Dealing with Uncertainty in Inference

Steven Gratton

Kavli Institute for Cosmology Cambridge

30 Sep 2021

Bayes' Theorem

- $p(theory|data) = \frac{p(data\,|\,theory)\ p(theory)}{p(data)}$

But what should one do if?

- $p(theory|data) = \dfrac{??????????????????\;\; p(theory)}{p(data)}$

- i.e. one doesn't know what the likelihood is

When do such situations arise?

- Lacking something…
  - Knowledge about the experiment
  - Mathematical knowledge
  - CPU time
  - …

Options?

- Guess a potential likelihood

Options?

- Guess a potential likelihood

- Simulations

Options?

- Guess a potential likelihood

- Simulations
- Simulations with Neural Networks ("likelihood free" approaches…)

Options?

- Guess a potential likelihood
- Simulations
- Simulations with Neural Networks (“likelihood free” approaches…)
- …

GLASS (arXiv:1708.08479)

- Calculate/simulate what you can about what you most trust

- Use the *principle of maximum entropy* to “fill in the gaps” and effectively *construct a likelihood*

What is the principle of maximum entropy?

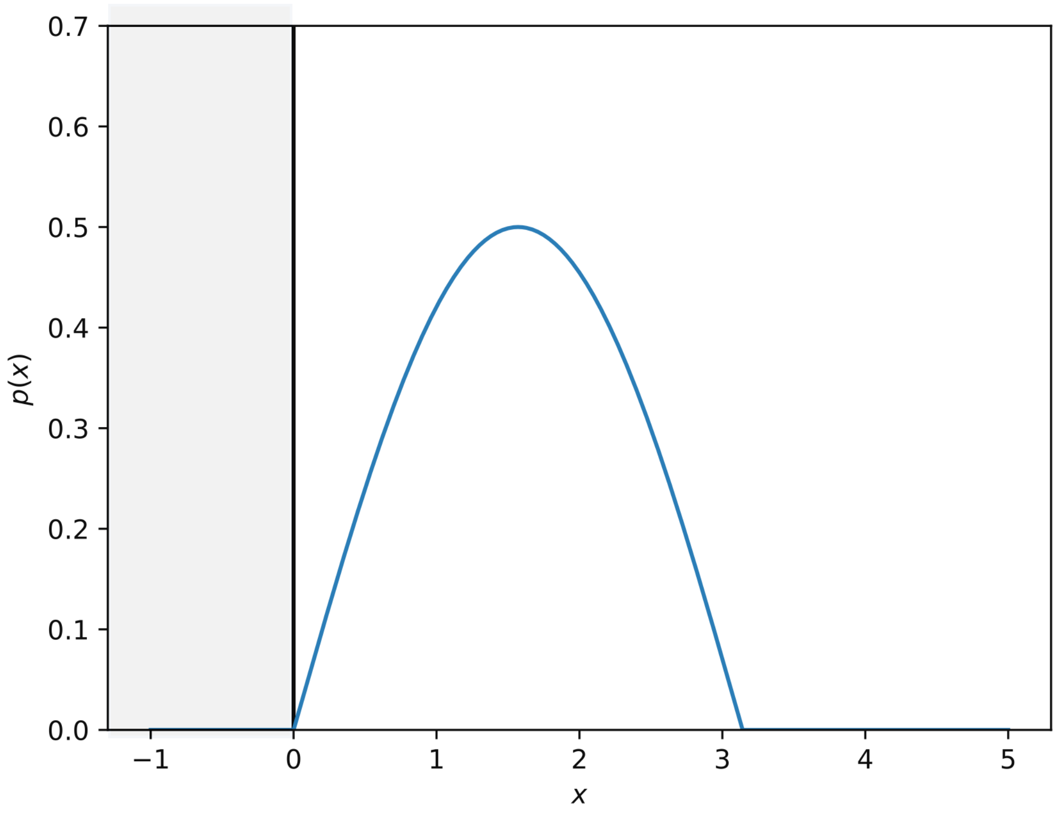
- Entropy (Shannon, 1948) measures the uncertainty of a probability distribution:

$$S = -\int p(x) \log p(x)\, dx$$

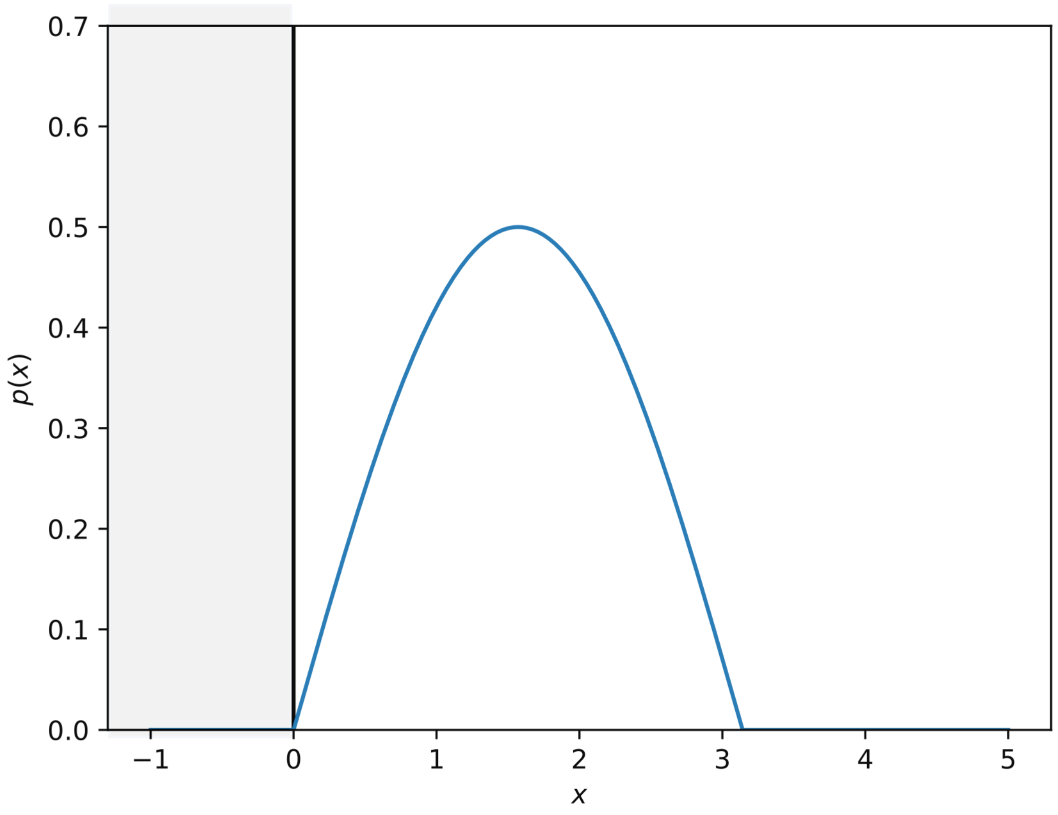
- So, if we *maximize* the entropy, subject to the constraints of what we *do* know, we generate the *broadest* or *most conservative* distribution consistent with those constraints (Jaynes, 2003 book)

- One shouldn’t go wrong if one uses this for inference! (One hopes…)

Example: $p(x) = \frac{1}{2}\sin(x)$



Example: $p(x) = \frac{1}{2}\sin(x)$

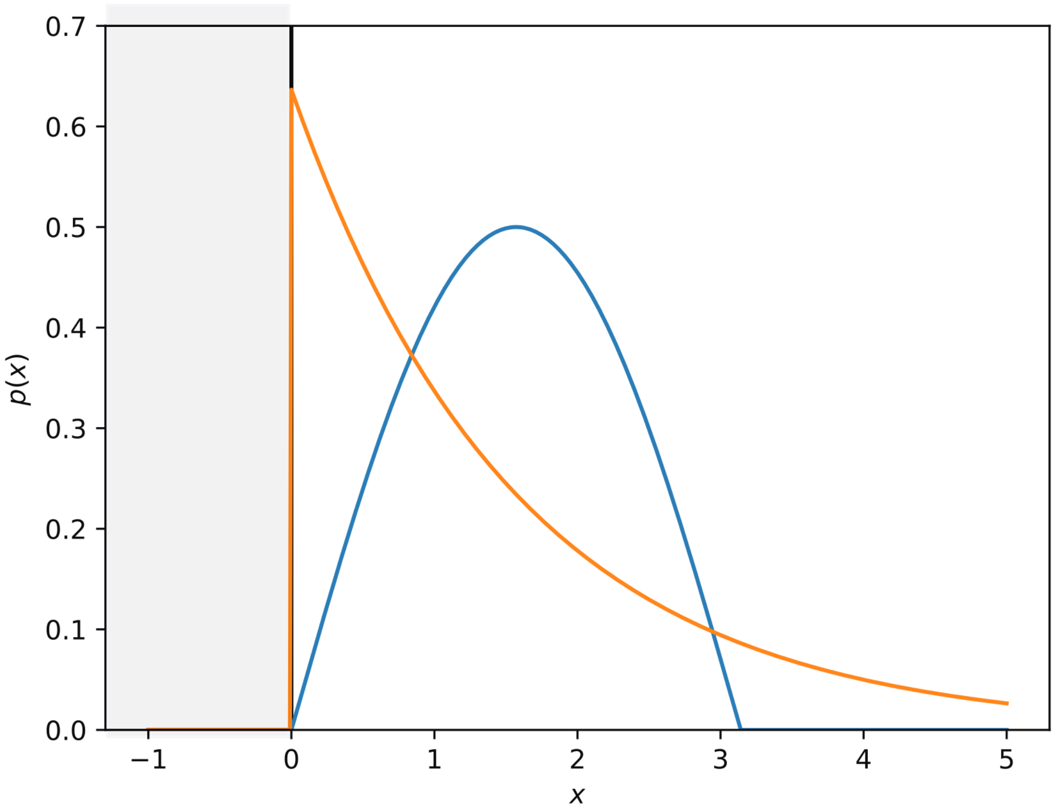


$$< x > = \frac{\pi}{2}$$

$$< x^2 > = \frac{\pi^2}{2} - 2$$

…

Fitting just the first moment:

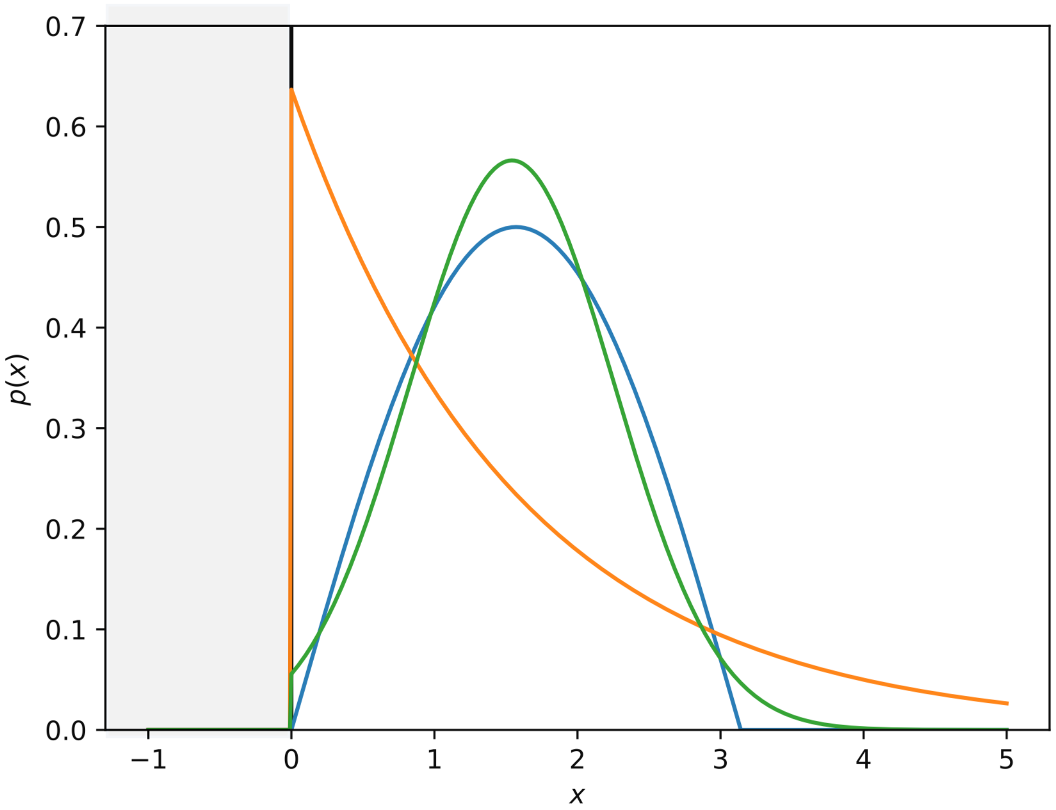


$$< x > = \frac{\pi}{2}$$

$$< x^2 > = \frac{\pi^2}{2} - 2$$

…

Fitting the first and second moments:

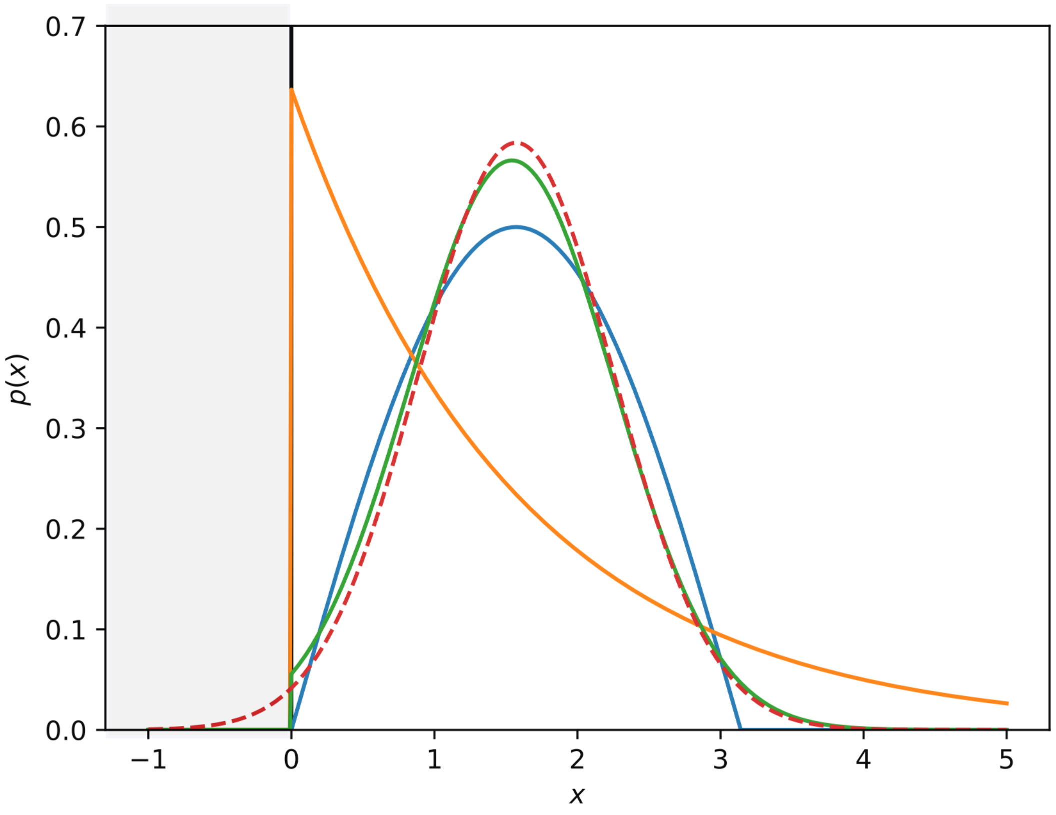


$$< x > = \frac{\pi}{2}$$

$$< x^2 > = \frac{\pi^2}{2} - 2$$

…

Cf. a gaussian with the same mean and var



$$< x > = \frac{\pi}{2}$$

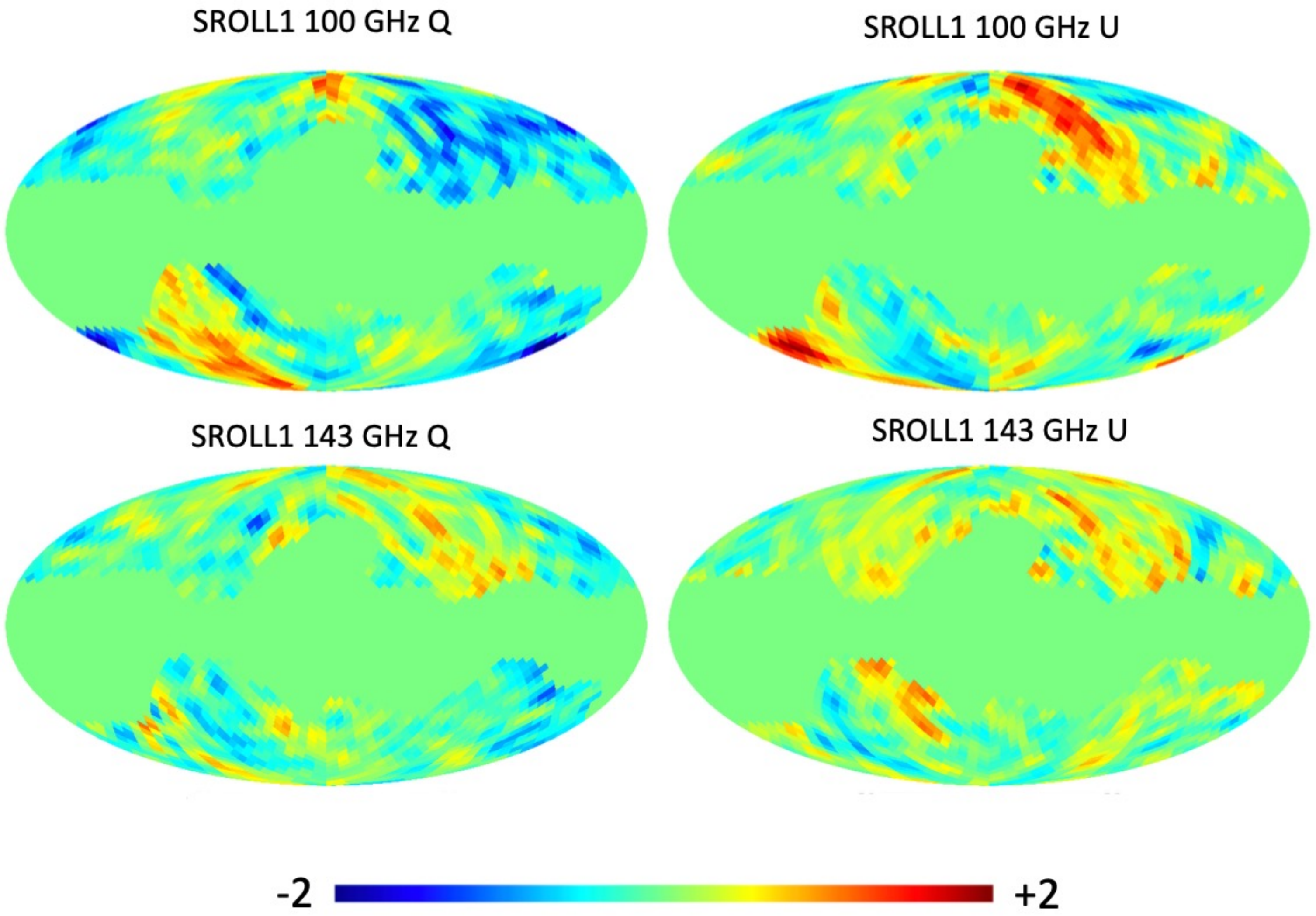
$$< x^2 > = \frac{\pi^2}{2} - 2$$

…

Application to Planck

- The target: inference of the optical depth to reionization via the height of the "bump" of the low multipole EE polarization power spectrum (also affects the TE spectrum but less significantly)

- The problem: large-scale systematic residuals in the polarization maps caused by non-linearities in the onboard analogue-to-digital converters, even with the special "SROLL1" processing developed to mitigate this effect for the 2018 Legacy Release…



SROLL1 100 GHz Q
SROLL1 100 GHz U
SROLL1 143 GHz Q
SROLL1 143 GHz U
-2
+2

- So we couldn’t use a pixel-based likelihood
- But, the actual form of a power-spectrum-based likelihood on a cut sky is not known…

Simulations?

- 300 (expensive) simulations with roughly representative ADCNL effects in them
- By subtracting the input CMB and foregrounds one thus has 300 approximately “noise-only” simulations

Planck Legacy Release: *SimLow*

- Use these 300 noise simulations, along with thousands and thousands of (easy) signal simulations with varying input taus, to build up the sample distributions of the cross power spectrum moments, *l* by *l*, as a function of tau

- Use fits to these distributions, evaluated at the measured values, as the likelihood

Limitations

- Not enough simulations to understand the correlations between the power spectrum elements
  - Either at the same multipole number l (e.g. between *TE* and *EE*)
  - Or between elements at different multipoles (Nb. we expect couplings between because of having to mask the stronger regions of galactic emission)

- Precluded a combined TT, TE & EE low-l likelihood using this method

Our reanalysis

- De Belsunce, Gratton, Coulton, and Efstathiou MNRAS 2021

- First step: we use the 300 noise simulations to inform the construction of a *model* of the noise

Three methods

- C-SimLow
  - Revised SimLow approach; main difference is that now that we have modelled the noise, we can easily make thousands and thousands of noise as well as signal realizations in estimating the distributions
- pydelfi
  - Similar to the above, but using a neural-network-based approach to learn the distribution of the cross spectra rather than assuming a functional form
- Momento
  - Application of the GLASS methodology

Momento

- Assuming our noise model, we are able to *calculate* on the fly moments of and between all cross spectrum elements (up to quartic order) of interest for any value of tau
- Using the GLASS procedure, we effectively build a likelihood consistent with these moments, and evaluate it at the data cross spectrum values, for each theory model we wish to consider

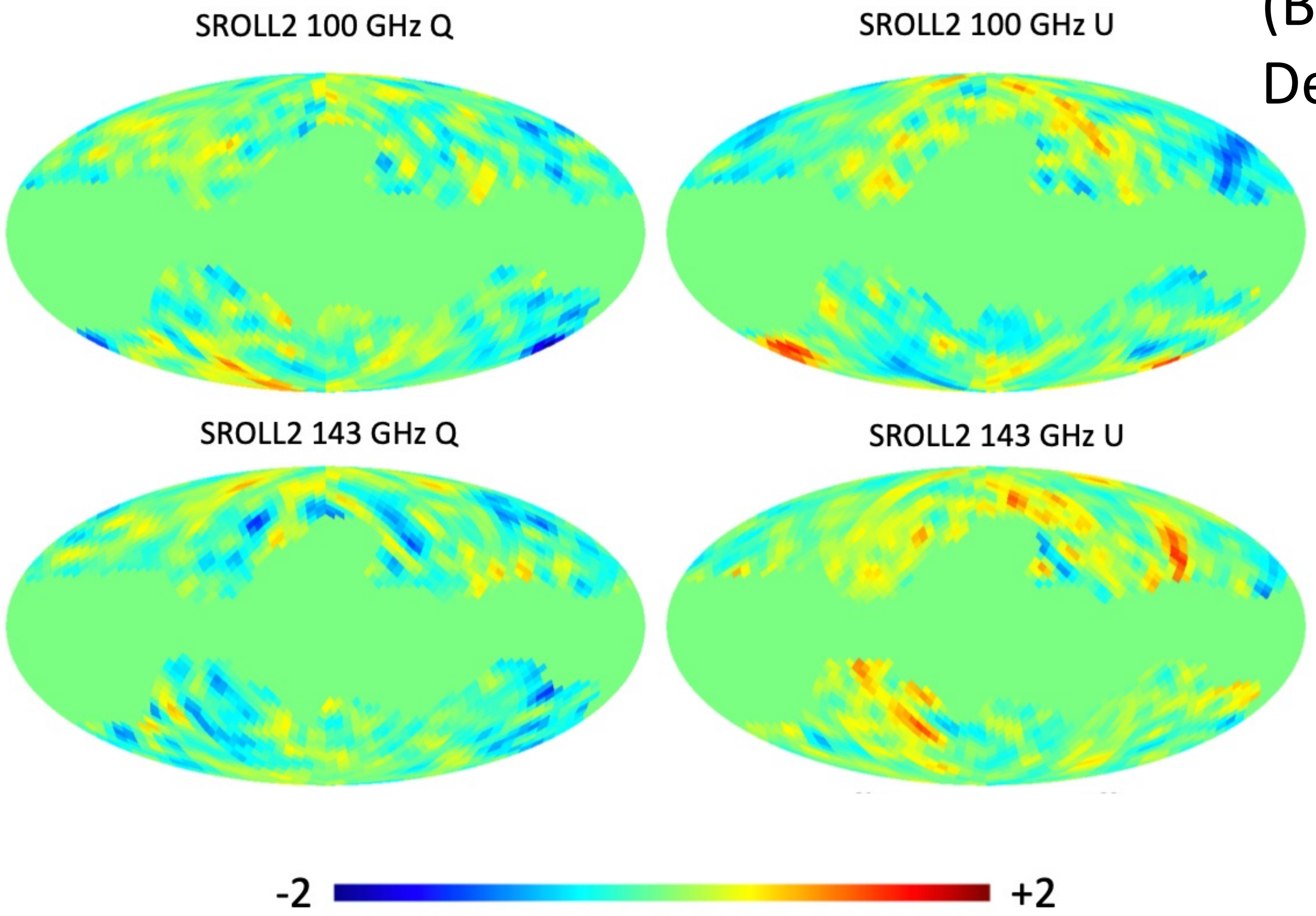
Results

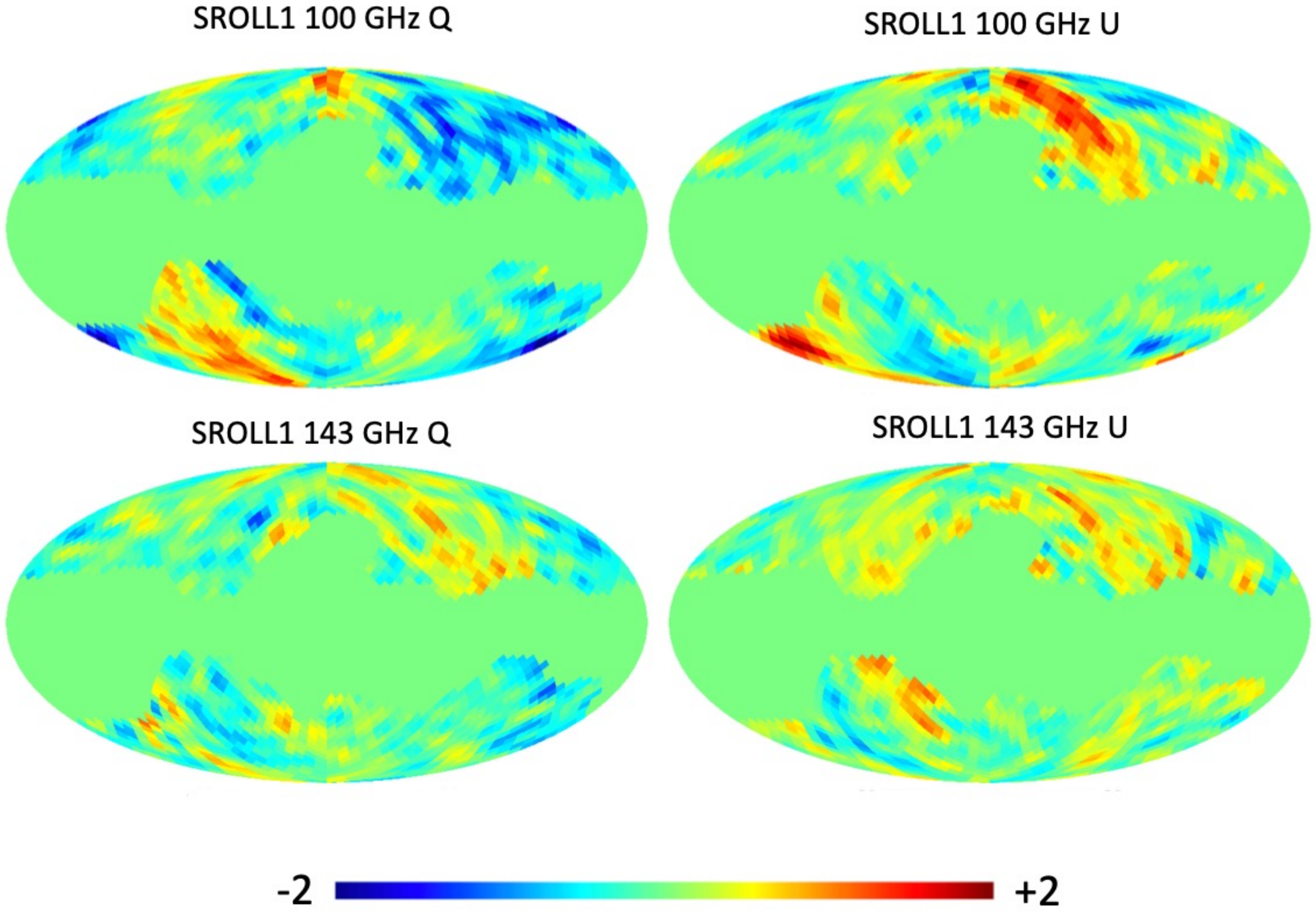
- All methods were consistent with each other (and with the legacy analysis), giving:

Data Set	Likelihood	τ (EE)	τ $(TTTEEE)$
Planck 2018	`C-SimLow`	0.0530 ± 0.0071	...
	`momento`	0.0507 ± 0.0063	0.0527 ± 0.0058
	`pydelfi`	0.0517 ± 0.0070	0.0513 ± 0.0078

Further Improvement: Revised, SROLL2, maps

(BWARE team, Delouis et al. 2019)

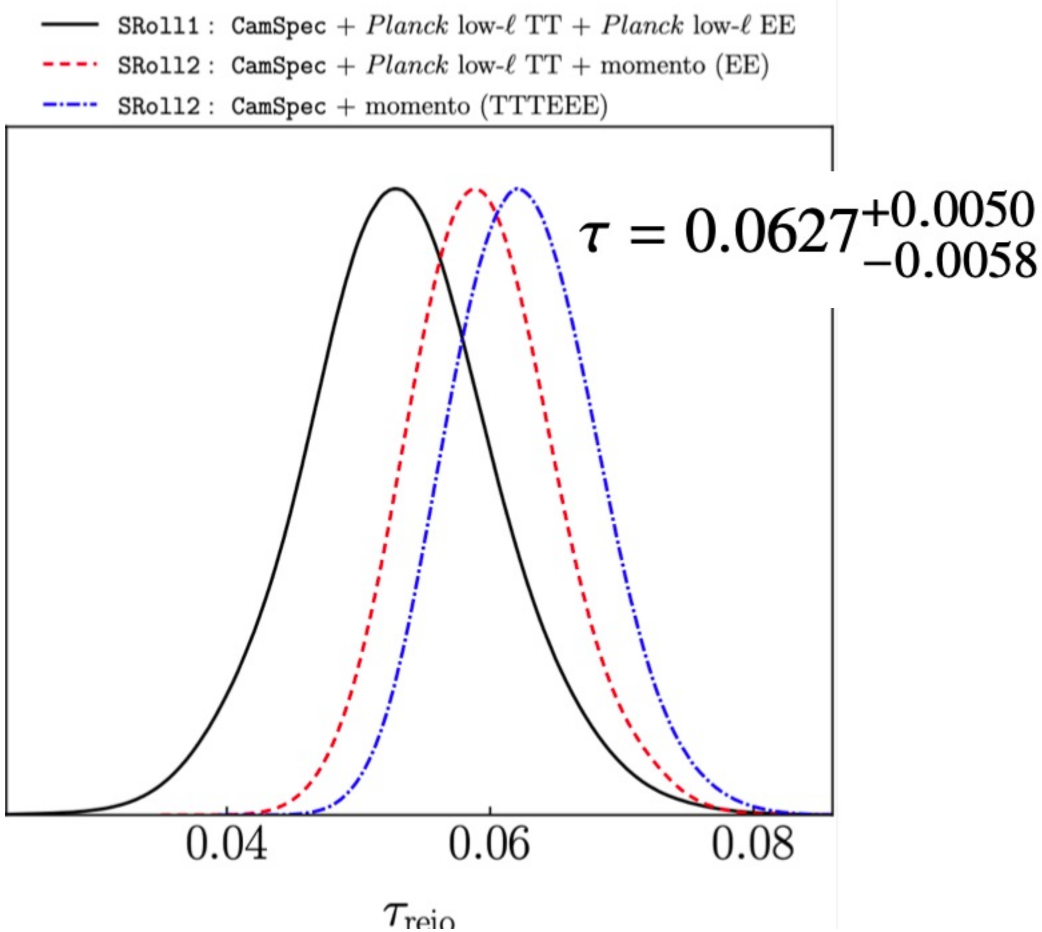



SROLL1 100 GHz Q
SROLL1 100 GHz U
SROLL1 143 GHz Q
SROLL1 143 GHz U
-2
+2

…give a slightly higher tau:

Data Set	Likelihood	$\tau\ (EE)$	$\tau\ (TTTEEE)$
Planck 2018	`C-SimLow`	0.0530 ± 0.0071	…
	`momento`	0.0507 ± 0.0063	0.0527 ± 0.0058
	`pydelfi`	0.0517 ± 0.0070	0.0513 ± 0.0078
`SRoll2`	`C-SimLow`	0.0582 ± 0.0057	…
	`momento`	0.0581 ± 0.0055	0.0604 ± 0.0052
	`pydelfi`	0.0588 ± 0.0054	0.0580 ± 0.0064

Combining with high-/ data…



Conclusions and Further Thoughts

- Dealing with these sorts of issues is tough!

- Relying on simulations is dangerous:
  - One probably won’t have enough of them
  - Even if one does, they might not be good enough in all details; one must be prepared to “guide” one’s analysis/neural network towards only using the things that are suitably trustable

- The GLASS procedure seems to have worked well and should be applicable not only to further CMB analysis but also more widely