Robust and Unbiased Analyses for Planned Low Radio Frequency Observations from the Lunar Surface

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

In collaboration with:
Joshua Hibbard, Neil Bassett, Johnny Dorigo Jones, Jack Burns
University of Colorado Boulder

GLOBAL 21-CM PIPELINE FOR INSTRUMENT DESIGN & DATA ANALYSIS

- The pipeline generates and uses sets of realizations for each systematic and the signal based on data and/or modelling (analytical; simulations) to fit observations
- This is particularly important when:
  - These components are difficult to physically measure or simulate at the level required, such as for the sky foreground and the antenna beam
  - The signal model to constrain cosmological and astrophysical parameters is unknown
- Proper goodness-of-fit evaluation is crucial to determine if the modeling sets are valid, specially when signal and systematics overlap and can compensate each other

Publicly available pipeline software for global 21-cm experiments:
Systematics removal & parameter constraints, pylinex https://github.com/CU-NESS/pylinex
Beam-weighted foreground modeling, perses https://github.com/CU-NESS/perses
Global 21-cm signal models, ares https://github.com/mirochaj/ares

EXPLORING SYSTEMATICS: LUNAR SUBSURFACE, ANTENA BEAM, SKY FOREGROUND

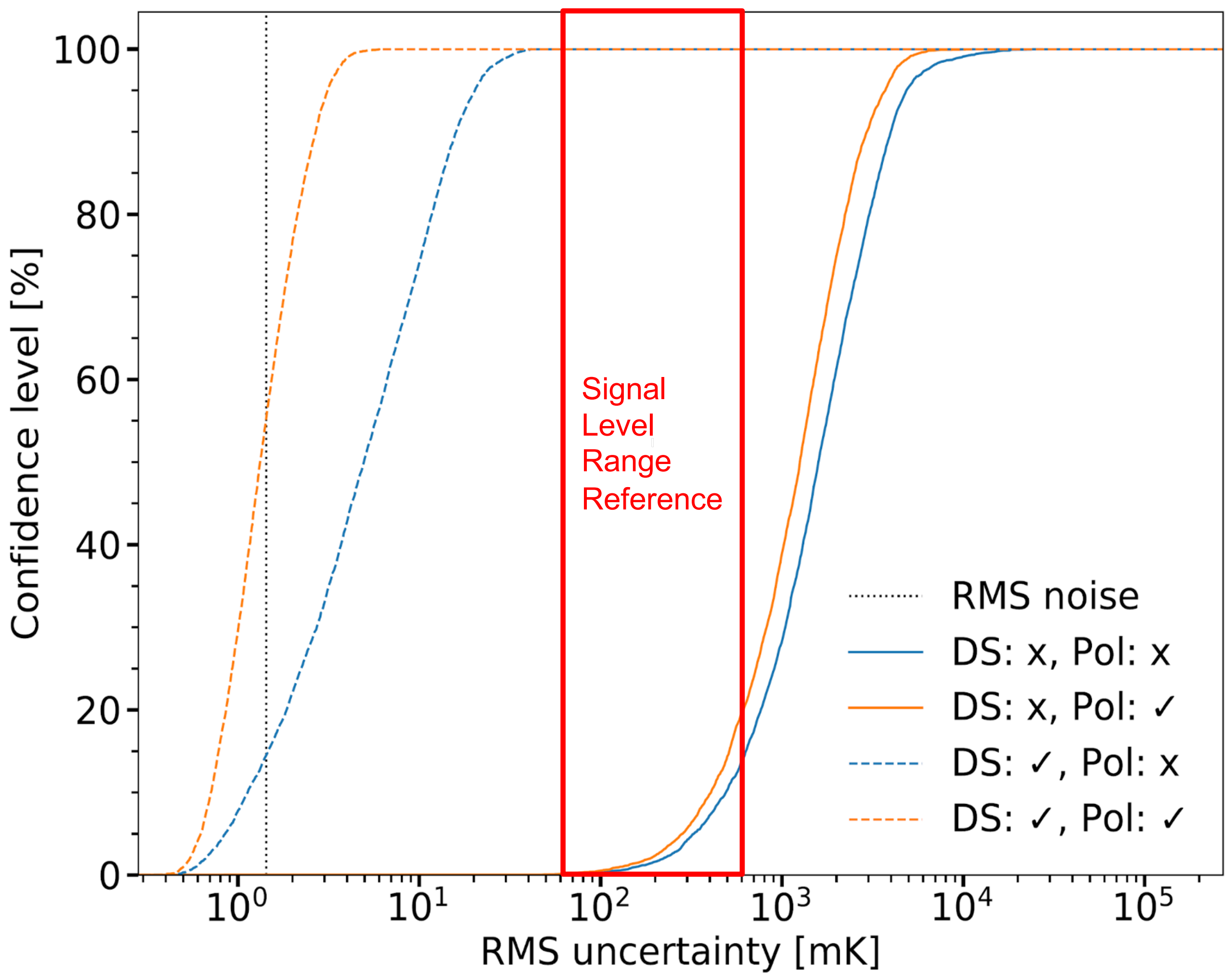
Radio wave Observations at the Lunar Surface of the photo-Electron Sheath (ROLSES)
NASA PI: Nat Gopalswamy
Illustration credit: Intuitive Machines

(See e.g., Burns et al. 2021, Planetary Science Journal, 2, 44B)

Lunar Surface Electromagnetics Experiment (LuSEE-Night)
NASA PI: Stuart Bale
Illustration credit: Firefly Aerospace

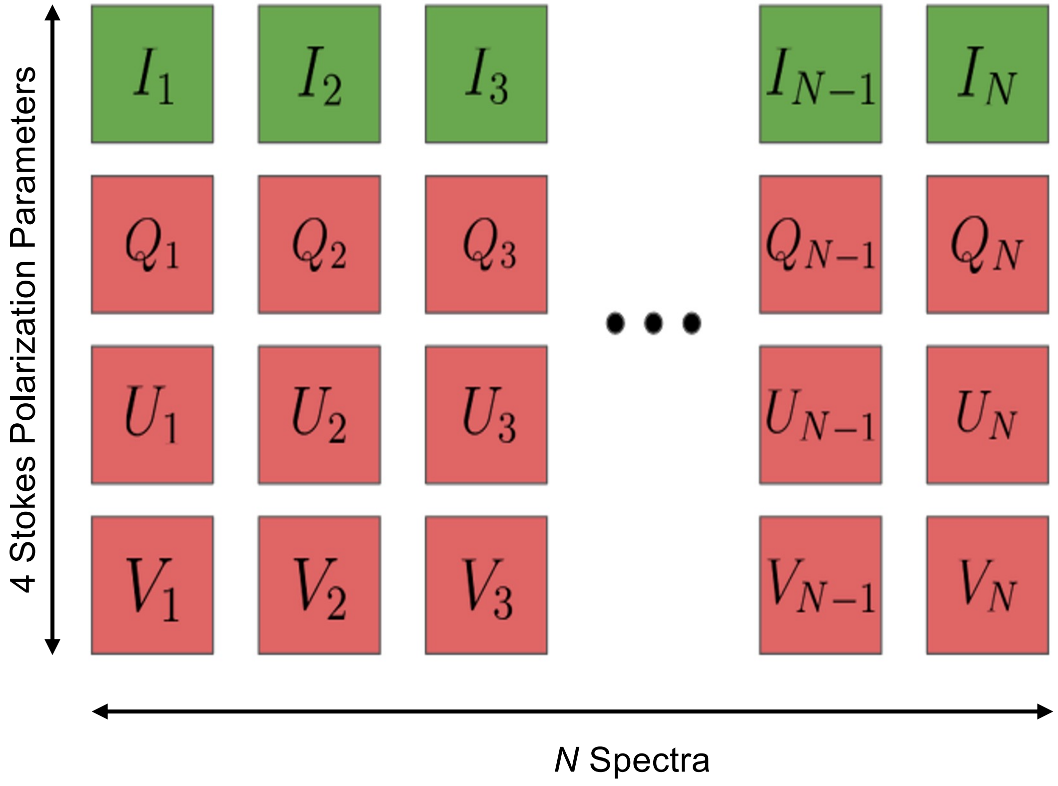
(See Bale et al. 2023, arXiv:2301.10345)

USING CORRELATED SPECTRA TO ROBUSTLY AND TIGHTLY CONSTRAIN UNCERTAINTIES

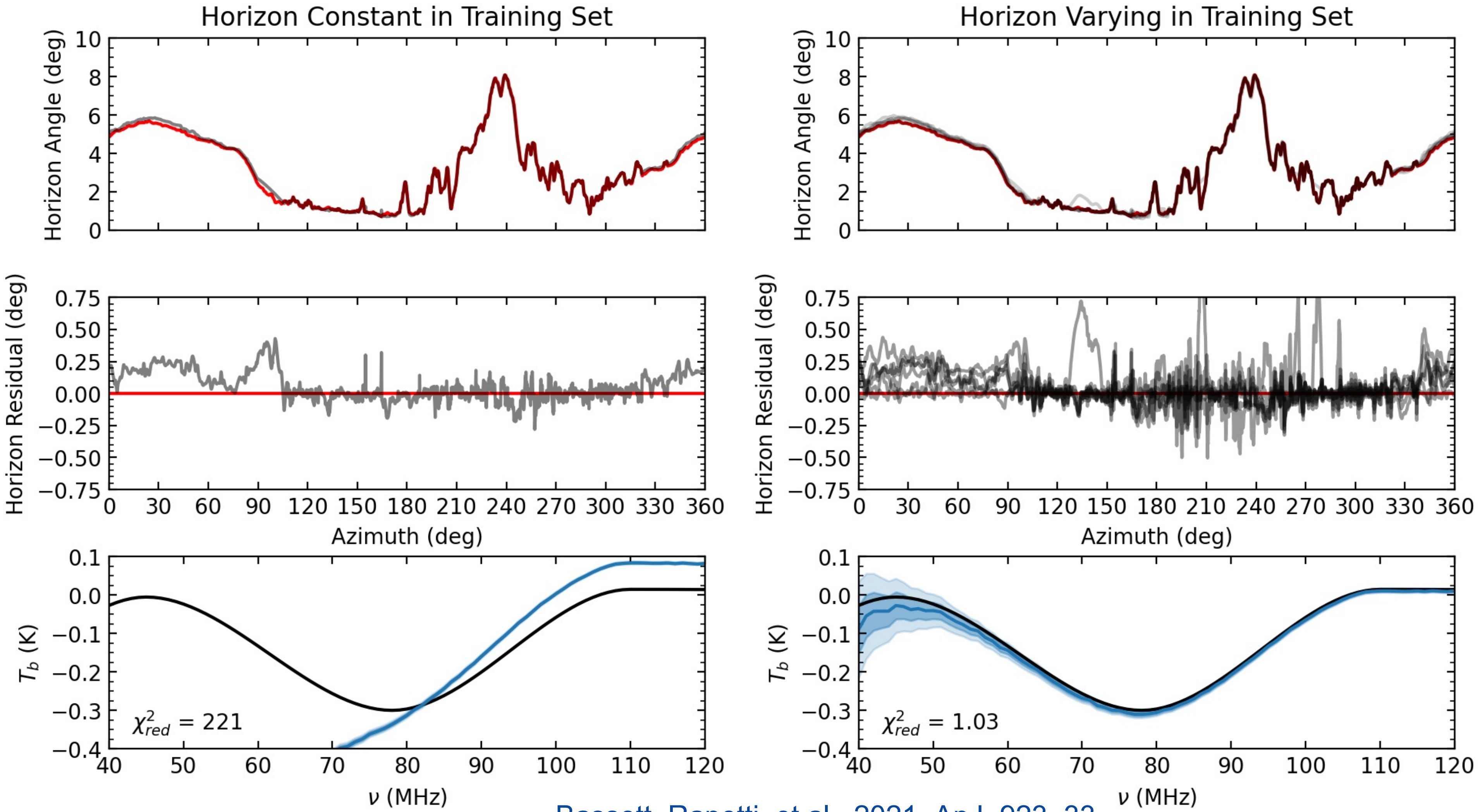


Tauscher, Rapetti, Burns (2020; Paper III)

Uncertainty of signal extraction depends primarily on overlap between foreground and signal models. This overlap can be decreased by using: 1) Polarization and 2) Many Correlated Spectra



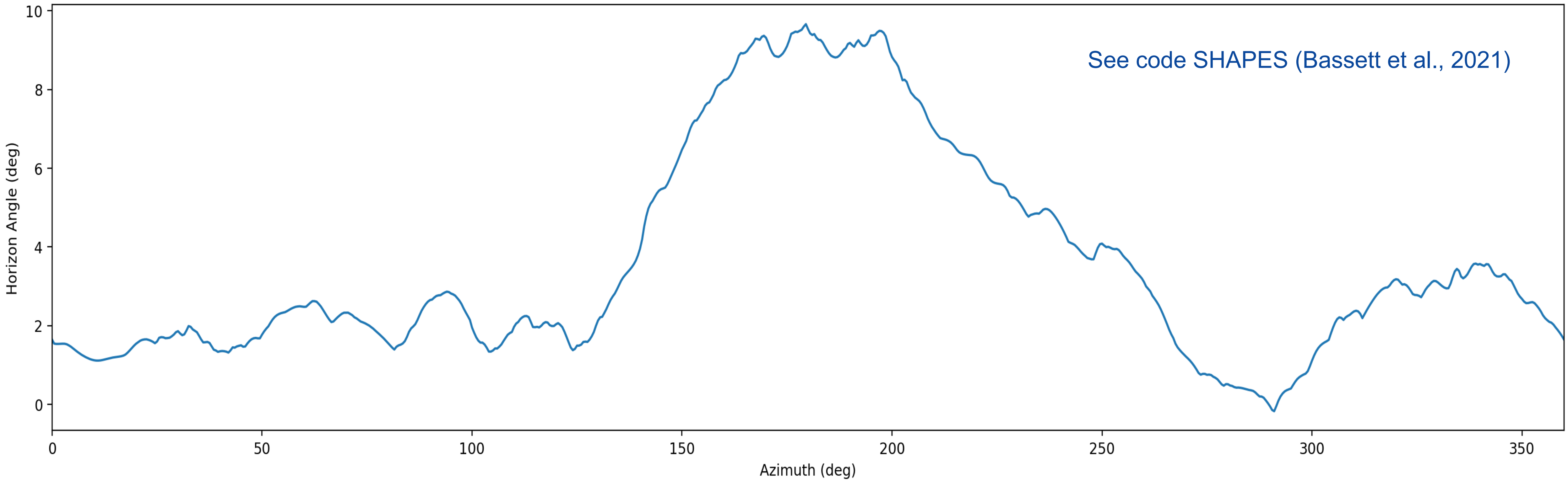
EXAMPLE OF A REALIZATION SET TO ENCAPSULATE LUNAR HORIZON UNCERTAINTIES



Bassett, Rapetti, et al., 2021, ApJ, 923, 33

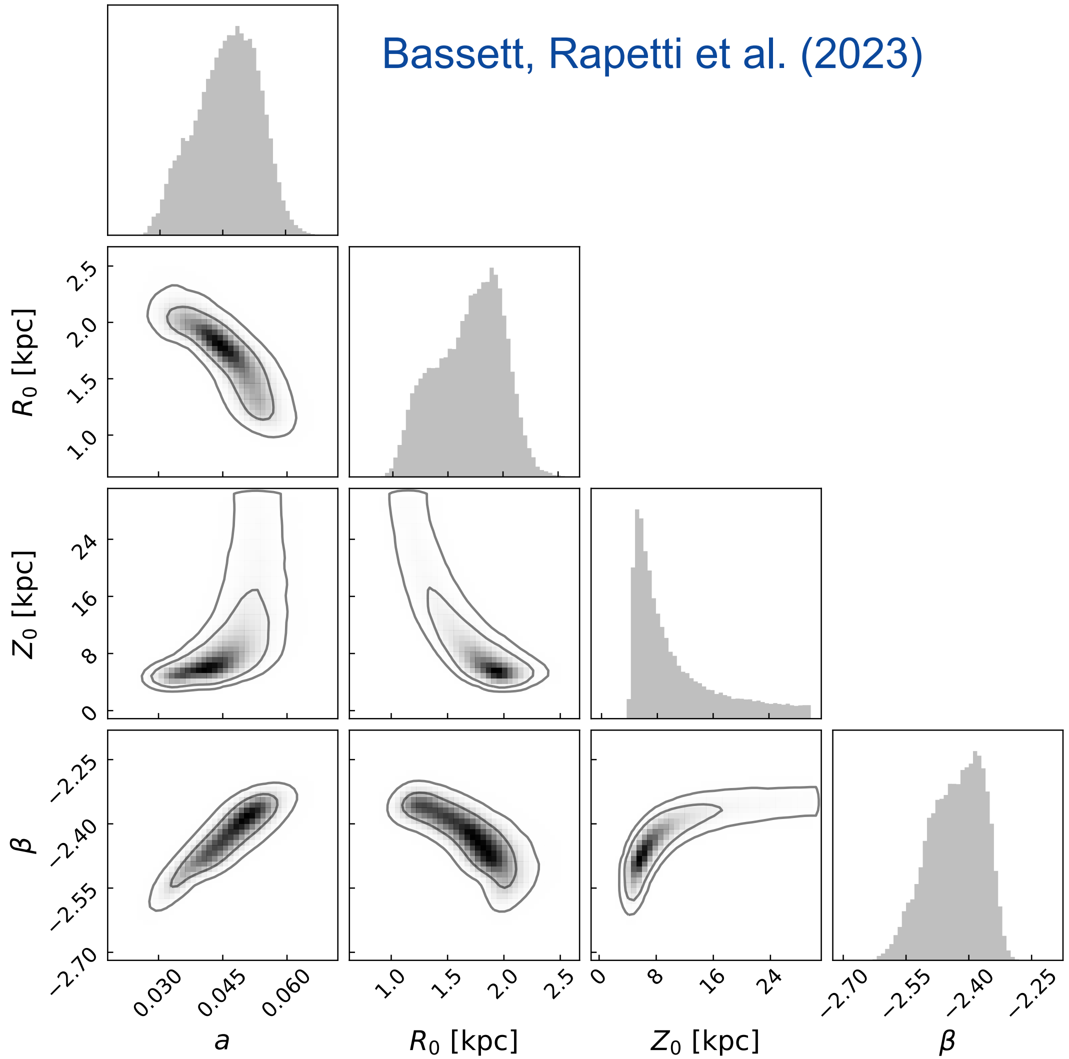
- Inevitably, there will be uncertainty in the horizon profile due to measurement error, uncertainty in location of instrument, etc.
- If horizon is assumed to have a shape that is incorrect, we cannot accurately extract the signal (left panels)
- But, if several realizations are included in a training set that encompass uncertainty, extraction greatly improves (right panels)
- Simulated global 21-cm signal extractions in the bottom panels

HORIZON FROM IM-1 LANDING SITE AT THE MOON'S SOUTH POLE AND ROLSES GOALS



- Determine the electron sheath density from ~1 to ~3 m above the lunar surface by measuring electron plasma frequency.
- Demonstrate detection of solar, planetary, & other radio emission from lunar surface.
- Explore Galaxy radio spectrum at <30 MHz.
- Aid development of lunar radio arrays.
- Measure the local EM environment, including that from the lander.
- Measure reflection of incoming radio emission from lunar surface and below.

MEASURING SYSTEMATICS: BEAM-WEIGHTED FOREGROUND



Bassett, Rapetti et al. (2023)

- Parker Solar Probe/FIELDS consists of 4 antennas
- We used FIELDS spectral measurements over rotation maneuvers to investigate the low frequency sky between 1 and 6 MHz
- Phase of rotation can be compared to time (LST) for ground-based experiments
- The figure shows roll average simultaneous fit posterior parameter constraints from 5 different days
- Modelling foreground emission using a modified version of the ULSA model from Cong et al. (2021), and free-free absorption using the Yao, Manchester & Wang (2017) model of the free electron density from pulsar measurements
- The constraints on the filling factor/absorption parameter *a* are in good agreement with the Gaensler et al. (2008) estimate of *a* = 0.04 +/- 0.01
- Based on a fit to the Haslam map at 408 MHz, Cong et al. (2021) estimated R_0 = 3.41 kpc, Z_0 = 1.12 kpc
- Our fit prefers larger values of Z_0 and Z_0 > R_0
- The spectral index is consistent with the fiducial value of -2.5

EXPLORING SYSTEMATICS: BEAM-WEIGHTED FOREGROUND

Model	Symbol	LSTs	Generating Equation	Θ	N_Θ	Constraints	Priors
Nonlinear	$\boldsymbol{\mathcal{M}}_{\rm nl}$	1,2, 5,10	$\sum_{j}^{N_r} K_j(\nu) A_j \left(\frac{\nu}{\nu_0}\right)^{\beta_j+\gamma_j \ln(\nu/\nu_0)} + T_{\rm CMB}$ (8)	β_j A_j γ_j	$(1,2,3)$ $\times N_r$	–	$\beta_j \sim U(-4.5,-2.0)$ $A_j \sim U(0.1,10)$ $\gamma_j \sim U(-0.1,0.1)$
Linear	$\boldsymbol{\mathcal{M}}_{\rm lin}$	1,2, 5,10	$\boldsymbol{F}_{\rm fg}\boldsymbol{x}_{\rm fg}$ (9)	$x^k_{\rm fg}$	N_x	$\boldsymbol{F}_{\rm fg} = {\rm cols}(\boldsymbol{U})$, $\boldsymbol{B}_{\rm fg} = \boldsymbol{U\Sigma V}^T$ and ${\rm col}(\boldsymbol{B}_{\rm fg})_i = \boldsymbol{\mathcal{M}}_{\rm nl}(\boldsymbol{\theta}_i)$	$\boldsymbol{\pi}_{\rm fg} \sim \mathcal{N}(\boldsymbol{\nu};\boldsymbol{\Lambda})$
LinLogPoly	$\boldsymbol{\mathcal{M}}_{\rm LinLogPoly}$	1	$\left(\frac{\nu}{\nu_0}\right)^{-2.5} \sum_{k=1}^{N_{py}} a_k \left[\ln \frac{\nu}{\nu_0}\right]^k$ (10)	a_k	N_{py}	–	$\boldsymbol{\pi}_{\rm poly} \sim \mathcal{N}(0;\sigma^2_{\rm poly})$
LinPoly	$\boldsymbol{\mathcal{M}}_{\rm LinPoly}$	1	$\left(\frac{\nu}{\nu_0}\right)^{-2.5} \sum_{k=1}^{N_{py}} a_k \left(\frac{\nu}{\nu_0}\right)^k$ (11)	a_k	N_{py}	–	$\boldsymbol{\pi}_{\rm poly} \sim \mathcal{N}(0;\sigma^2_{\rm poly})$
LinPhys	$\boldsymbol{\mathcal{M}}_{\rm LinPhys}$	1	$\left(\frac{\nu}{\nu_0}\right)^{-2.5} \sum_{k=1}^{2} a_k \left(\ln \frac{\nu}{\nu_0}\right)^k + a_3 \left(\frac{\nu}{\nu_0}\right)^{-4.5} + a_4 \left(\frac{\nu}{\nu_0}\right)^{-2}$ (12)	a_k	5	–	$\boldsymbol{\pi}_{\rm poly} \sim \mathcal{N}(0;\sigma^2_{\rm poly})$
MSF DiffPoly	$\boldsymbol{\mathcal{M}}_{MDP}$	1	$\sum_{k=1}^{N_{MSF}} a_k(\nu-\nu_0)^k$ (13)	a_k	N_{MSF}	$\boldsymbol{Ga} \leq \boldsymbol{0}$	–
MSF LogLogPoly	$\boldsymbol{\mathcal{M}}_{MLLP}$	1	$10^{\sum_{k=1}^{N_{MSF}} a_k(\log_{10}\nu)^k}$ (14)	a_k	N_{MSF}	$\boldsymbol{Ga} \leq \boldsymbol{0}$	–

- Seven commonly employed foreground models: 2 forward-models, one nonlinear & the other linear; 5 phenomenological models, three polynomials (linear) and 2 maximally-smooth polynomials (nonlinear)
- Used to fit simulated mock spectra built from intrinsic foregrounds with realistic spatial and spectral structure, chromatic beams, horizon profiles, and discrete time-sampling

Hibbard, Rapetti, et al., 2023, submitted to ApJ

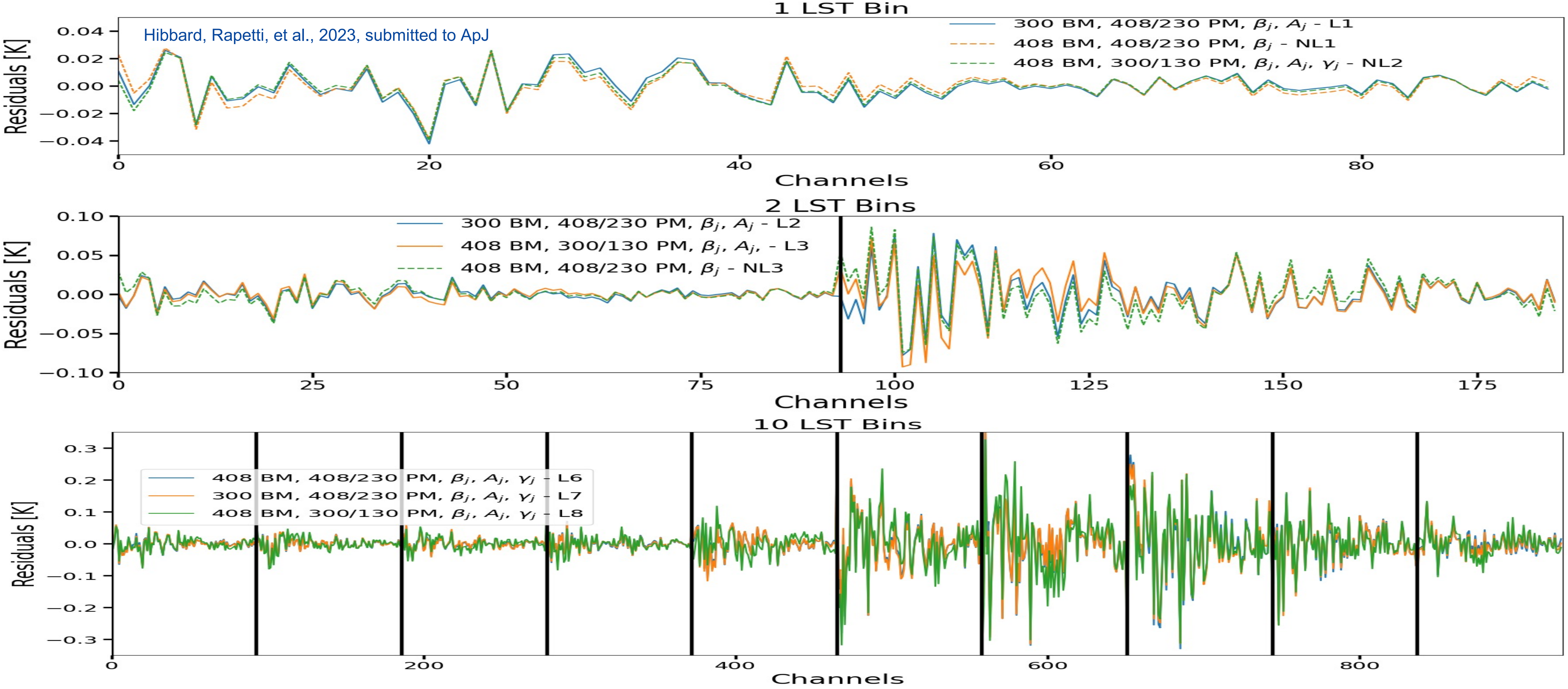
EXPLORING SYSTEMATICS: BEAM-WEIGHTED FOREGROUND

Model			Linear							Nonlinear					
LSTs	Input	θ_j	N_r	χ^2_{red}	σ	ln Z	N_x	p_{ks}	BF	N_r	χ^2_{red}	ln Z	N_θ	p_{ks}	BF
1	IDEAL	β_j	4	0.96	1	-615.71	7	0.94	—	4	1.02	-81.05	4	0.99	NL1
		β_j, A_j	3	1.14	1	-625.24	6	0.97	—	4	1.02	-105.78	8	0.97	—
		β_j, A_j, γ_j	2	1.14	1	-628.66	6	0.90	—	4	1.18	-119.07	12	0.88	—
	300 BM	β_j	4	0.96	1	-614.26	7	0.93	—	4	1754	-7.8e4	4	1.36e-24	—
		β_j, A_j	3	1.10	1	-623.23	6	0.99	L1	4	1.02	-107.39	8	0.98	—
		β_j, A_j, γ_j	2	1.12	1	-626.91	6	0.95	—	4	1.10	-98.22	12	0.93	—
	(300/130) PM	β_j	4	0.99	1	-613.8	5	0.998	—	4	3.9e10	-1.74e12	4	~ 0	—
		β_j, A_j	3	1.06	1	-618.83	8	0.998	—	4	1.03	-112.26	8	0.94	—
		β_j, A_j, γ_j	2	1.09	1	-623.53	6	0.88	—	4	1.09	-112.71	12	0.999	NL2
2	IDEAL	β_j	5	1.00	1	-1129.88	9	0.68	—	8	1.13	-154.24	8	0.81	—
		β_j, A_j	4	1.02	1	-1160.86	13	0.81	—	8	1.03	-335.72	16	0.83	NL3
		β_j, A_j, γ_j	15	1.03	1	-1192.70	18	0.80	—	8	1.53	-307.33	24	0.09	—
	300 BM	β_j	5	1.03	1	-1130.74	13	0.51	—	8	7.83e4	-6.96e6	8	5.6e-45	—
		β_j, A_j	4	1.02	1	-1155.56	13	0.89	L2	8	1.15	-364.17	16	0.82	—
		β_j, A_j, γ_j	4	1.08	1	-1189.71	14	0.78	—	8	1.27	-234.77	24	0.50	—
	(300/130) PM	β_j	6	1.13	1	-1134.67	21	0.71	—	8	2.1e10	-1.82e12	8	~ 0	—
		β_j, A_j	4	1.00	1	-1142.86	13	0.97	L3	8	3.43	-500.73	16	0.002	—
		β_j, A_j, γ_j	4	1.04	1	-1162.65	14	0.96	—	8	1.34	-220.24	24	0.26	—
5	IDEAL	β_j	8	0.99	1	-2553.1	25	0.85	—	9	8.54	-2034.15	9	1.1e-23	—
		β_j, A_j	9	1.47	7	-3178.06	30	0.14	—	9	1.10	-391.24	18	0.87	—
		β_j, A_j, γ_j	18	1.01	1	-2746.97	37	0.85	—	9	1.10	-506.06	27	0.96	NL4
	300 BM	β_j	8	1.07	1	-2570.87	24	0.97	L4	9	1.43e5	-3.27e7	9	2.5e-127	—
		β_j, A_j	9	1.45	7	-3110.48	30	0.19	—	9	1.14	-467.83	18	0.58	—
		β_j, A_j, γ_j	18	1.01	1	-2789.27	35	0.73	—	9	1.36	-657.18	27	0.19	—
	(300/130) PM	β_j	14	1.03	1	-2642.16	30	0.86	—	9	8.1e9	-1.83e12	9	~ 0	—
		β_j, A_j	10	1.31	5	-2773.33	30	0.68	—	9	914.59	-2.04e5	18	1.2e-96	—
		β_j, A_j, γ_j	12	1.01	1	-2662.62	33	0.93	L5	9	7.9	-2221.81	27	8.6e-20	—
10	IDEAL	β_j	14	1.08	2	-4833.53	36	0.80	—	9	477.92	-2.20e5	9	1.4e-198	—
		β_j, A_j	13	2.58	33	-6783.17	46	1.91e-8	—	9	64.25	-2.9e4	18	7.1e-148	—
		β_j, A_j, γ_j	19	1.04	1	-5059.1	58	0.82	L6	9	27.6	-1.29e4	27	2.5e-107	—
	300 BM	β_j	12	1.40	8	-5004.78	39	0.03	—	9	76973	-3.54e7	9	2.4e-252	—
		β_j, A_j	13	2.39	29	-6696.14	46	1.3e-7	—	9	32.01	-1.48e4	18	1.7e-122	—
		β_j, A_j, γ_j	19	1.07	2	-5183.84	56	0.94	L7	9	19.66	-9134.45	27	8.6e-80	—
	(300/130) PM	β_j	17	2.34	28	-6408.78	54	2.7e-9	—	9	4.0e9	-1.84e12	9	~ 0	—
		β_j, A_j	20	2.02	5	-6060.94	73	1.0e-4	—	9	41717	-1.90e7	18	2.0e-251	—
		β_j, A_j, γ_j	20	1.34	7	-5104.81	79	0.12	L8	9	18109	-8.18e6	27	6.6e-223	—

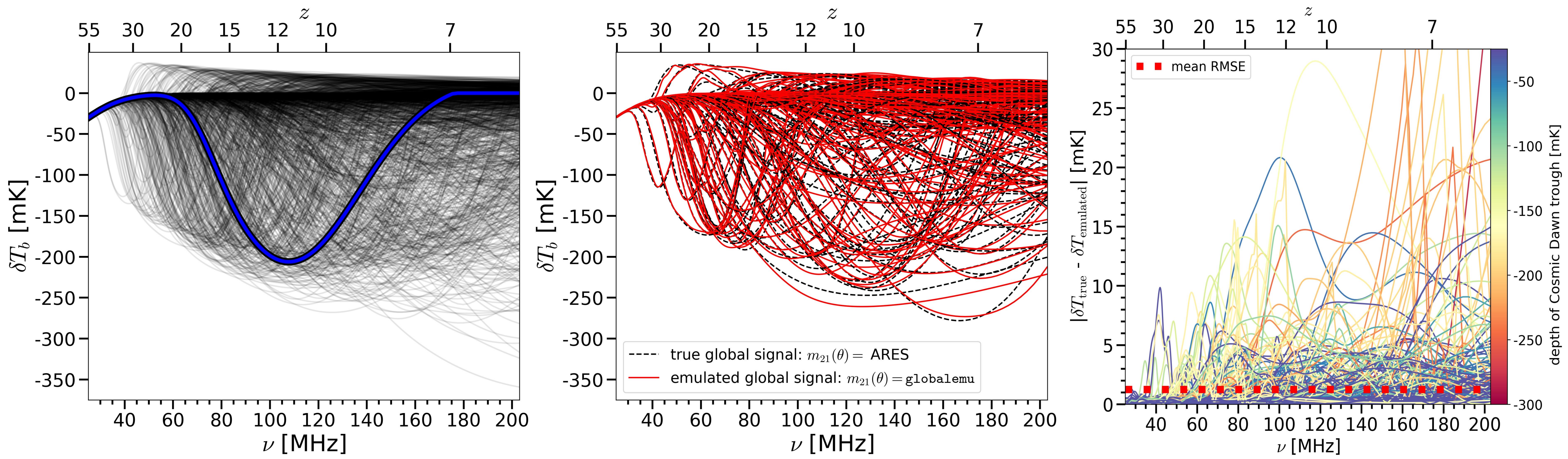
- In each category, the best-fit model based on the KS-test p-value, p_{ks}, is in gold. Model fits which do not pass the null hypothesis exhibit $p_{ks} < 0.05$ and are in gray.
- See Joshua Hibbard's talk yesterday.

Hibbard, Rapetti, et al., 2023, submitted to ApJ

EXPLORING SYSTEMATICS: BEAM-WEIGHTED FOREGROUND



EMULATING ARES GLOBAL SIGNALS WITH GLOBALEMU



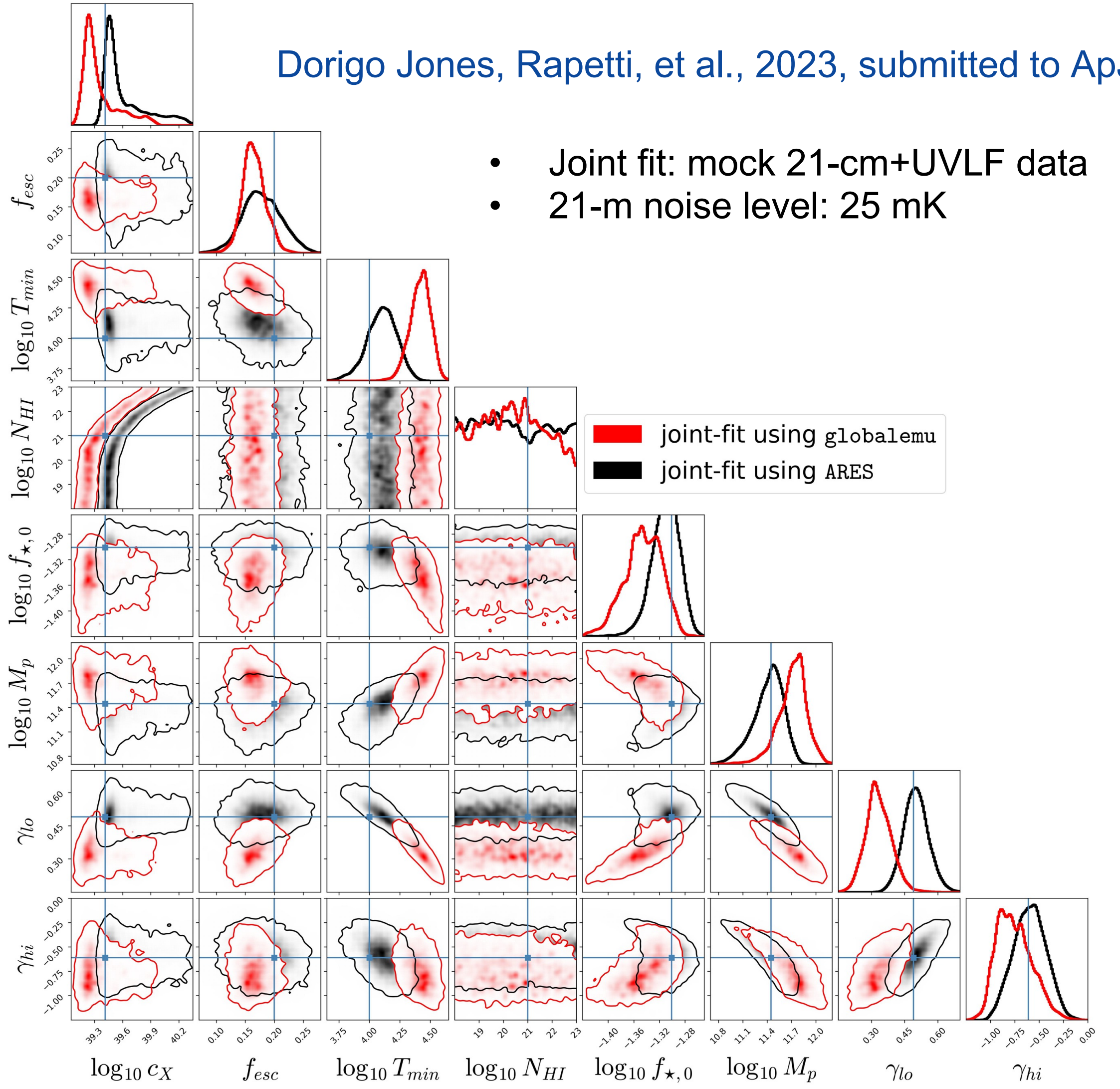
- **Left:** Subset of the training set (10% out of 24,000 total) containing mock global 21-cm signals generated by ARES when varying eight astrophysical parameters. The full training set was used to train globalemu (Bevins et al. 2021). Shown in bolded blue is the fiducial global 21-cm signal.
- Middle: Subset of the test set (200 out of 2,000) generated by ARES ('true' global signals; black, dashed curves) and the corresponding subset of emulations from the globalemu network (solid, red curves) trained on the ARES training set.
- **Right:** Emulation residuals, with color depicting the depth of the Cosmic Dawn trough of the respective signal. The dotted, red line indicates the mean RMSE of 1.25 mK between the emulated and 'true' signals in the full test set.

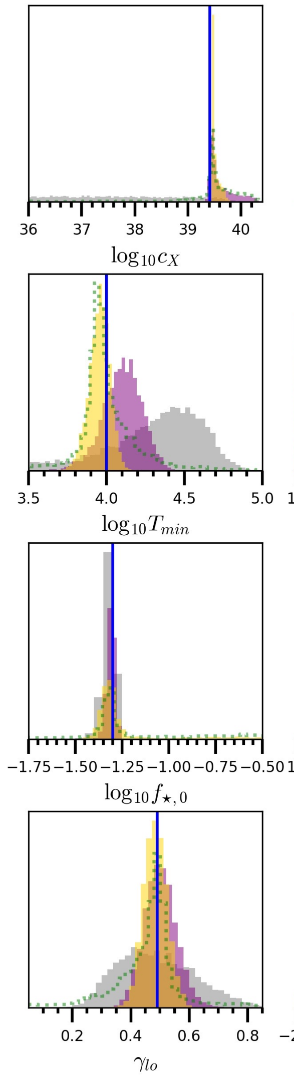
Dorigo Jones, Rapetti, et al., 2023, submitted to ApJ

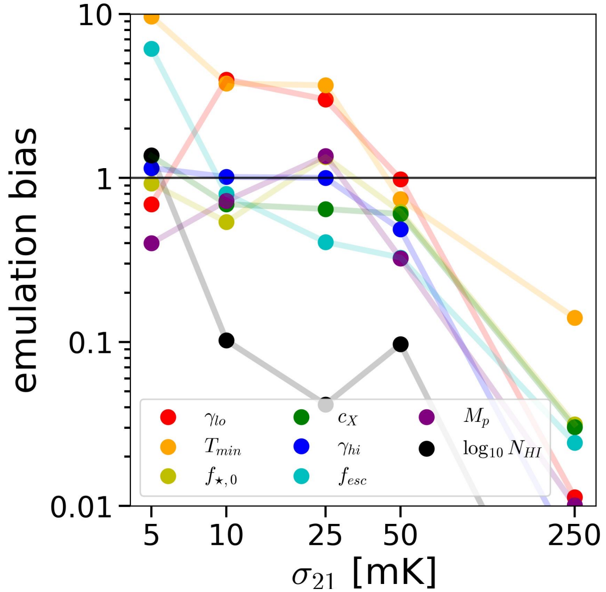
EMULATING ARES GLOBAL SIGNALS WITH GLOBALEMU

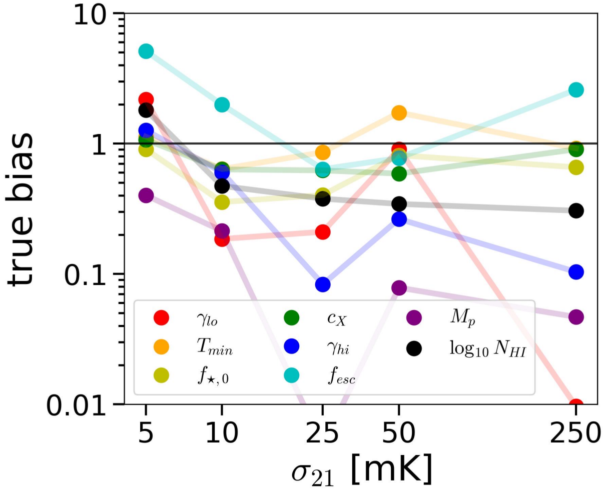
Dorigo Jones, Rapetti, et al., 2023, submitted to ApJ

- Joint fit: mock 21-cm+UVLF data
- 21-m noise level: 25 mK





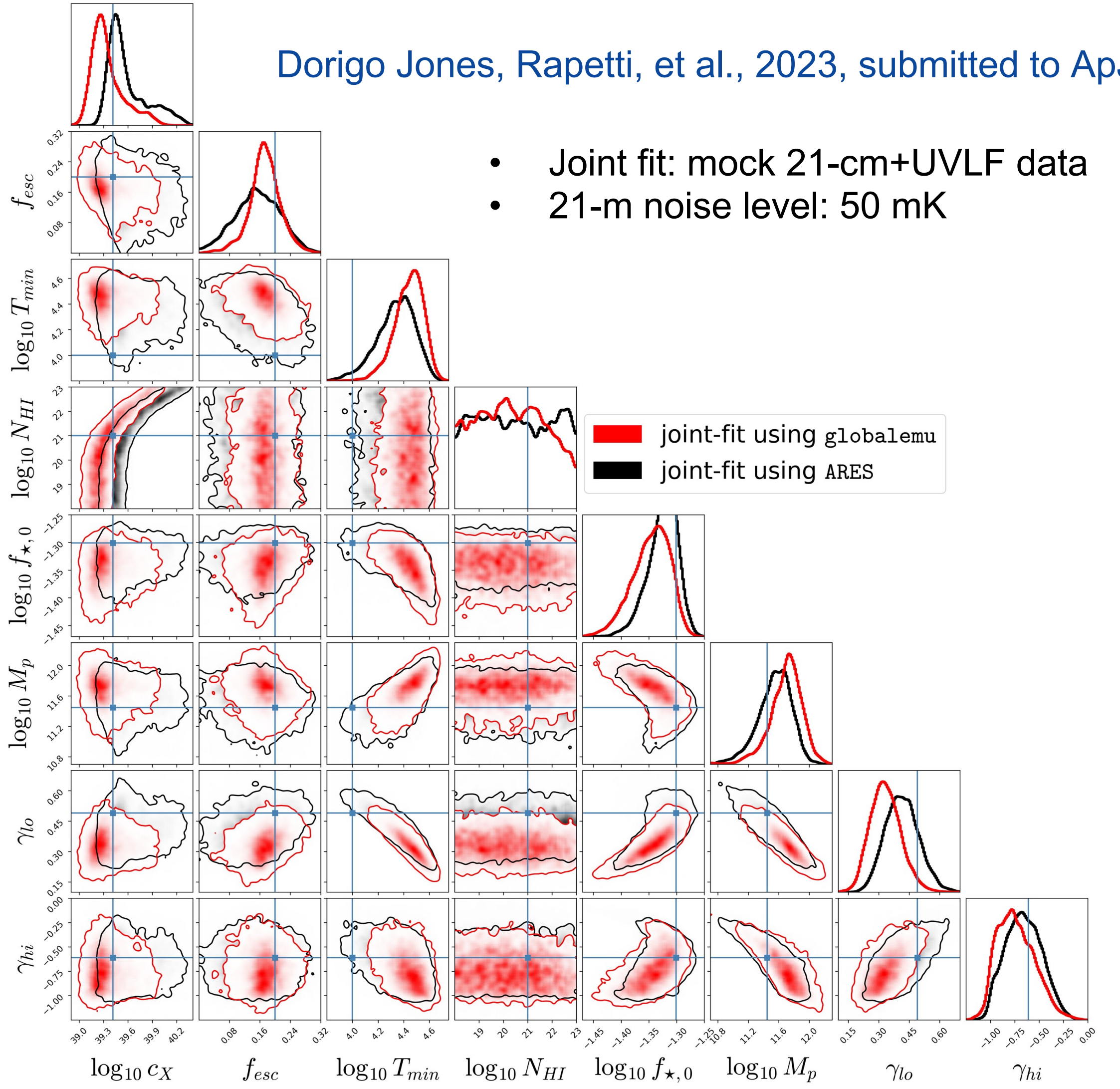


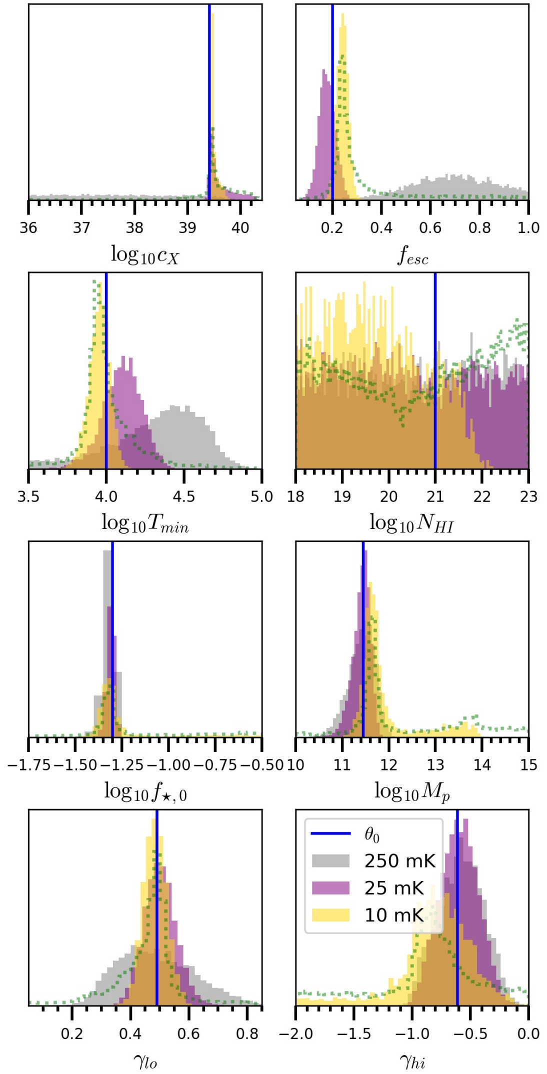


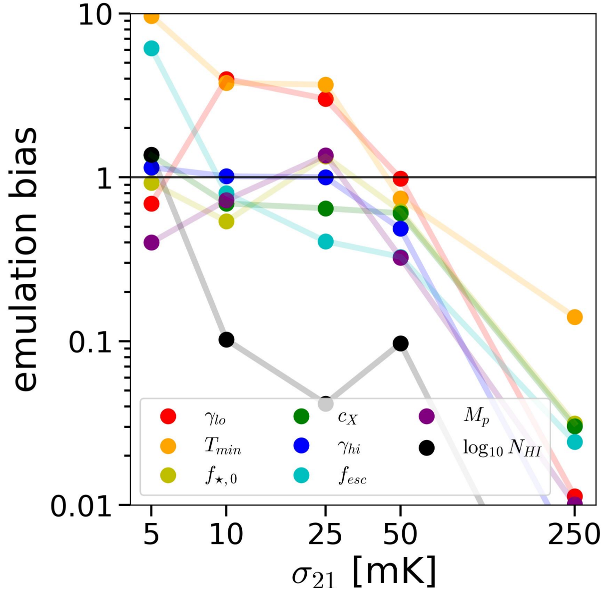
EMULATING ARES GLOBAL SIGNALS WITH GLOBALEMU

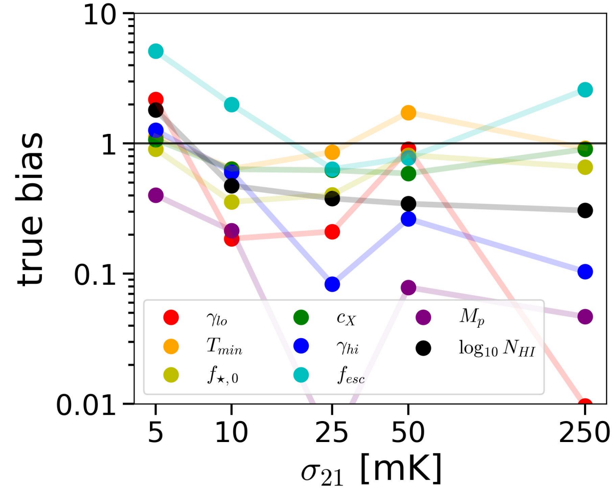
Dorigo Jones, Rapetti, et al., 2023, submitted to ApJ

- Joint fit: mock 21-cm+UVLF data
- 21-m noise level: 50 mK

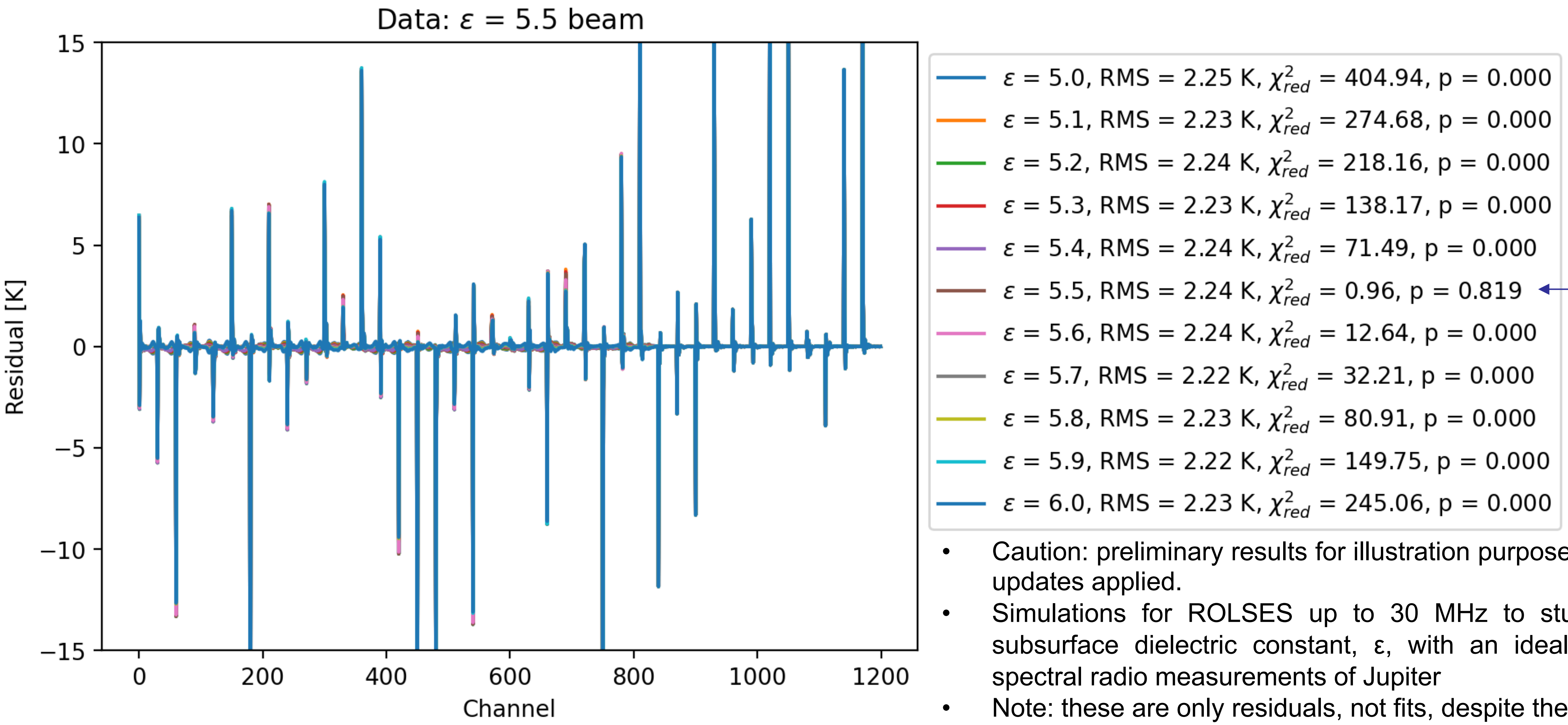








EXPLORING SYSTEMATICS: LUNAR SUBSURFACE (PRELIMINARY)



Credit: Neil Bassett, Joshua Hibbard, Valerie Wong

- Caution: preliminary results for illustration purposes only; recent updates applied.
- Simulations for ROLSES up to 30 MHz to study the lunar subsurface dielectric constant, ε, with an ideal antenna via spectral radio measurements of Jupiter
- Note: these are only residuals, not fits, despite the goodness-of-fit measures employed for guidance

CONCLUSIONS: BENEFITS FROM OUR METHODS

- Modeling sets obtainable from theory, simulations, lab measurements, and observations used to describe and encompass uncertainties

- These modeling sets can be specifically suited for a given experiment, allowing for instance the direct inclusion of complex systematics models, such as from observed foreground maps weighted with detailed beam simulations, avoiding the need for smooth, phenomenological models

- End-to-end simulations and data analyses for ROLSES and LuSEE-Night can thus be carried out via this pipeline. Specific modeling sets for each experiment are required for this purpose

- Accurate models for systematics such as the beam-weighted foreground and properties of the lunar subsurface are critical to describe the data at the required level

- Goodness of fit statistics and strategies to determine the validity of the modeling sets