MANCHESTER
1824
The University of Manchester

Chiborg: a Bayesian jackknife framework for testing consistency of multiple measurements

Dr. Mike Wilensky
Jodrell Bank Centre for Astrophysics
michael.wilensky@manchester.ac.uk
SKA Cosmology SWG meeting 2023

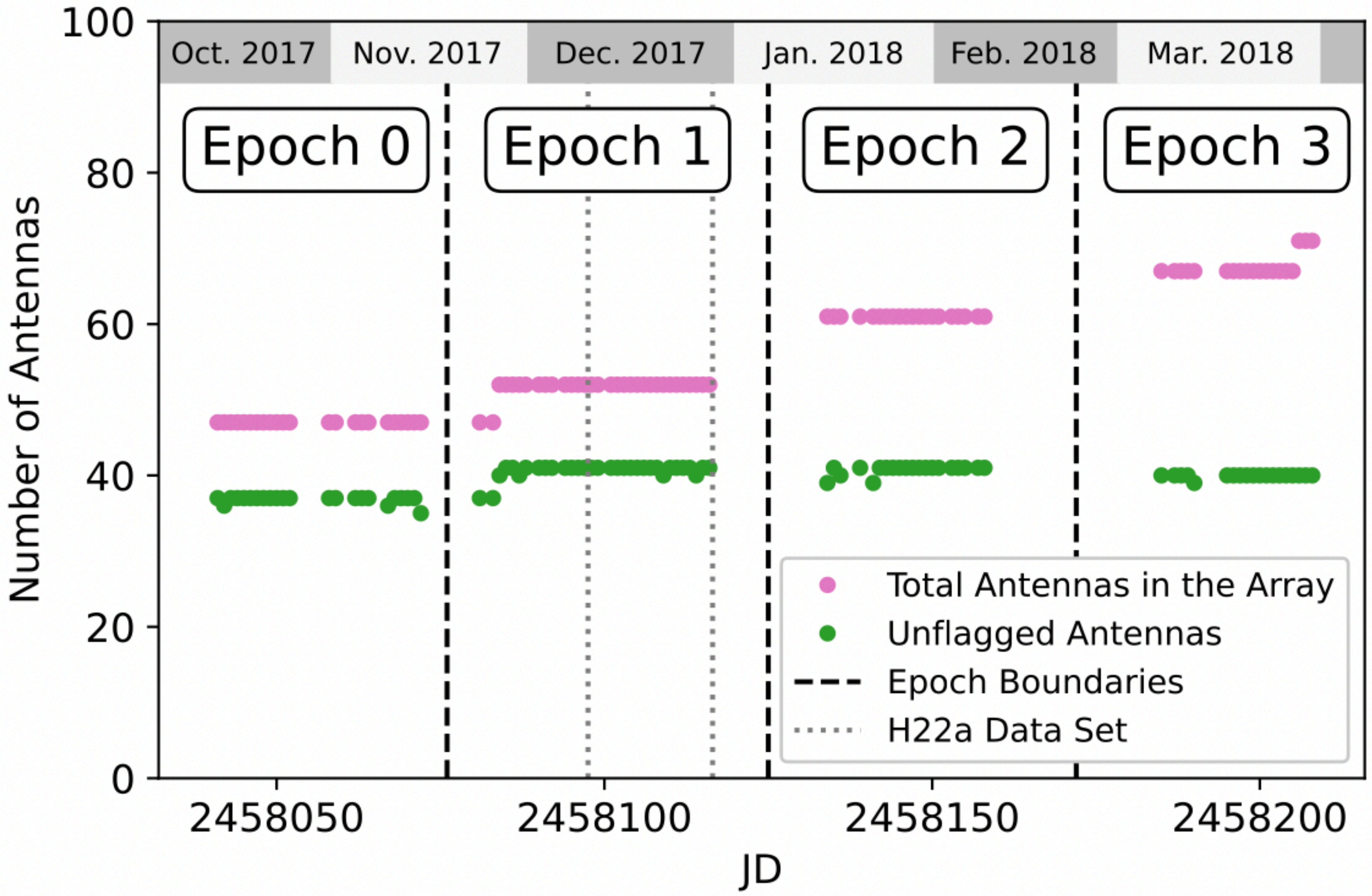
MANCHESTER 1824
The University of Manchester

Talk Layout

- Intro to chiborg via a toy example
- Quick formalism overview
- Revisiting the toy example
- Example with HERA
- Future directions and other concluding remarks

The University of Manchester

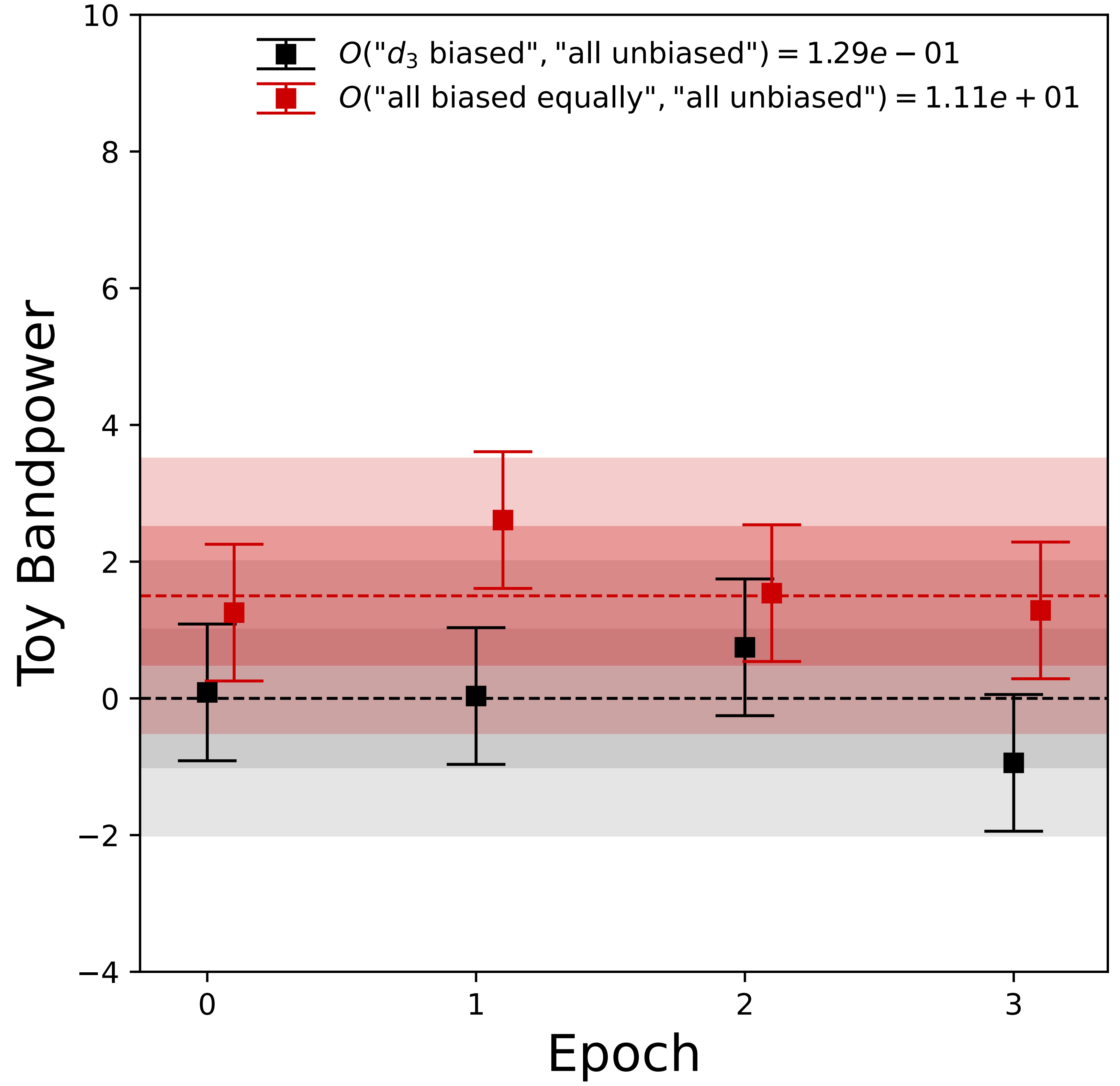
Problem Setup

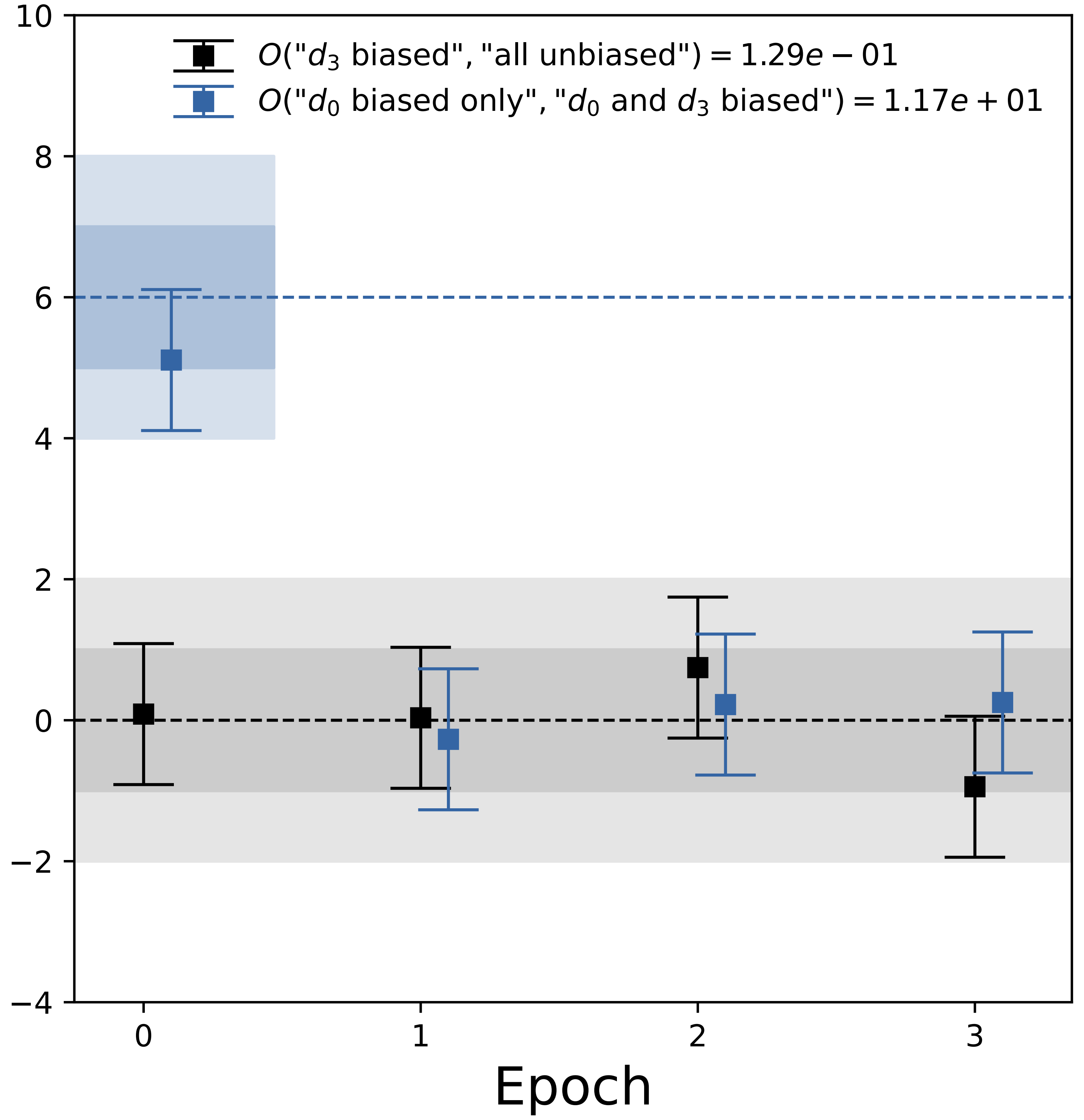


- Array was in construction during data-taking
- Can we combine all these data for a full-season limit?
- Make power spectra for each epoch independently and see if they are statistically consistent.

Source: https://arxiv.org/pdf/2210.04912

χ-by-cyborg aka *chiborg*[1,2]





- One meaning for inconsistent is that there is a systematic *bias* in some data but not in others
- Sometimes we can identify such data by eye by performing a "χ-by-eye"
- Somewhat misleading term since a χ^2 test considers whether the data are globally a good fit i.e. *it does not point to specific outliers*
- *chiborg* uses Bayesian hypothesis testing to determine to identify the problematic data points

[1] https://github.com/mwilensky768/chiborg

[2] https://arxiv.org/abs/2210.17351

Basic Formalism

Hypothesis to test e.g. "the third datum is biased"

Bias distribution under the hypothesis

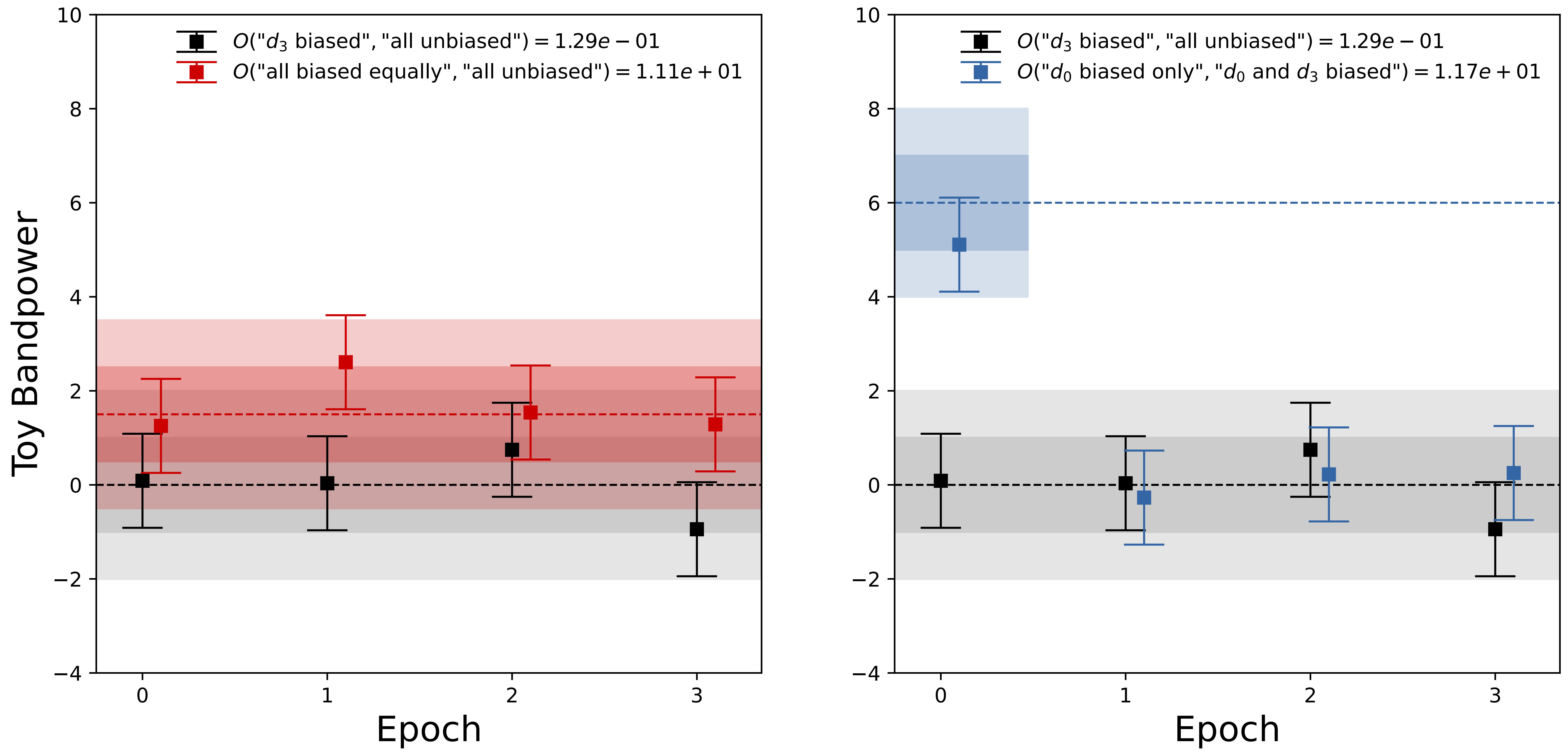
$$P(H\,|\,\mathbf{d}) = \int d\boldsymbol{\varepsilon} \frac{P(\mathbf{d}\,|\,\boldsymbol{\varepsilon}, H)P(\boldsymbol{\varepsilon}\,|\,H)P(H)}{P(\mathbf{d})}$$

Some data may be more suspicious a priori

unknown bias

Measurements (including uncertainty)

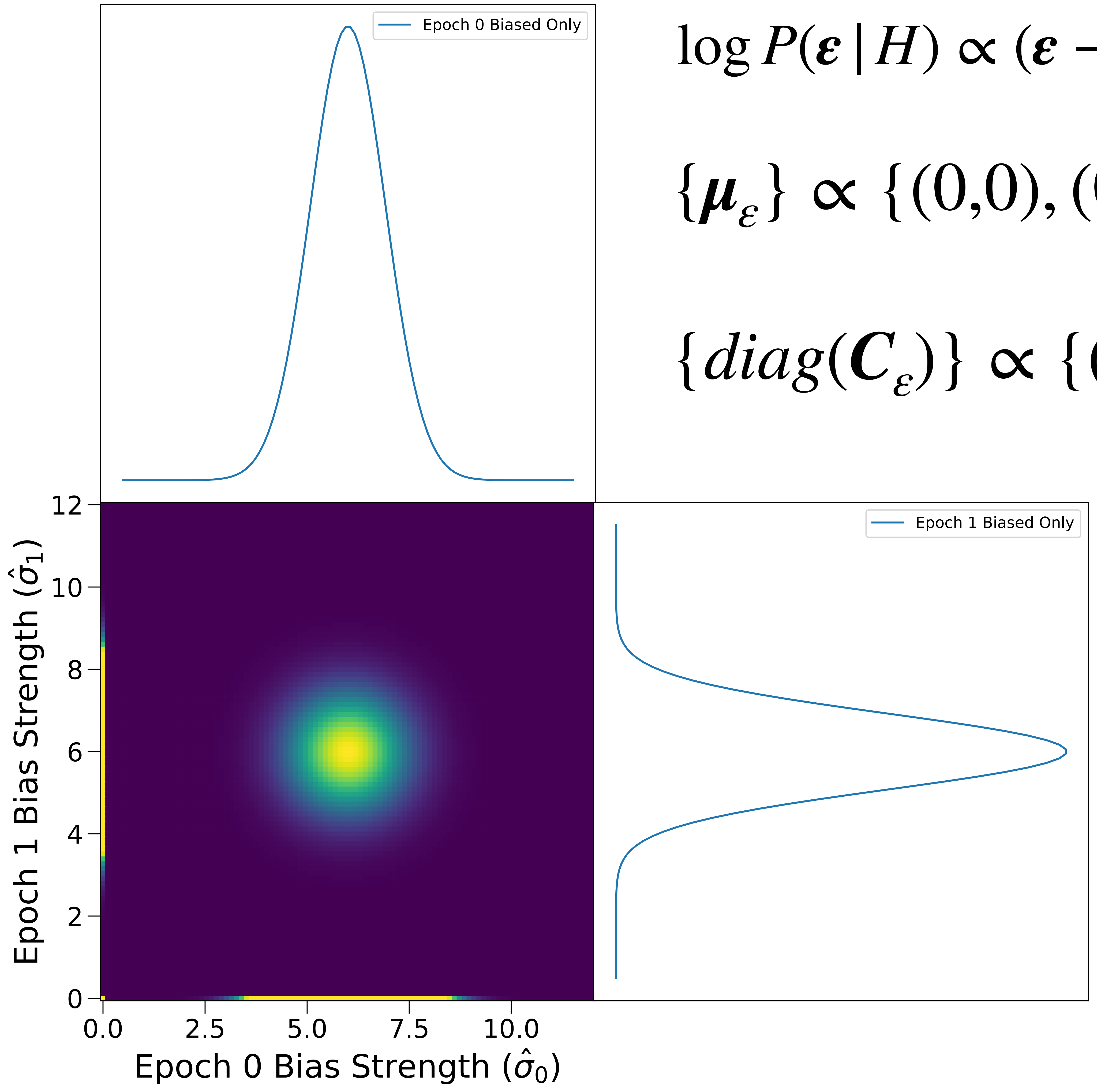
χ-by-cyborg aka *chiborg*[1]



- Has tools to help users to formulate hypothesis sets
- Does not need homoschedastic data or even independent data (correlations are allowed!)
- Meaning of statistical significance can be defined in terms of a loss function, i.e. it is embedded in a decision-making framework

[1] https://github.com/mwilensky768/chiborg

‘All diagonal’ hypothesis set



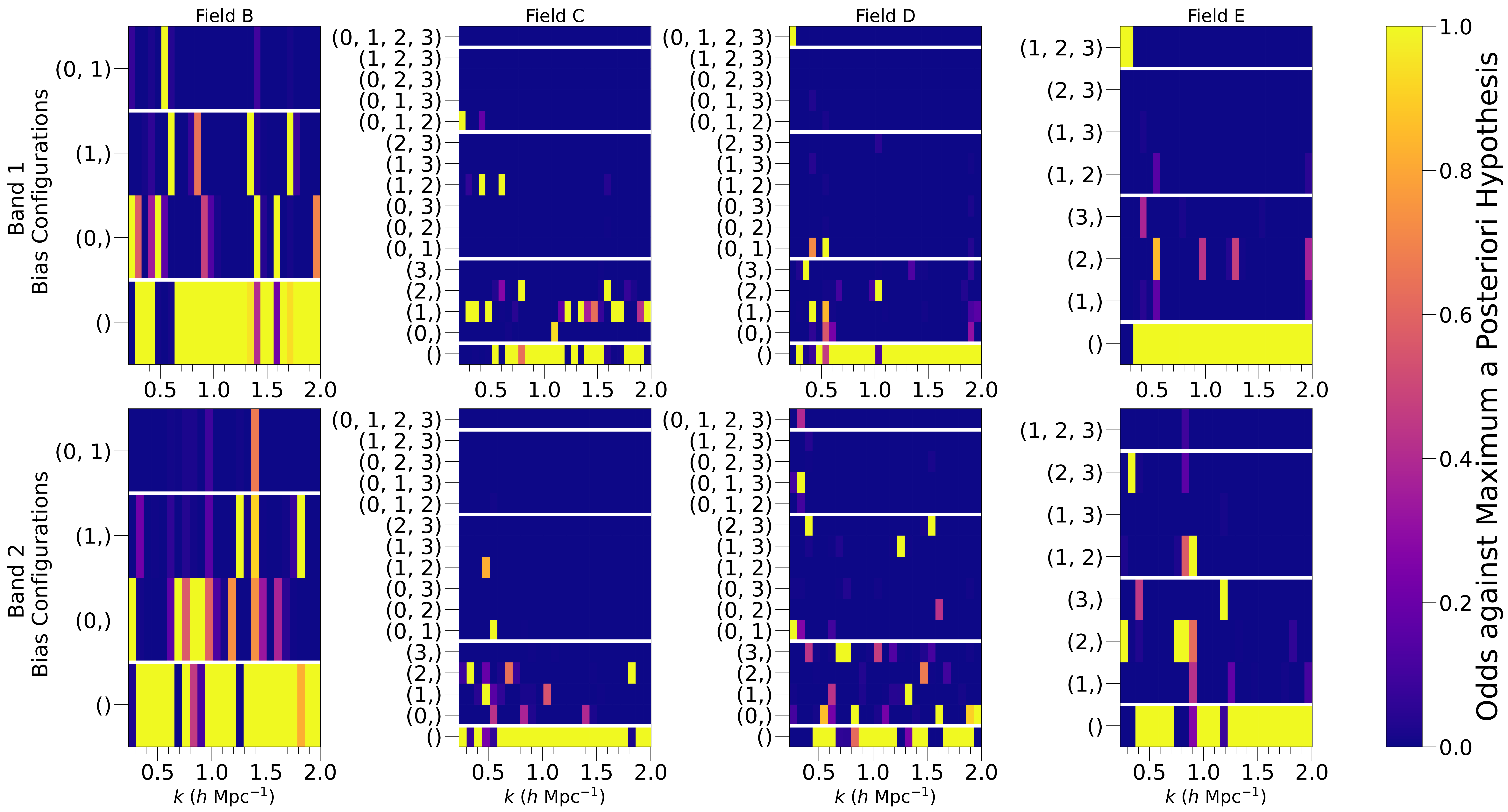
$$\log P(\boldsymbol{\varepsilon} \,|\, H) \propto (\boldsymbol{\varepsilon} - \boldsymbol{\mu}_\varepsilon)^T \mathbf{C}_\varepsilon^{-1} (\boldsymbol{\varepsilon} - \boldsymbol{\mu}_\varepsilon)^T$$ (Gaussian prior on bias)

$$\{\boldsymbol{\mu}_\varepsilon\} \propto \{(0,0), (0,1), (1,0), (1,1)\}$$ (all 2-bit sequences)

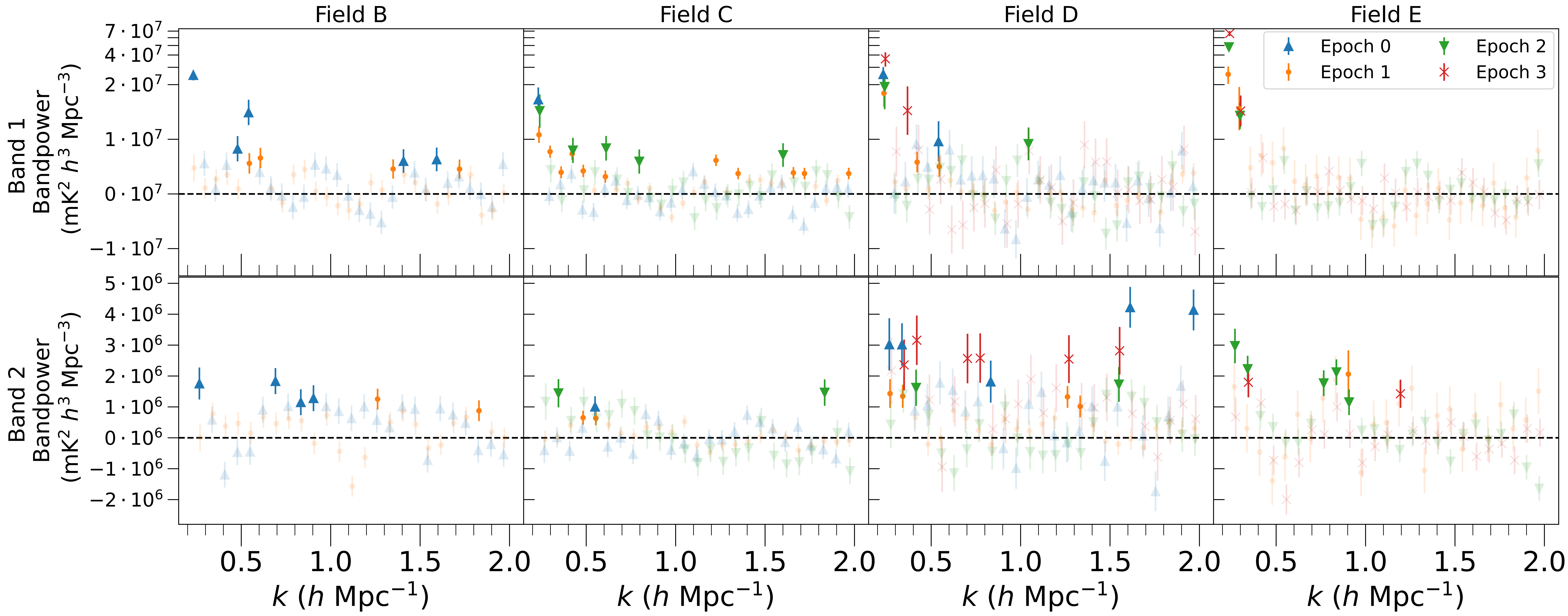
$$\{diag(\boldsymbol{C}_\varepsilon)\} \propto \{(0,0), (0,1), (1,0), (1,1)\}$$ (all 2-bit sequences on diagonals, 0 on off-diagonals)

- Somewhat exhaustive hypothesis set, but scales like 2^N. Suited for small data sets.
- Proportionality constant must be chosen so that the test returns meaningful results (have tools for this).

Results with HERA pt. 2



Results with HERA pt. 1



MANCHESTER 1824
The University of Manchester

Conclusions

- Chiborg is a rigorous Bayesian method for determining statistical consistency

- It was battle tested on HERA data and proved successful at its purpose

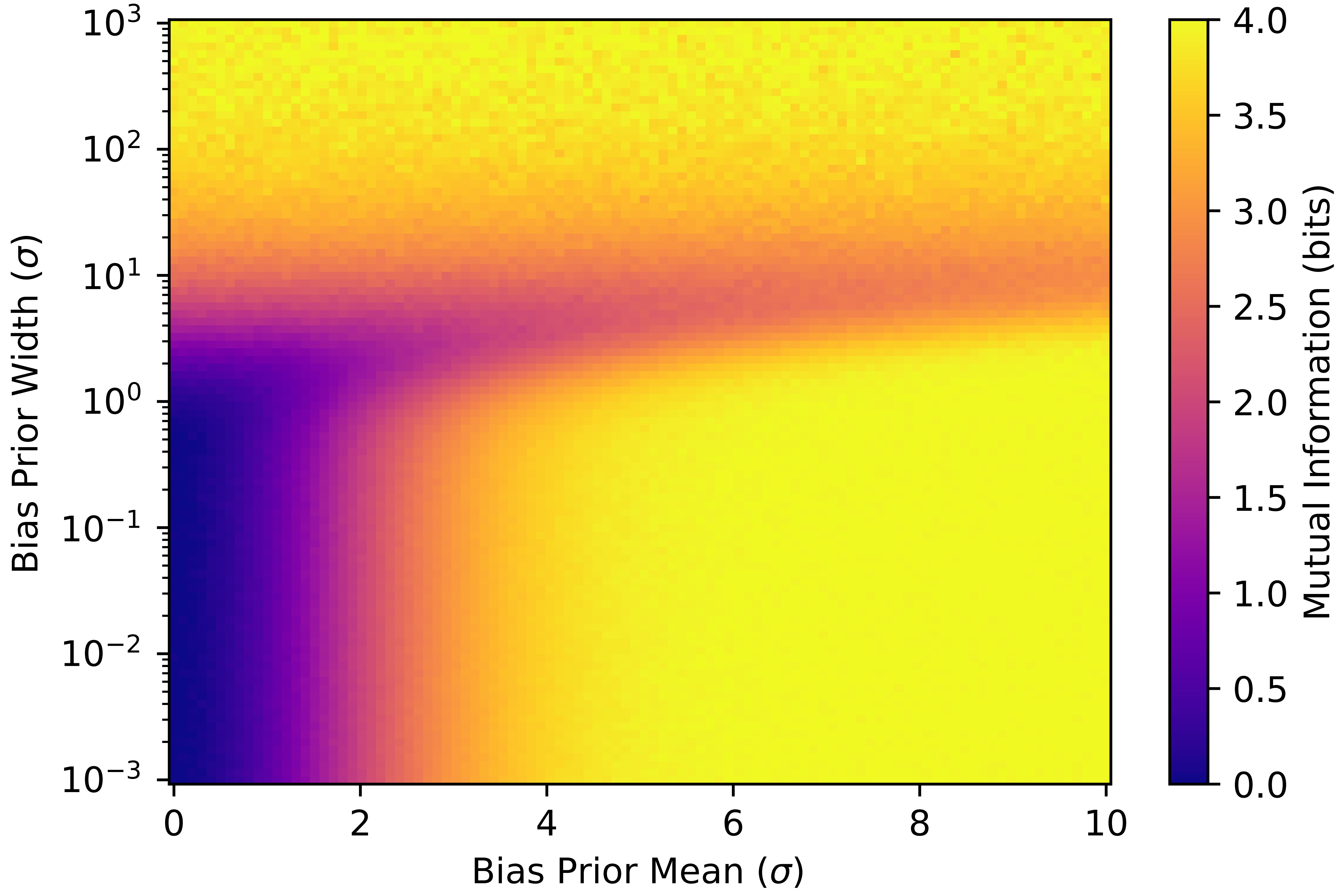
- In general, it could be used for determining consistency of any set of measurements of a given quantity, e.g. power spectra from different seasons of a given instrument, power spectra from different instruments, different global 21-cm measurements, measurements of H_0…

- All implemented in an open-source python package: https://github.com/mwilensky768/chiborg

- We also wrote a paper: https://arxiv.org/abs/2210.17351

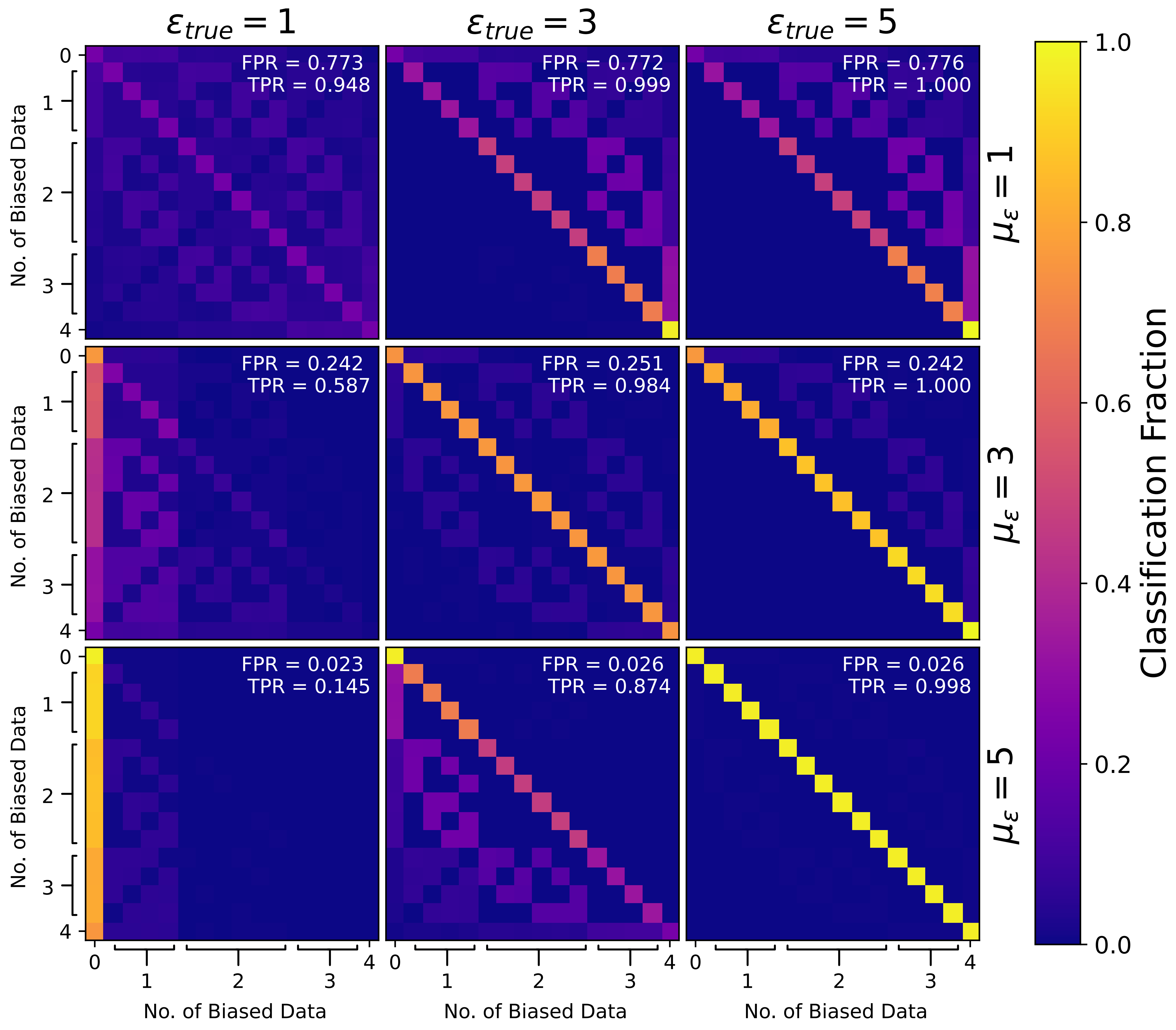
Backup Slides Past this Point

Mutual information (region of discernability)



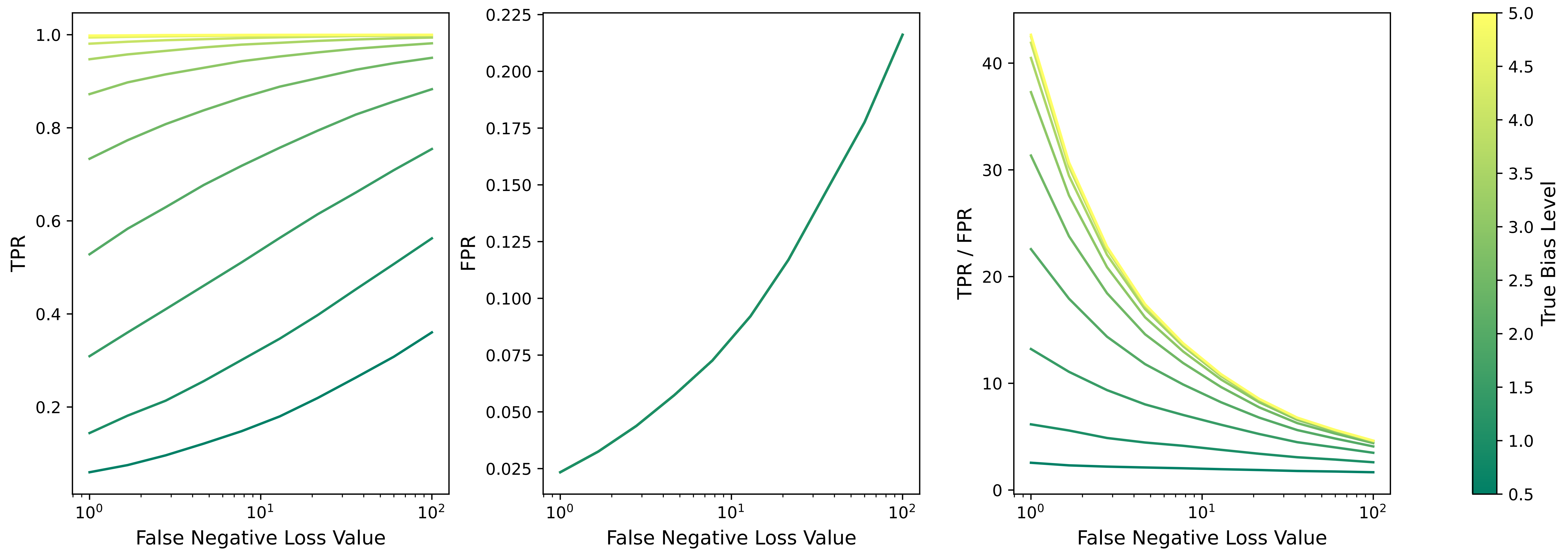
- Mutual information tells how much the Shannon entropy of a random variable is reduced by knowing another
- We can ask for the mutual information between the data and the hypotheses, assuming the prior generating distribution is correct
- Region of high mutual information tells us that the data will *on average* specify which hypothesis is most likely, whatever it is. Therefore, we want ot choose hyperparameters where mutual information is high, otherwise the test results will be confusing.

How it does with Gaussian samples



- Diagonal elements show correct # of biases identified
- Upper triangle shows overidentifications (false pos.)
- Lower triangle shows underidentifications (false neg.)
- Setting bias prior too close to noise level causes lots of false positives (top row)
- Setting bias prior too far causes lots of false negatives
- Can predict which hypotheses are distinguishable using information theory, which can inform this setting
- Can also use simulations like this in addition to a loss function to develop desired settings

Tuning with a loss function



$$\langle L(H_i) \rangle = \sum_{j=0}^{Q-1} L(H_i, H_j) P(H_j \,|\, \mathbf{d})$$

$$\mathbf{L}_{\textbf{low FNR}} = \begin{pmatrix} 0 & L & L & \dots & L \\ 1 & 0 & 1 & \dots & 1 \\ 1 & 1 & 0 & \dots & 1 \\ \vdots & & & \ddots & \end{pmatrix}$$

- If one has some means of setting a loss function, then this relieves them of the need to arbitrate statistical significance
- On the other hand, one could use the loss function to tune an acceptable false positive/negative rate for a given bias strength