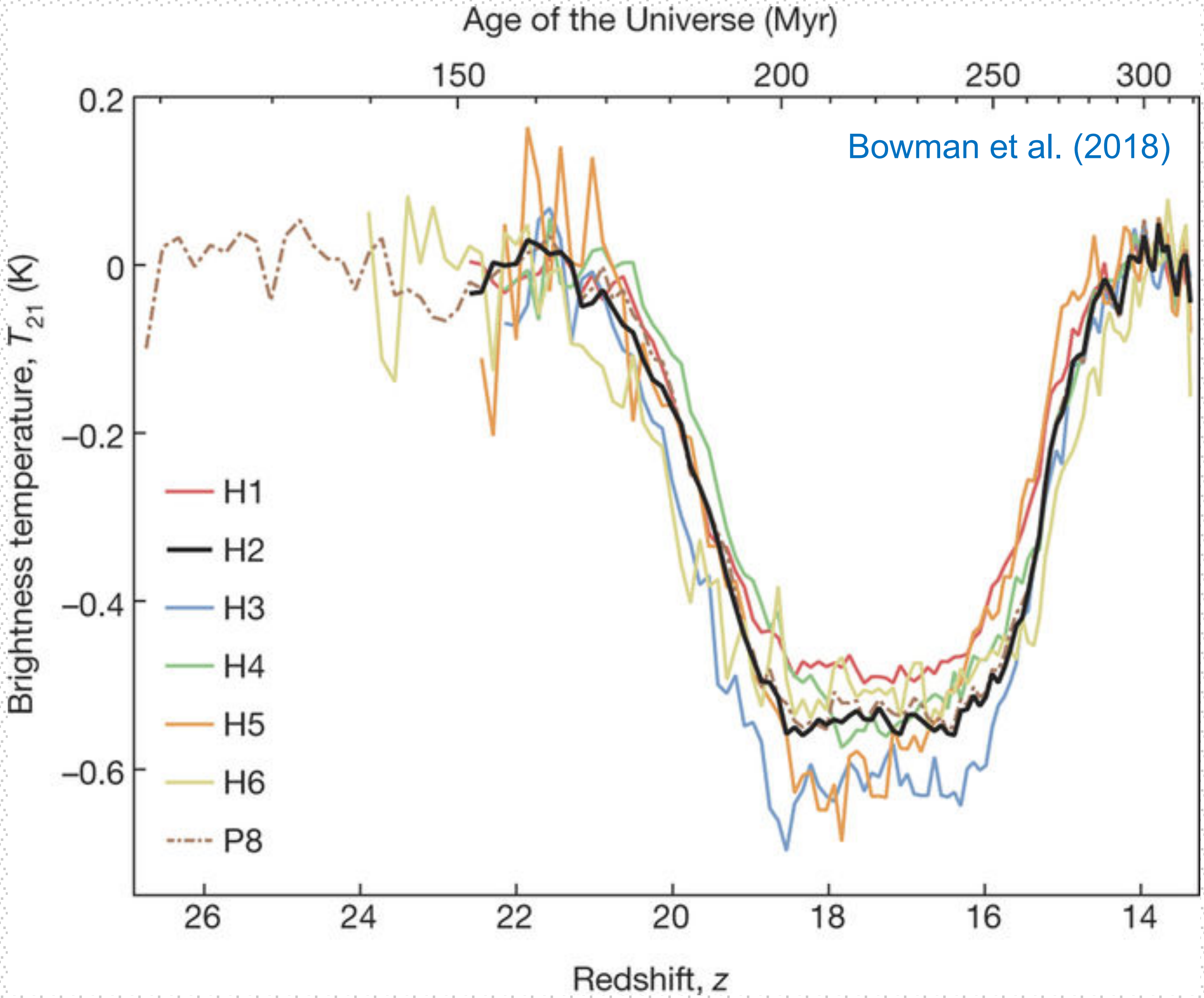

Tauscher et al. (2018)




**Foreground Characteristics**
- Spectrally smooth
- Spatial structure
- Polarized

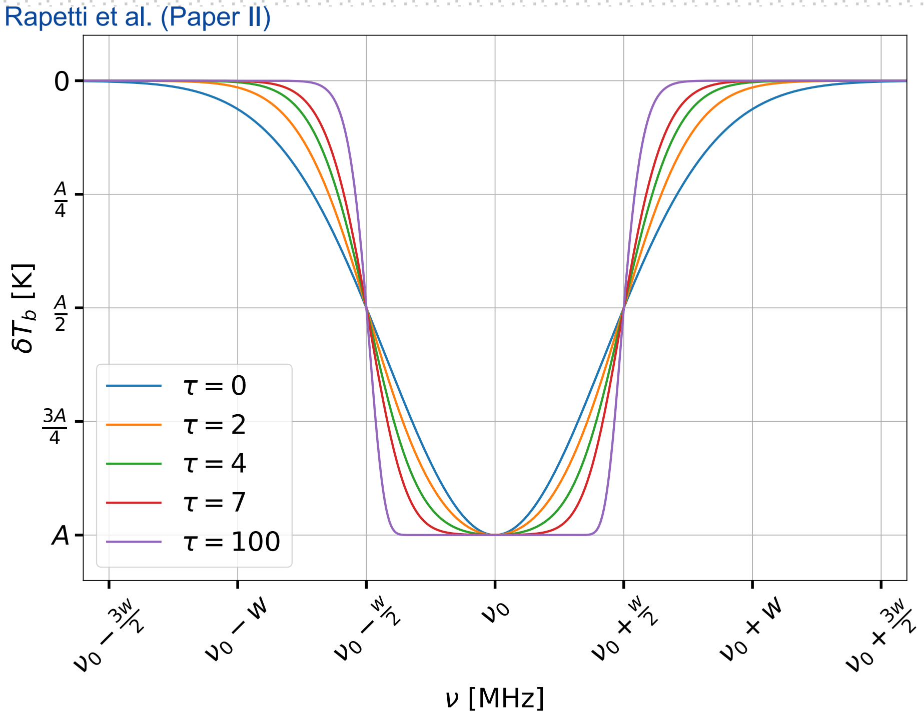
**Signal Characteristics**
- Spectral structure
- Spatially isotropic
- Unpolarized

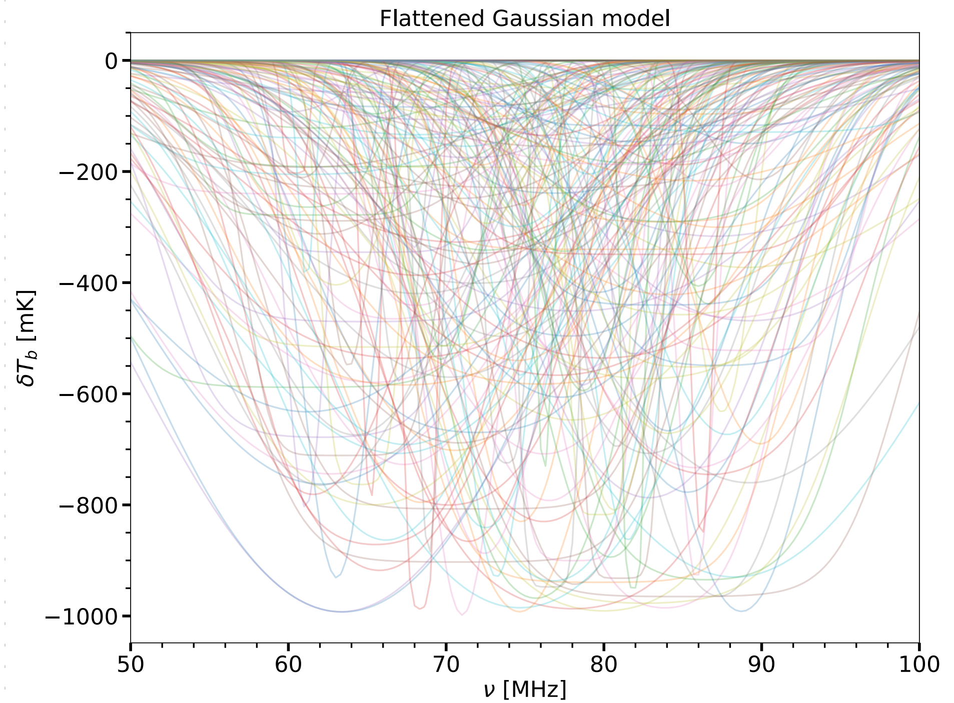
# MOTIVATION FOR MODELS: OBSERVATIONAL MEASUREMENTS




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 using a flattened Gaussian model.

# GLOBAL 21-CM MODELS: FLATTENED GAUSSIAN




The flattened Gaussian model depends on parameters: A (amplitude), $\nu_0$ (central freq.), w (FWHM), and τ (flattening). The first three shift and scale the signal while τ (at constant w) determines how long around $\nu_0$ the signal stays near its maximum depth.

Sample of 200 curves from the training set for the flattened Gaussian model.

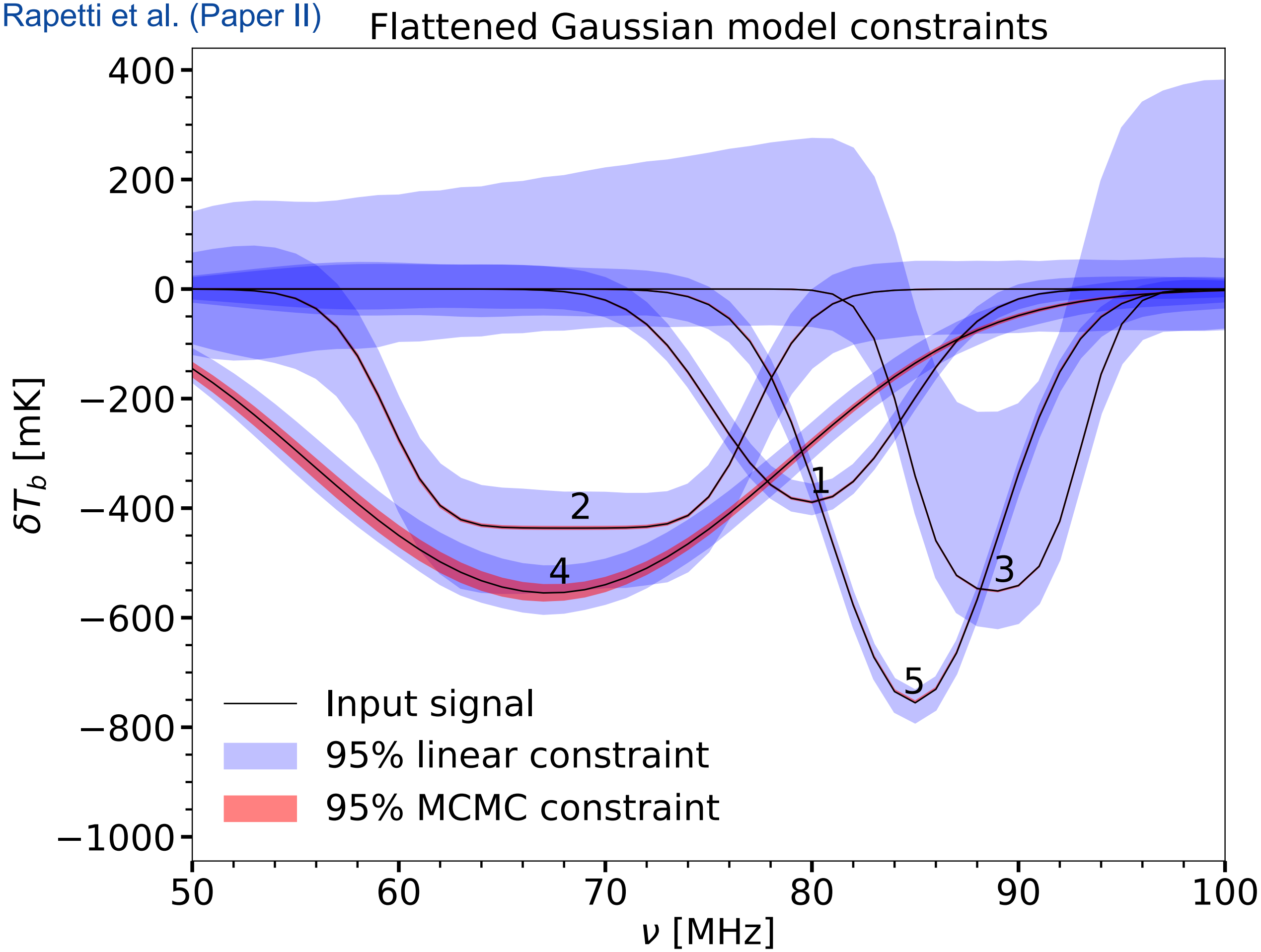
- A uniform (-1, -0.1) K
- $\nu_0$ uniform (60, 90) MHz
- w uniform (1, 30) MHz
- τ exponential (1)

# MCMC CONDITIONAL FIT OF 21-CM SIGNAL MODEL OVER SVD FOREGROUND TERMS

- Our conditional MCMC marginalizes over the SVD foreground modes at each step, while efficiently exploring only the physical parameter space.
- This calculation is exact (not an approximation) and allows for the natural separation of linear nuisance parameters without a need for a parametric model and nonlinear signal parameters to be MCMC sampled.
- This properly accounts for overlaps between the signal and the systematics (beam-weighted foreground, receiver, etc.).
- Other experiments in physics/astrophysics/cosmology could benefit from our current set of novel solutions.

# FLATTENED GAUSSIAN MODEL: LINEAR AND MCMC SPECTRAL CONSTRAINTS

Rapetti et al. (Paper II)



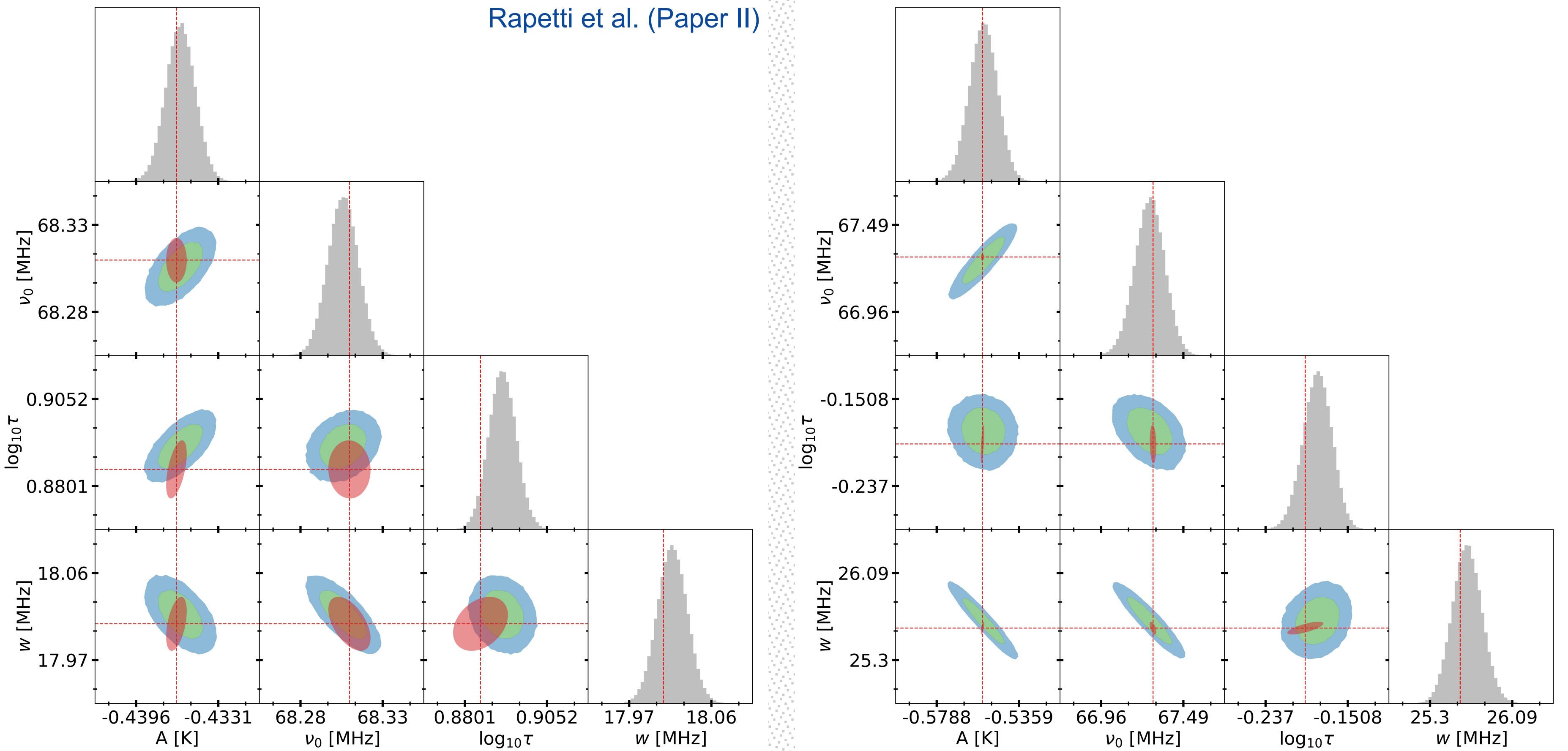

Pipeline spectral constraints for five random flattened Gaussian cases successfully recovered:

Blue bands: 95% confidence intervals from the linear fit, with SVD signal and foreground modes.

Red bands: 95% confidence intervals from the MCMC fit, with the full nonlinear signal model and SVD foreground modes.

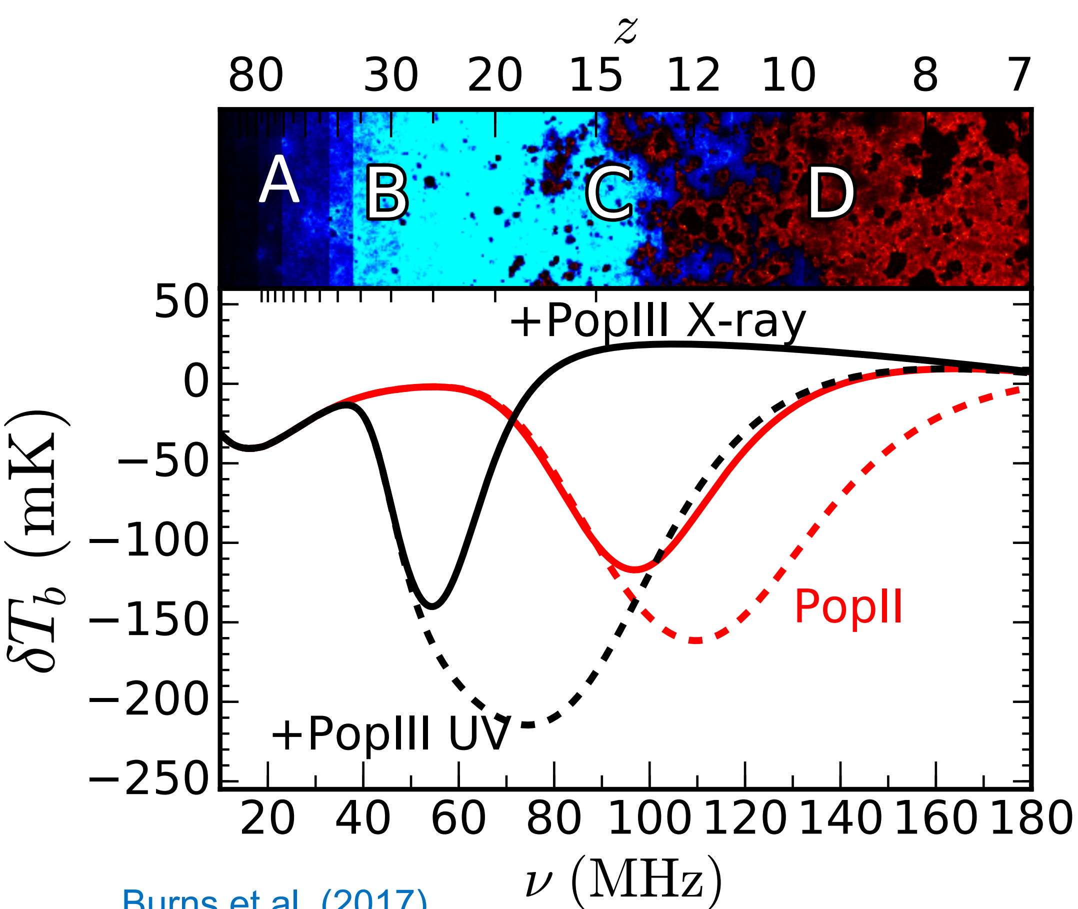
Note that for the linear fit, the 95% confidence intervals correspond to 8.75σ.

# FLATTENED GAUSSIAN MODEL: FULL MCMC PARAMETER CONSTRAINTS




1D (gray) and 2D (68/95%) MCMC posterior parameter constraints. The red, dashed lines mark the input parameters. The left (right) plot corresponds to case 2 (4) before. The red contours represent 95% errors for statistical noise alone. In case 4, systematics clearly play a larger role than in case 2.

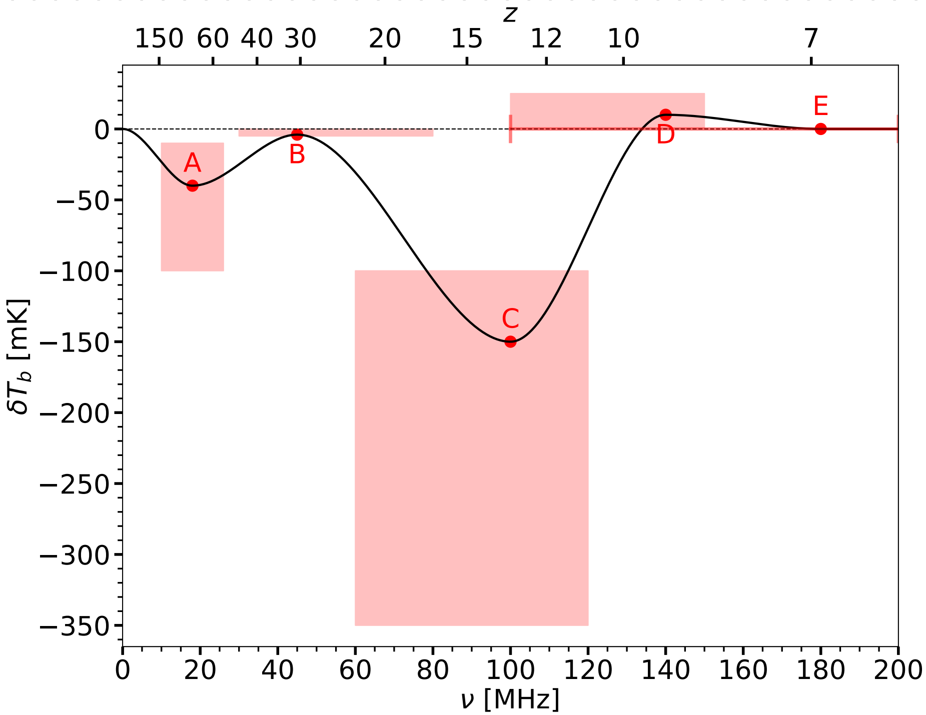
# MOTIVATION FOR MODELS: HYDROGEN COSMOLOGY THEORY



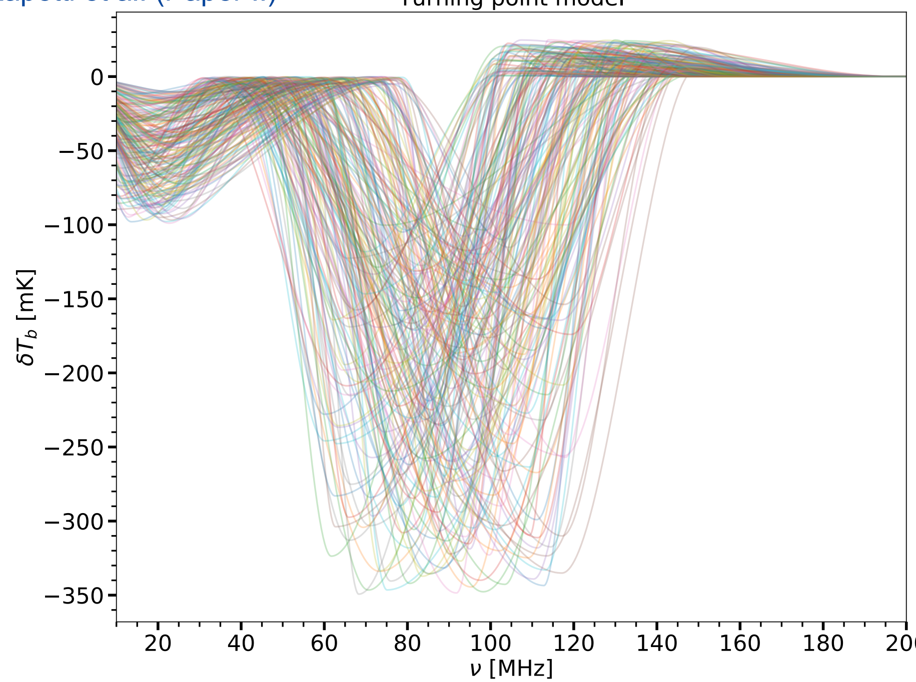

Burns et al. (2017)

- d$T_b$ is a combination of temperatures: $T_S$ spin, $T_k$ kinetic, $T_\alpha$ Lyman-a, $T_\gamma$ background (CMB).
- A: Expansion recouples $T_S \rightarrow T_\gamma$
- B: First stars Ly-a emission couples back $T_S \rightarrow T_k$
- C: Heating sources including initial black hole accretion drive $T_k \rightarrow T_\gamma$
- D: Reionization removes signal ($x_{HI} \rightarrow 0$).

# GLOBAL 21-CM MODELS: TURNING POINT

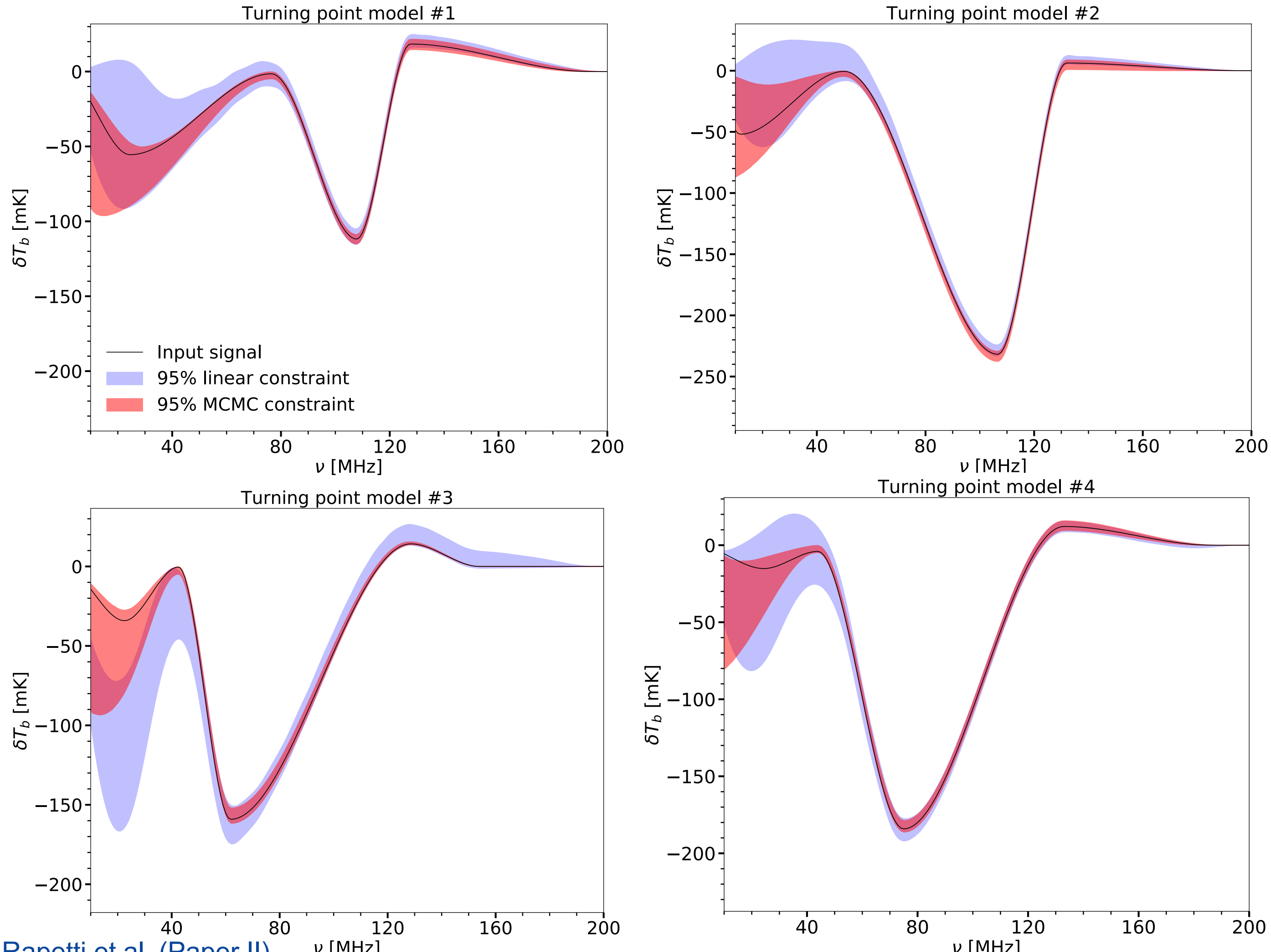



Rapetti et al. (Paper II)




- **Black line**: typical model with movable red dots defining a spline interpolation, for which $\delta T_b$ and its derivative are also fixed to 0 at ν= 0.
- Broadly speaking, A represents Dark Ages, B Cosmic Dawn, C-D the epoch of heating & D-E the epoch of reionization.
- The red filled regions around A-D show the allowed positions of the points to build the training set (200 samples, right panel). The red horizontal line with vertical bars on its ends marks the same for E.
- Adjacent frequencies are forced to be at least 10 MHz apart.

# TURNING POINT MODEL:
# LINEAR AND MCMC SPECTRAL CONSTRAINTS




Rapetti et al. (Paper II)

Pipeline spectral constraints for four random cases successfully recovered:

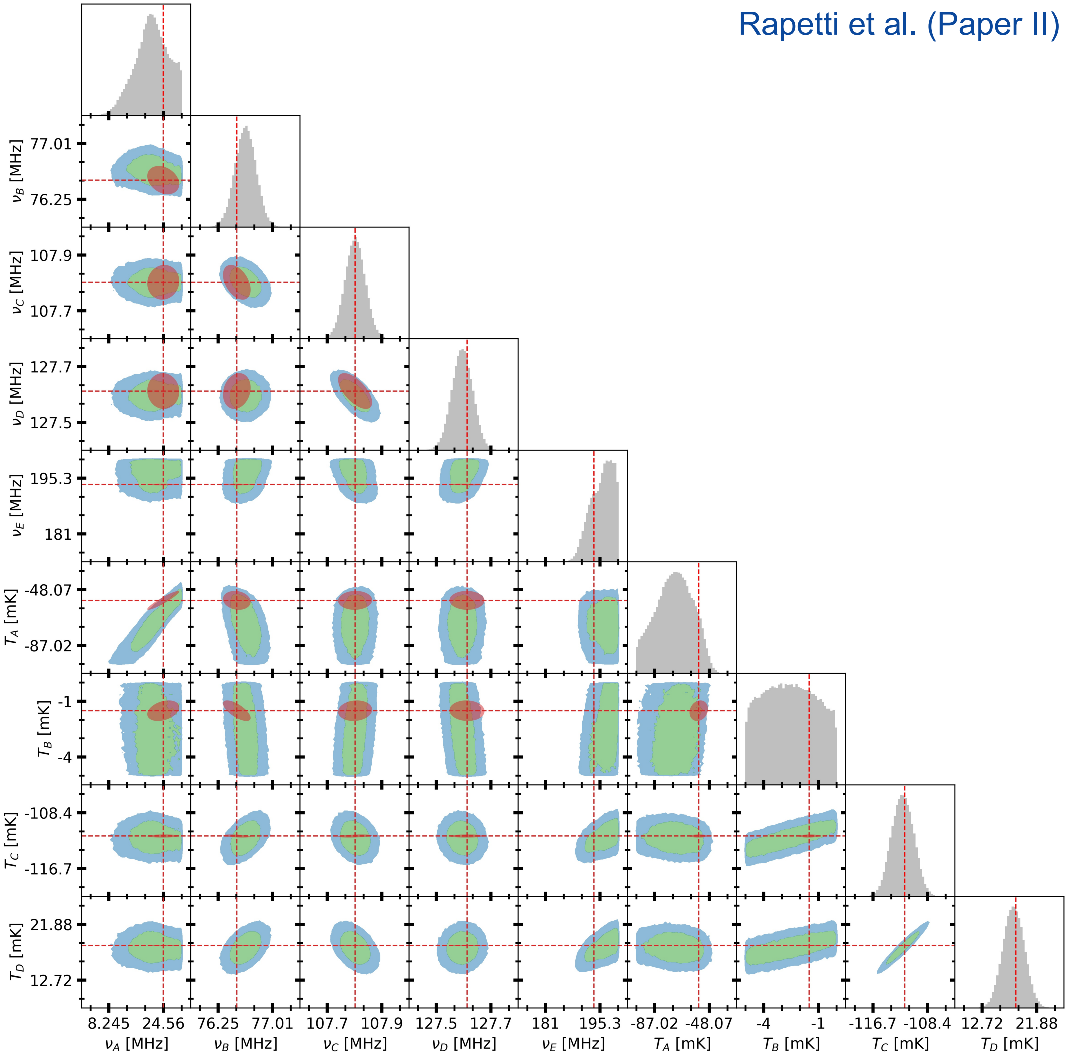
Blue bands: 95% confidence intervals from the linear fit, with SVD signal and foreground modes.

Red bands: 95% confidence intervals from the MCMC fit, with the full nonlinear signal model and SVD foreground modes.

For the linear fit, the 95% intervals correspond to 2.5σ.

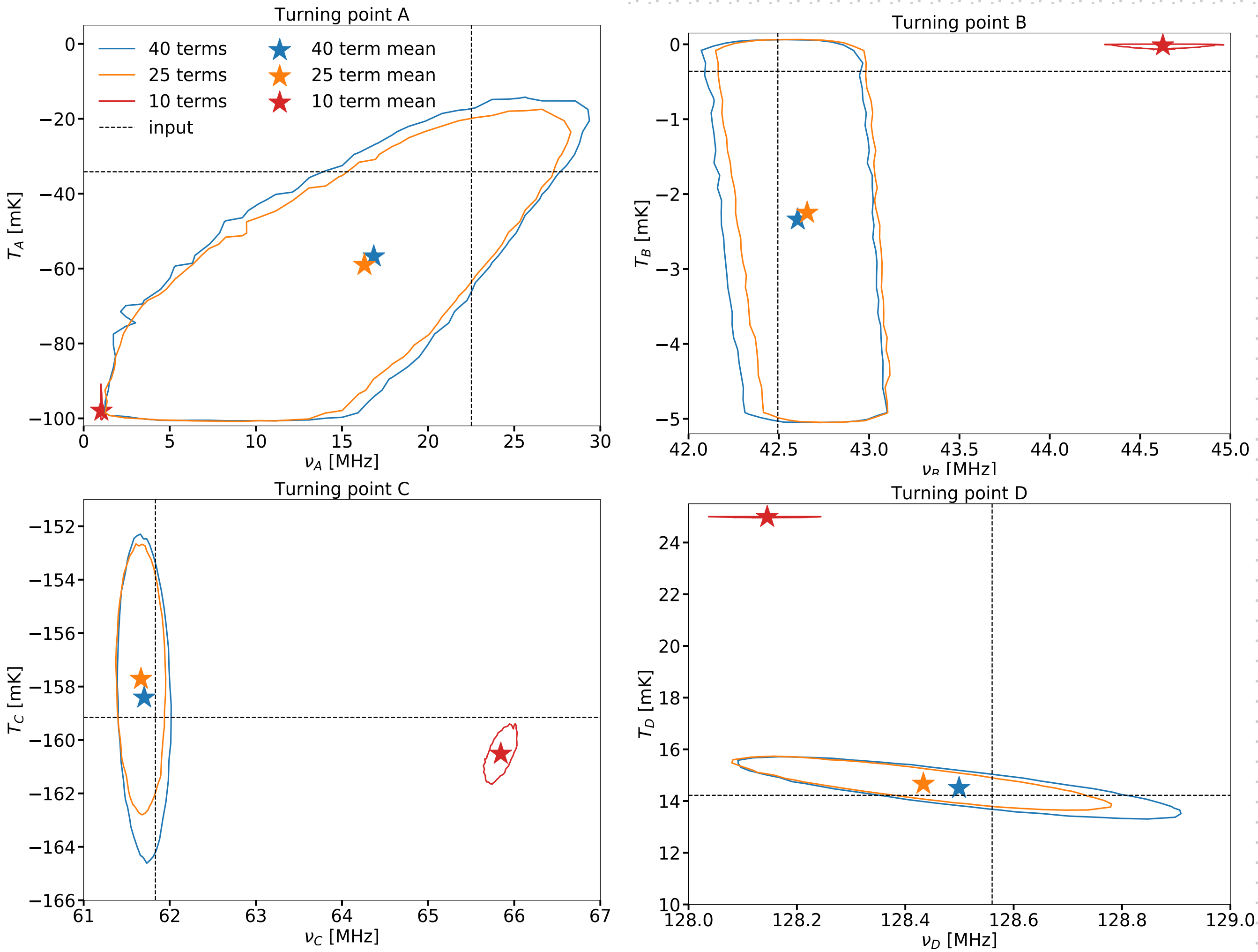
# TURNING POINT MODEL: FULL MCMC PARAMETER CONSTRAINTS

Rapetti et al. (Paper II)




- 1D (gray) and 2D (68/95%) MCMC posterior parameter constraints.
- The red, dashed lines mark the input parameters.
- The red ellipses represent 95% confidence contours when only the statistical noise (Fisher-matrix estimated) obscures the signal.
- All intervals assume 800 hours of integration.
- Note e.g. that the temperature of turning point B, allowed to only vary from -5 to 0 mK, is not constrained within the prior, and the temperature C is well constrained.

# TURNING POINT MODEL: NUMBER OF MARGINALIZED FOREGROUND PARAMETERS




Rapetti et al. (Paper II)

- 95% constraints on A-D if marginalizing over 10 (red), 25 (orange) & 40 (blue) terms in the MCMC fit.

- The dashed lines indicate the input parameters.

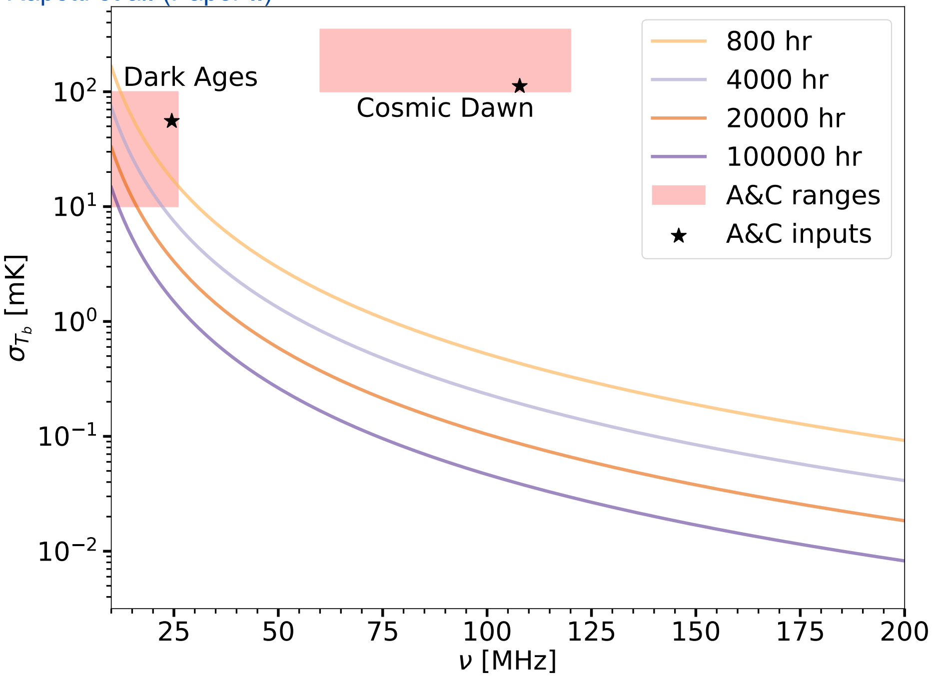
- 10 terms are not sufficient to explain the foreground (in the linear fit, 24 were chosen), so the signal is biased with spuriously tight constraints.

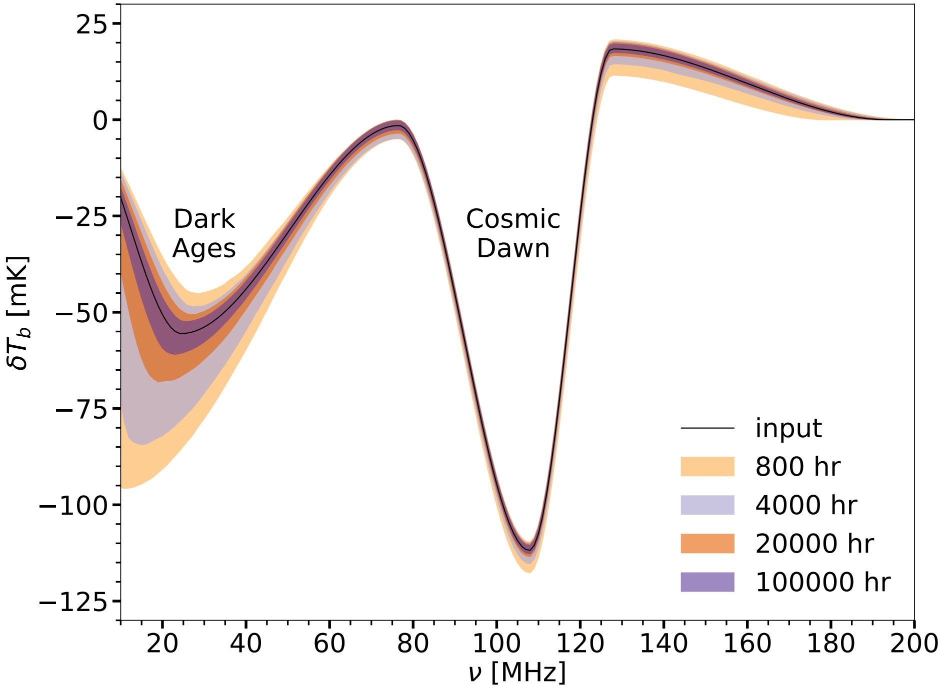
- For 25 terms, the signal is recovered with realistic errors.

- For 40 terms, there is no qualitative change thanks to the use of foreground priors.

# TURNING POINT MODEL: INCREASING THE INTEGRATION TIME

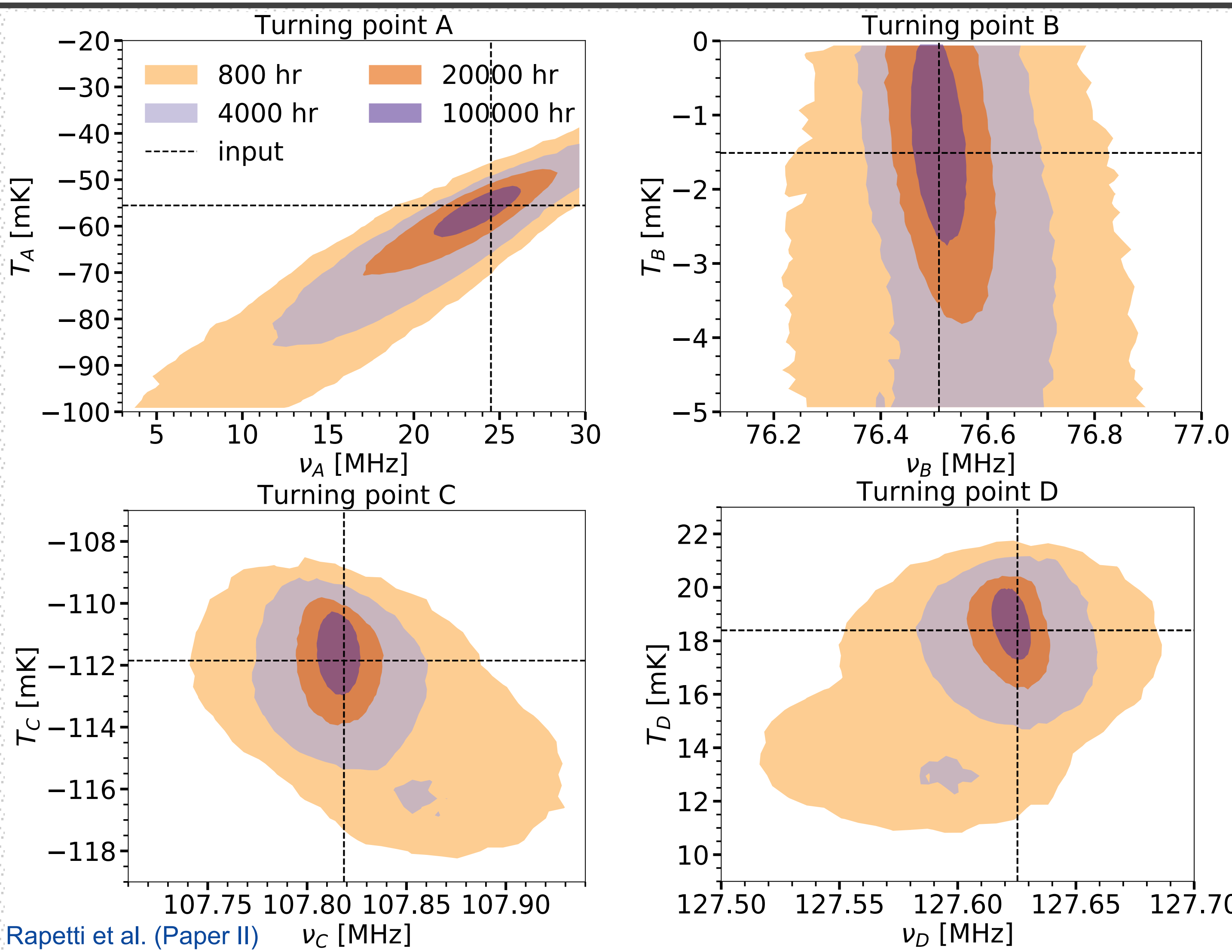
Rapetti et al. (Paper II)




- Left: 1σ noise levels for the factor of 5 increases of integration times from our reference of 800 hours. The red rectangles indicate the allowed values for turning points A (Dark Ages) and C (Cosmic Dawn) and the **black stars** the **input values**.
- Right : Full uncertainties in frequency space for four different integration times with the same random seed for noise generation.

# TURNING POINT MODEL: INCREASING THE INTEGRATION TIME




Rapetti et al. (Paper II)

# SUMMARY

- We employ a linear, fast, analytic methodology to separate the global 21-cm signal from systematics, with which it can have large overlaps, to estimate the starting point of a full MCMC search of any selected physical signal model.
- We utilize the linear SVD foreground terms to properly and efficiently (in terms of convergence) account for this modeling by marginalizing over these generally large number of parameters at each step of our MCMC signal calculation.
- We test our novel pipeline on two physically motivated signal models, flattened Gaussian (observationally based) and turning point (theoretical), and successfully recover the input parameters for multiple random cases.
- EDGES, CTP, DAPPER, FARSIDE, etc. measurements should benefit from this statistically rigorous, robust pipeline which is able to extract the 21-cm signal while modeling the systematics using detailed training sets from theory, simulations and observations.