Class 5: Plot workshop cont., worries, more stats

Miguel F. Morales
Bryna Hazelton

Plot workshop

- Briefly describe science
- What is the question you want to answer with this plot?
  - Identify key information & information that is less important
  - Identify key comparisons
  - What can make plot easier to absorb?
  - Are there red herrings?
- Is there additional information you need to answer the question?
  - Can it be added to the plot (in a digestible way)?
  - Is there a partner plot (plot story)?

Data density examples

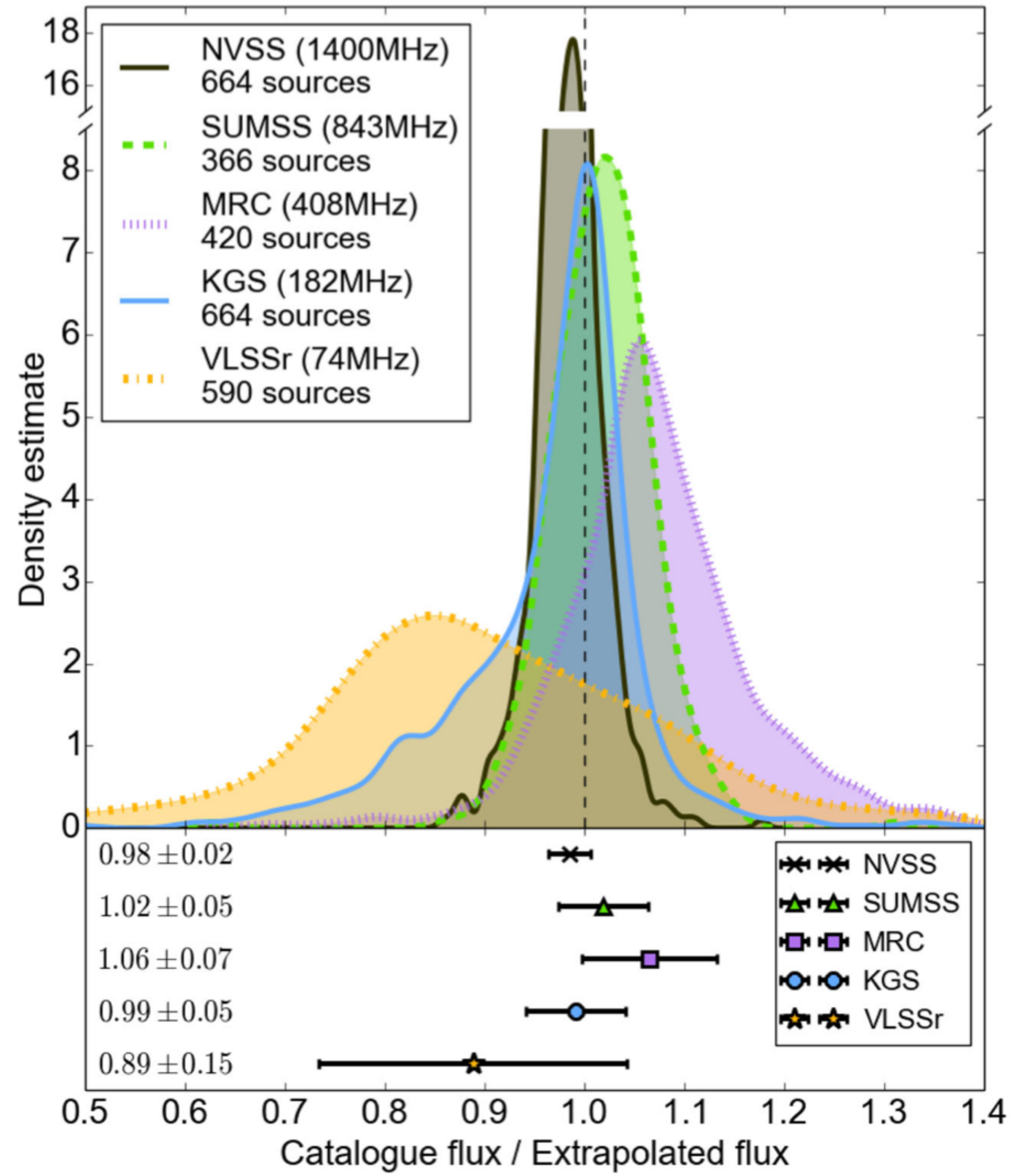
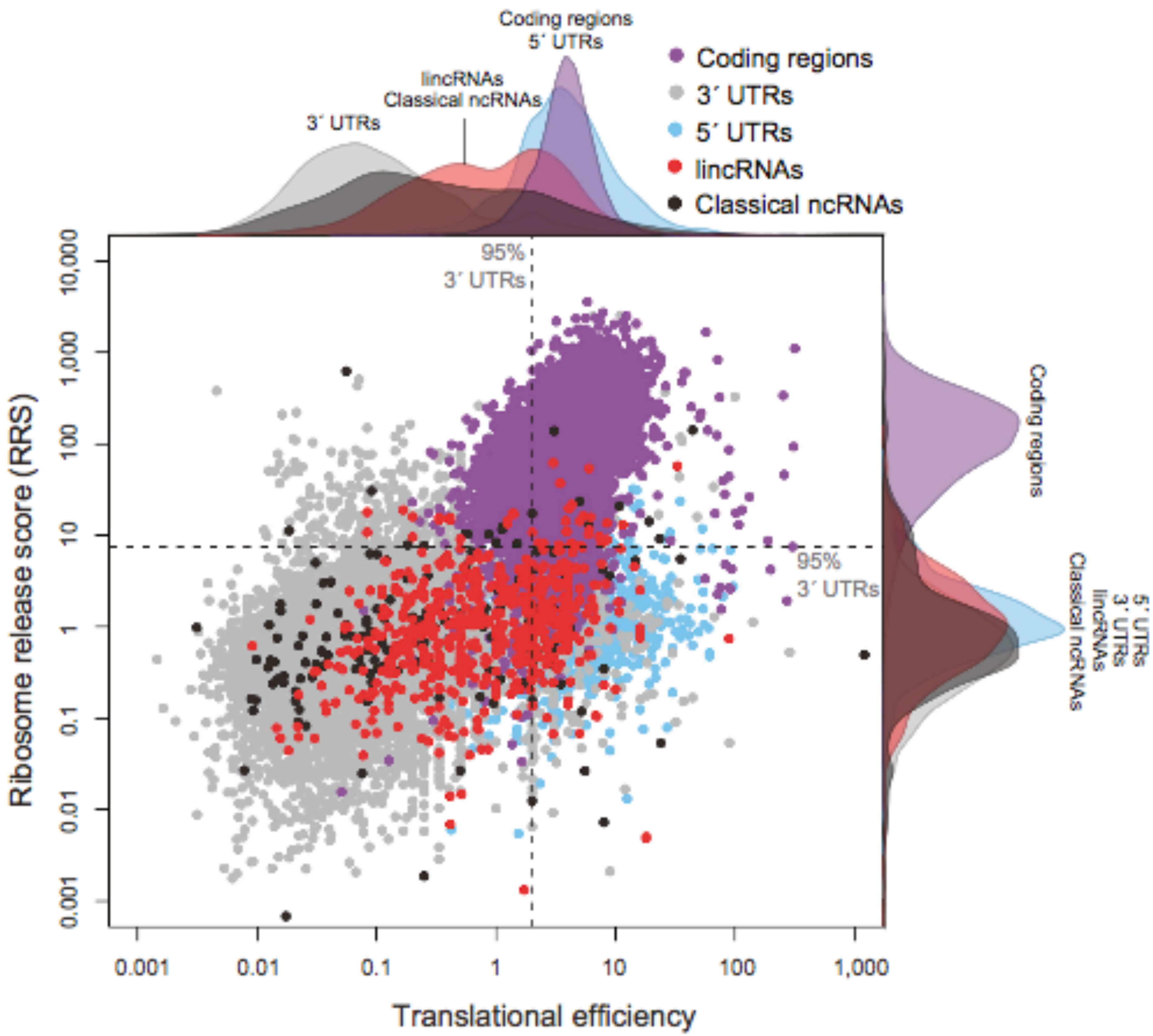
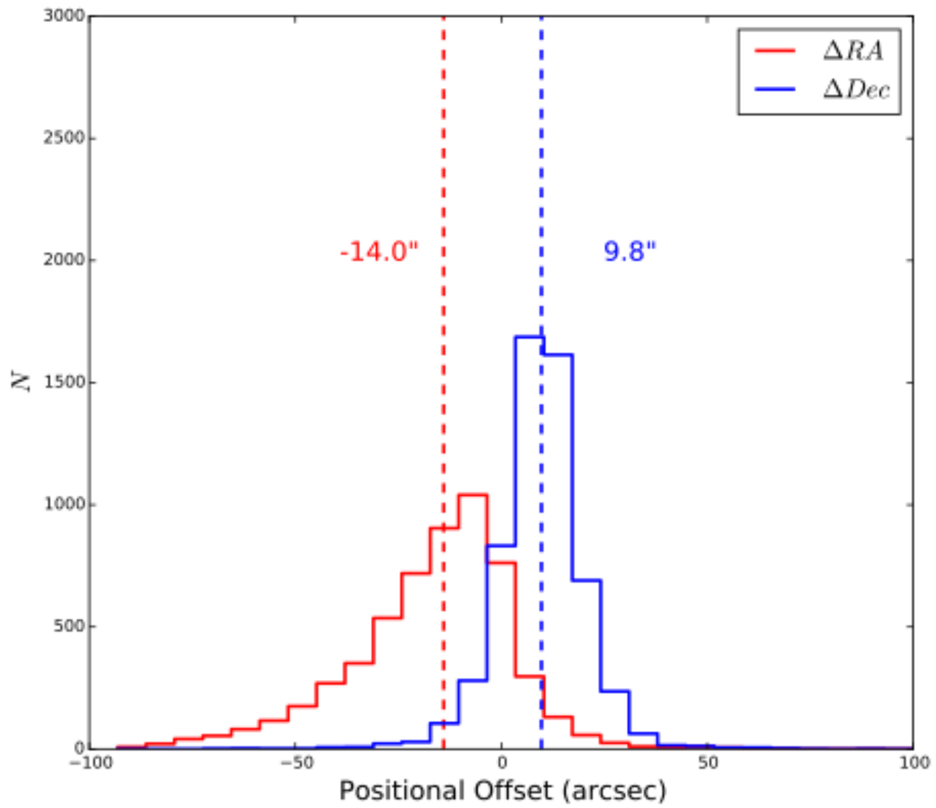


Figure 2. The ratio between observed flux density and extrapolated flux density from a fit to the SED is shown for every time a catalogue appeared in a match with at least two other catalogues for `isolated` sources. The upper panel shows a univariate kernal density estimation of each distribution (note broken y axis due to the sharp peak in the NVSS ratio distribution), while the lower panel shows the median and median absolute deviation of each distribution. The KGS spectral index agrees very well with no indication of flux bias on average.

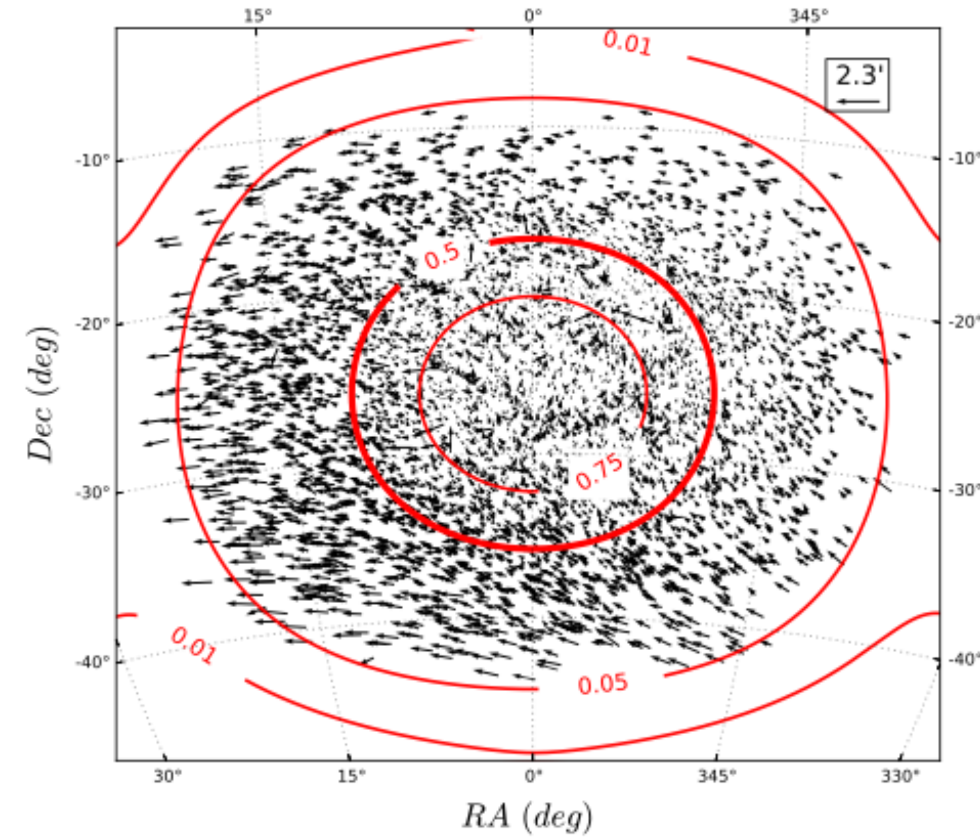


Coding regions
5′ UTRs
lincRNAs
Classical ncRNAs
3′ UTRs
Coding regions
3′ UTRs
5′ UTRs
lincRNAs
Classical ncRNAs
95%
3′ UTRs
95%
3′ UTRs
Ribosome release score (RRS)
10,000
1,000
100
10
1
0.1
0.01
0.001
0.001
0.01
0.1
1
10
100
1,000
Translational efficiency
Coding regions
5′ UTRs
3′ UTRs
lincRNAs
Classical ncRNAs



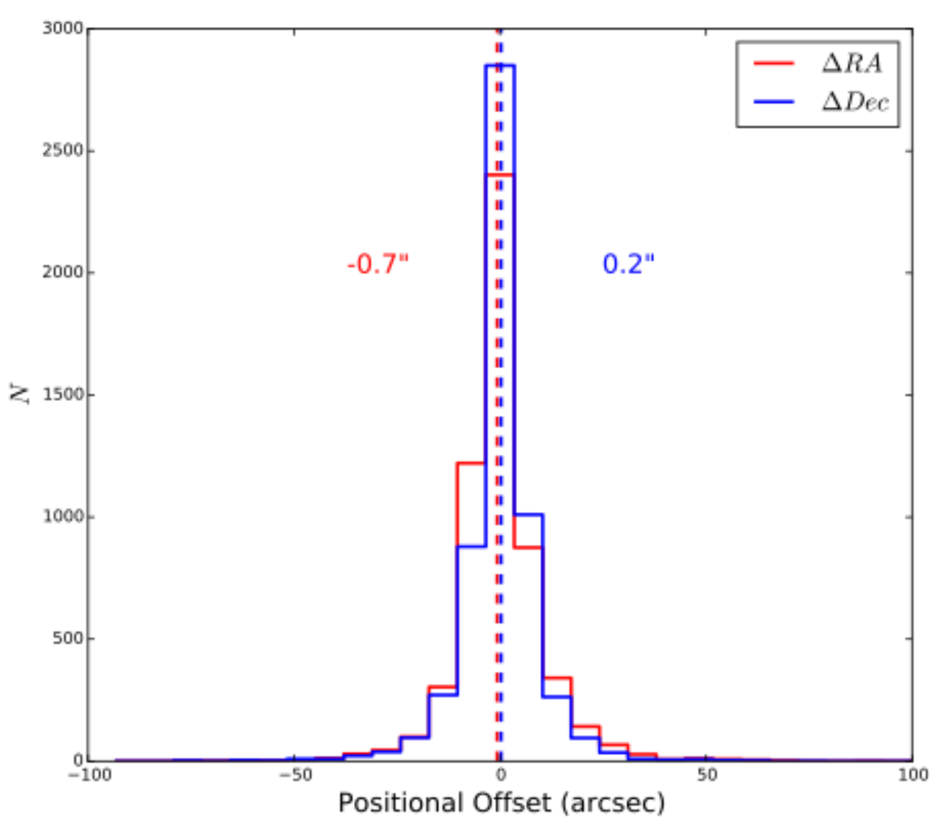
ΔRA
ΔDec
-14.0"
9.8"
N
Positional Offset (arcsec)

(a)



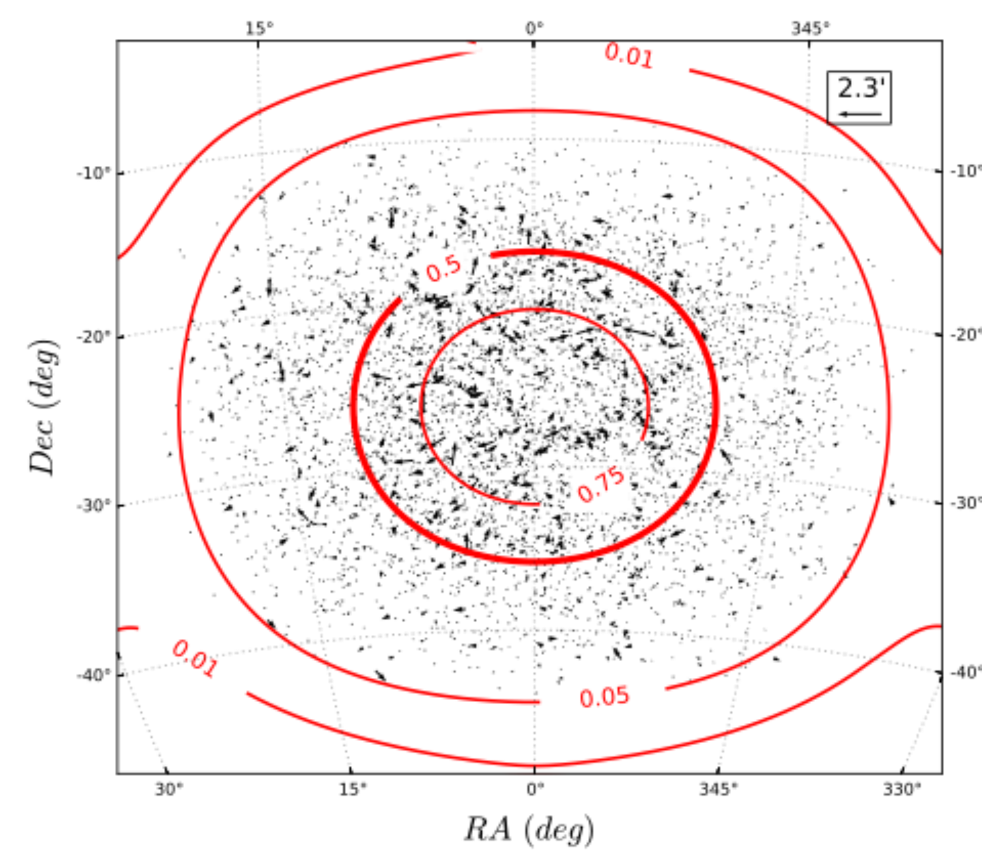
2.3'
0.01
0.05
0.5
0.75
Dec (deg)
RA (deg)

(b)



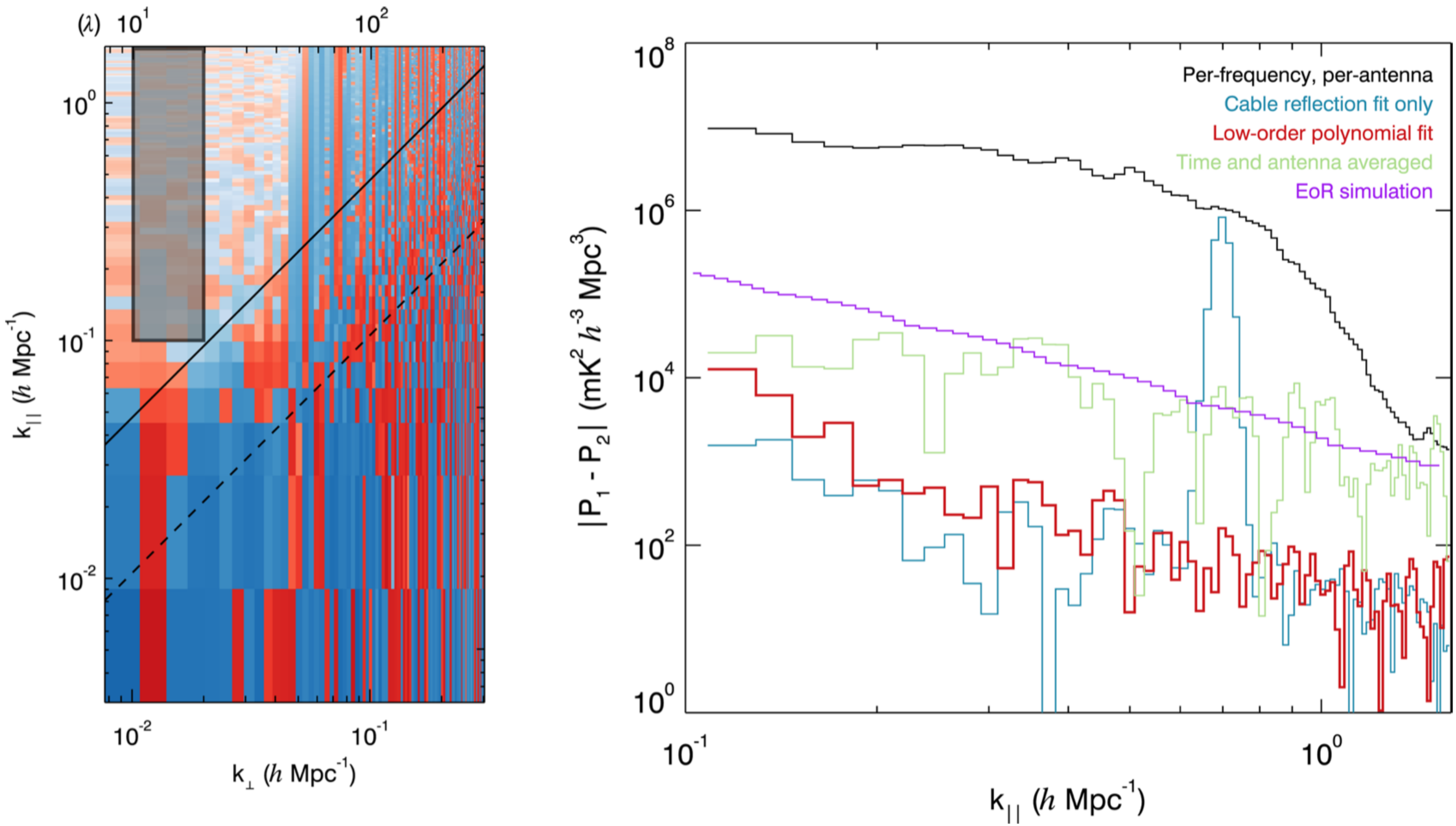
ΔRA
ΔDec
-0.7"
0.2"
N
Positional Offset (arcsec)

(c)



2.3'
0.01
0.05
0.5
0.75
Dec (deg)
RA (deg)

(d)



(λ) 10^1 10^2
10^0
10^-1
10^-2
k|| (h Mpc^-1)
10^-2 10^-1
k⊥ (h Mpc^-1)
10^8
10^6
10^4
10^2
10^0
|P1 - P2| (mK^2 h^-3 Mpc^3)
Per-frequency, per-antenna
Cable reflection fit only
Low-order polynomial fit
Time and antenna averaged
EoR simulation
10^-1
10^0
k|| (h Mpc^-1)

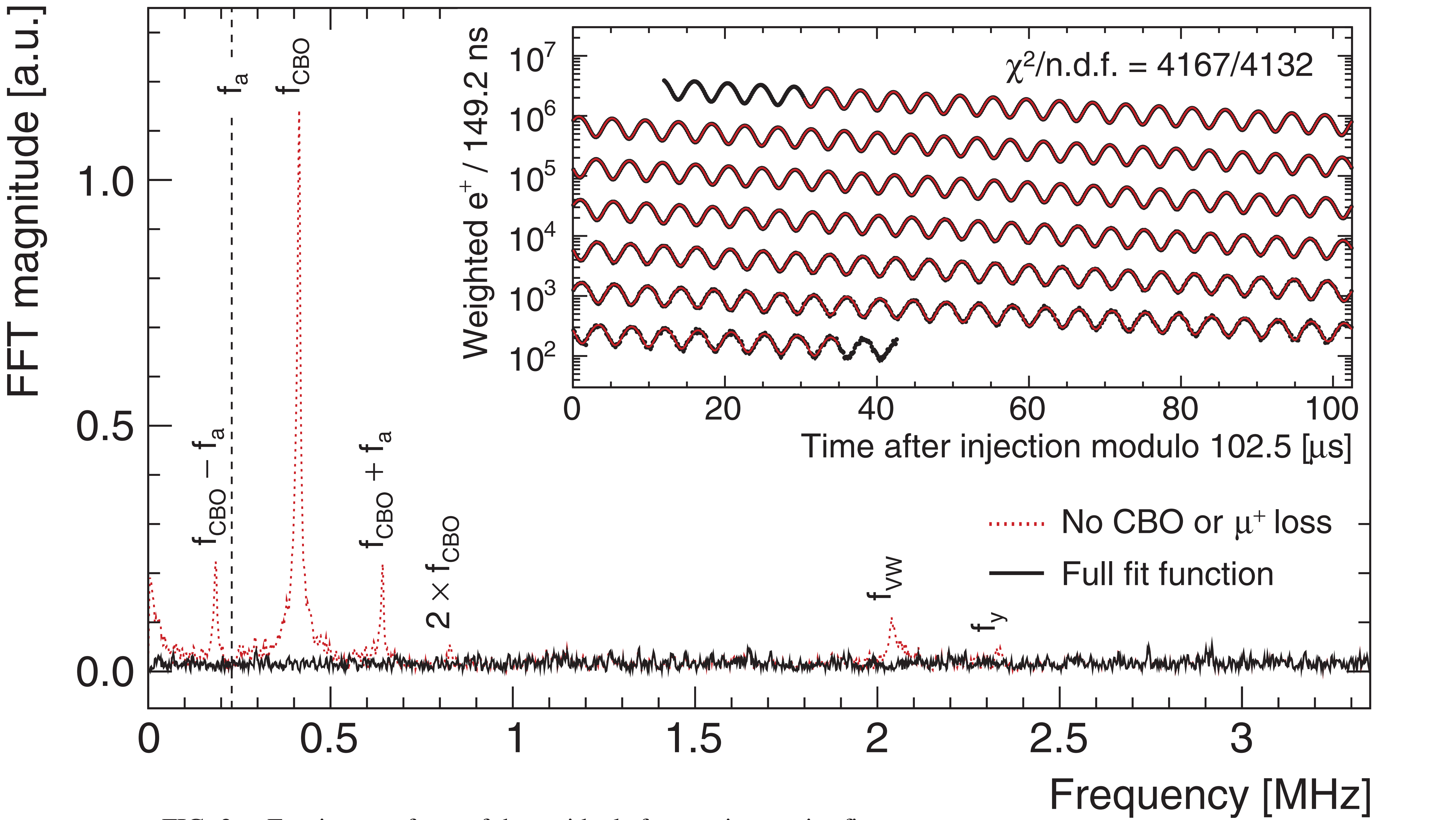
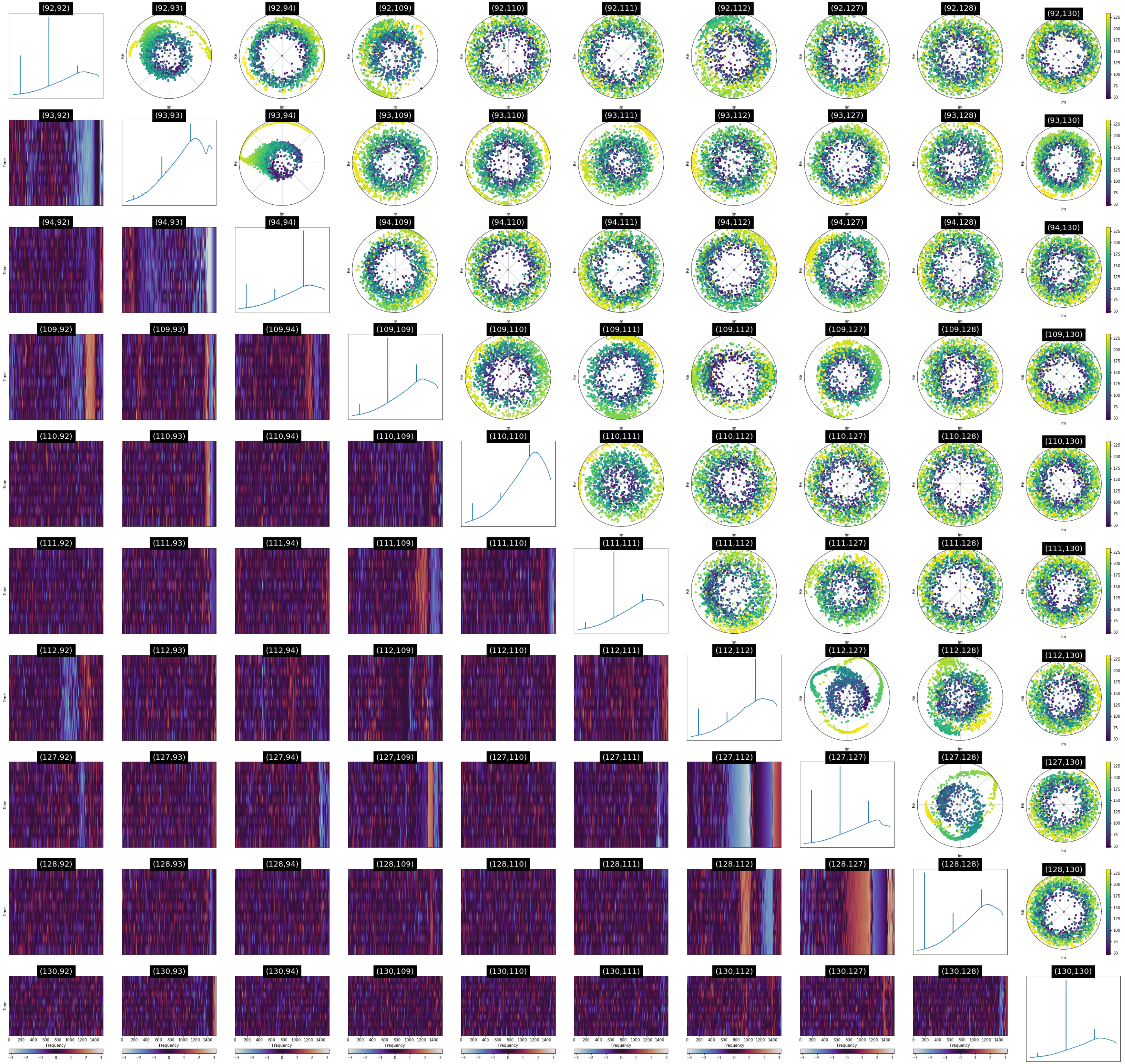


FIG. 2. Fourier transform of the residuals from a time-series fit following Eq. (5) but neglecting betatron motion and muon loss (red dashed), and from the full fit (black). The peaks correspond to the neglected betatron frequencies and muon loss. Inset: asymmetry-weighted e^+ time spectrum (black) from the Run-1c run group fit with the full fit function (red) overlaid.

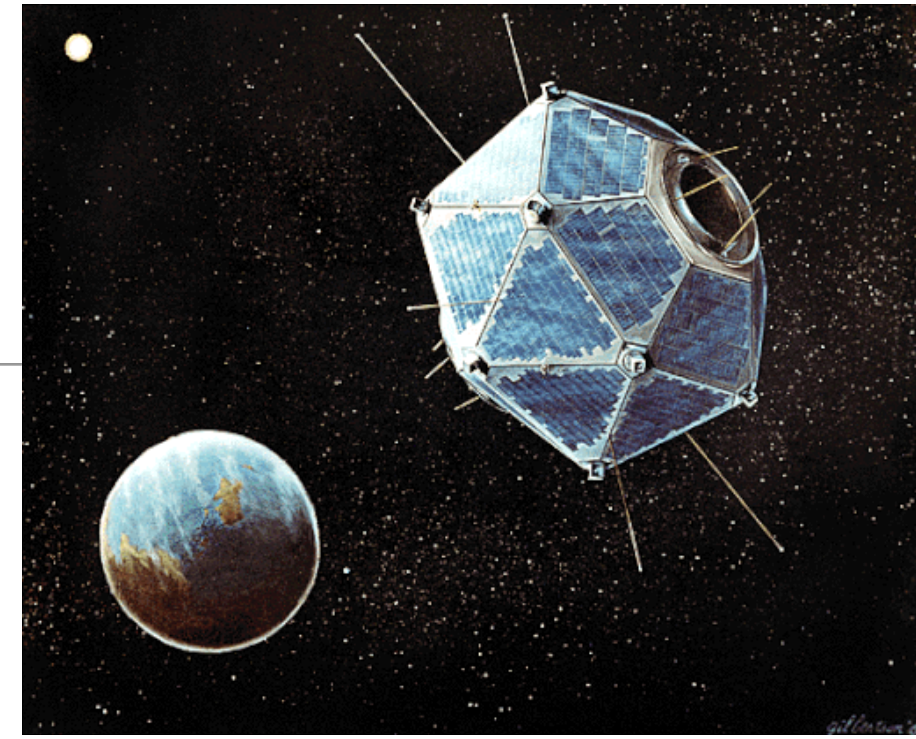


(92,92)
(92,93)
(92,94)
(92,109)
(92,110)
(92,111)
(92,112)
(92,127)
(92,128)
(92,130)
(93,92)
(93,93)
(93,94)
(93,109)
(93,110)
(93,111)
(93,112)
(93,127)
(93,128)
(93,130)
(94,92)
(94,93)
(94,94)
(94,109)
(94,110)
(94,111)
(94,112)
(94,127)
(94,128)
(94,130)
(109,92)
(109,93)
(109,94)
(109,109)
(109,110)
(109,111)
(109,112)
(109,127)
(109,128)
(109,130)
(110,92)
(110,93)
(110,94)
(110,109)
(110,110)
(110,111)
(110,112)
(110,127)
(110,128)
(110,130)
(111,92)
(111,93)
(111,94)
(111,109)
(111,110)
(111,111)
(111,112)
(111,127)
(111,128)
(111,130)
(112,92)
(112,93)
(112,94)
(112,109)
(112,110)
(112,111)
(112,112)
(112,127)
(112,128)
(112,130)
(127,92)
(127,93)
(127,94)
(127,109)
(127,110)
(127,111)
(127,112)
(127,127)
(127,128)
(127,130)
(128,92)
(128,93)
(128,94)
(128,109)
(128,110)
(128,111)
(128,112)
(128,127)
(128,128)
(128,130)
(130,92)
(130,93)
(130,94)
(130,109)
(130,110)
(130,111)
(130,112)
(130,127)
(130,128)
(130,130)
Re
Im
Time
Frequency

Trials factor

Look elsewhere effect

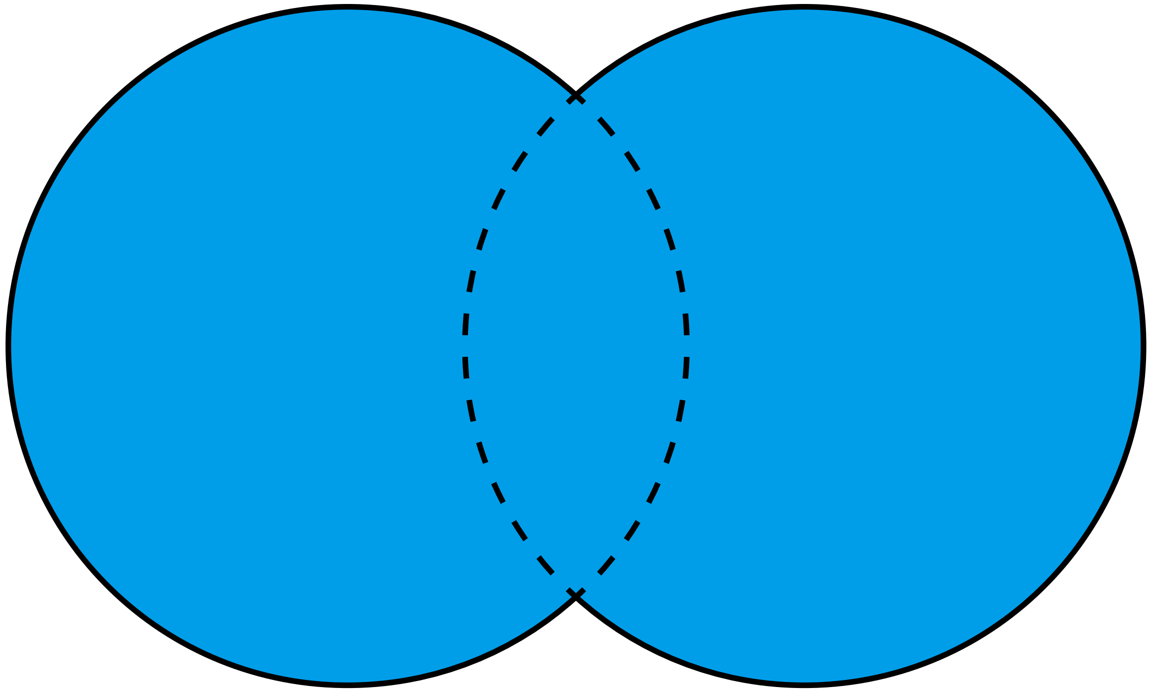
Trials



The Vela satellites were launched in the mid 1960’s to look for atmospheric nuclear explosions. Assume the internal gamma-ray detector had a background of 0.85 events per second. After scanning through 120 days of data it has looked at 10.4 million 1 second intervals. How many ‘5 sigma’ 1 second events will it have seen due to the background?

OR

$$P(A \text{ or } \mathrm{B}) = \mathrm{P(A)} + \mathrm{P(B)} - \mathrm{P(A \text{ and } B)}$$

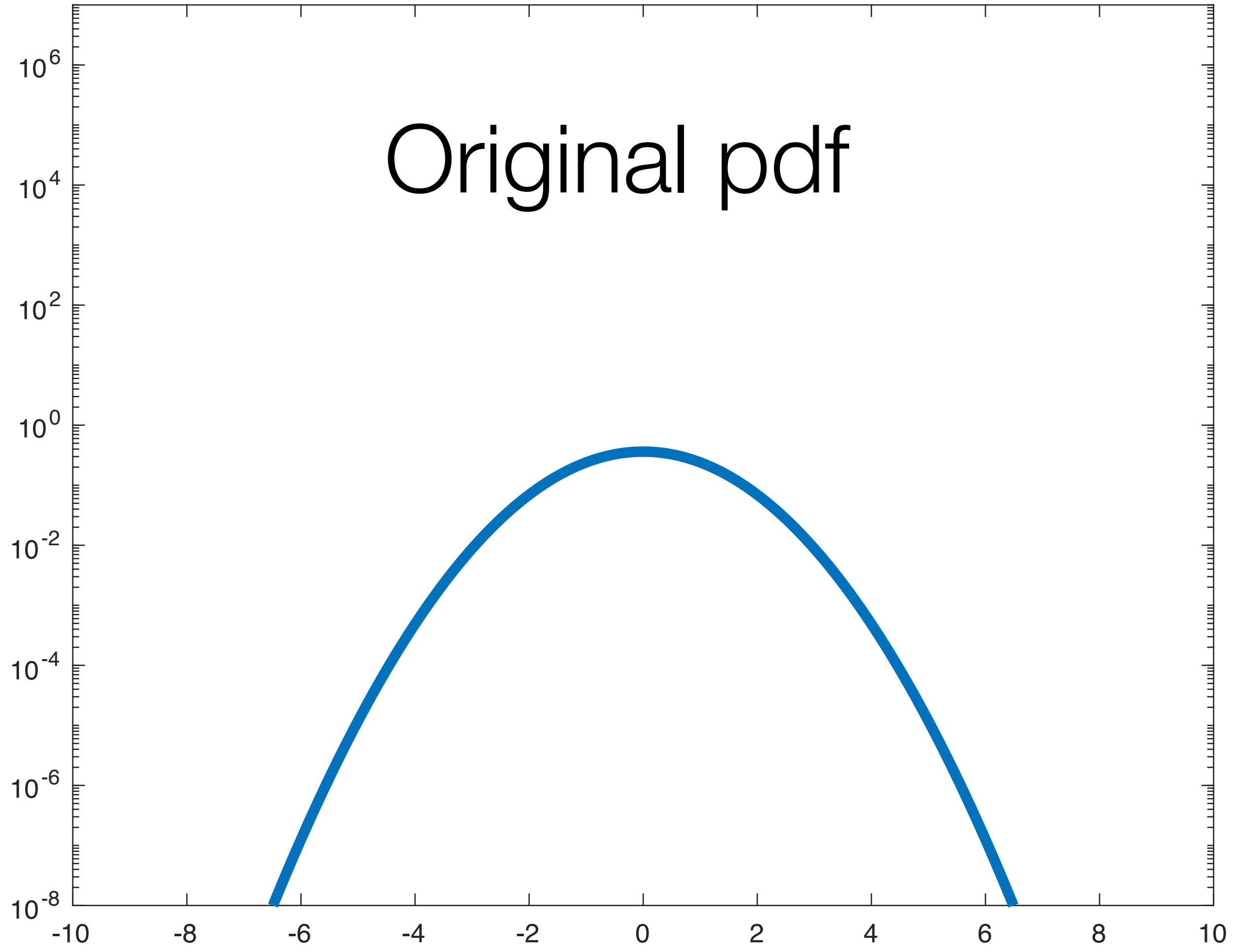


- When small prob

$$P(A \text{ or } \mathrm{B}) \approx \mathrm{P(A)} + \mathrm{P(B)}$$

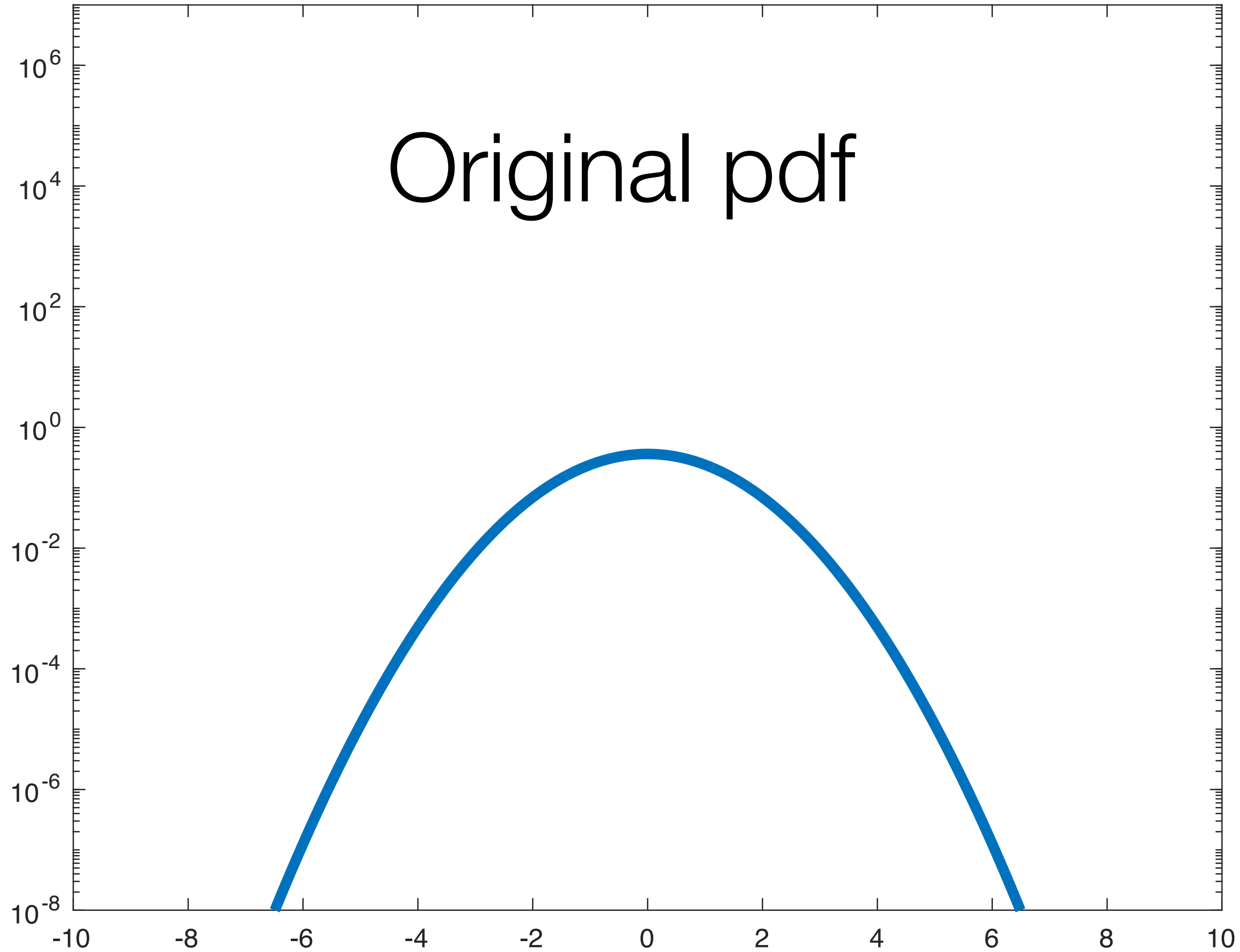
Three closely related questions

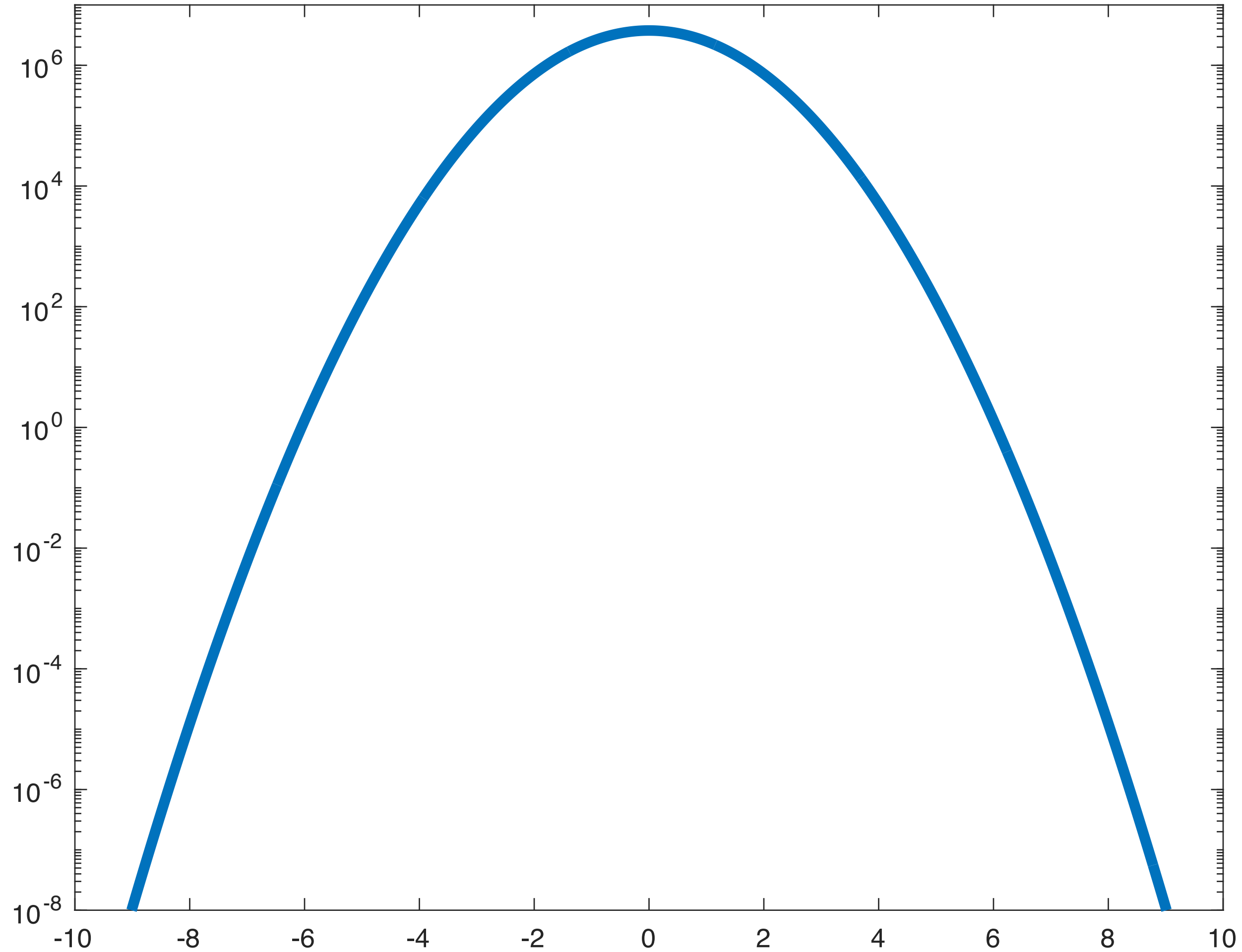
- How many events above X would I expect to see?
- What is the probability of seeing <u>one</u> event above X?
- What is the probability of seeing one or more events above X?



How to deal with trials

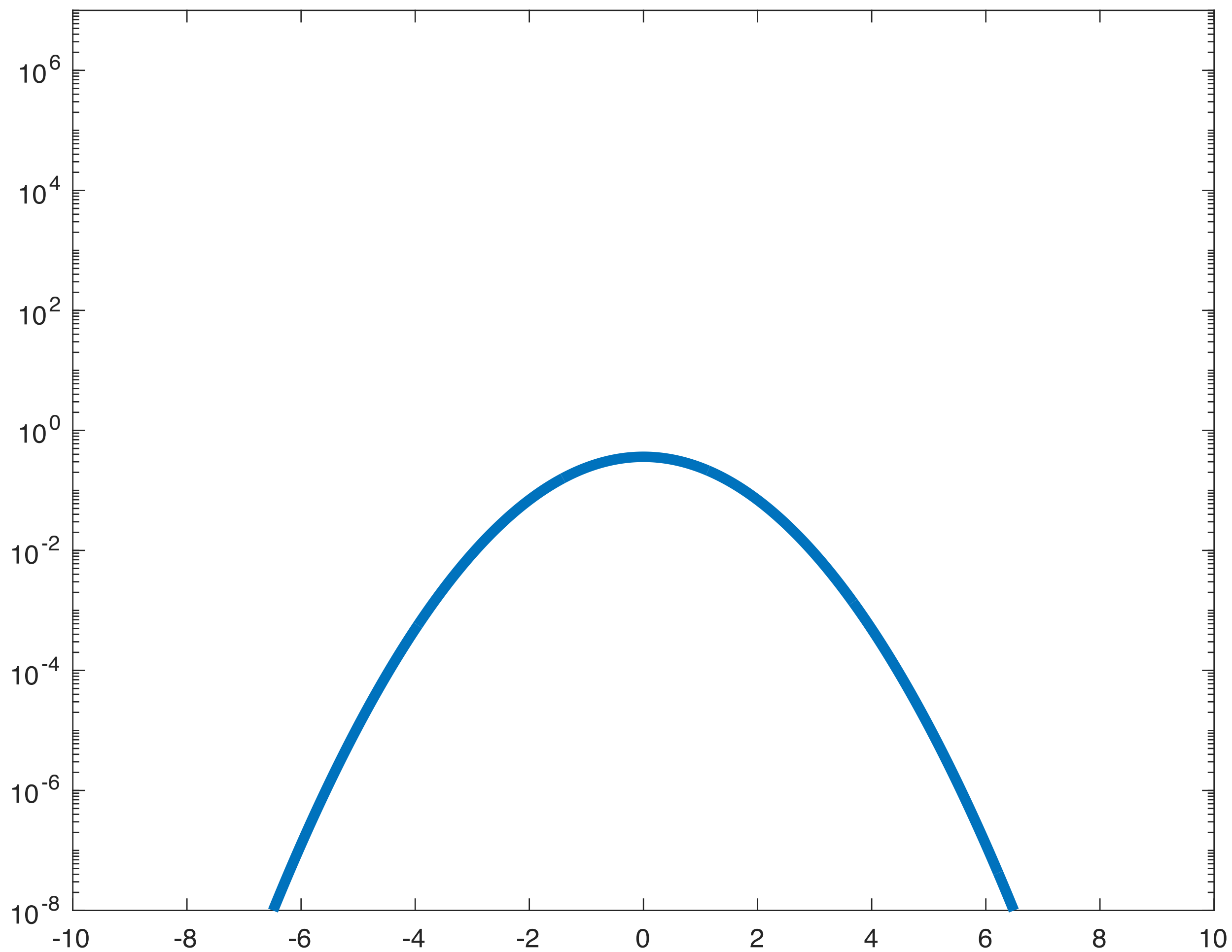
- Multiplying pdf() by number of independent locations searched gives number of events of strength X over ensemble; integral is number of expected events stronger than X
- When expected number is small; expected number ≈ Prob of 1 event stronger than X

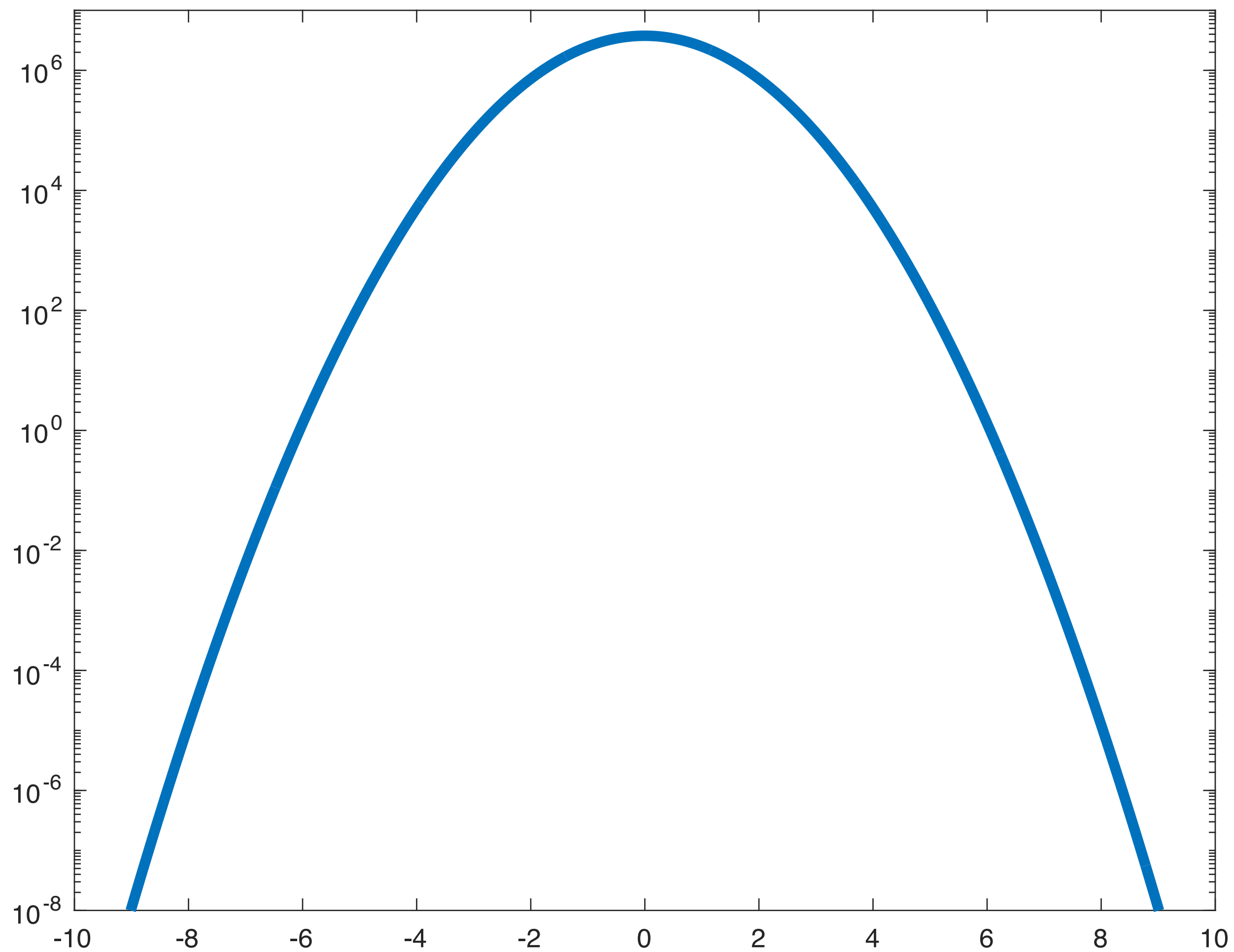




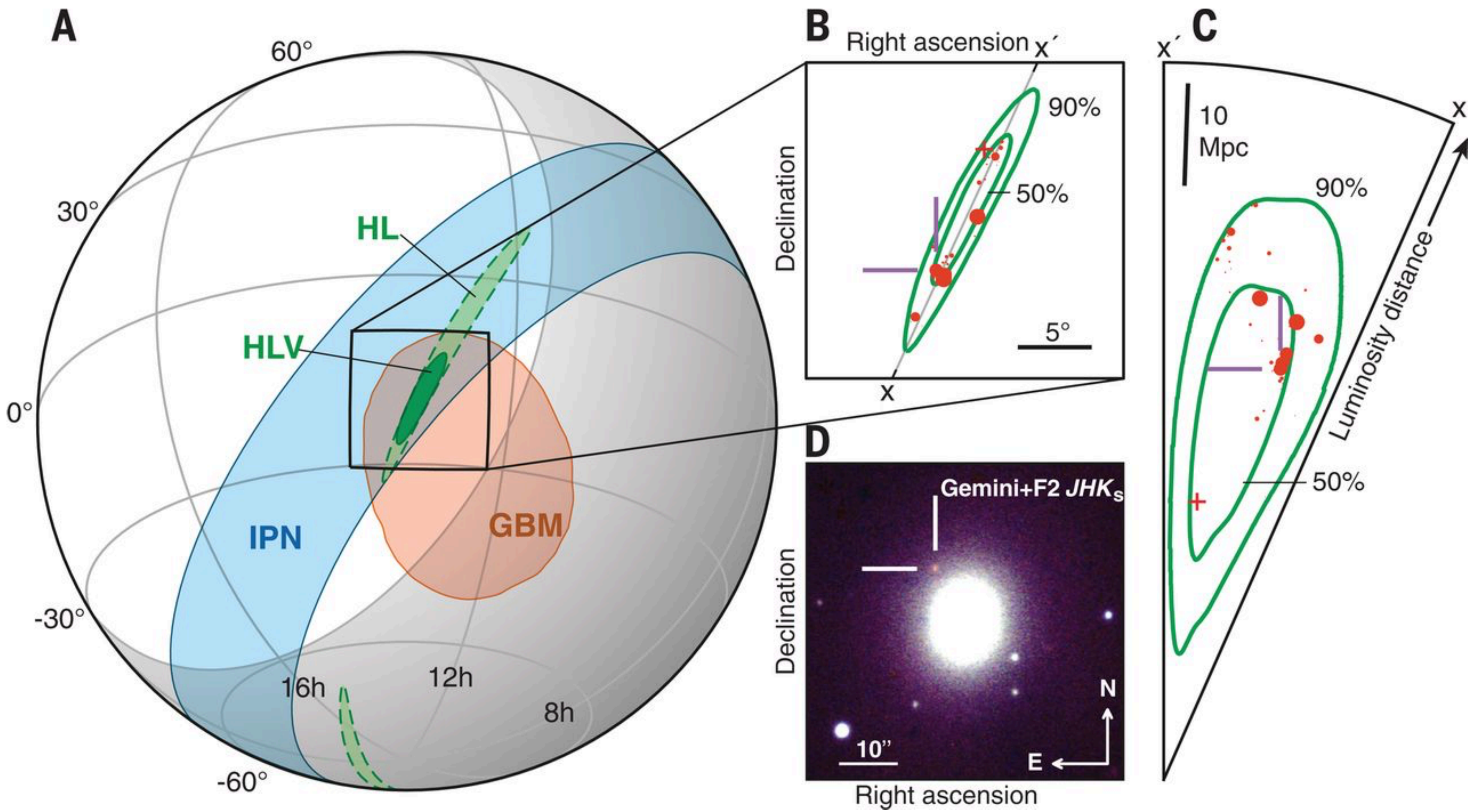
How to deal with trials

- There is a sensitivity penalty
- Really must get tails right
- Sensitivity penalty is small (~3x)

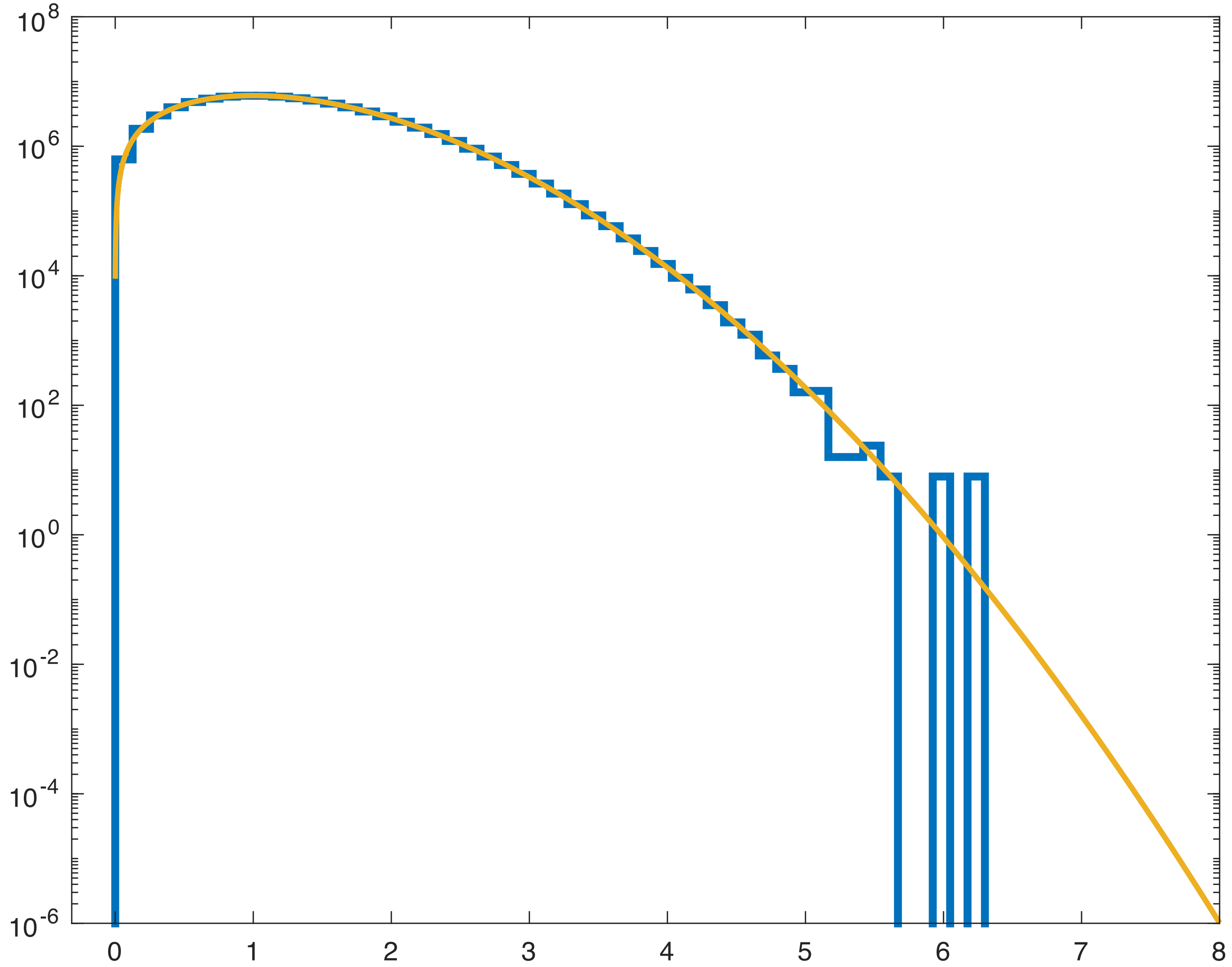




LIGO NS-NS counterpart search



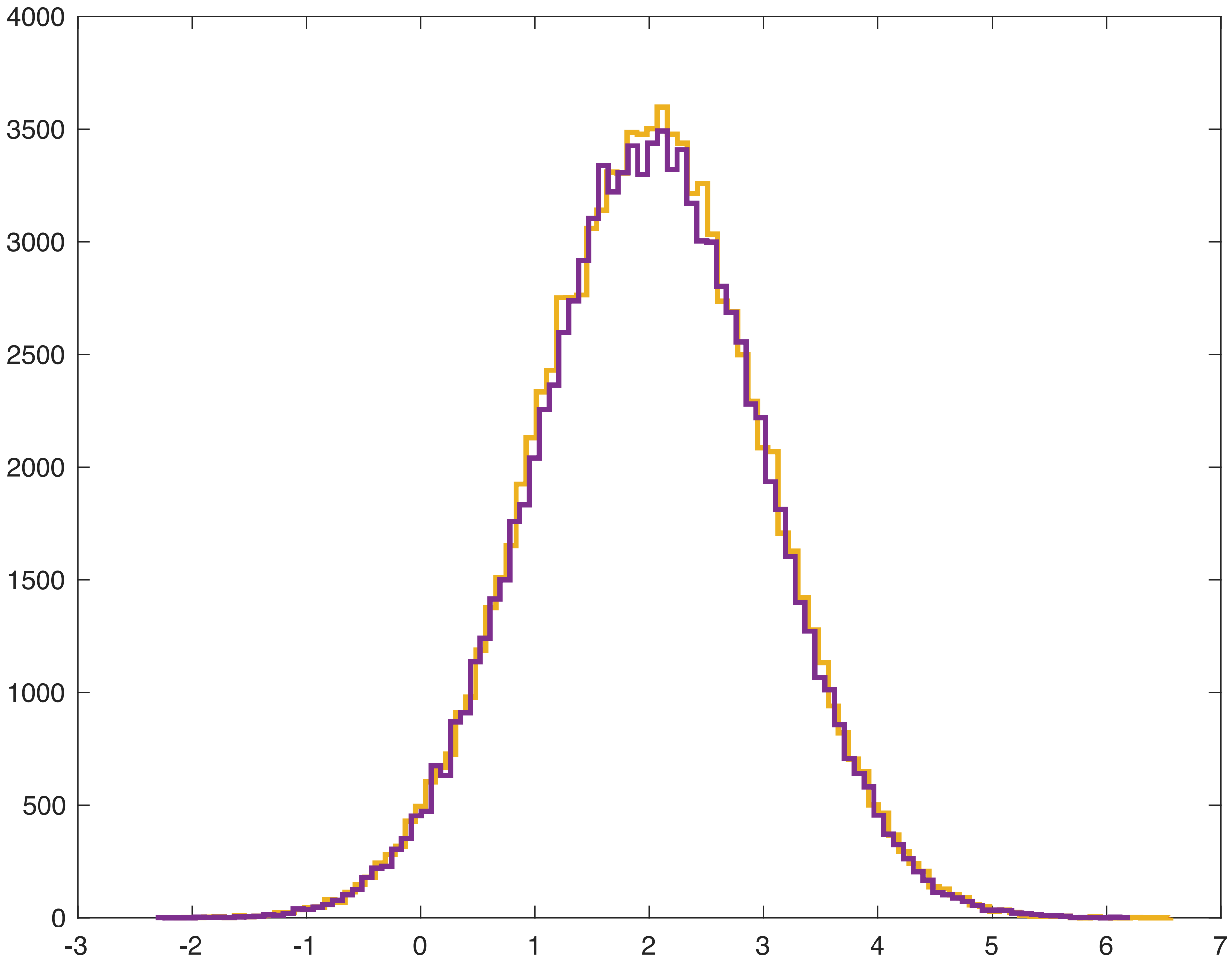
Data can drive trials



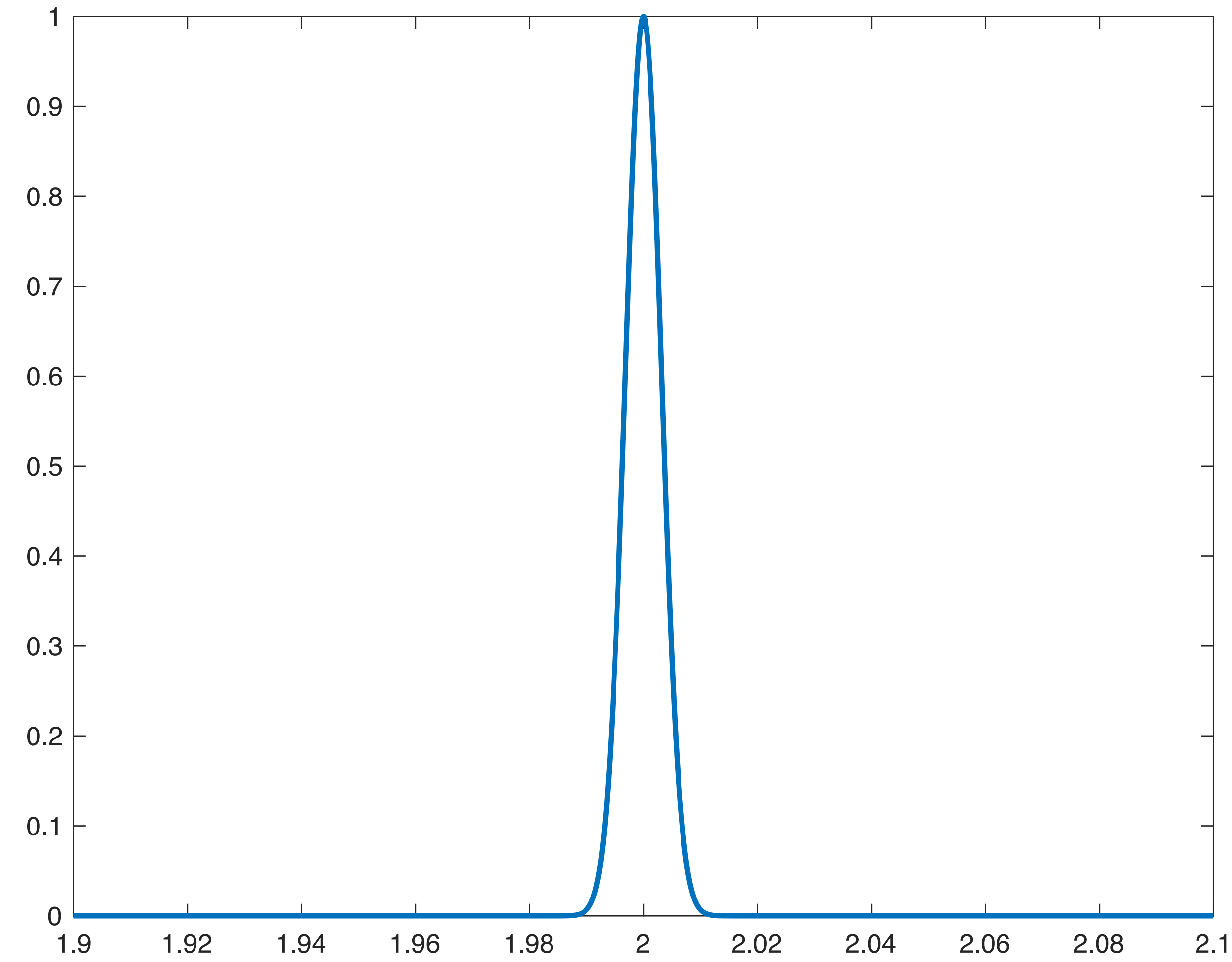
Parameters

Single parameter distribution

Population distribution
(10^5 measurements)



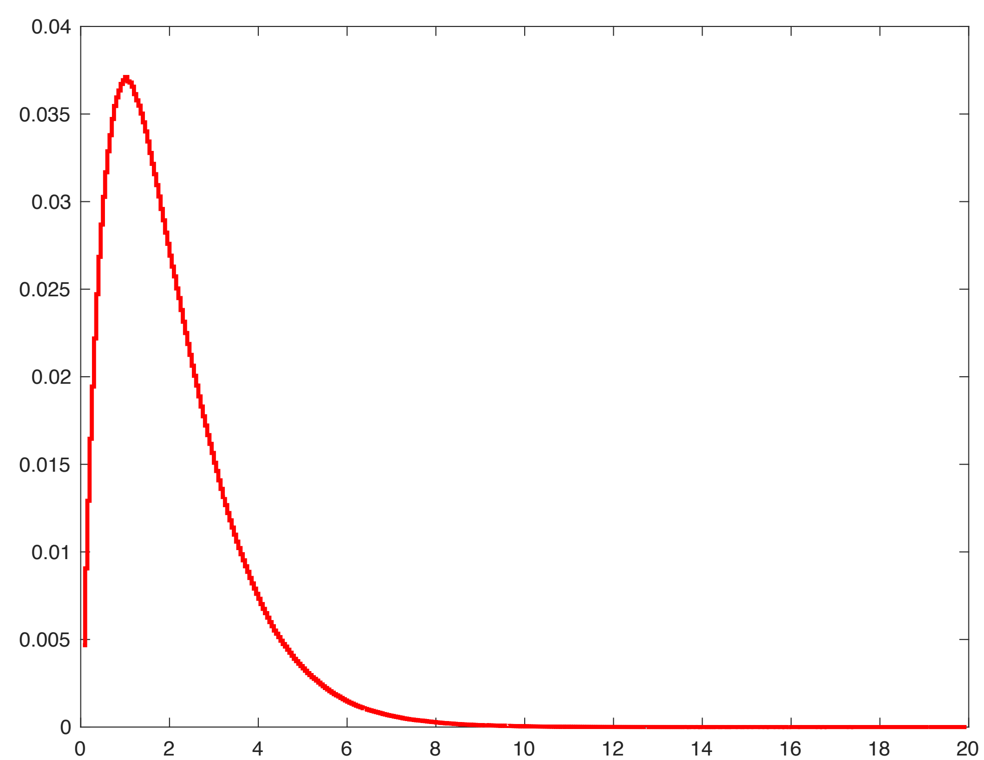
Distribution of the mean



$$\hat{m} = \langle \text{data} \rangle$$

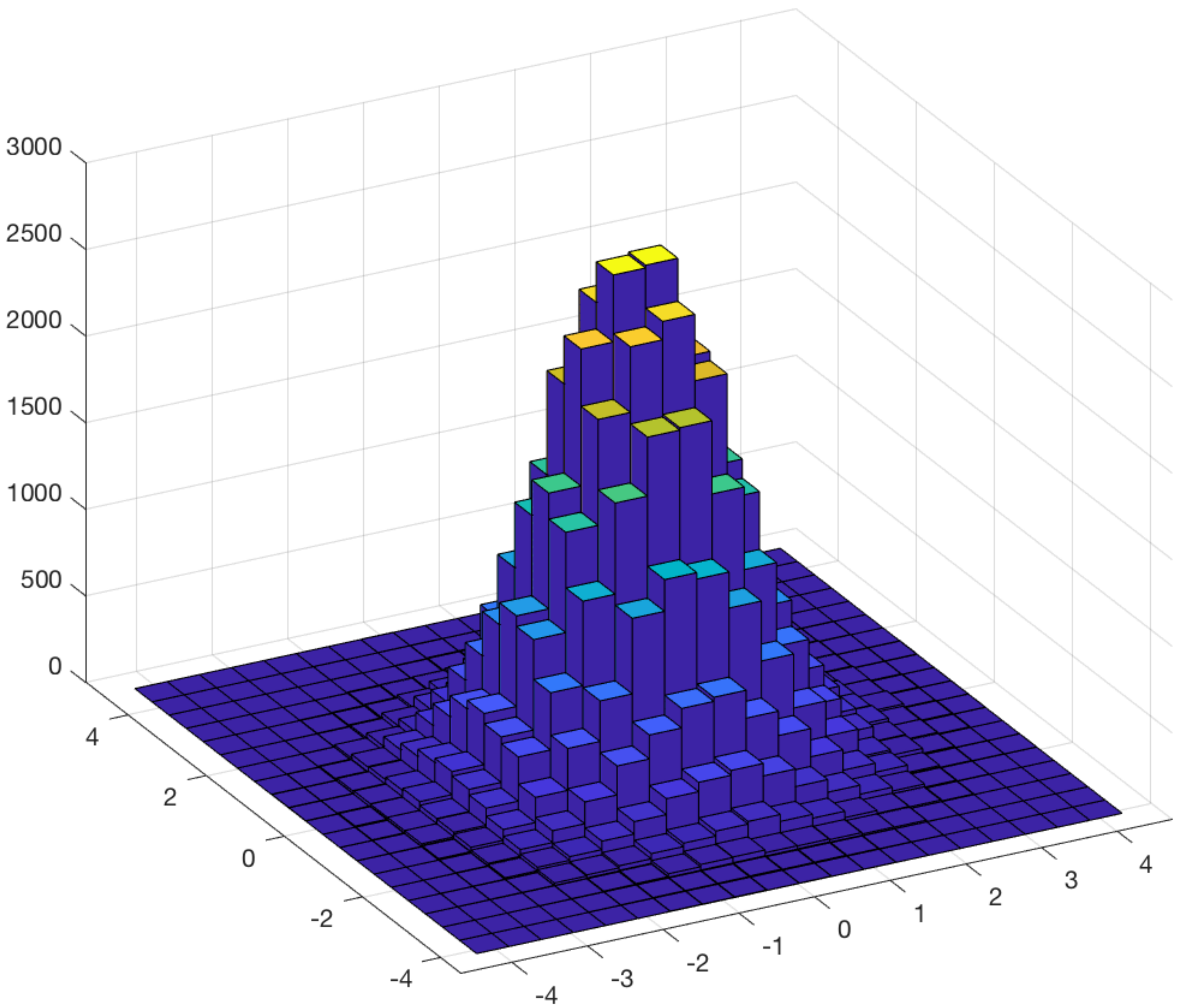
Parameters

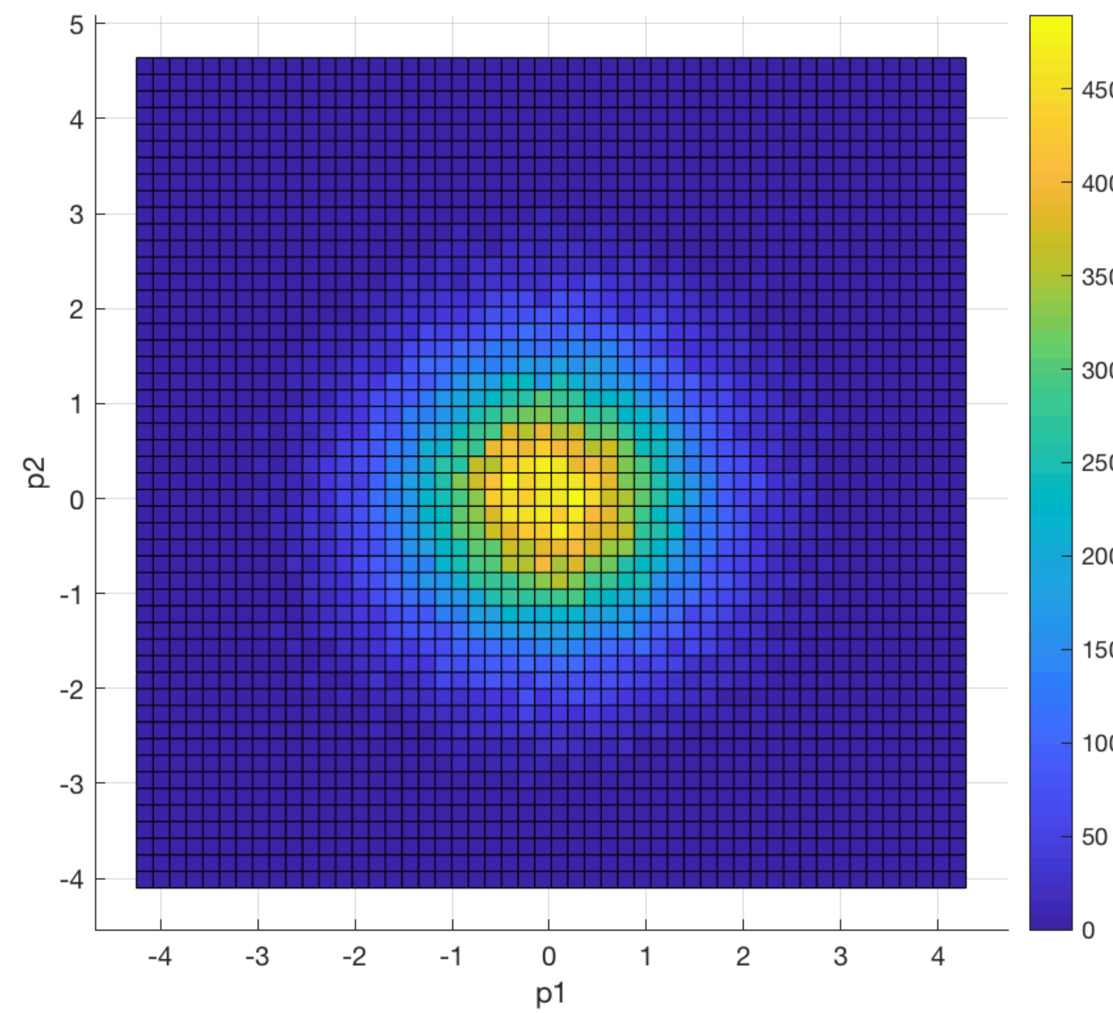
- Treat like a measurement—a value with some distribution
- Because the underlying is Gaussian does not mean the parameter is Gaussian, but can often propagate answer



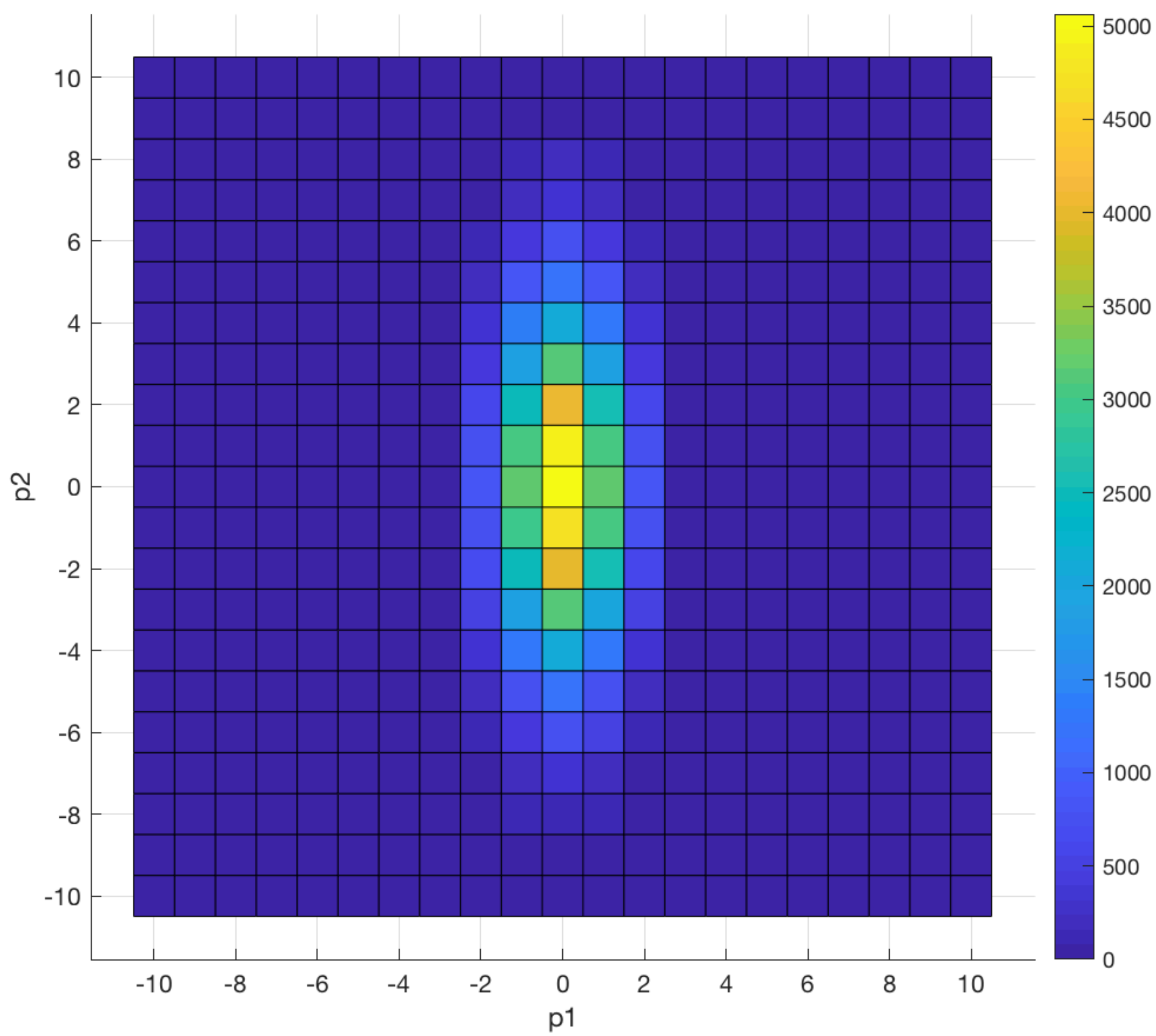
chi-square

Multiple parameters



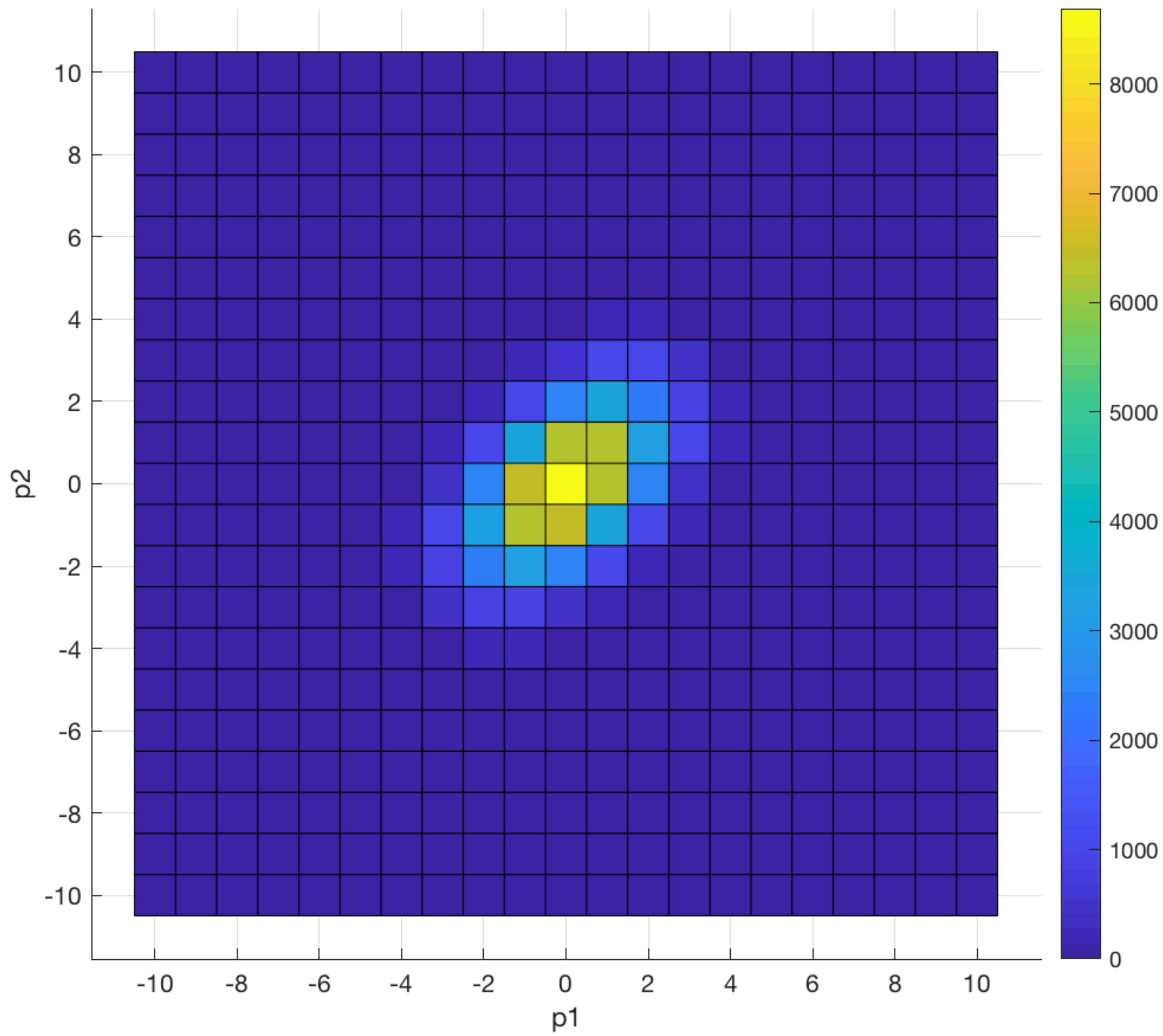


Non-equal variance



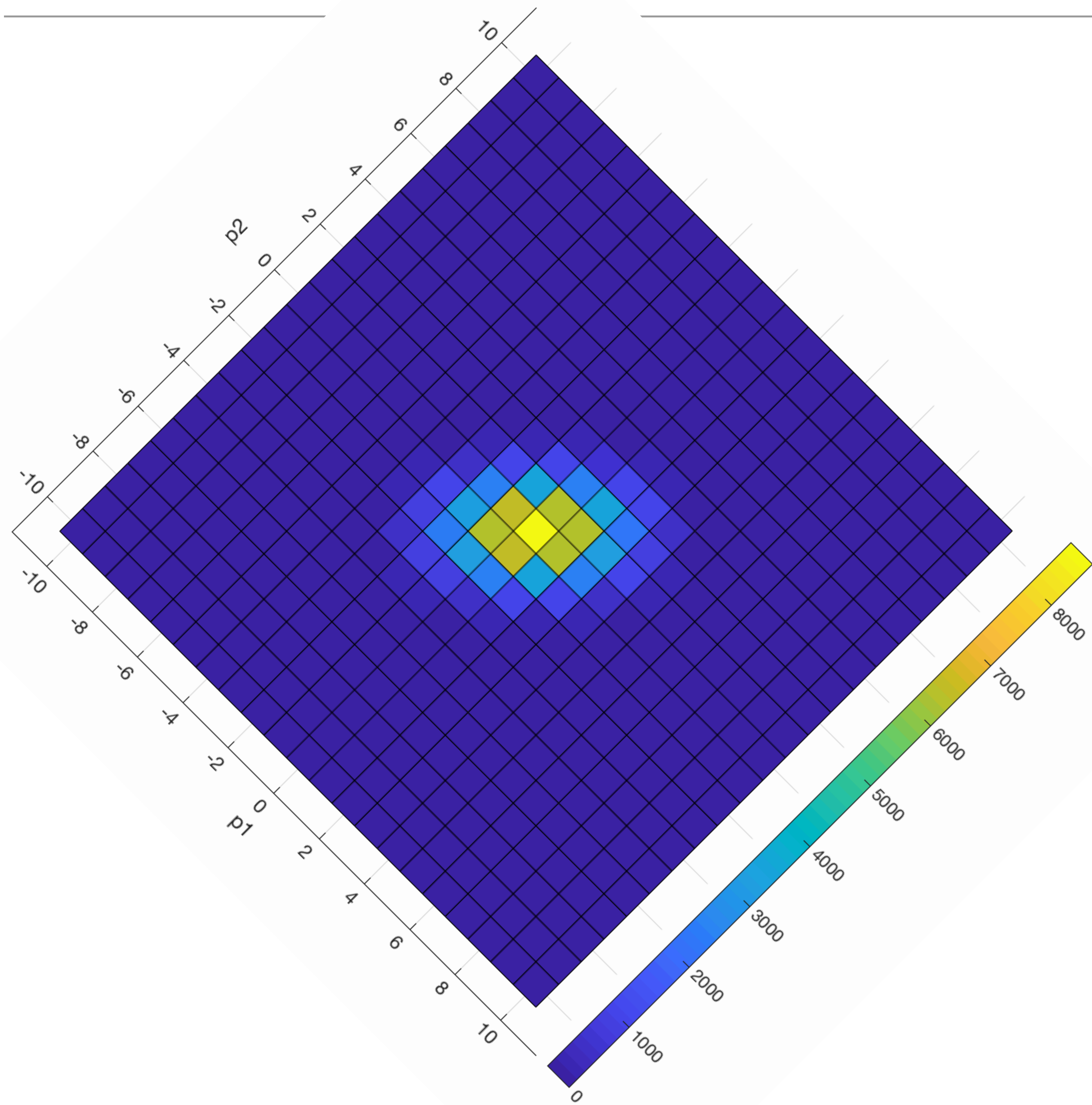
$$\sigma_1^2; \sigma_2^2$$

Covariance



$$C(p_1, p_2) = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}$$

Covariance



$$C(p_1', p_2') = \begin{bmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{22}^2 \end{bmatrix}$$

Aside on Fisher information matrix

- Dark Energy Task Force technical appendix is a good resource
- $C = F^{-1}$
- F is the second derivative of the log likelihood at peak
  $$F_{ij} = -\left\langle \frac{\partial^2 \ln \mathscr{L}}{\partial p_i \partial p_j} \right\rangle$$
- Conceptually is the Gaussian width $(1/\sigma_i^2)$ of the $\ln \mathscr{L}$
- Assumes Gaussian statistics(!)
- Easy to marginalize over nuisance parameters by converting to C, dropping rows/columns, inverting back

Triangle plot

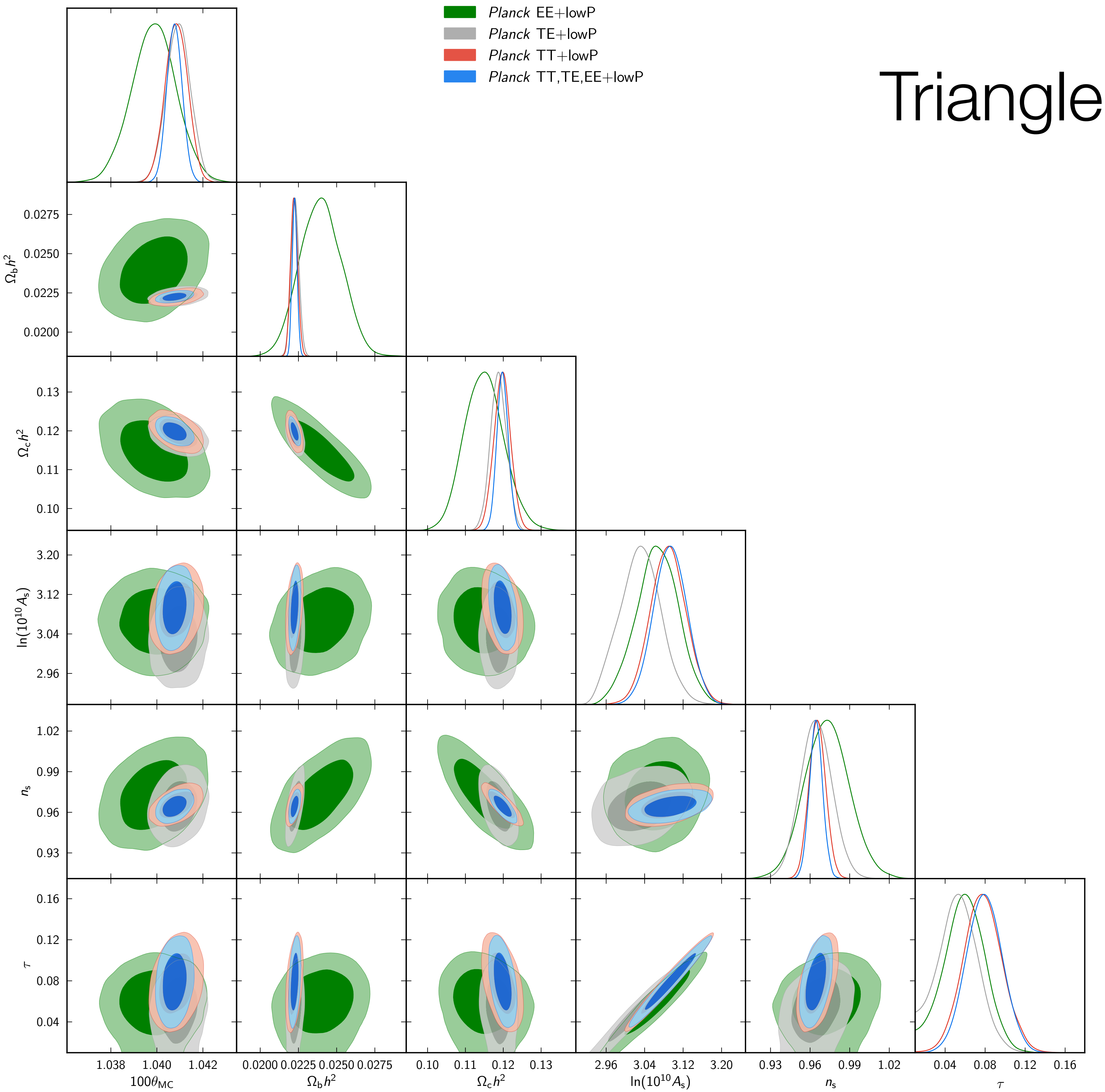
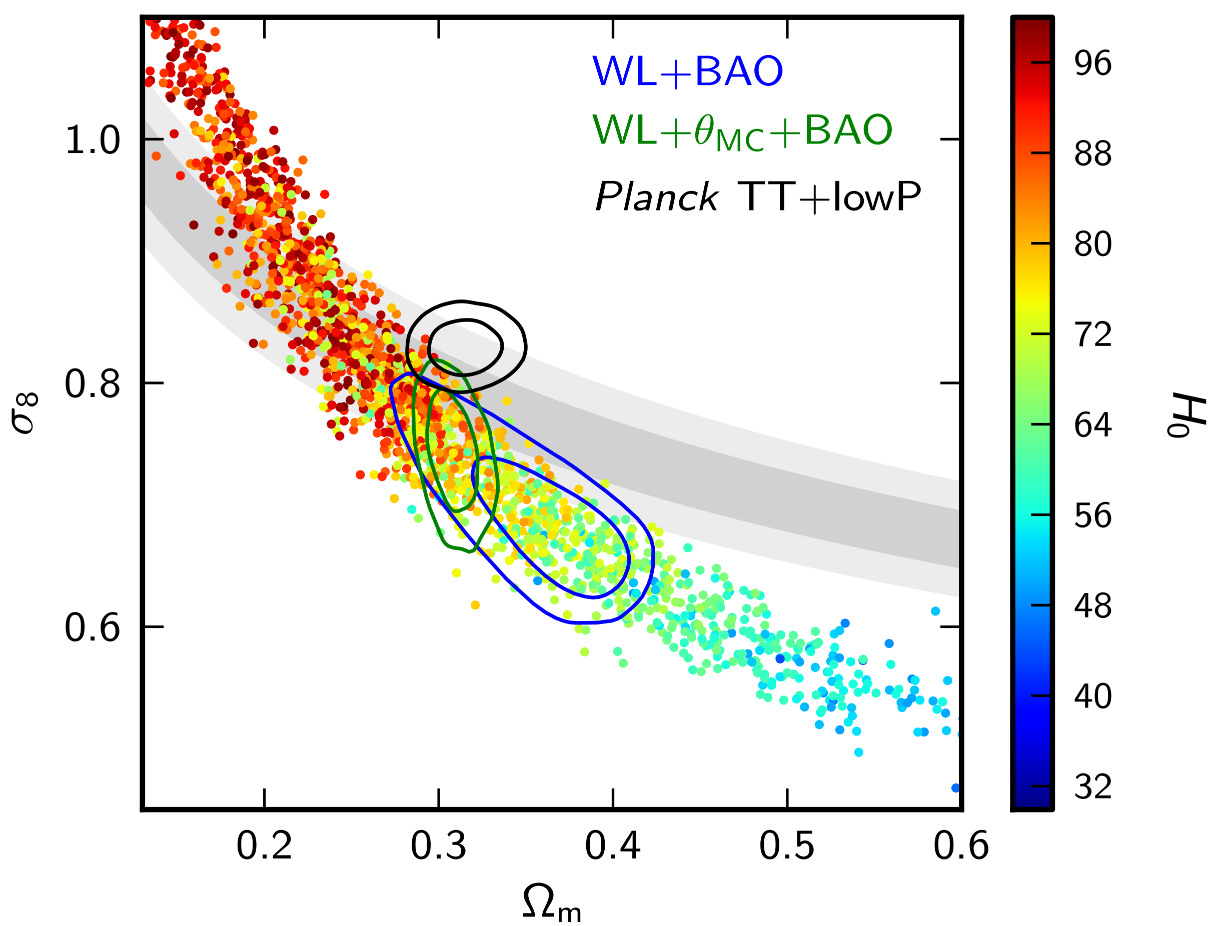
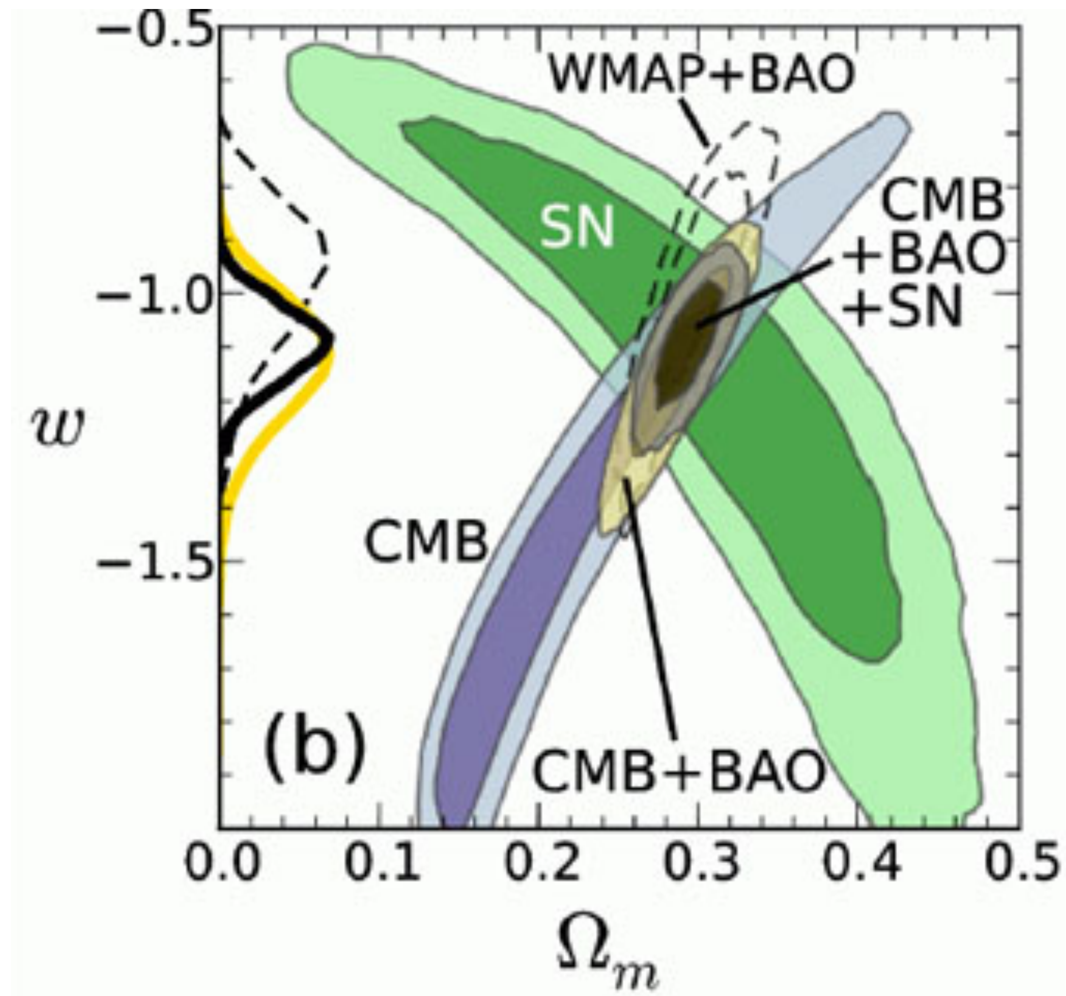


Fig. 6. Comparison of the base ΛCDM model parameter constraints from *Planck* temperature and polarization data.

Planck

Complicated interactions lead to curved multi-parameter plots (can't be described with C or F)



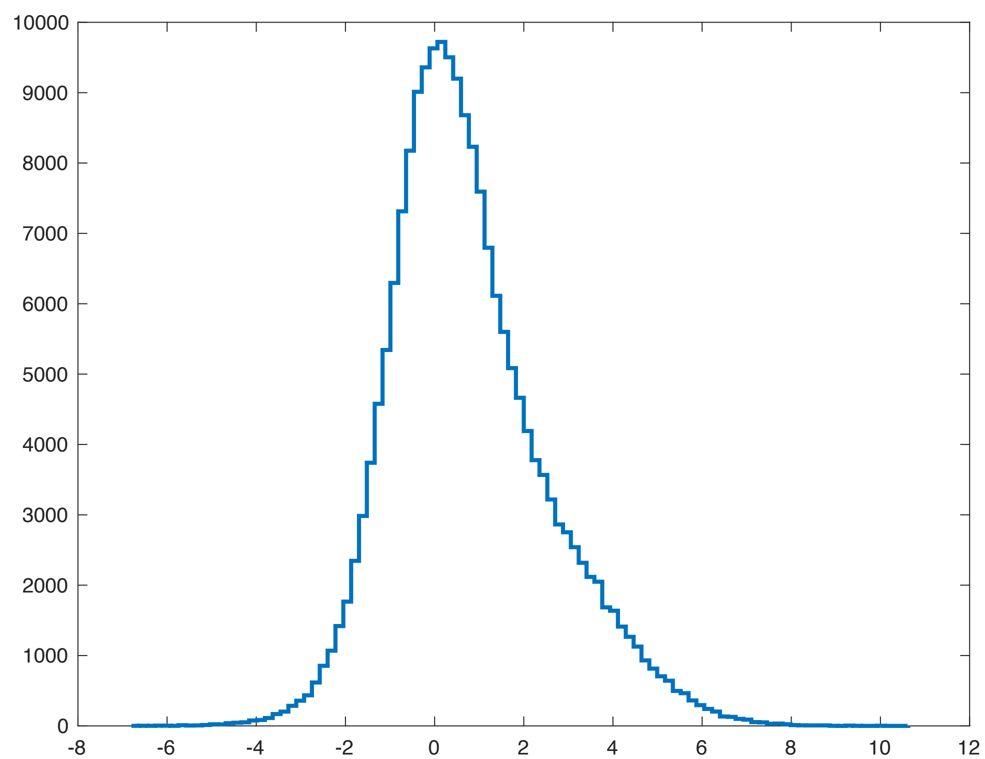


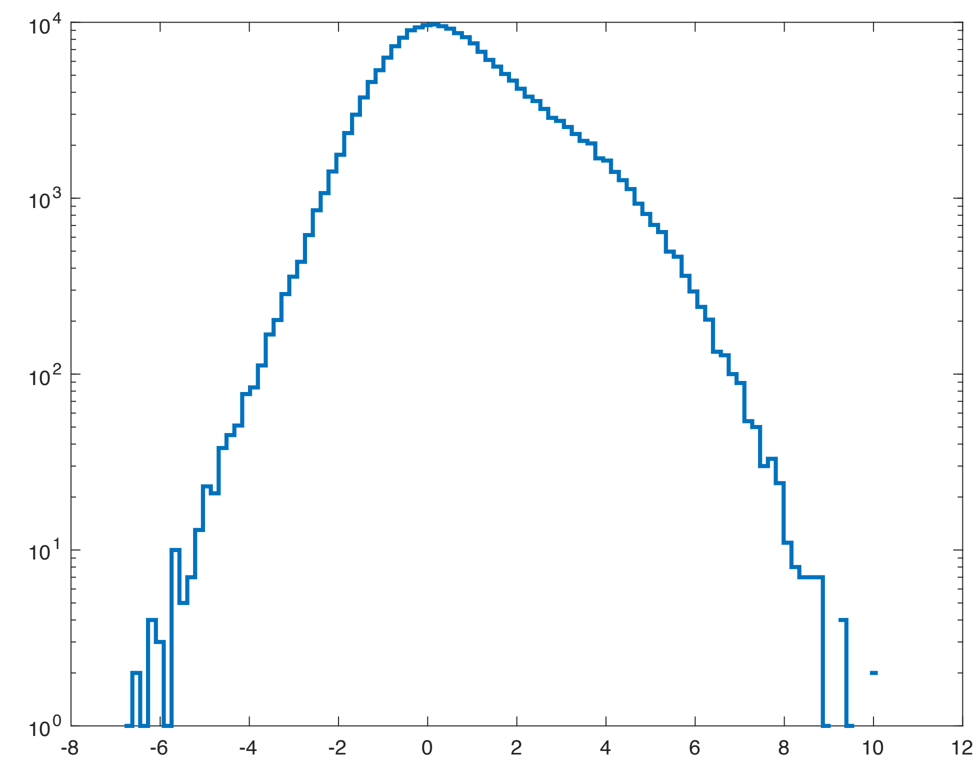
Take home message

- Multiple parameters can depend on each other in complicated ways
- Don’t assume they are independent of one another
- Can be due to Theory, your model, instrument, nature, or interactions

The art of parameterization

So you have a distribution to parametrize…





How do you pick ‘good’ parameters?

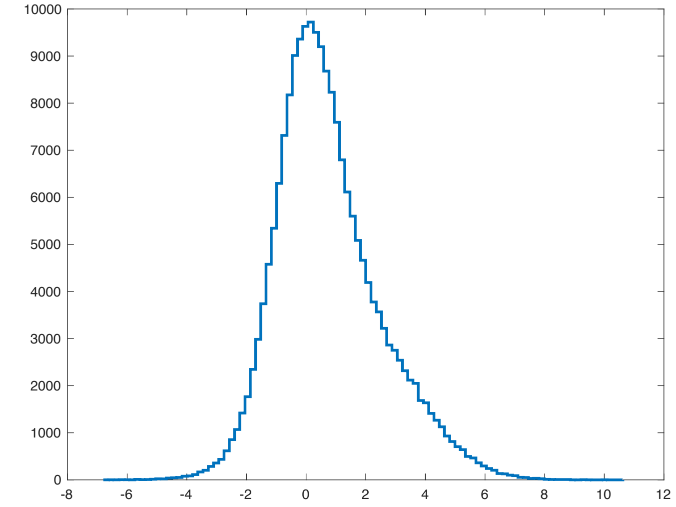
No good answer, topic of the remainder of the course

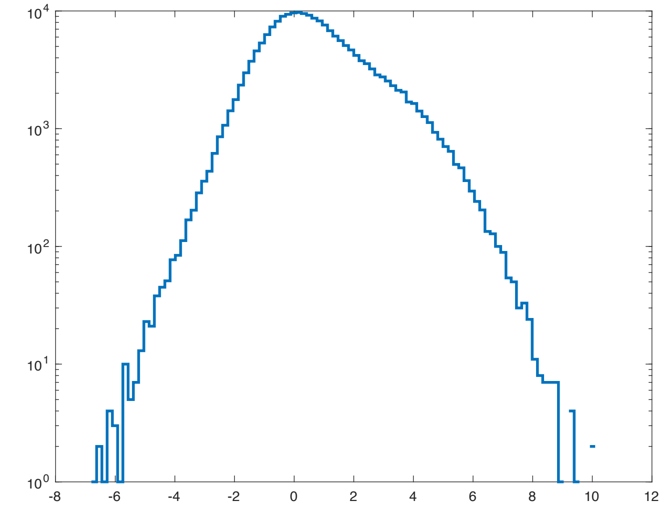
Failure modes

- Too few parameters of the wrong type don't describe the distribution ⇛ significance calculations wrong

- Too many parameters just 'fit the elephant'

    poorly constrained and/or highly covariant ⇛ unstable





Maybe it’s not new?

- Sometimes you can propagate the distribution from an earlier one that you understood.

Do you have a physical model?

- Physical models, particularly of an instrument, tend to work better.
- If they don’t work, you often learn something

Maybe there is a systematic?

- Often weirdness is caused by systematics
- Finding systematics is really adding to a physical model

Tests

- What are the covariances of the parameters on subsets of the data?
- Can I take special data to prove/explore a proposed physical model?
- Do parameters respond to my worries? (Does it pass jackknife tests?)

Love your data,
This is the value you add