Data analysis and foreground removal for *DARE* and its prototypes

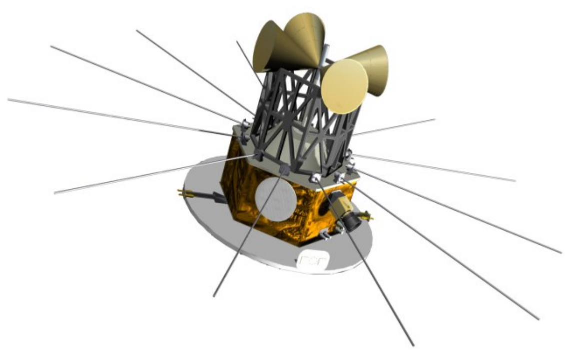
Geraint Harker

Center for Astrophysics and Space Astronomy

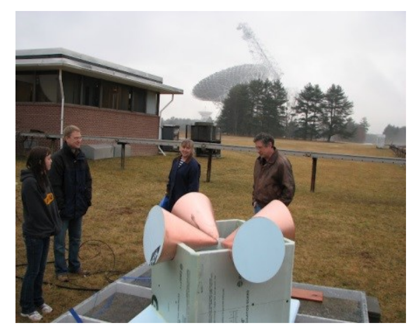
University of Colorado Boulder

The Dark Ages Radio Explorer

- Proposed:
  – Global 21-cm mission.
  – Low lunar orbit, collects science data over the far side at 40-120 MHz, shadowed from RFI from Earth.
  – Deemed selectable in the last Explorer round, will be reproposed for the next one.



- Current status:
  – An initial field test of a *DARE*-like instrument in March 2012 showed effects of RFI and ionosphere.
  – We are about to deploy a next-generation prototype with an updated antenna and system, ready to test our new calibration approach, and study the ionosphere and low-frequency foregrounds in more detail.

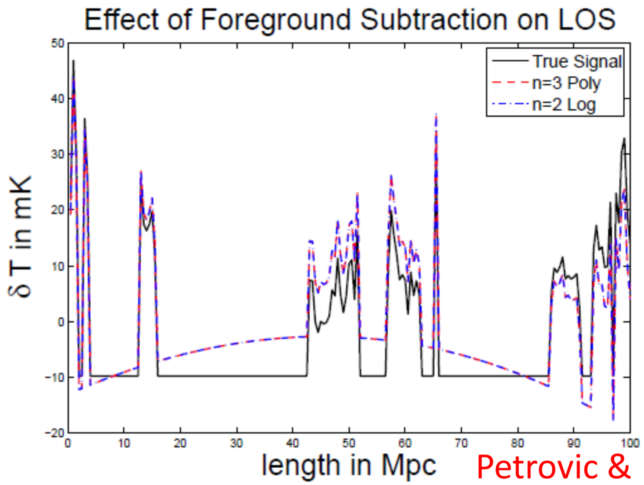


Outline

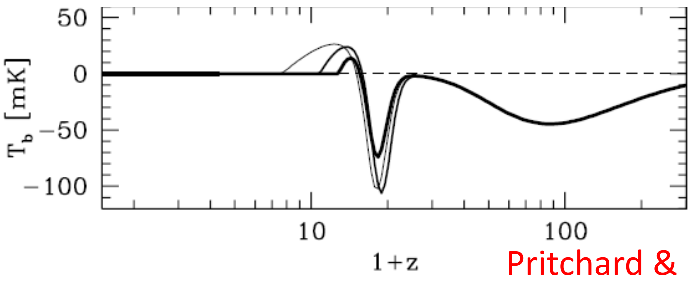
- Unique challenges for global signal experiments.
- Our approach to data analysis for *DARE*.
- Development of our analysis pipeline so far.
- Application to synthetic data, and the effects of the ionosphere.
- Future developments.

Interferometric and sky-averaged 21-cm foregrounds

- For a single pixel in an interferometric map, the signal probably looks quite 'rough':
  - Can exploit the smoothness of the foregrounds, either to remove them or to isolate the 21-cm signal in *k*-space.
  - May also make use of different angular correlations of signal and foregrounds.
- Global signal likely to be much smoother and more degenerate with the foregrounds.
  - Foreground removal definitely required: poor separation in k-space.
  - Some angular resolution may still help.
  - More assumptions required.



Petrovic & Oh (2011)



Pritchard & Loeb (2008)

Features of *DARE* which affect our approach to data analysis

- We want to escape RFI and the ionosphere (see later!), which drives us to propose a mission orbiting the Moon.
- This mandates a fairly compact, mechanically simple, single antenna.
  - Low resolution: expect a beam with an area of about $\pi/2$ sr.
  - Needs to be stable and extremely well-characterized: should be possible from space.
- We can point freely; whole sky accessible at all times.

Analysis pipeline - I

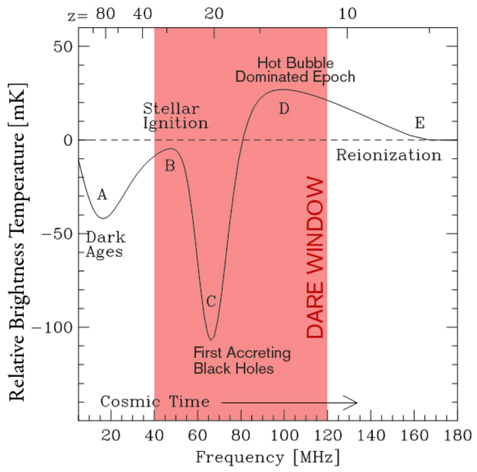
- We assume that *DARE* produces some number (≈8) of independent spectra from different pointing directions.
- These are modelled using parametrized descriptions of the signal, foregrounds (Galaxy, Sun, …), instrument and (for the ground-based prototypes) the ionosphere.
  - Hard to get non-parametric approaches to work with this small number of spectra.
  - However, the number of parameters (cases with up to ≈80 considered so far) is manageable with modern algorithms and computing power, allowing us to recover all parameters, with their errors (and covariances), simultaneously and rigorously.

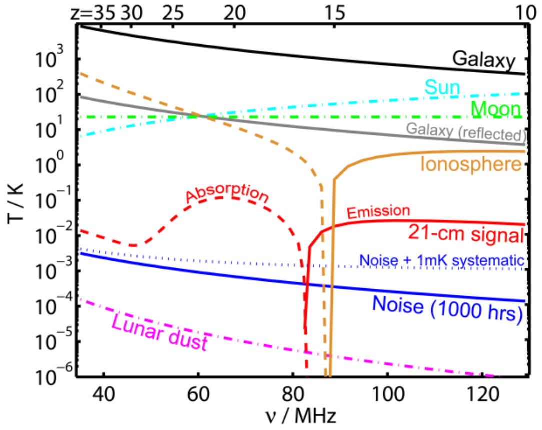
Analysis pipeline – II

- Explore parameter space to find the best-fitting model for a (synthetic or observed) data set.
- The code for fitting this comes in two parts:
  - Computing the likelihood given particular data and a set of parameters (currently have implementations in Matlab, Fortran 90 and python…).
  - A sampler for exploring parameter space and performing inference
    - Will show results using **`emcee`** and **`MultiNest.`**
    - Earlier implementations used a straightforward Metropolis-Hastings sampler written in Matlab (as in GH et al. 2012) and **`CosmoMC`** run as a generic sampler.

Signal and foreground modelling

- Signal is represented by a cubic spline with turning points labelled A-E; free parameters are the positions of turning points B, C and D.
- Galactic foregrounds and the Sun modelled as polynomials in log*T*-logν.
- The Moon is modelled as a thermal source with a certain temperature, that also reflects the other foregrounds with a certain reflectivity.
- We also have a parametrized model for the ionosphere (see later).



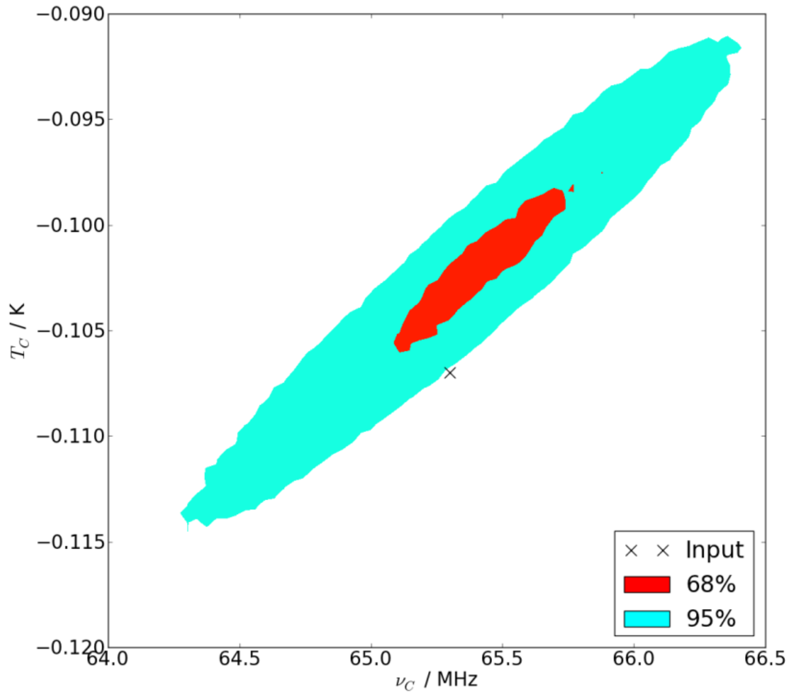


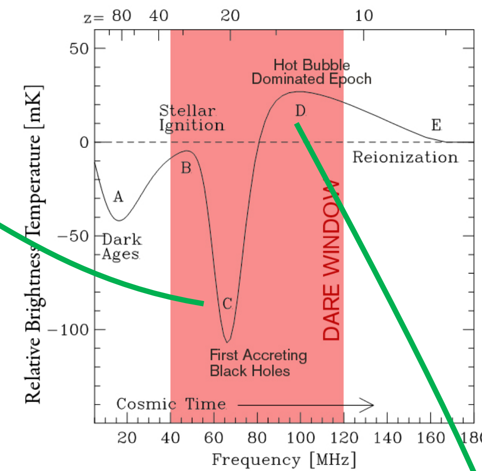
Instrument modelling

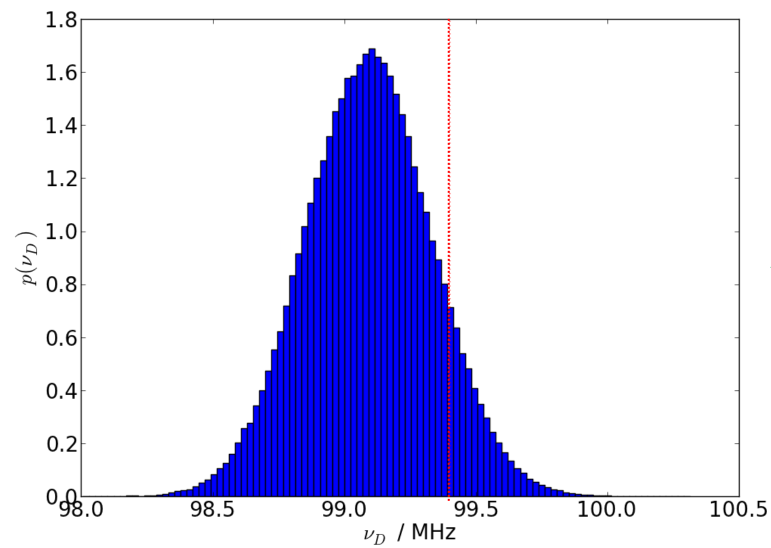
$$T_{Ant} = \frac{P_{ON}}{P_{OFF}} \frac{G_{amp0}}{G_{amp1}} \left[T_{amb} + T_{amp0} + \frac{T_{Rx}}{G_{amp0}} \right] - T_{amp1} - \frac{T_{Rx}}{G_{amp1}}$$

- Switch between a single load and the antenna.
- Gains and noise terms treated as functions of various parameters of the system (S-parameters, impedances, …) which themselves are (possibly complex) functions of frequency.
- To be superseded by a new calibration approach where the single load is an antenna emulator, and the systematic errors are randomized through switching; even then, the front-end gains will still have to be modelled in a similar way (as described in Abhi Datta's talk, J1-2, and a paper by Bradley et al. soon to be submitted to *Radio Science*).

Turning point constraints: 1000 hrs in lunar orbit

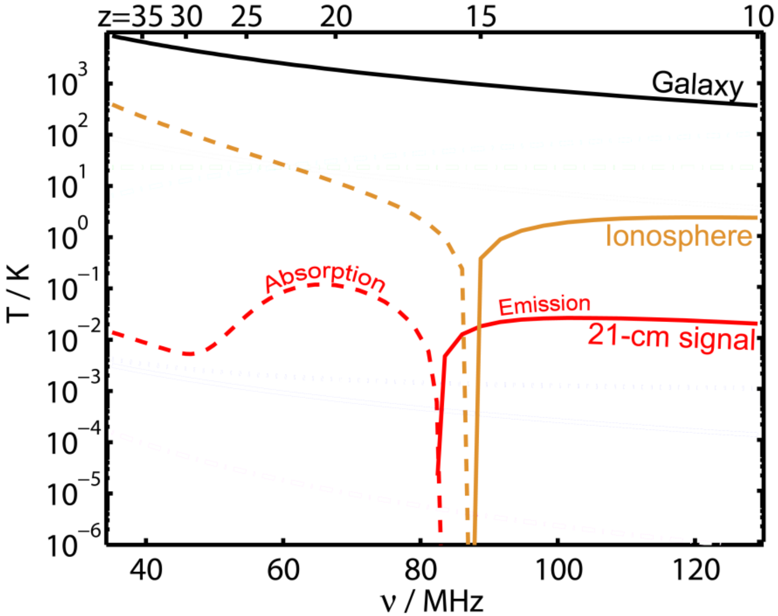




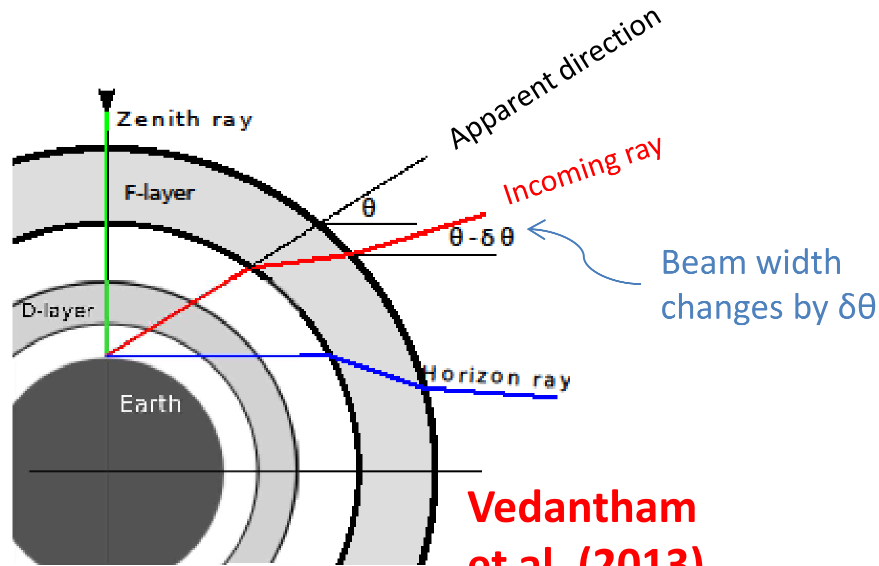


The ionosphere in global experiments

FGs, expected signal and ionospheric residuals



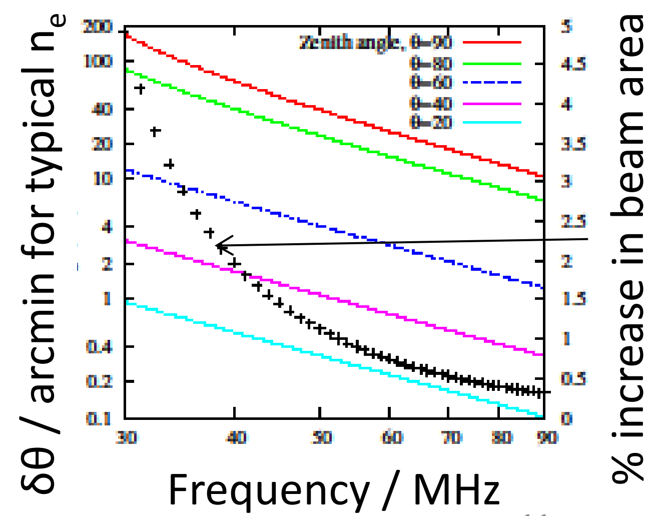
The ionosphere absorbs at low radio frequencies, while its hot electron population also produces emission. Spectral shape *can mimic our signal* and is *time-variable*: fitting with static models may be inadequate.



Vedantham et al. (2013)

Refraction →frequency- and time-variable beam

Horizon cutoff can also lead to sharp spectral features during moderately active conditions.



Including a *very* simplistic model of the ionosphere

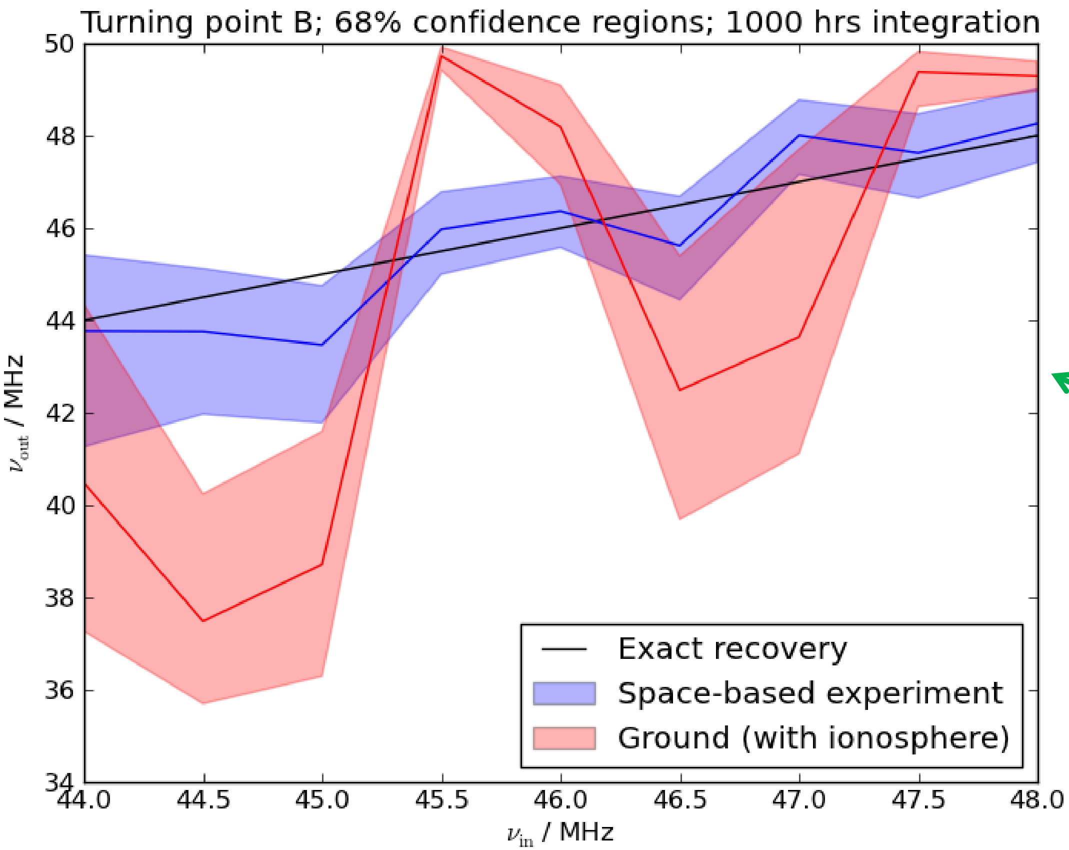
- Absorption and emission terms:

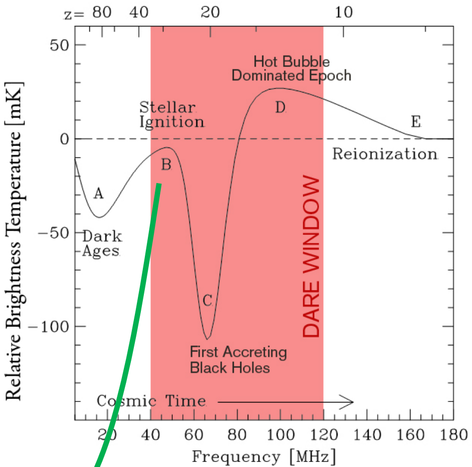
$$T = T_{\mathrm{sky}} + A_0\left(T_{\mathrm{e}} - T_{\mathrm{sky}}\right)\left(\frac{\nu}{\nu_0}\right)^{-2}$$

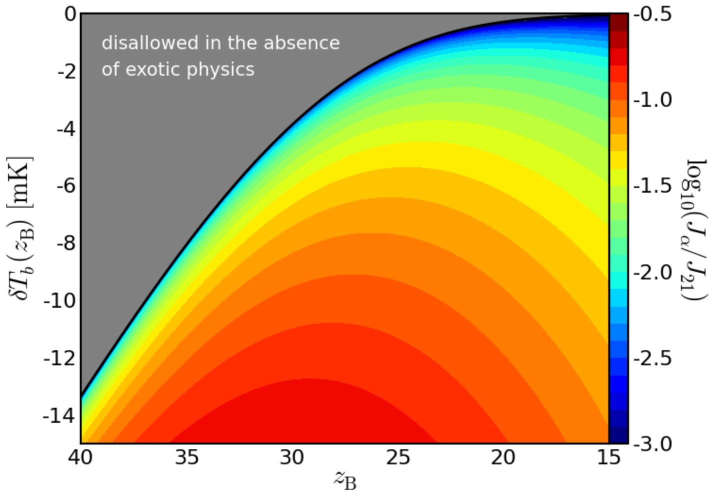
ν_0=80 MHz, A_0≈0.01, T_{e}≈1000 K (or more)

- Refractive effects:
  - Beam changes size, adding an area $\propto \nu^{-2}$.
  - We assume this mixes in an extra smooth foreground component, weighted by ν^{-2}.
  - We do not fit for the parameters of this additional foreground, which is equivalent to assuming it can be absorbed into the other smooth components.

Ionospheric results



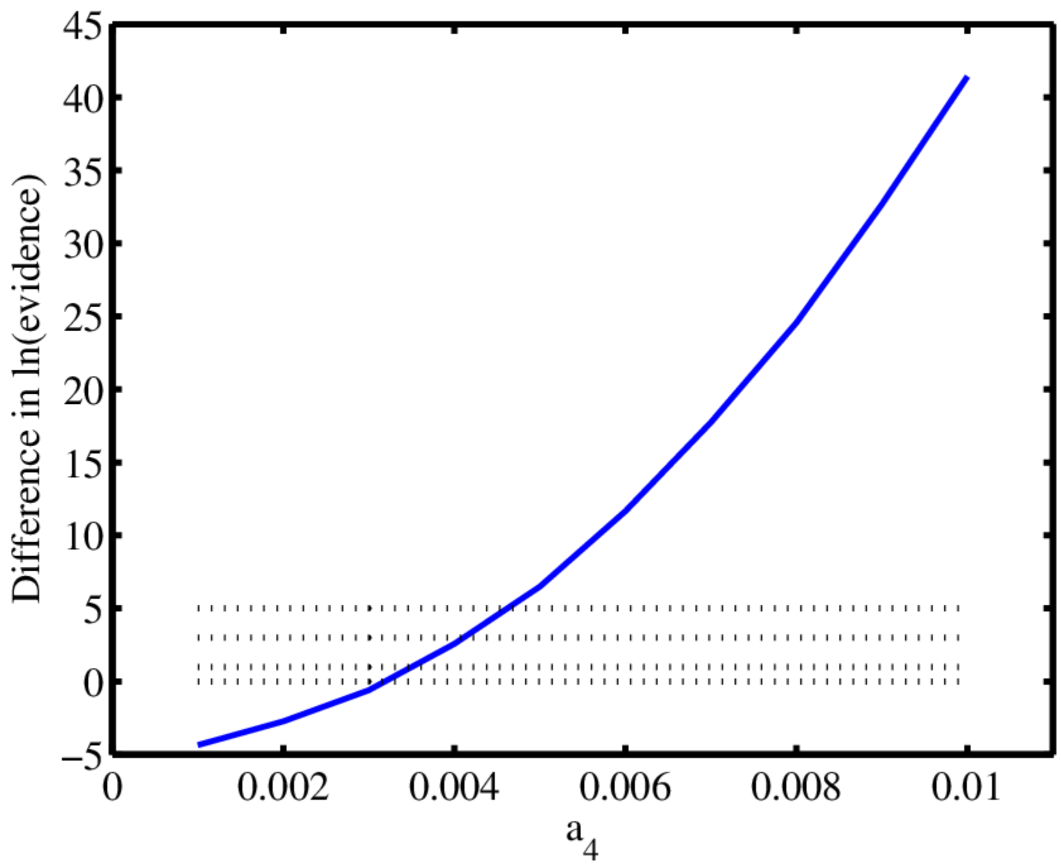




Mirocha, GH & Burns, 2013, ApJ, 777, 118.

Detecting the need for more complex foreground models

- Consider 1 hr of pristine data from a *DARE*-like experiment in lunar orbit.
- Can we detect the need to use a more complex foreground model than a third-order polynomial in log*T*-logν?
- Look at the Bayesian evidence ratio using **`MultiNest`.**
- $$\log T_{\mathrm{FG}} = \log T_0 + \sum_n a_n \left[\log(\nu/\nu_0)\right]^n$$
- How large does a_4 have to be before the data compel us to include it in our model?



Future developments

- Variable ionosphere (based on GPS measurements) and a more physically motivated model.
- Update with new calibration method, and allow the instrument parameters to be fit freely.
- Bayesian model selection with different 21-cm signal models.
- For the ground-based experiments, optimally construct the spectra of different sky regions that enter the analysis from the time-ordered data.
- Validate the pipeline and the foreground and ionospheric models with data from the prototype antenna.