Class 4: Introduction to quality plots

Miguel F. Morales
Bryna Hazelton

Outline

- Common purposes of plots
- Perception & accessibility basics
- Emphasis
- Where is the information?
- Drillable
- Developing a consistent language
- Examples

Kinds of plots

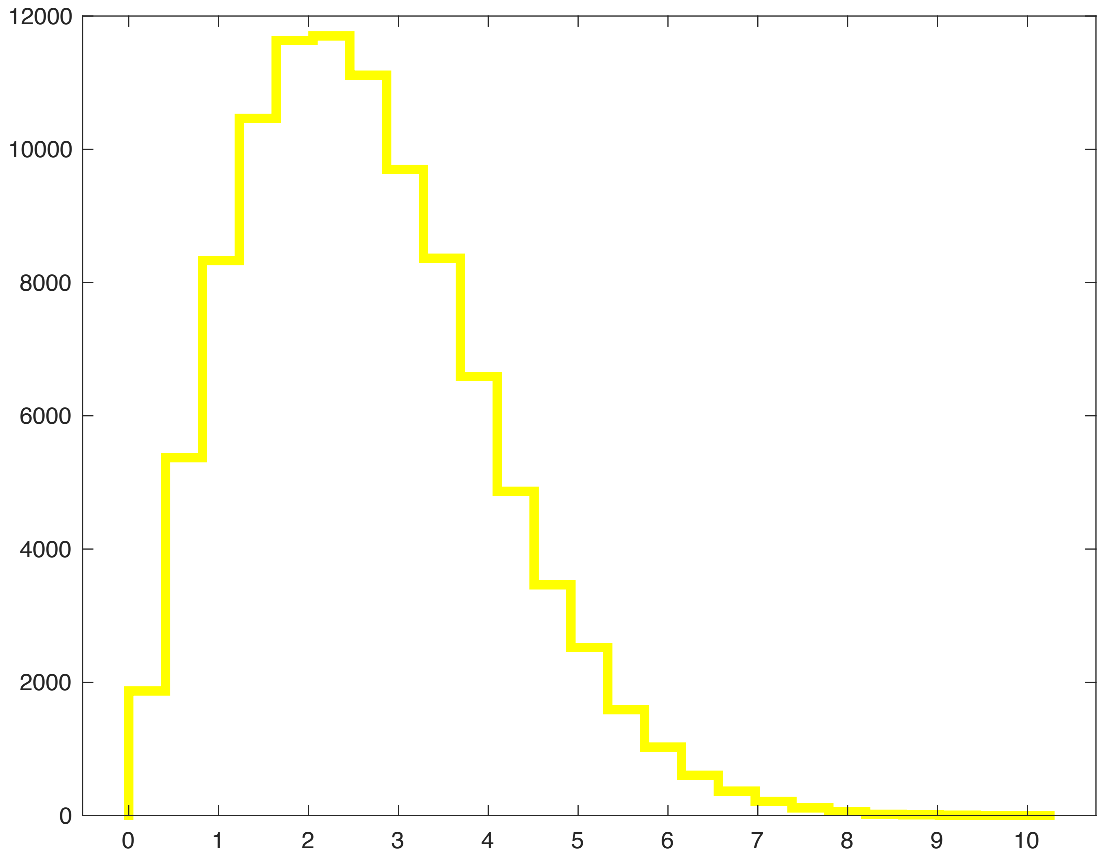
- Plots for answering a question (plots for you)
- Plots for monitoring data (readability)
- Plots for publication & presentation (pedagogical)

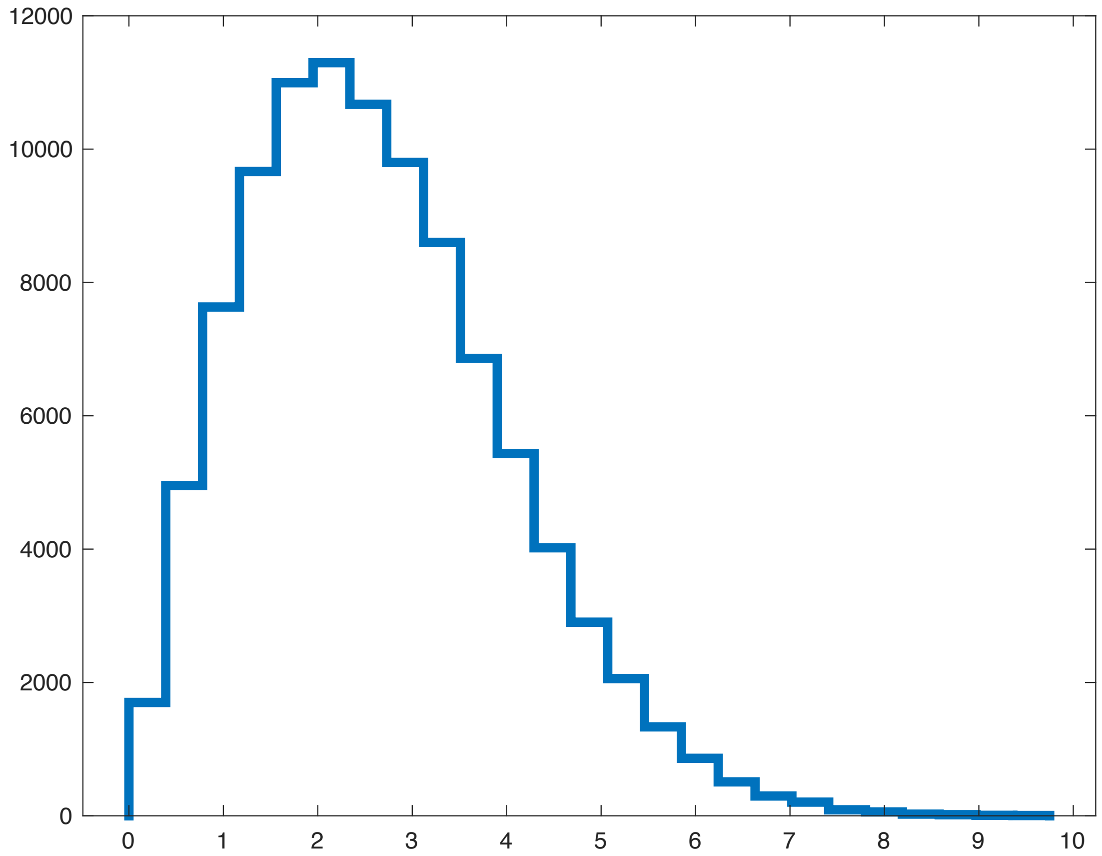
Scientific extensions to Tufte ideas

Perception basics

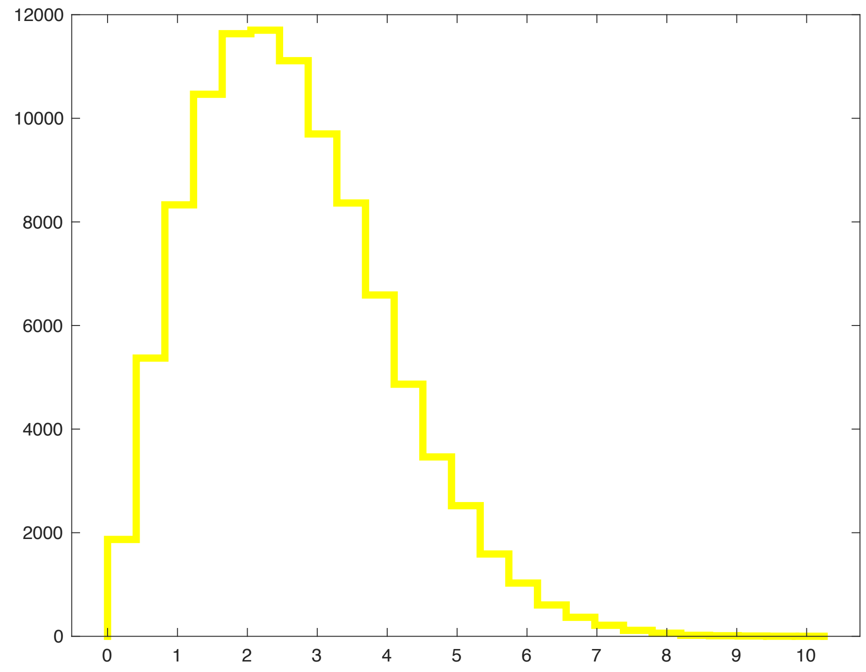
Object identification happens in black & white

Never use yellow on white

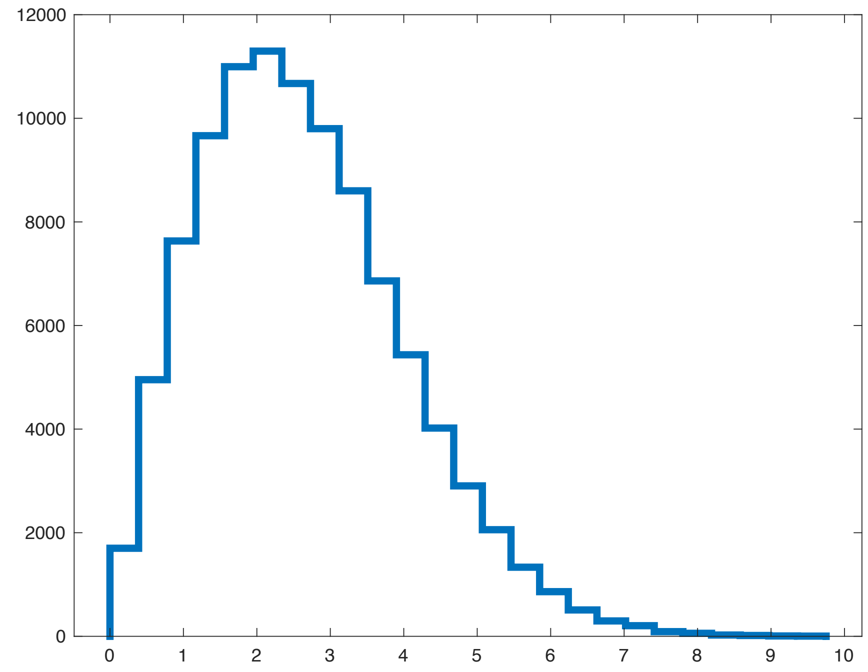




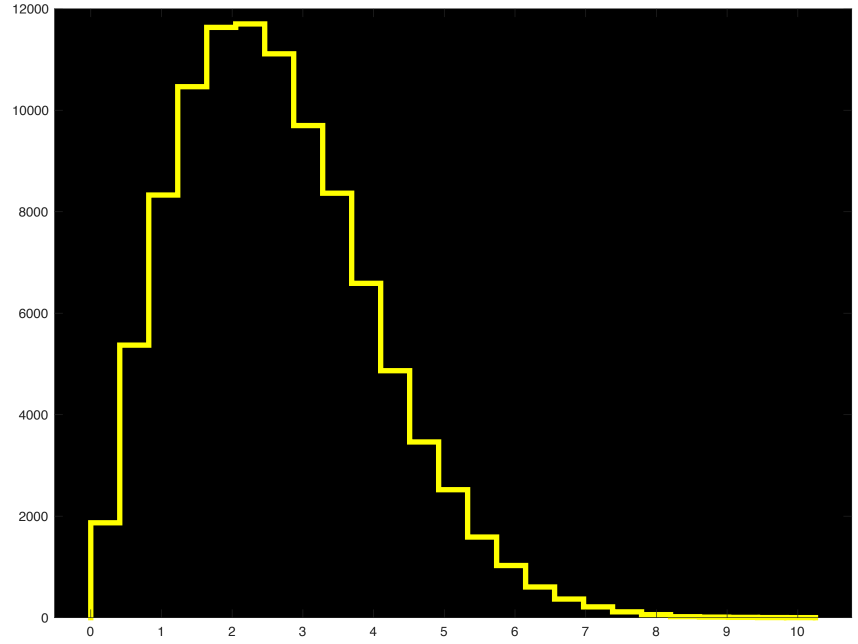
Contrast



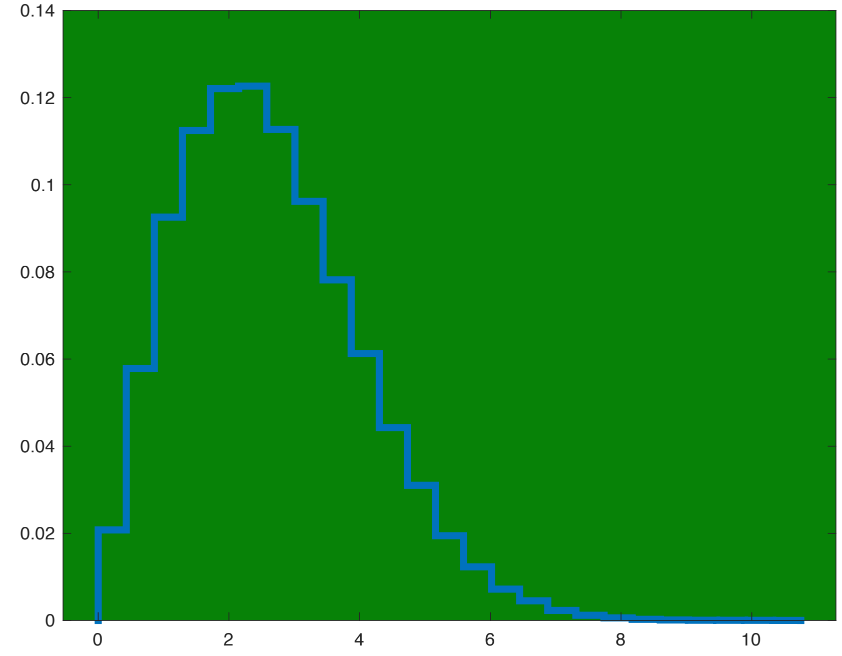
12000
10000
8000
6000
4000
2000
0
0 1 2 3 4 5 6 7 8 9 10



12000
10000
8000
6000
4000
2000
0
0 1 2 3 4 5 6 7 8 9 10



12000
10000
8000
6000
4000
2000
0
0 1 2 3 4 5 6 7 8 9 10



0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0 2 4 6 8 10

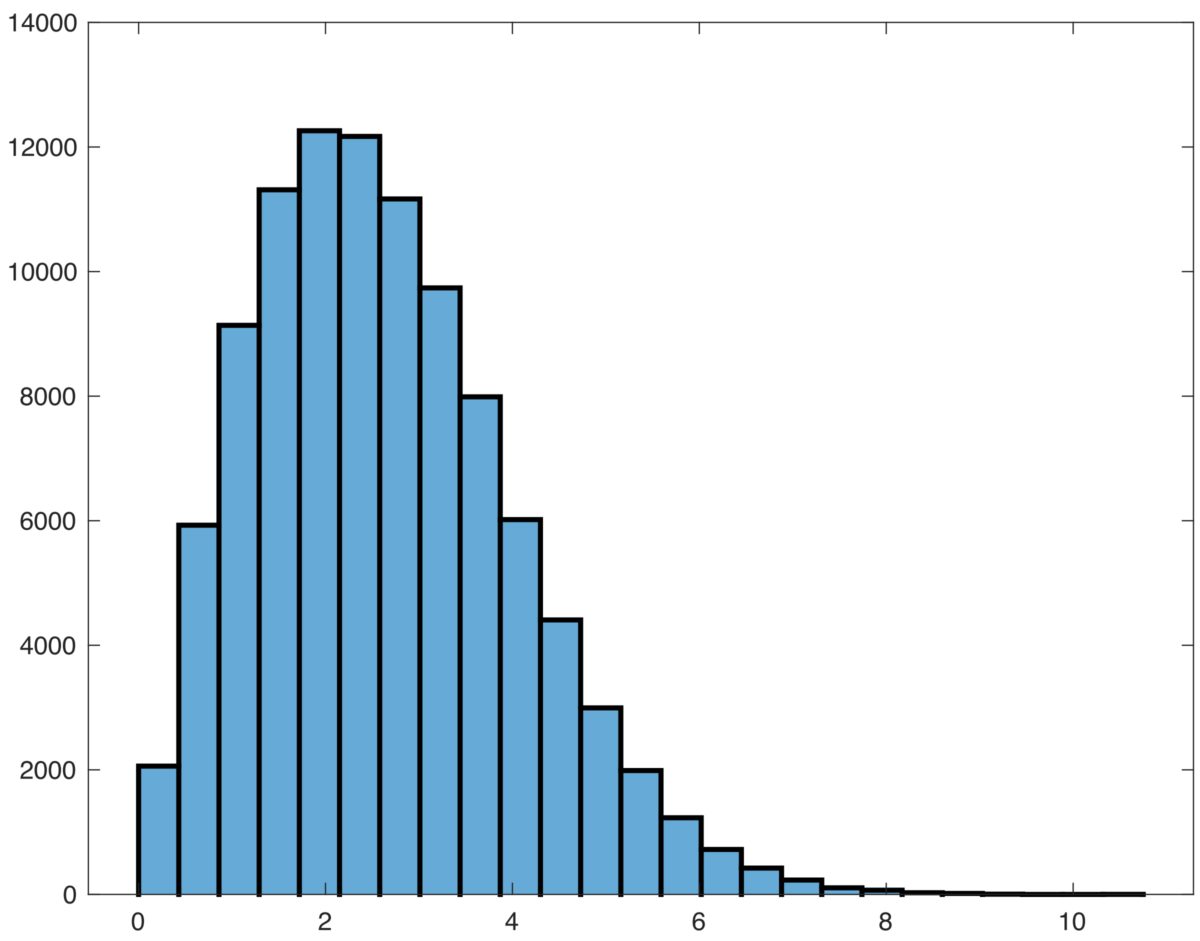
Accessibility

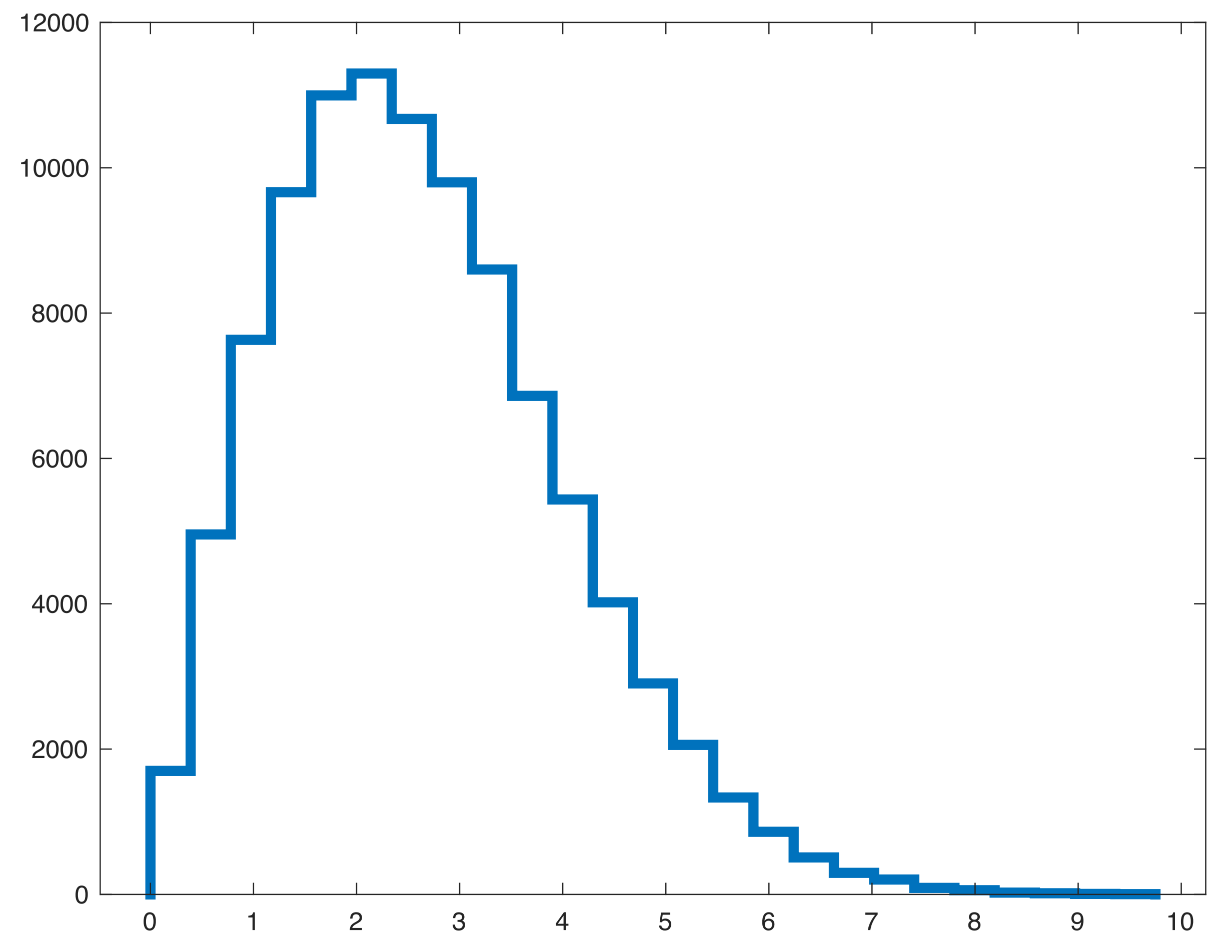
- Very different perceptions of color
- Use color blindness simulators

https://colororacle.org

Emphasis

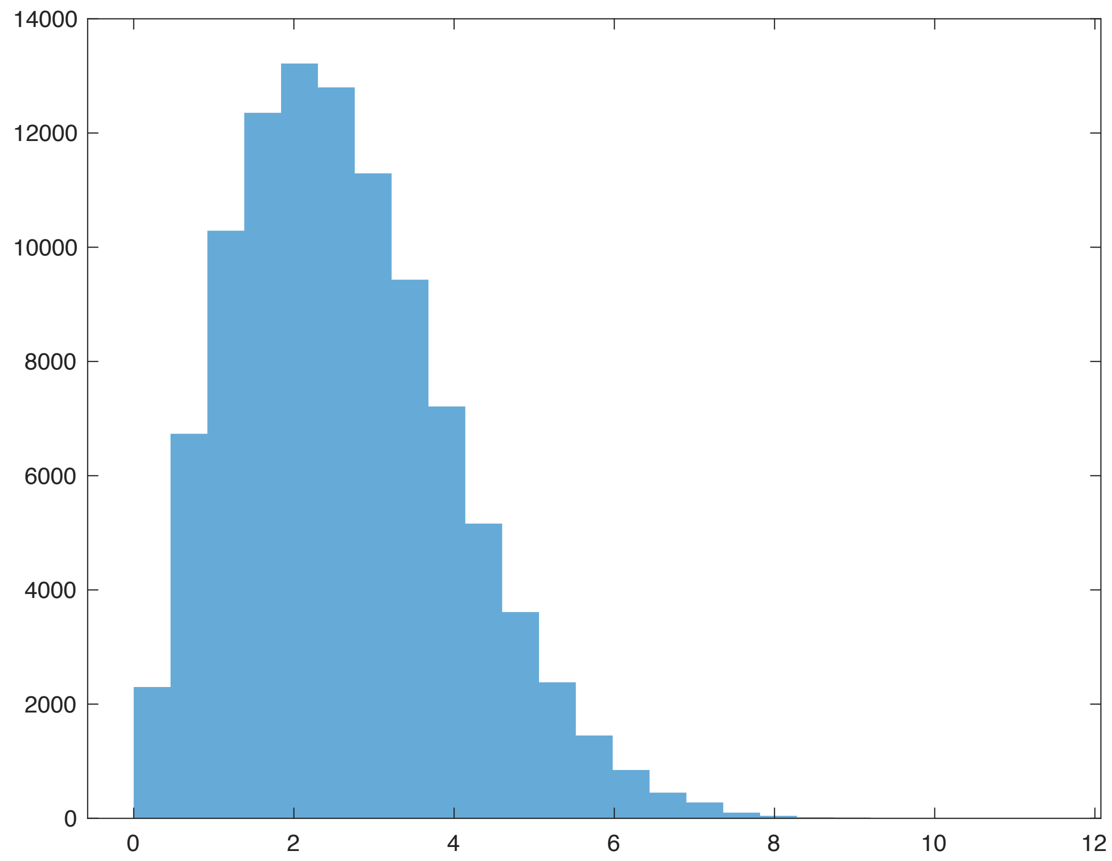
What is important?

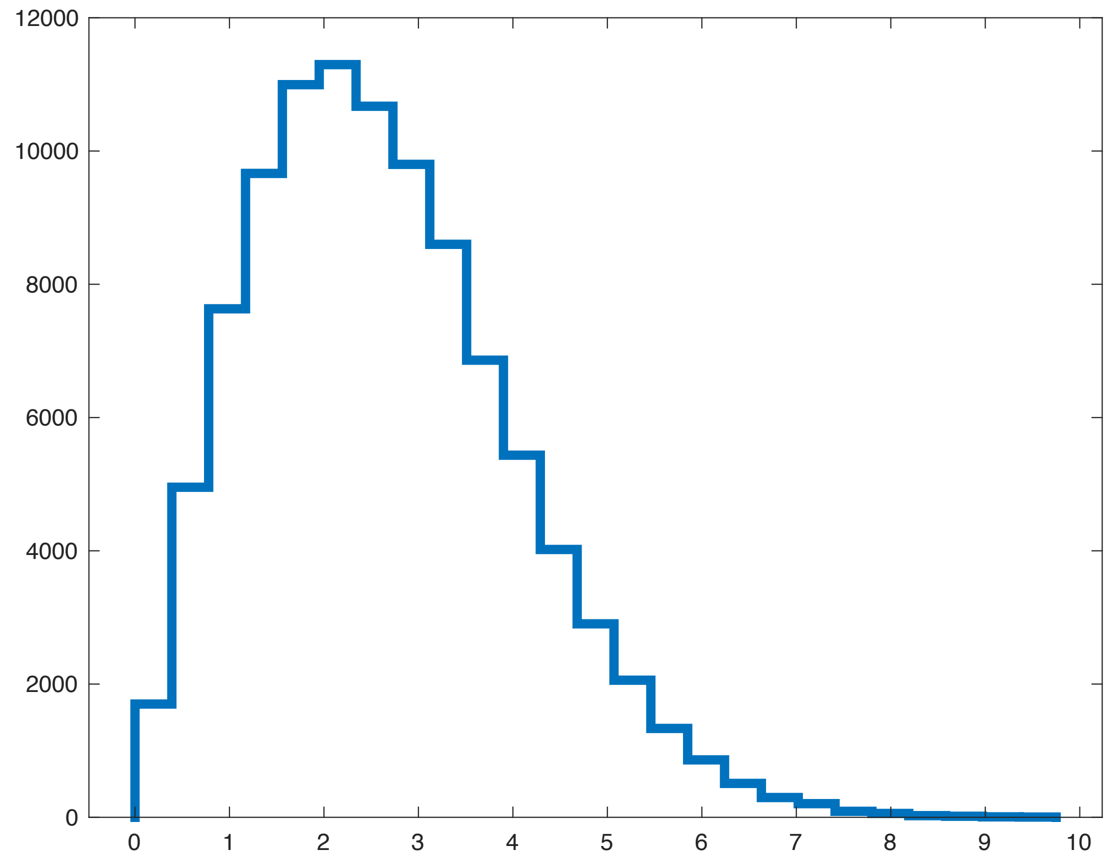




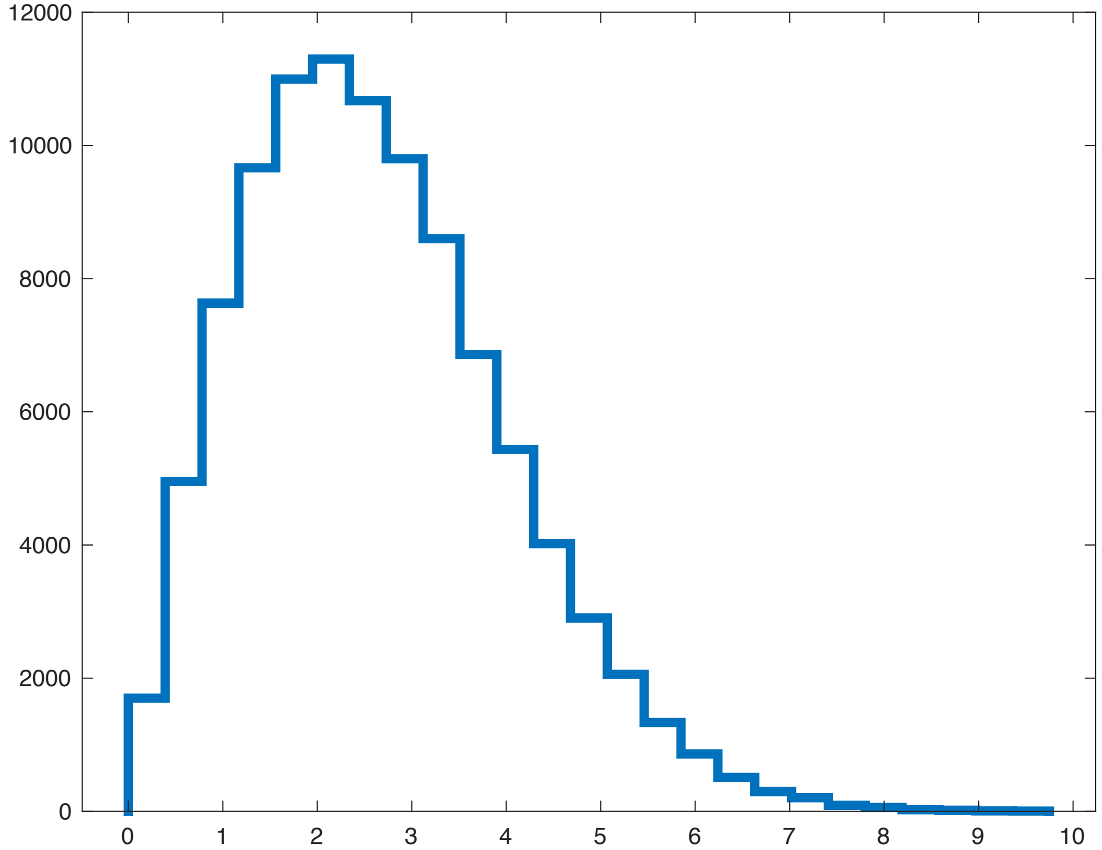
Are ink/lines communicating information?

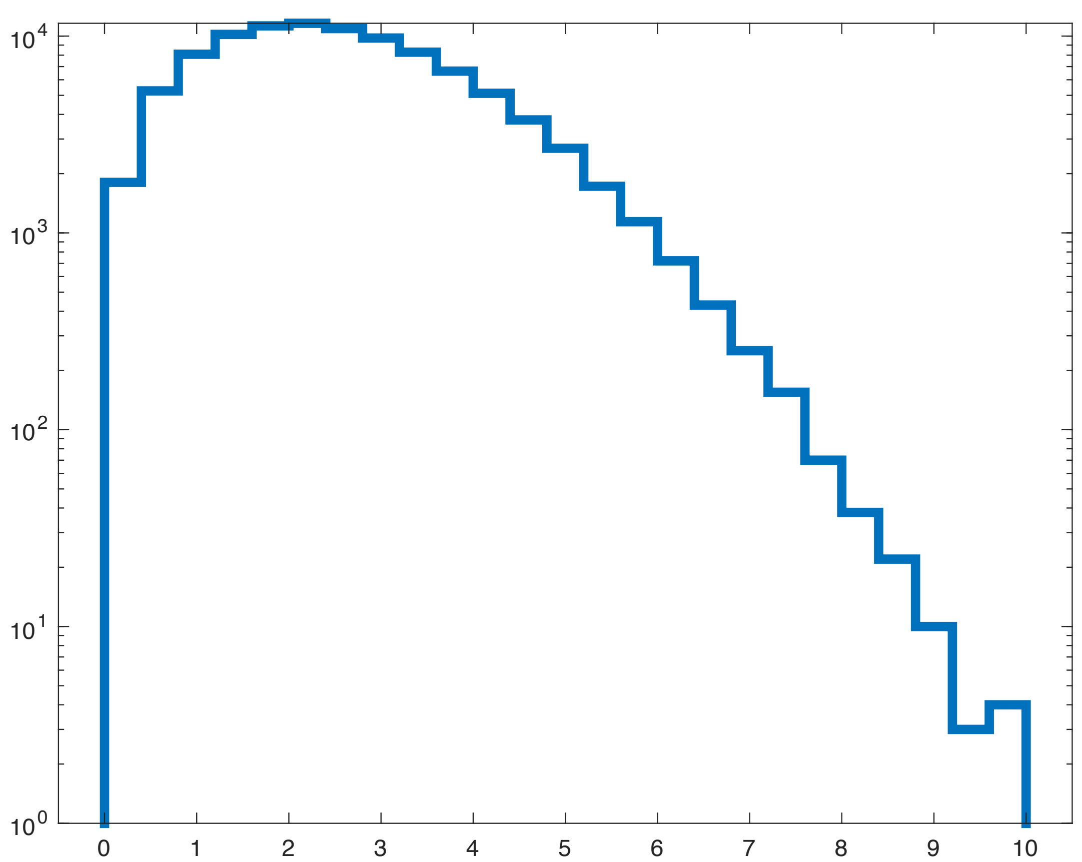
Emphasis


14000
12000
10000
8000
6000
4000
2000
0
0
2
4
6
8
10
12

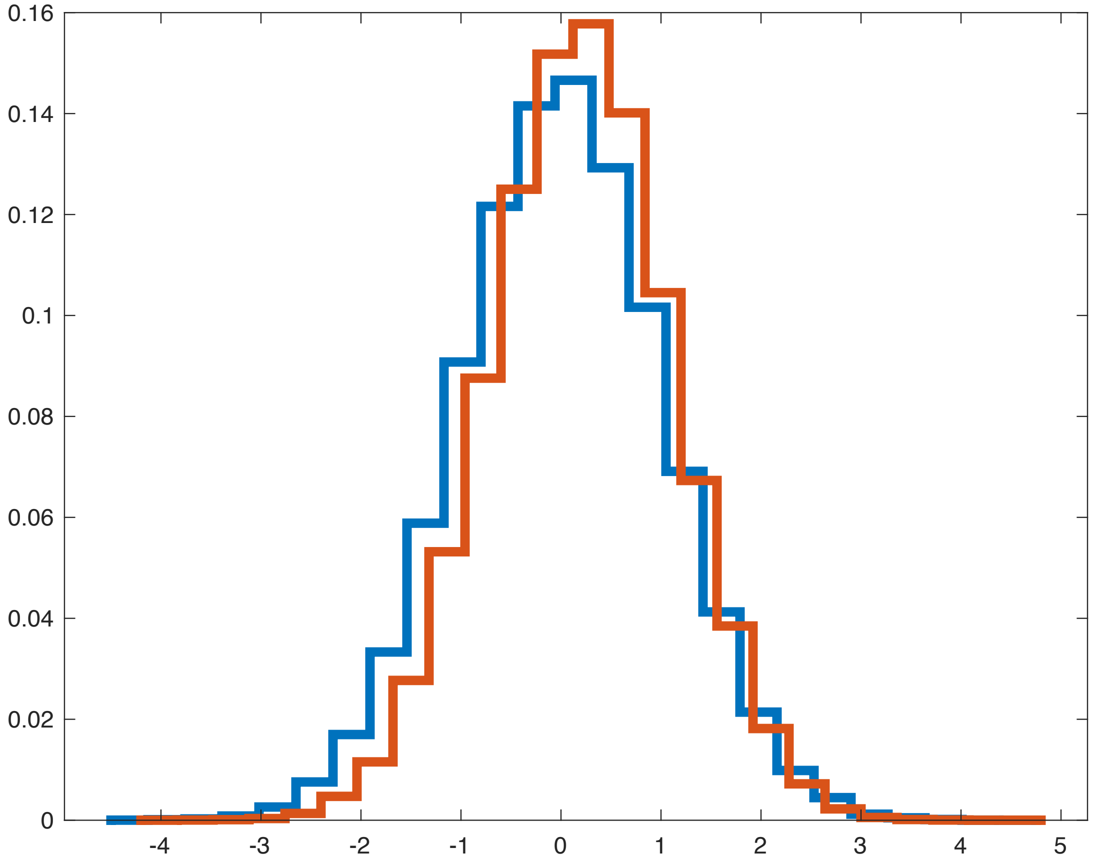

12000
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6
7
8
9
10

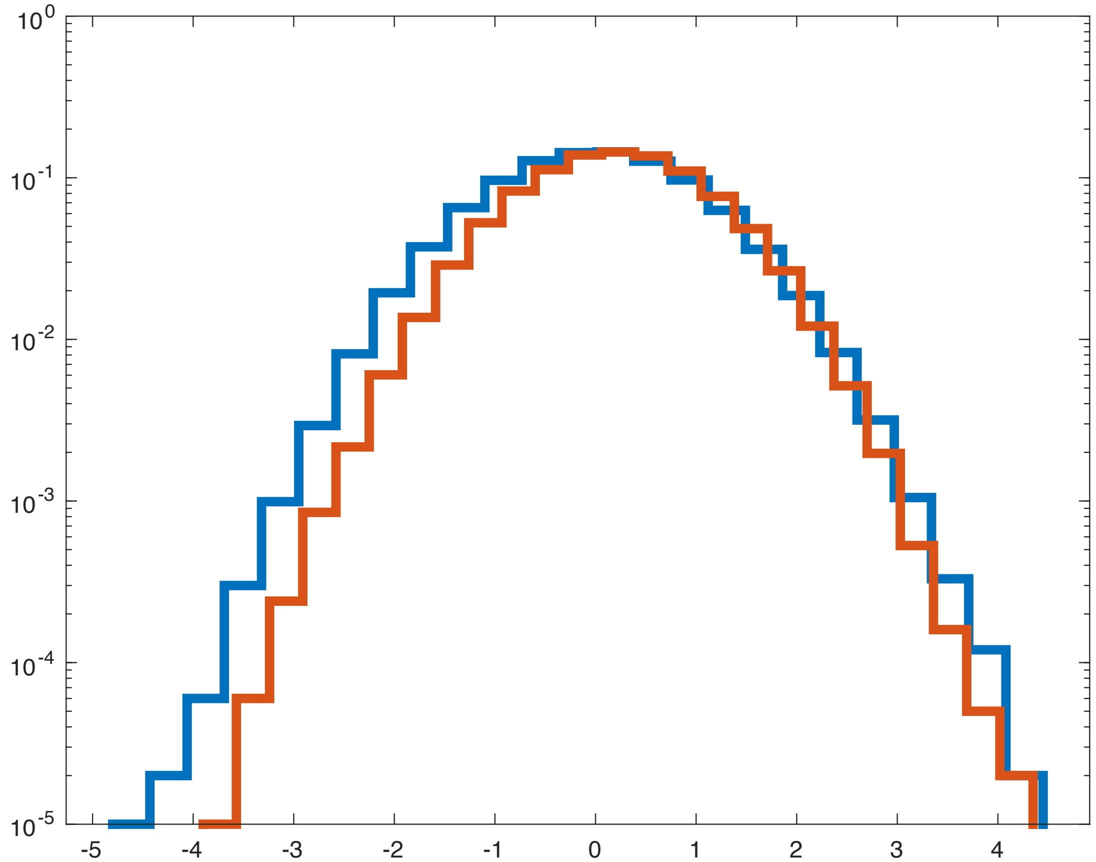
Emphasis with axes



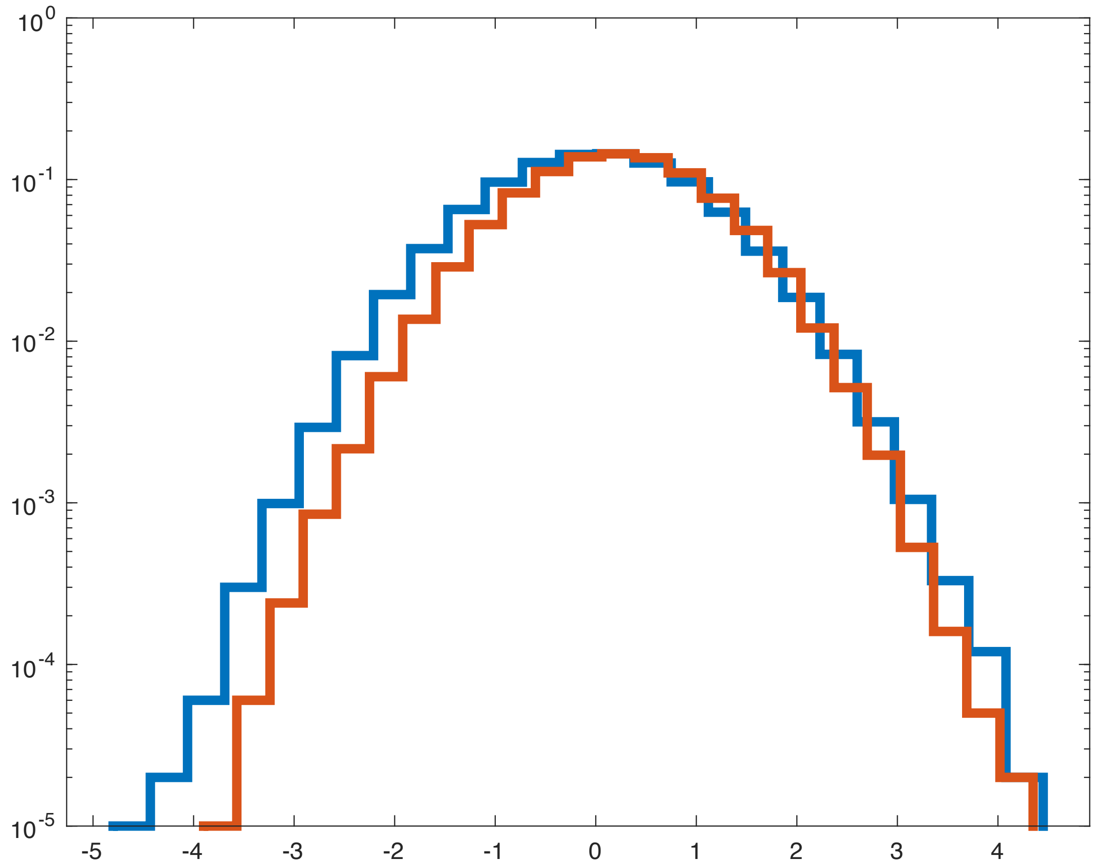


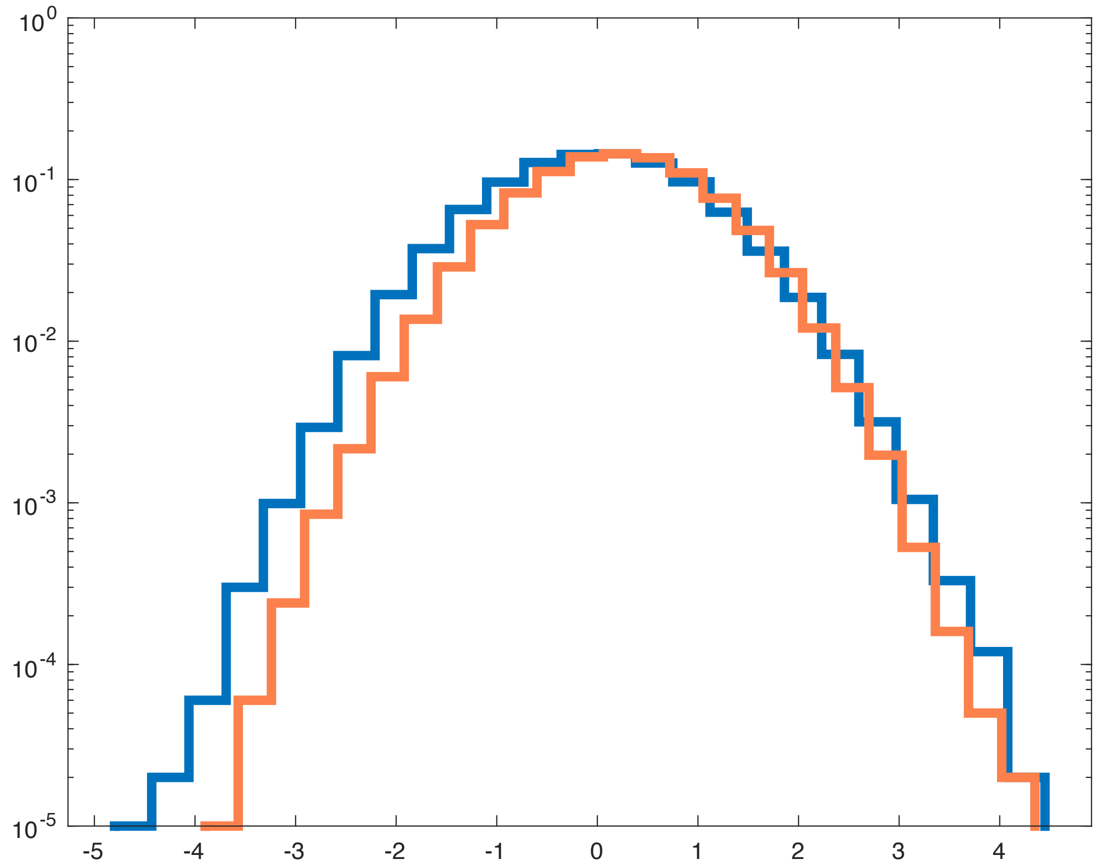
What are you looking for?





What are you looking for?

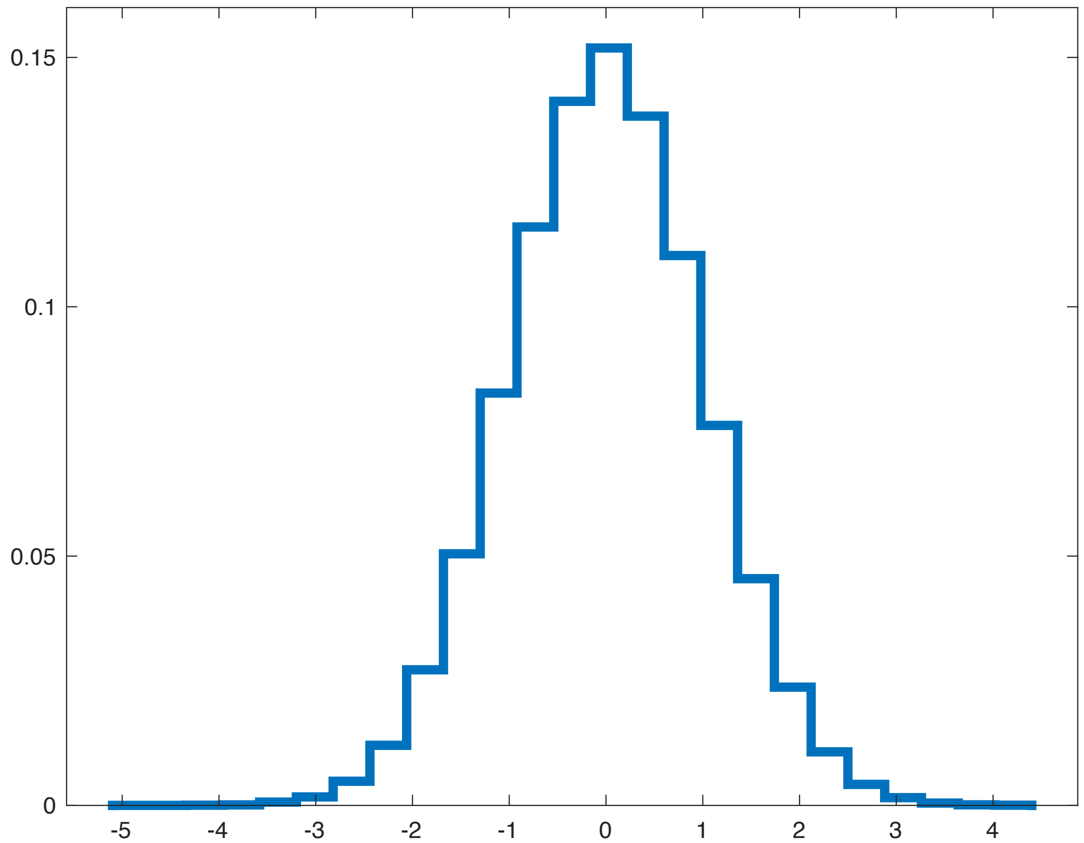


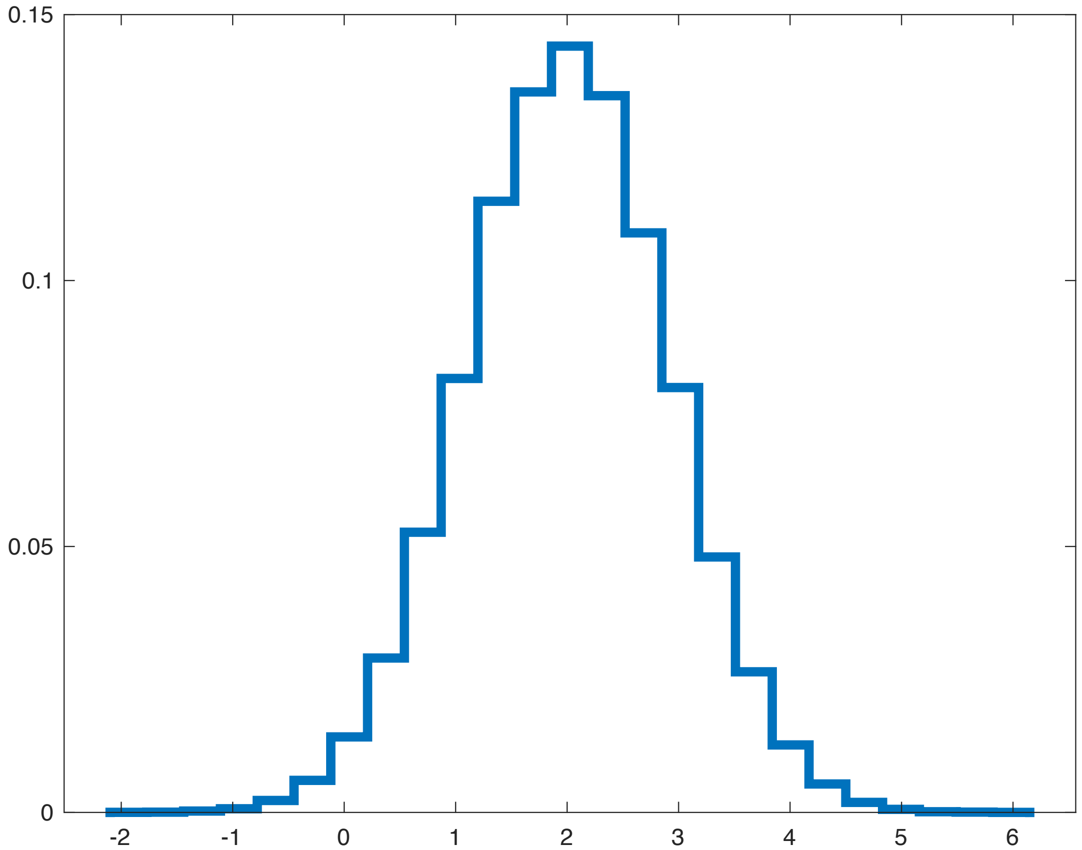


Where is the information?

Where is the information?

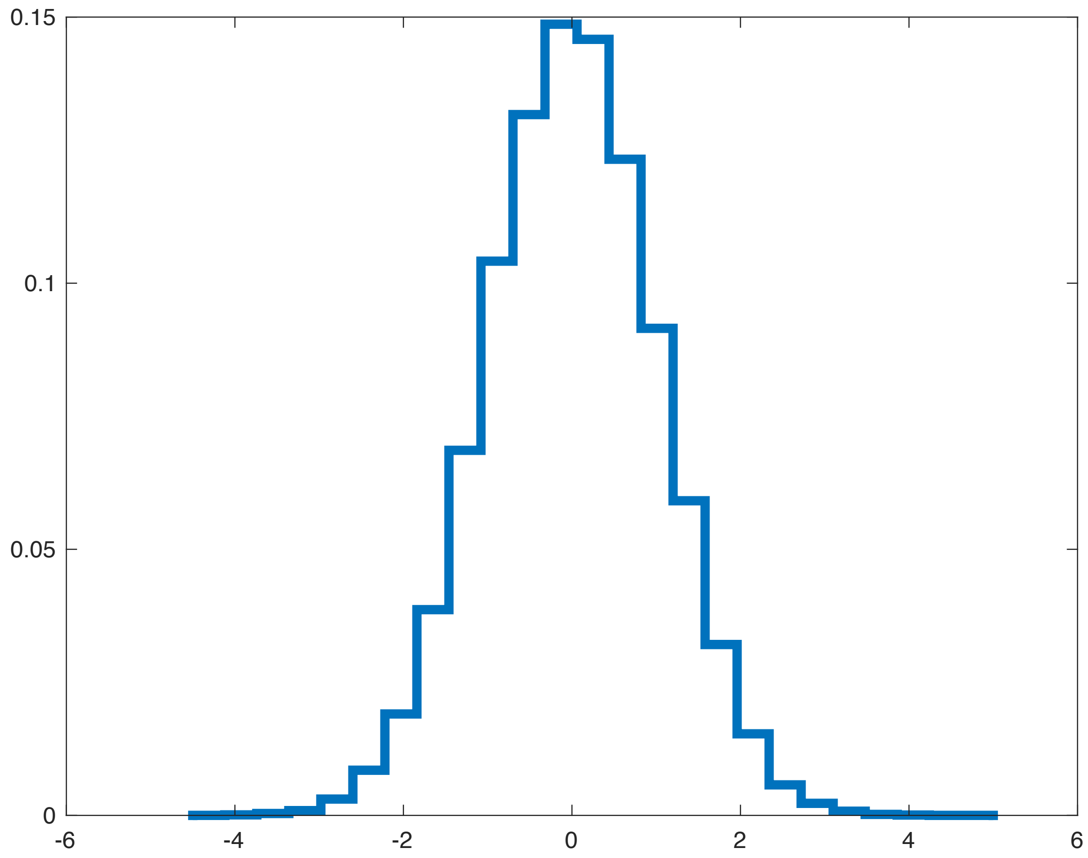
Compare two distributions

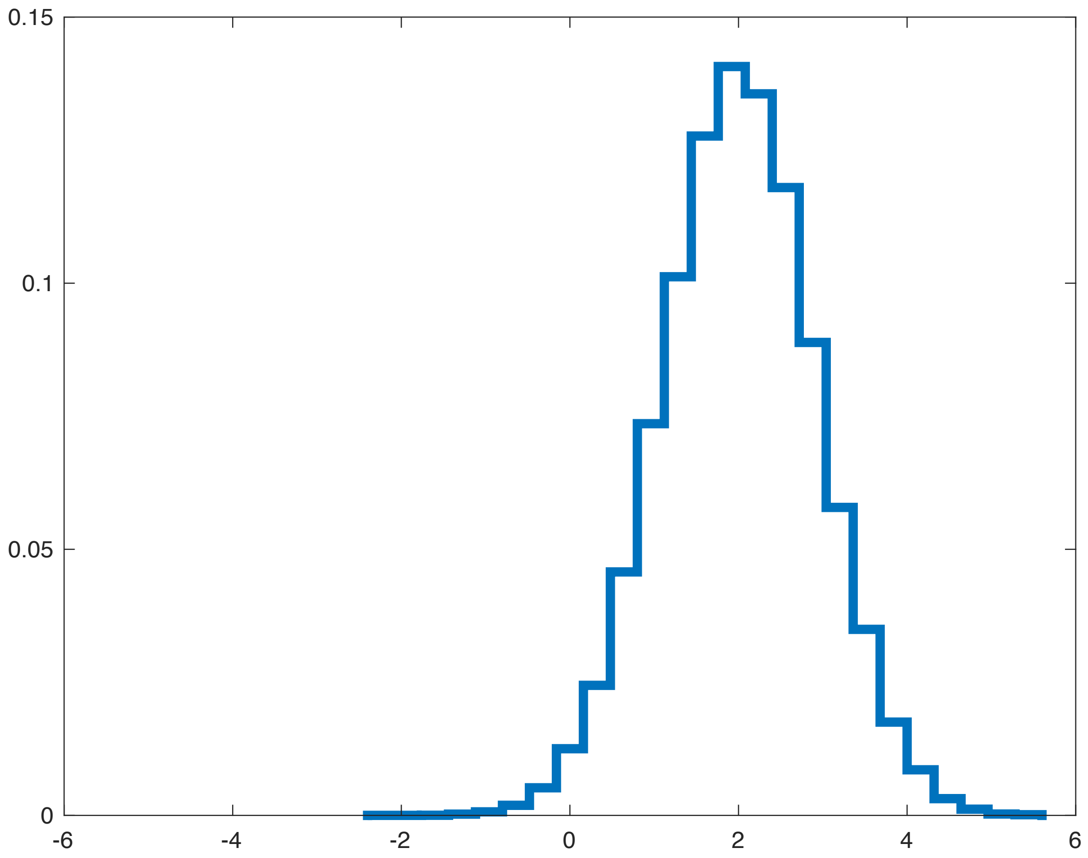




Where is the information?

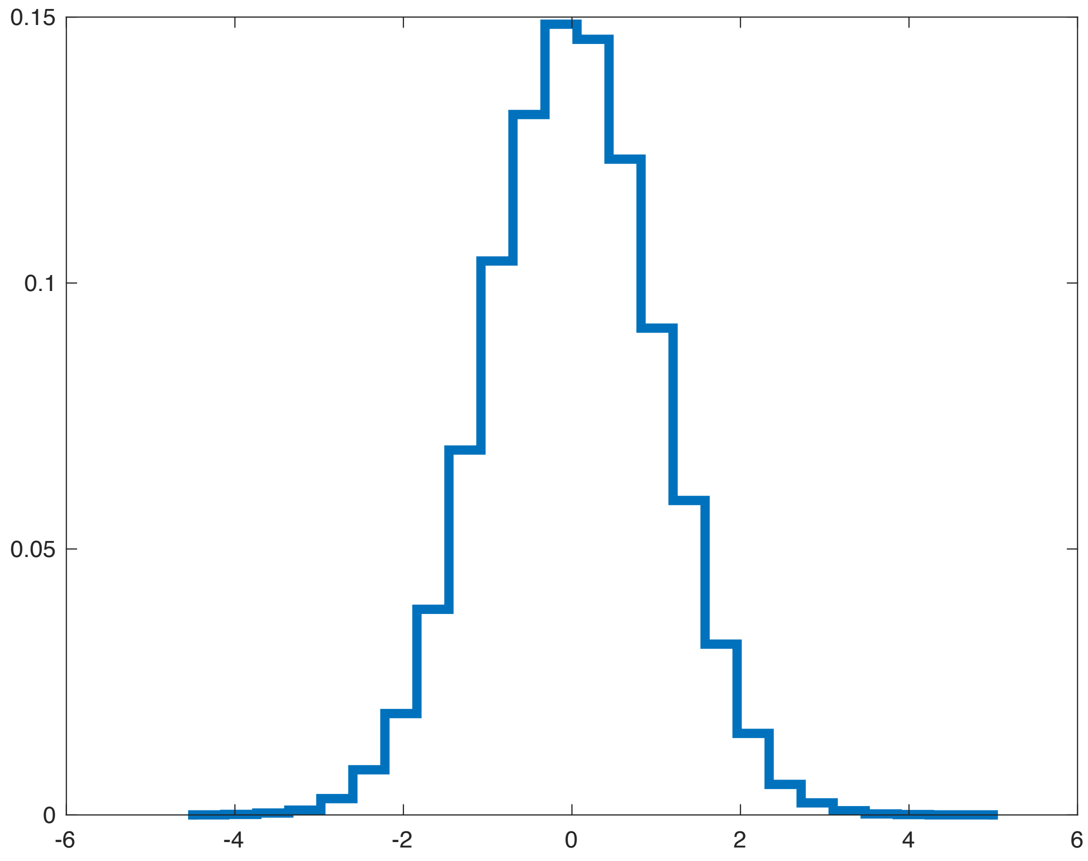
Compare two distributions

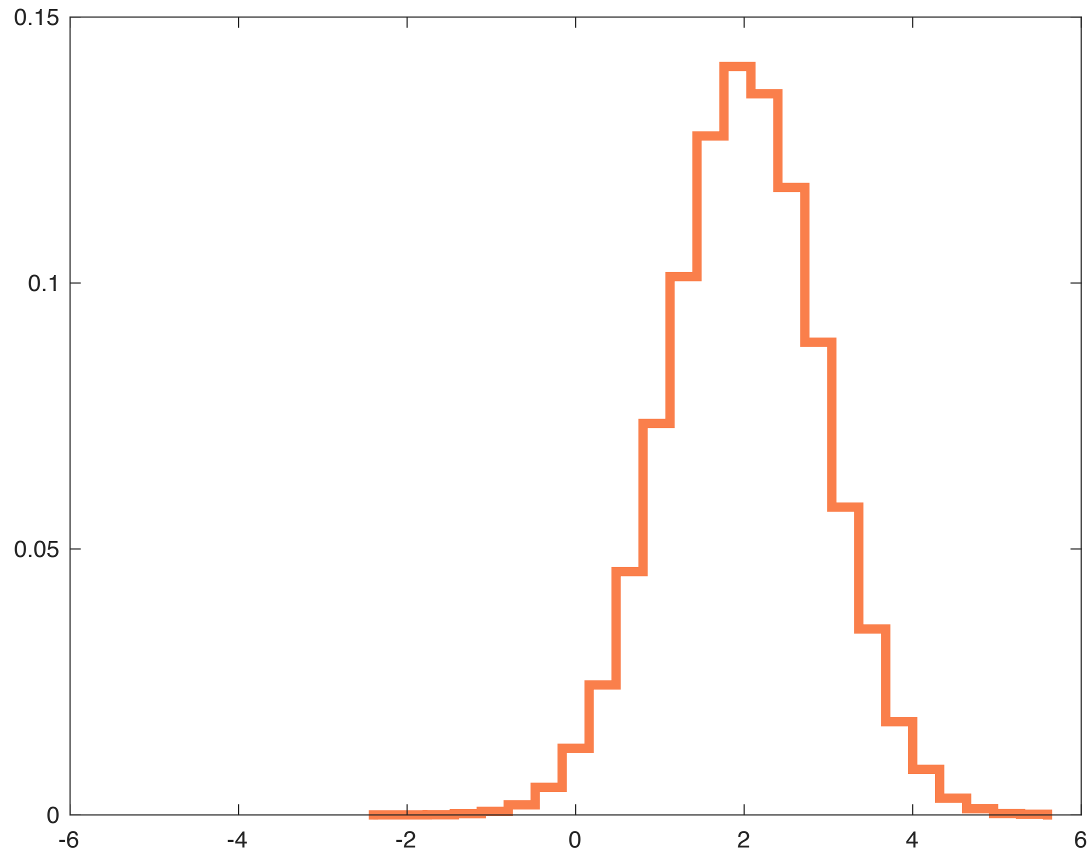




Where is the information?

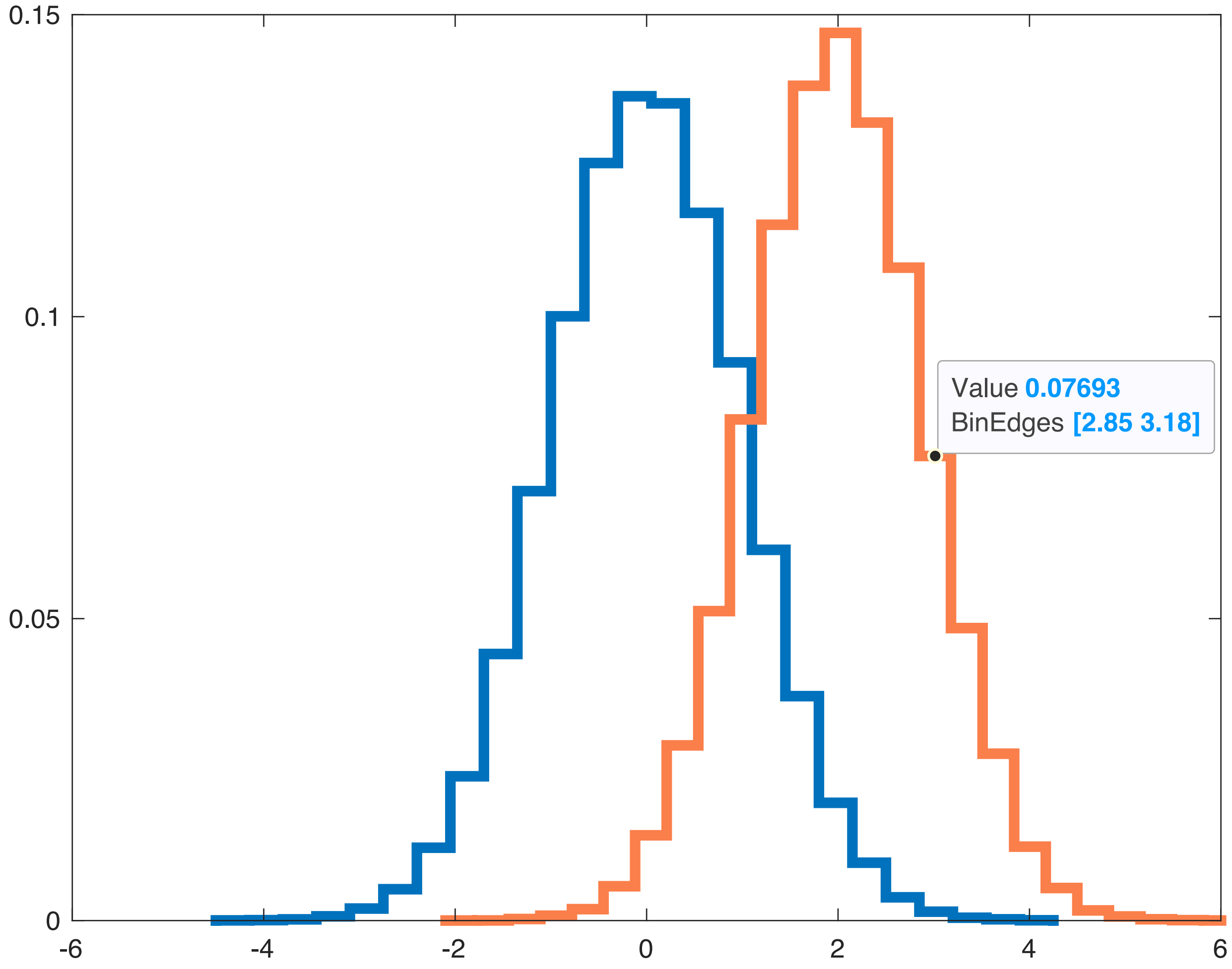
Compare two distributions





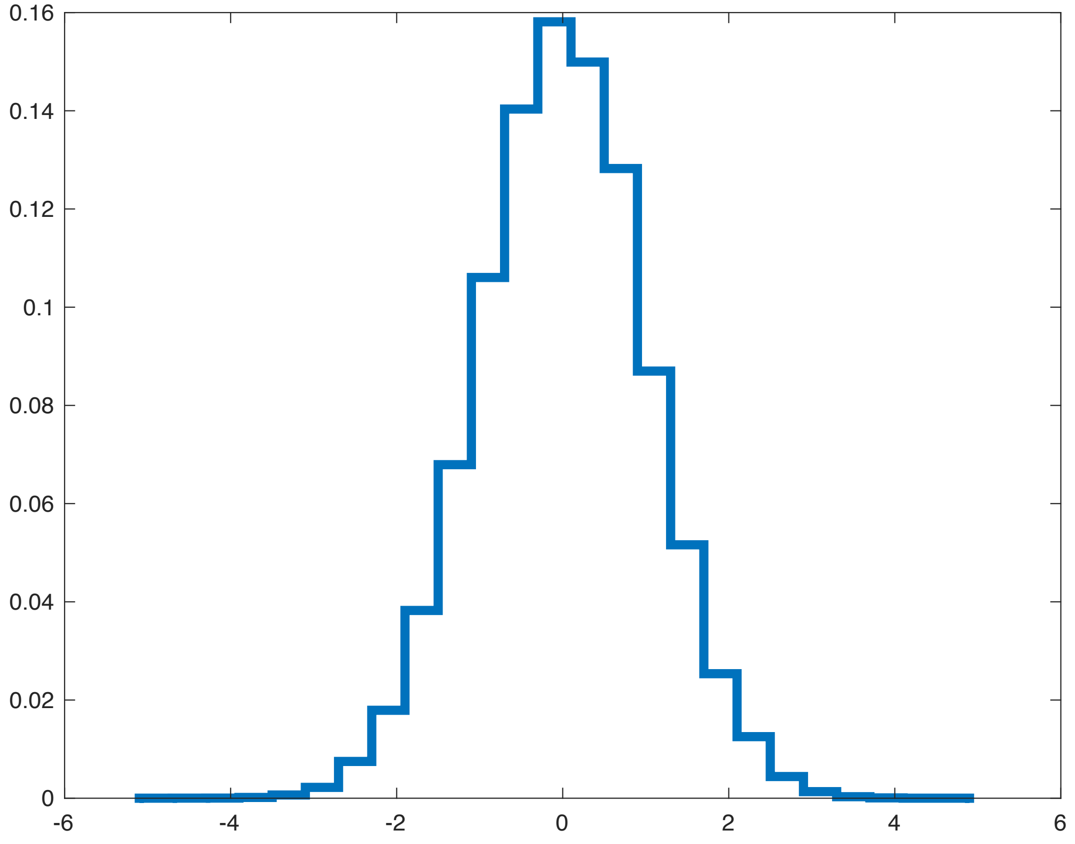
Where is the information?

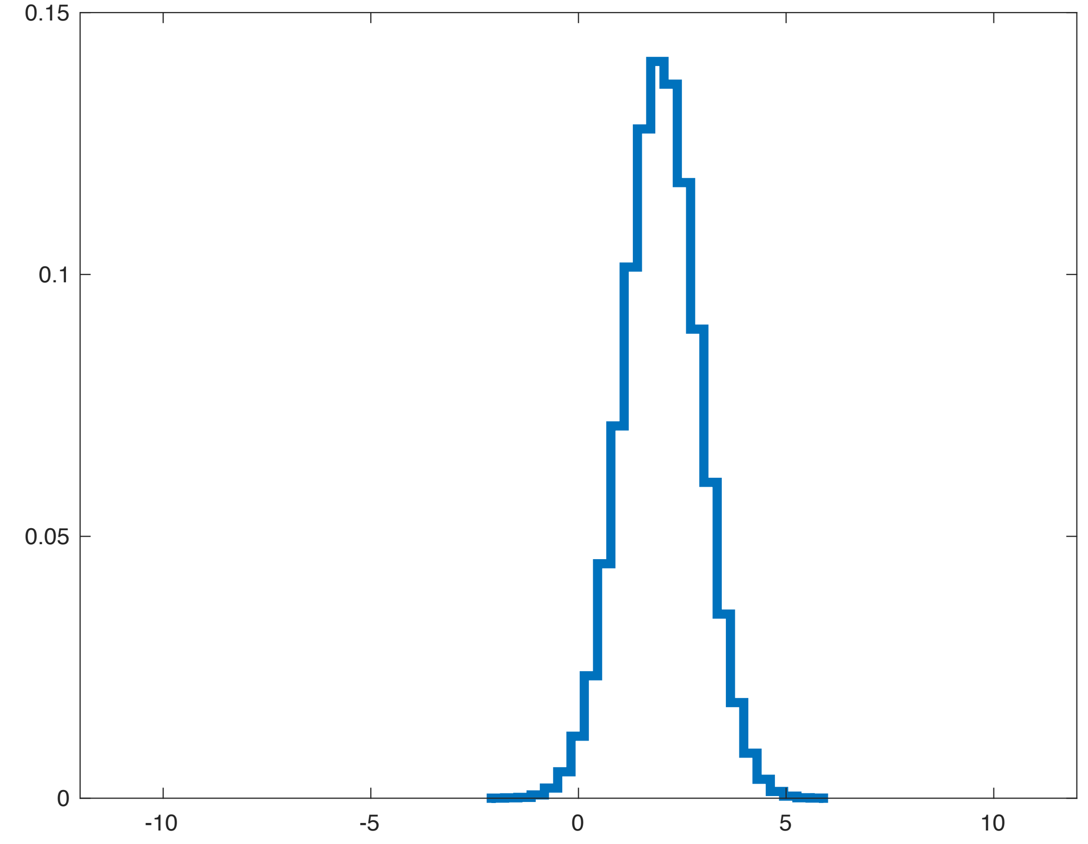
Compare two distributions



Avoid visual red herrings

Which distribution is wider?





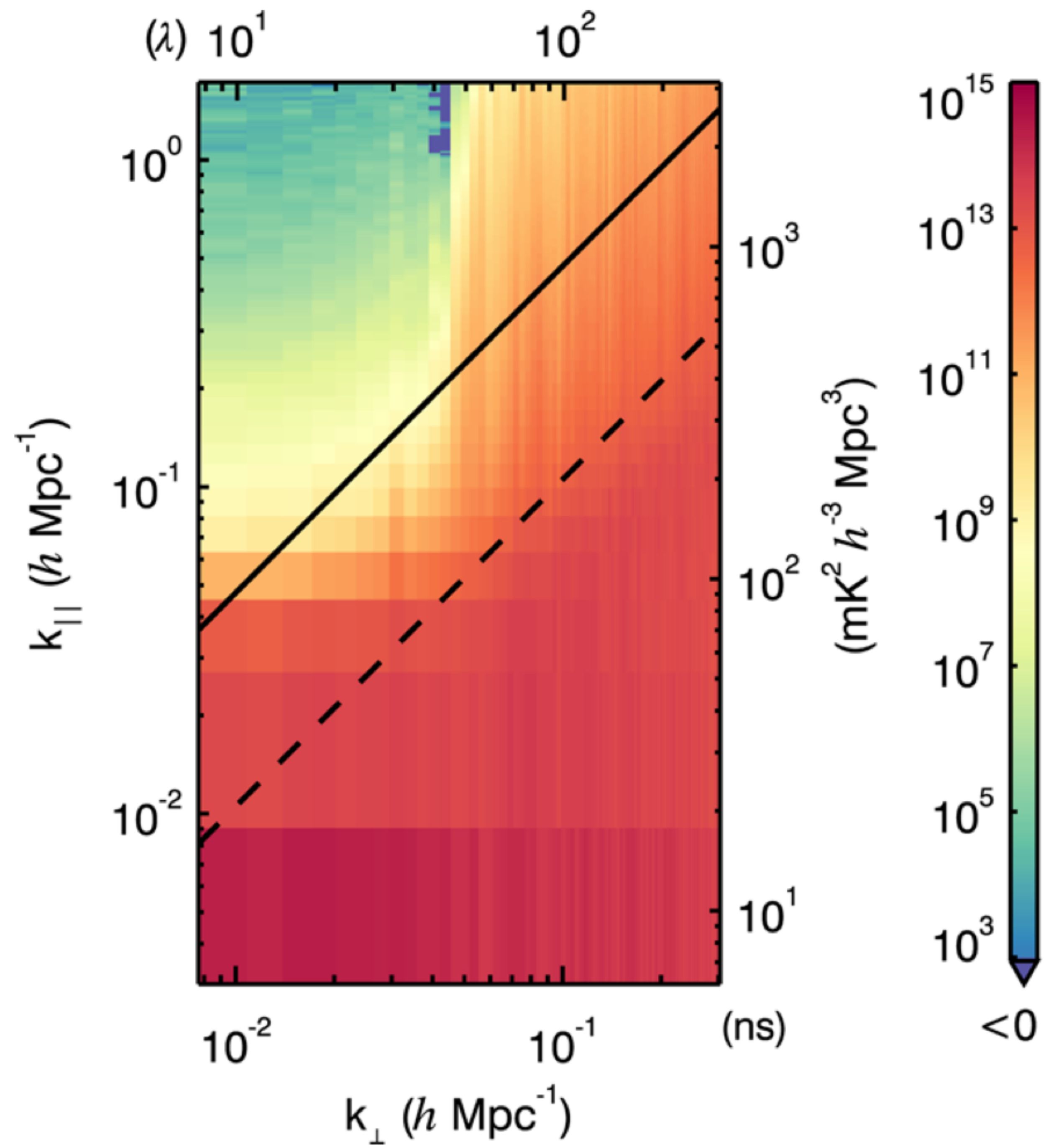
Where is the information?

- Try to never put key information in axis labels or colorbar label
- Make ‘natural’ visual interpretation true
  - Position, size, intensity, etc.
  - Avoid visual red herrings
- Reduce cognitive overhead

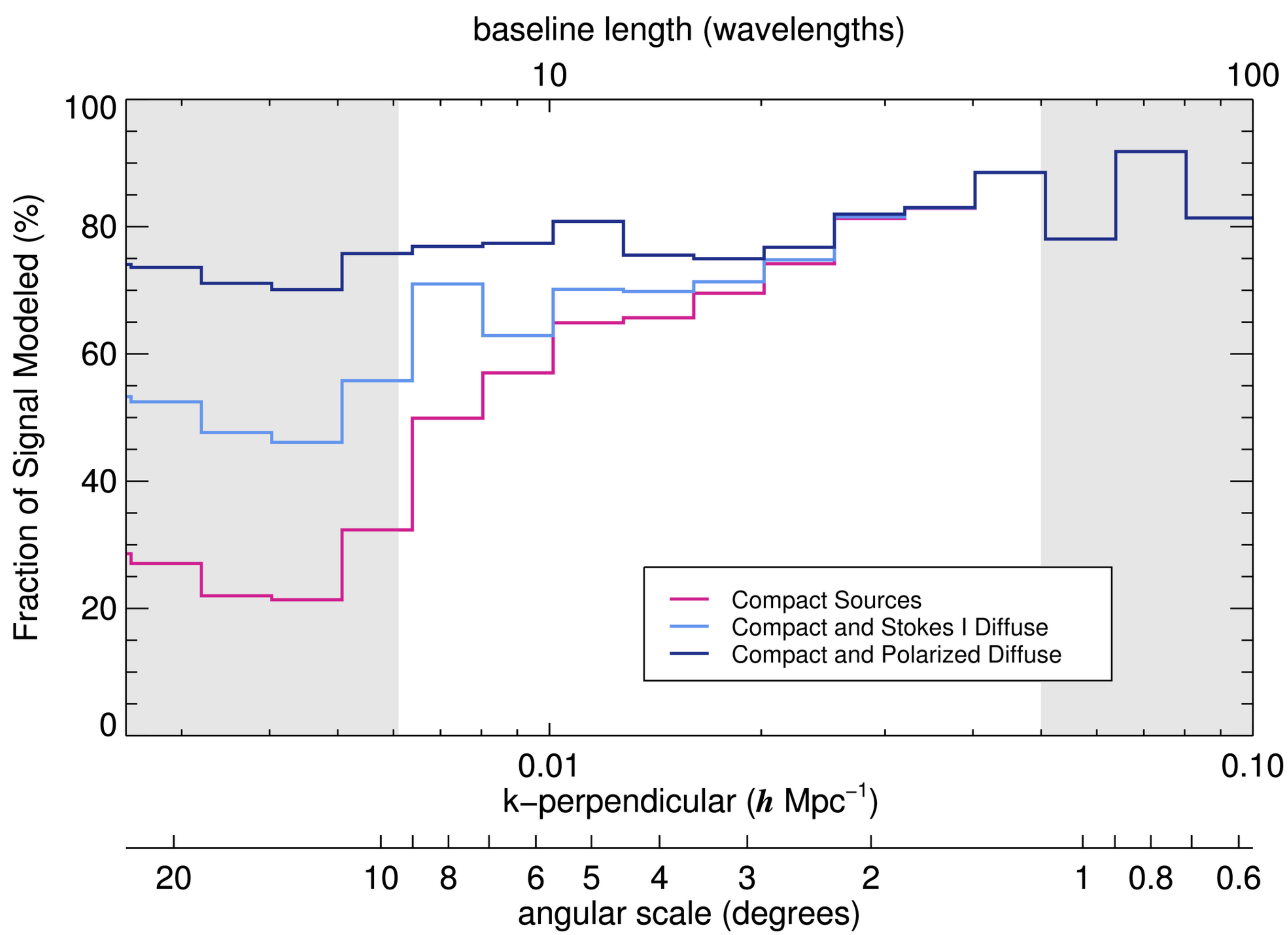
Listen to your subconscious

Drillable plots

Show optional information for careful study



Show optional information for careful study



(a) East-West polarization

Developing a consistent language

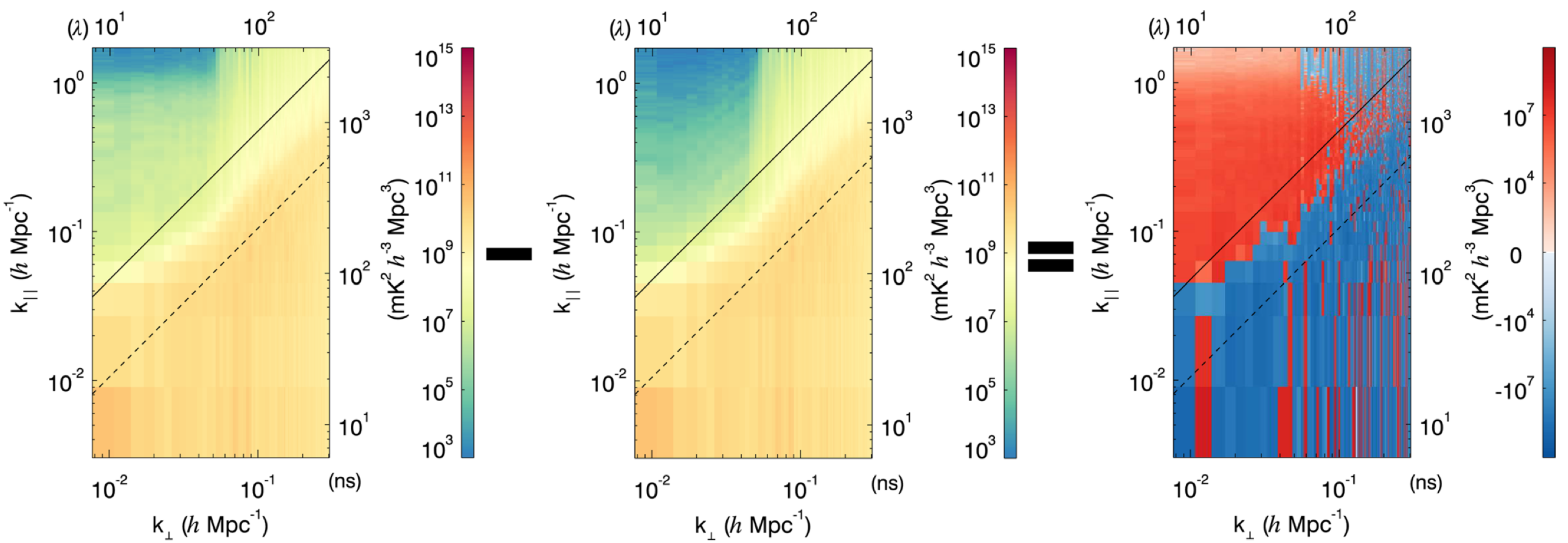
Consistent language

- Same color lines, marker styles, etc. within one paper
- Be very conscious of developing a reusable plot style
    - Helps you look through lots of data
    - Allows you to focus in on features because foundation is stable
    - Often become collaboration or field ‘standards’
    - Use git to codify

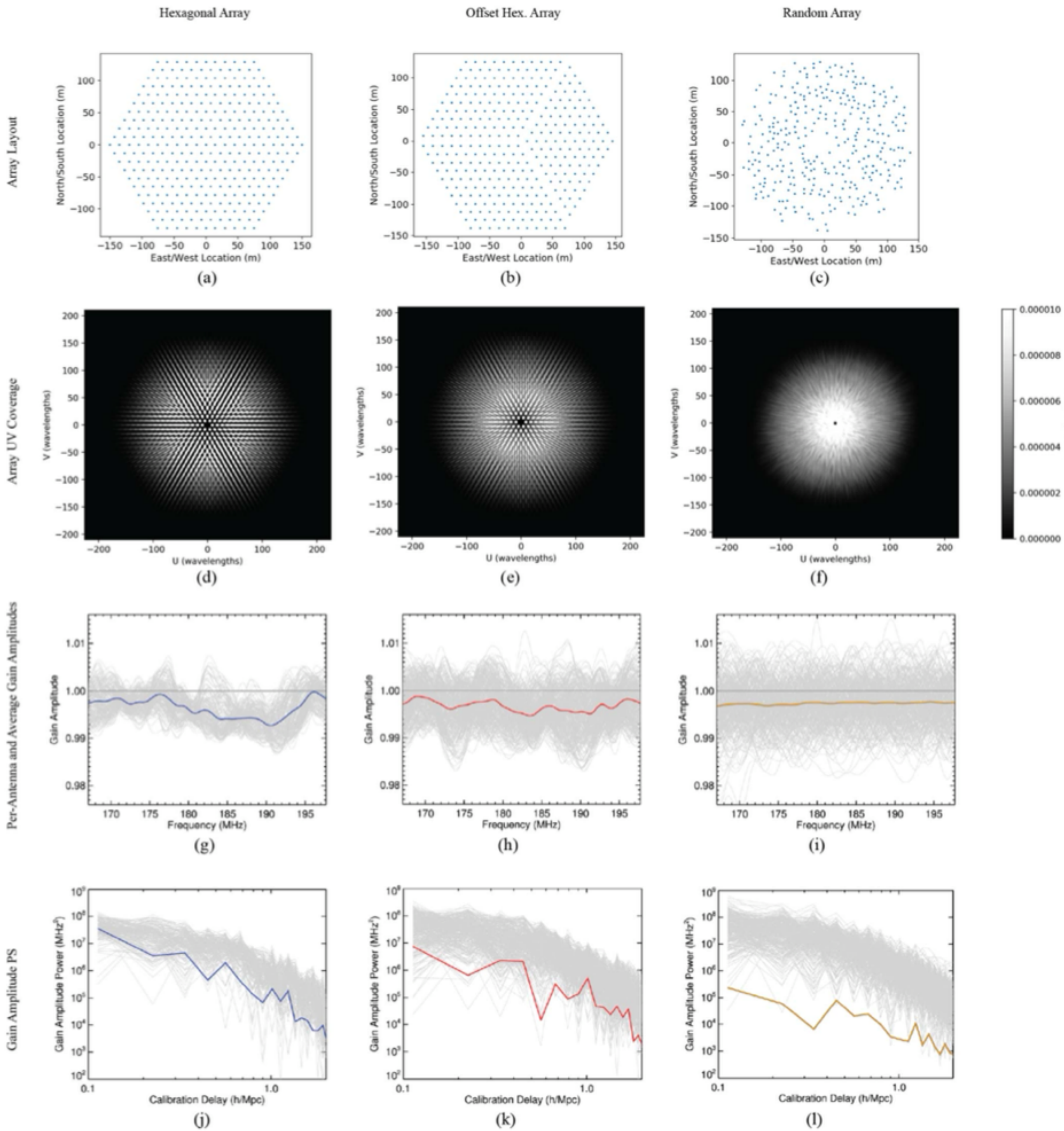
Consistent language

Plot stories:

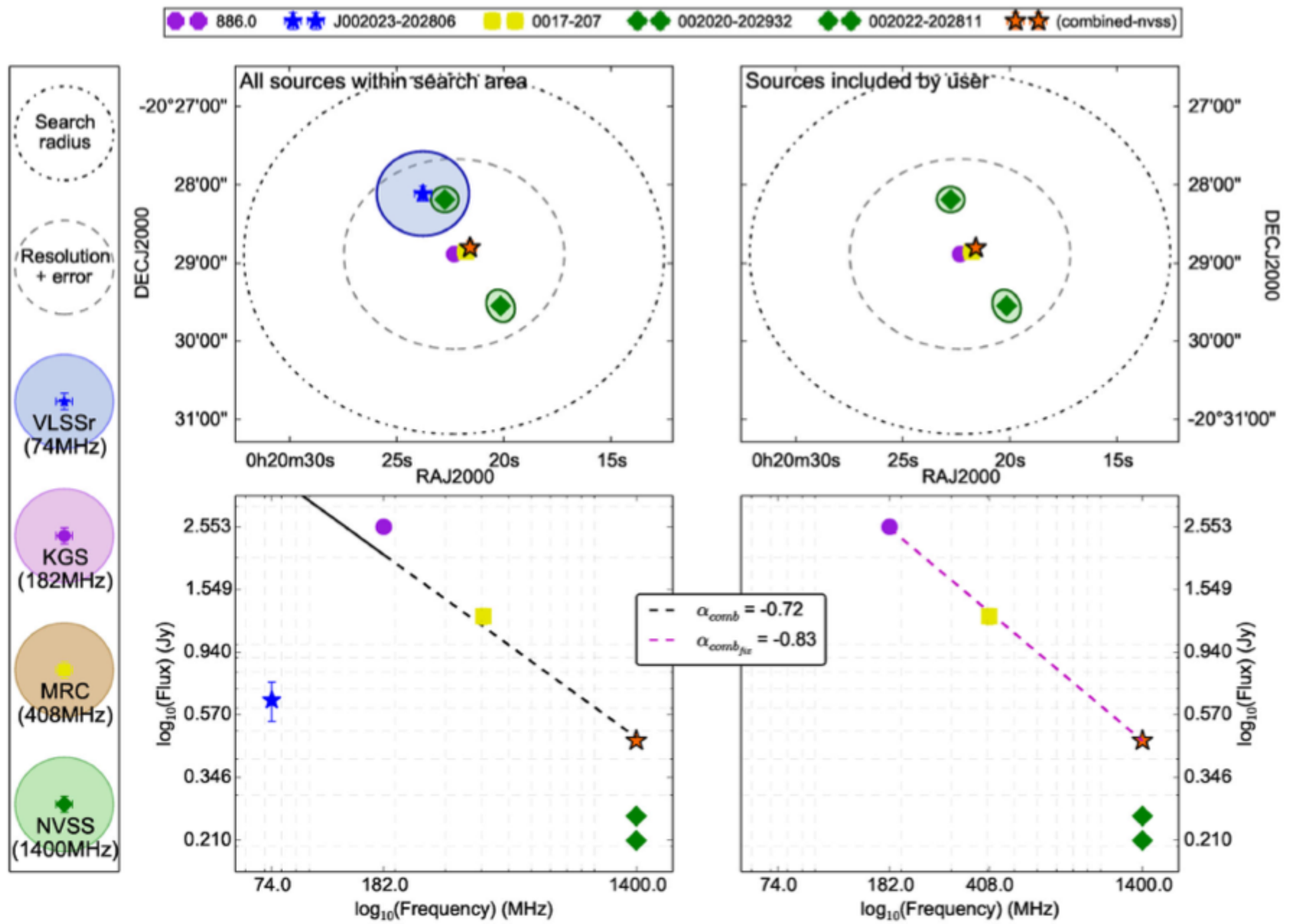
- Within a paper or notebook
- Within multi-panel plots



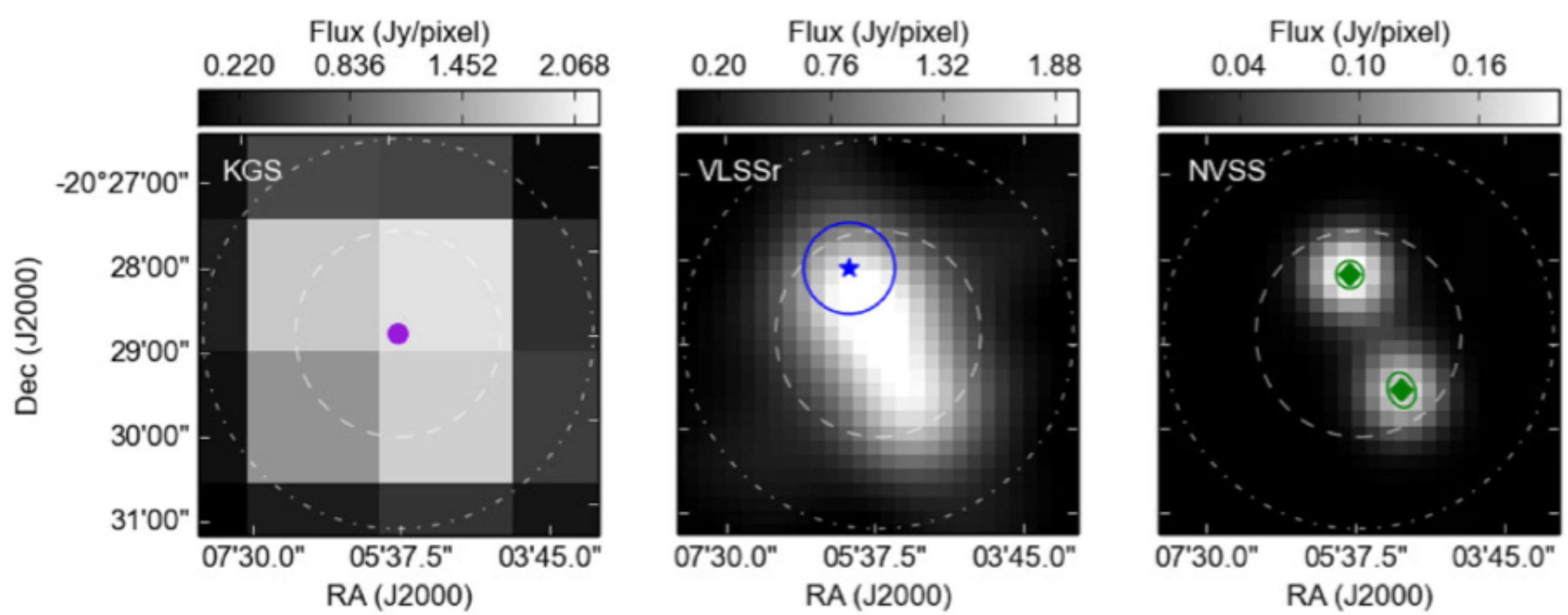
(λ) 10^1 10^2
k|| (h Mpc^-1)
k⊥ (h Mpc^-1)
(ns)
(mK^2 h^-3 Mpc^3)
10^15
10^13
10^11
10^9
10^7
10^5
10^3
10^7
10^4
0
-10^4
-10^7



Hexagonal Array
Offset Hex. Array
Random Array
Array Layout
North/South Location (m)
East/West Location (m)
(a)
(b)
(c)
Array UV Coverage
V (wavelengths)
U (wavelengths)
Weight (Jy/beam)
(d)
(e)
(f)
Per-Antenna and Average Gain Amplitudes
Gain Amplitude
Frequency (MHz)
(g)
(h)
(i)
Gain Amplitude PS
Gain Amplitude Power (MHz²)
Calibration Delay (h/Mpc)
(j)
(k)
(l)

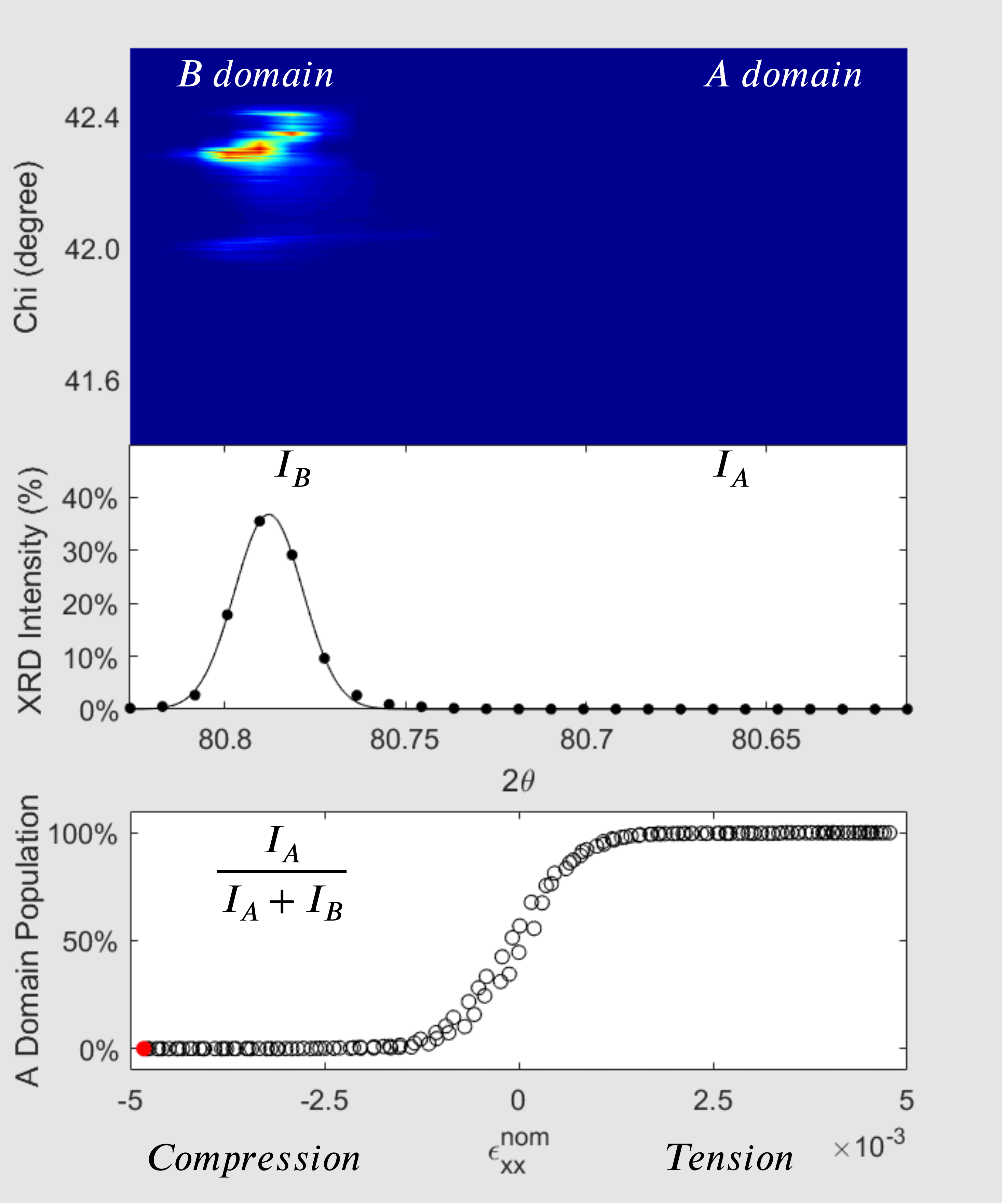


(a) Example of visualized position (top) and SED (bottom) information for a complicated match before (left) and after (right) manual modification. Ellipses indicate the reported major/minor axis and position angle.



(b) Example postage stamp images inspected for complicated matches. The white dash/dotted circles correspond to the search radius and resolution+error as indicated in (a)

- (Top plot) the intensity position on the detector gives the length values of a and b which change with strain
- (Middle plot) the intensity is summed vertically and fit to 2 Gaussians
- (Bottom plot) The relative intensity $\frac{I_A}{I_A + I_B}$ gives the relative A domain population which change vs strain.



<u>Interesting result!</u>
Lattice constants freeze in place during detwinning!

Implies that the domain pinning is much softer than the crystal lattice

Can smoothly detwin the sample from B to A and back

HERA notebook examples

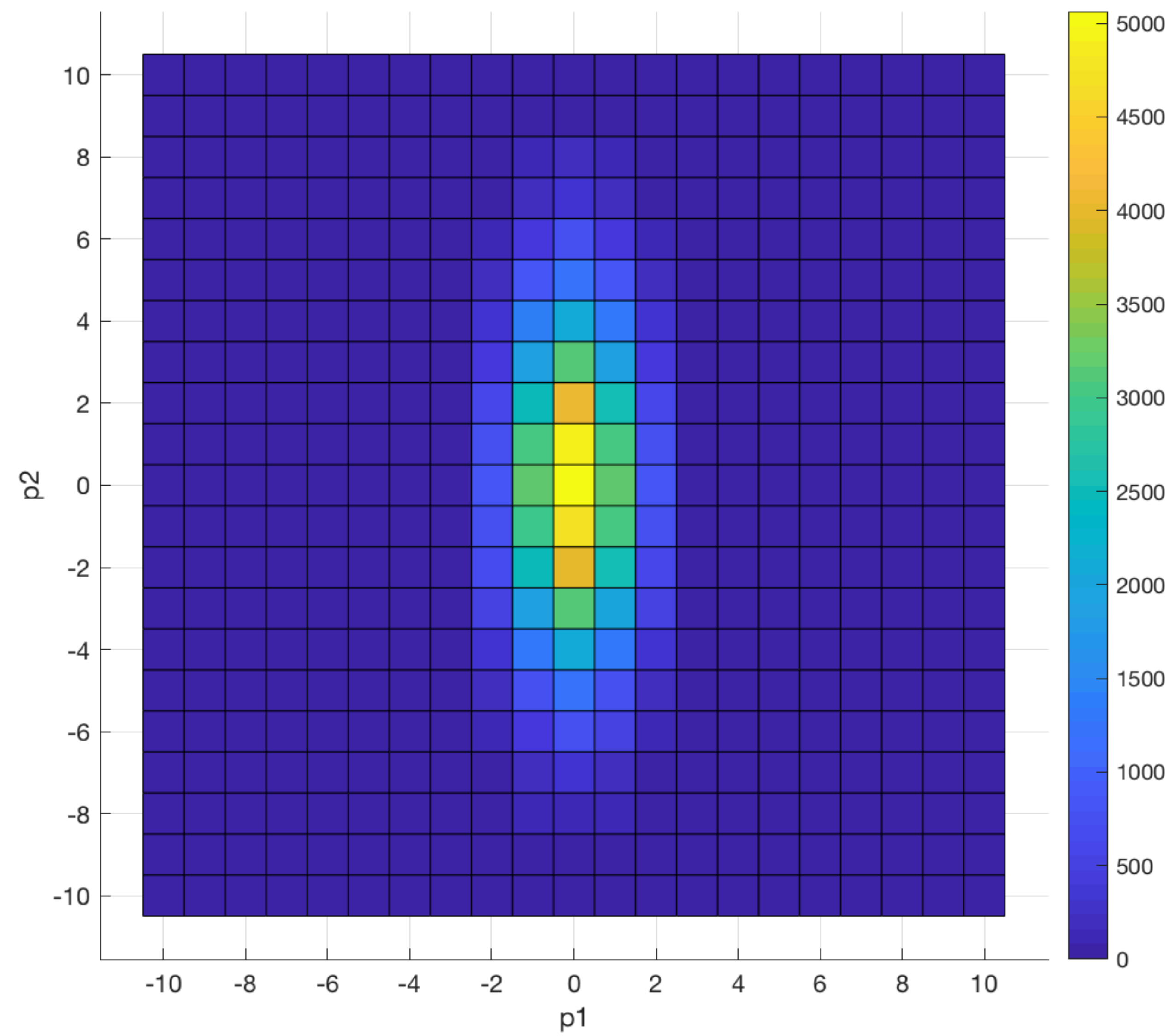
Data density

Data density

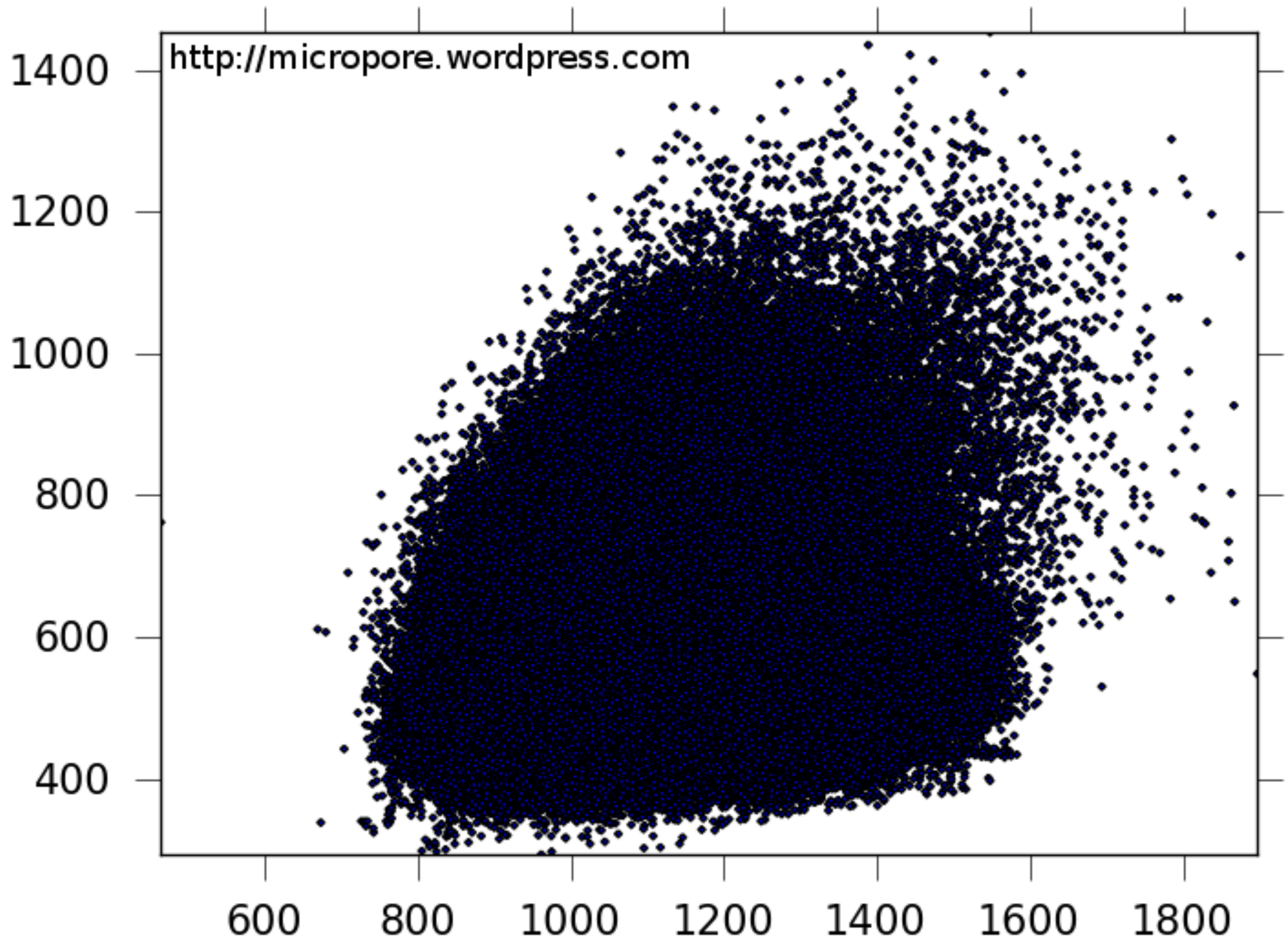
- All plots are averages or cuts through data
  - All plots choose what they emphasize
- When looking for features want to get data in your brain
  - Different styles have different pros & cons

Scatter, 2D histograms & images

- Many more points
- Lots of points overlap
- Colorbars hard to read

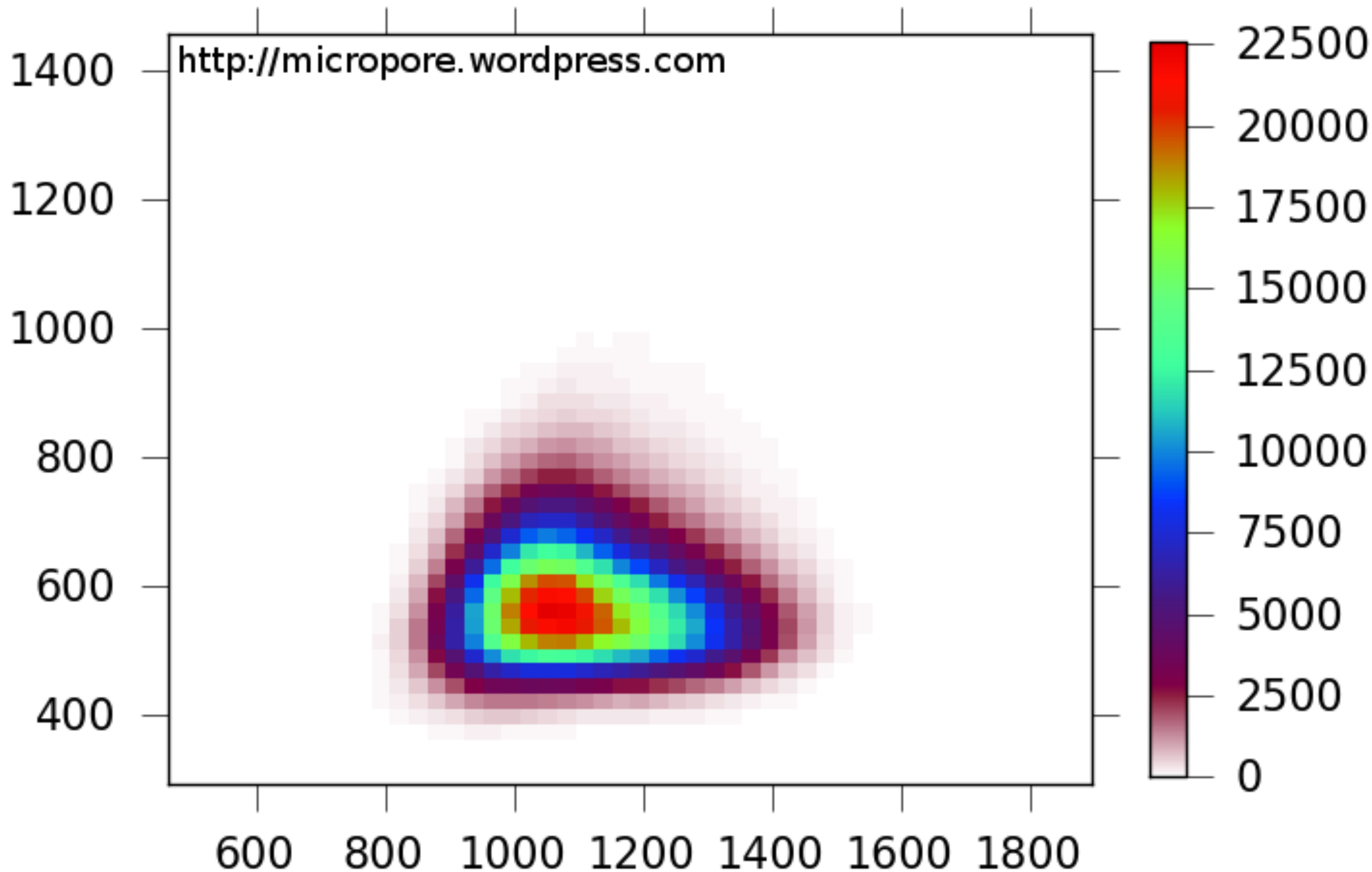


Scatter

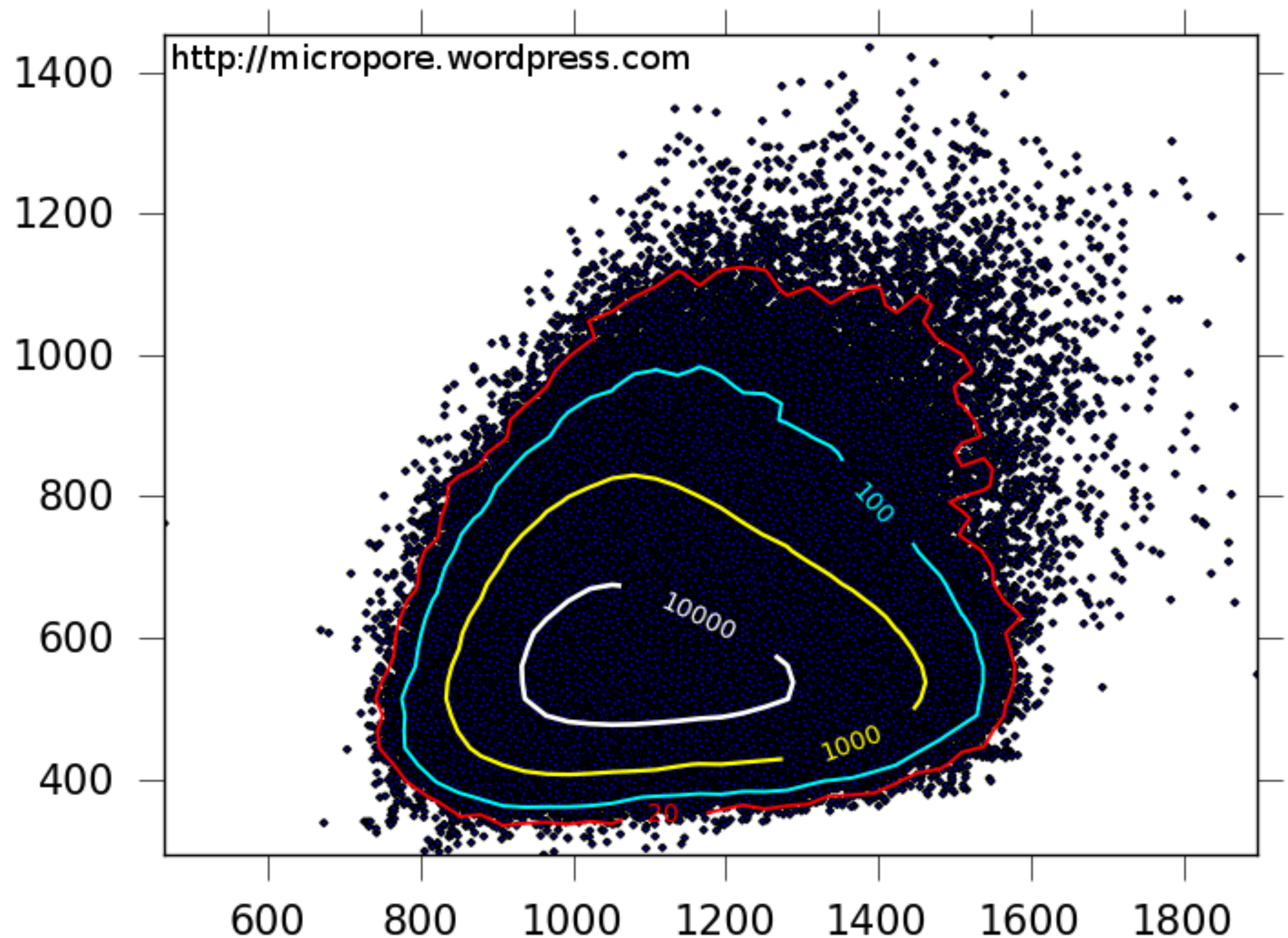


http://micropore.wordpress.com
1400
1200
1000
800
600
400
600
800
1000
1200
1400
1600
1800

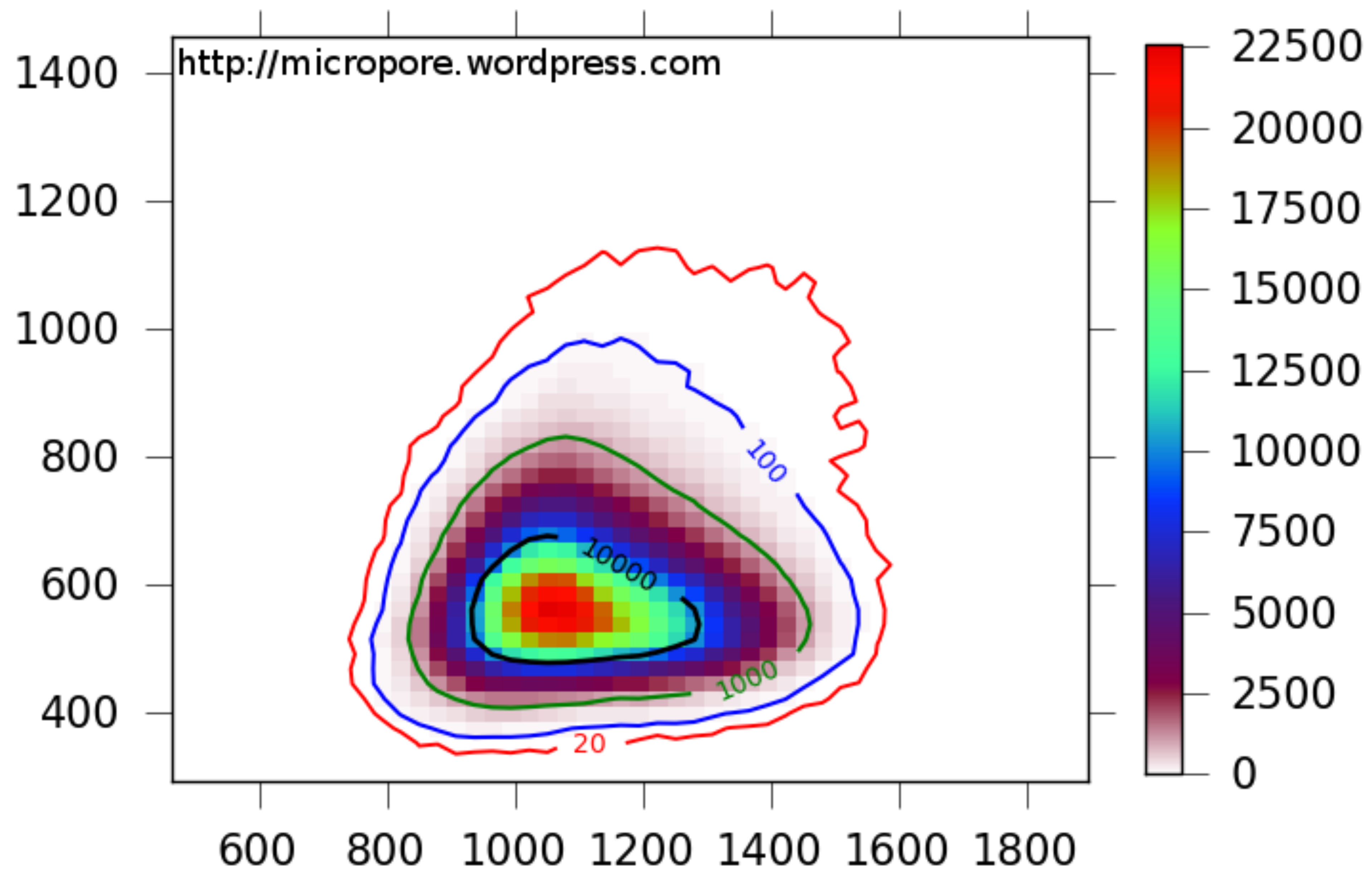
2D histogram



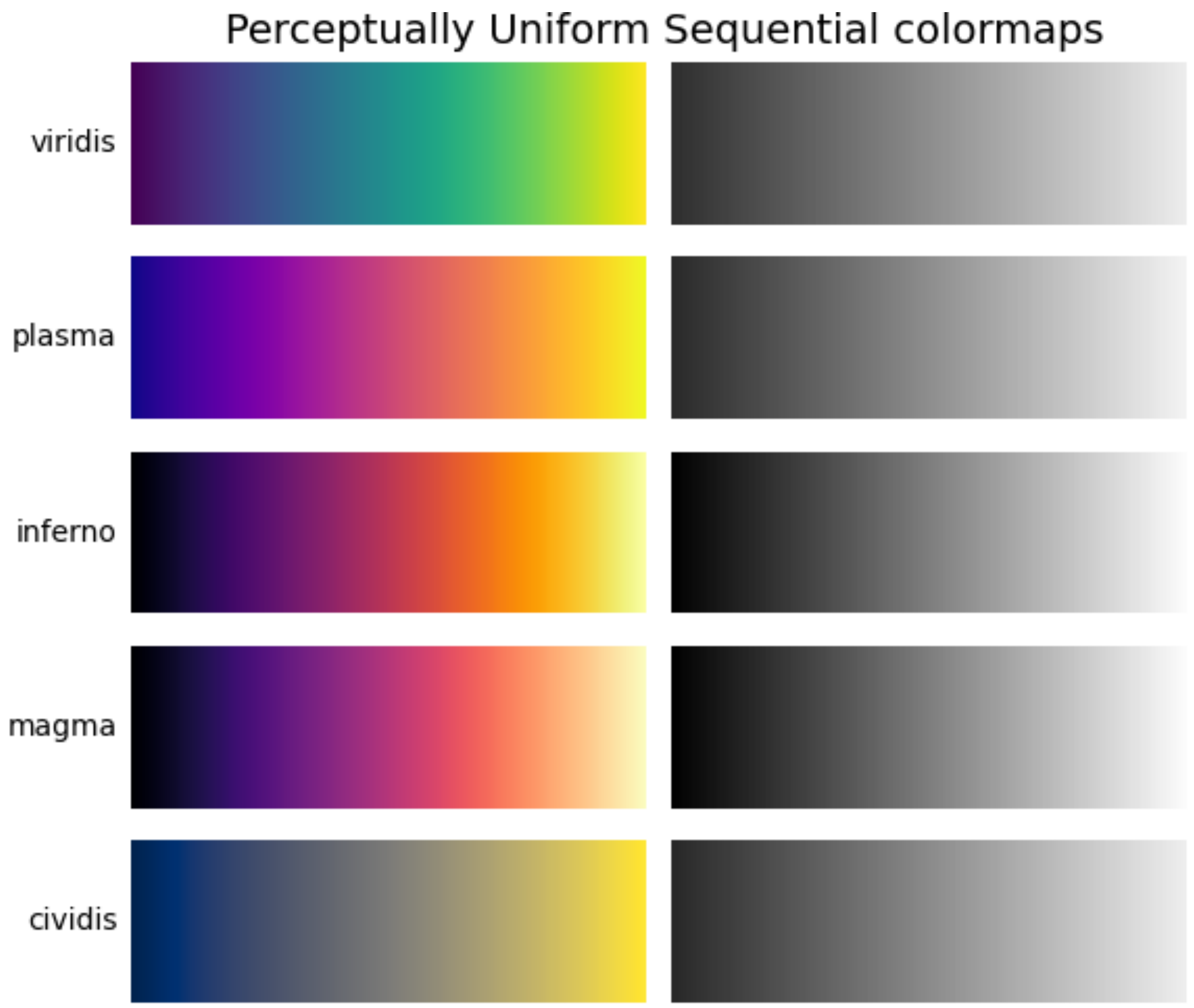
Scatter + contour



2D histogram + contour

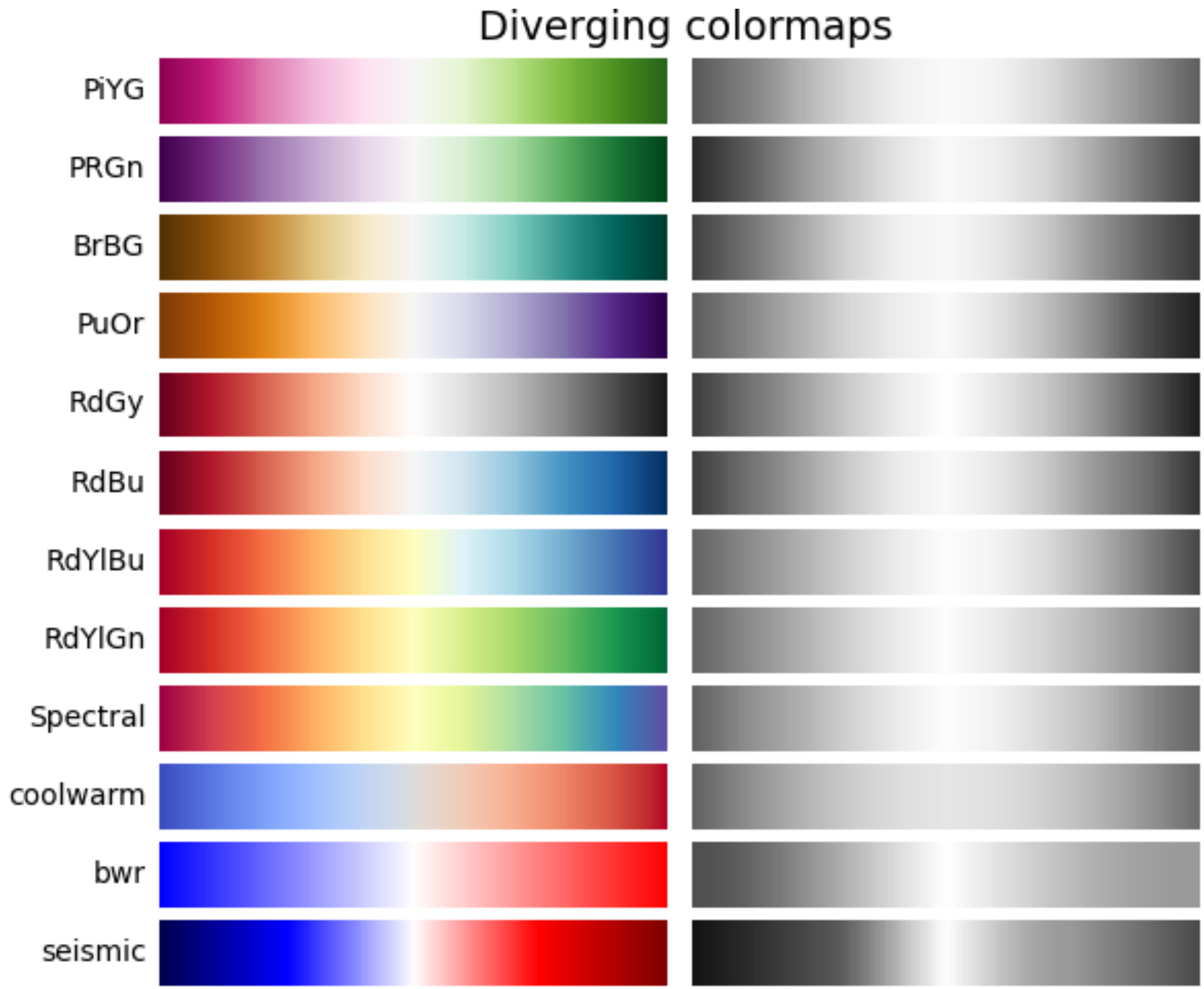


Colorbars



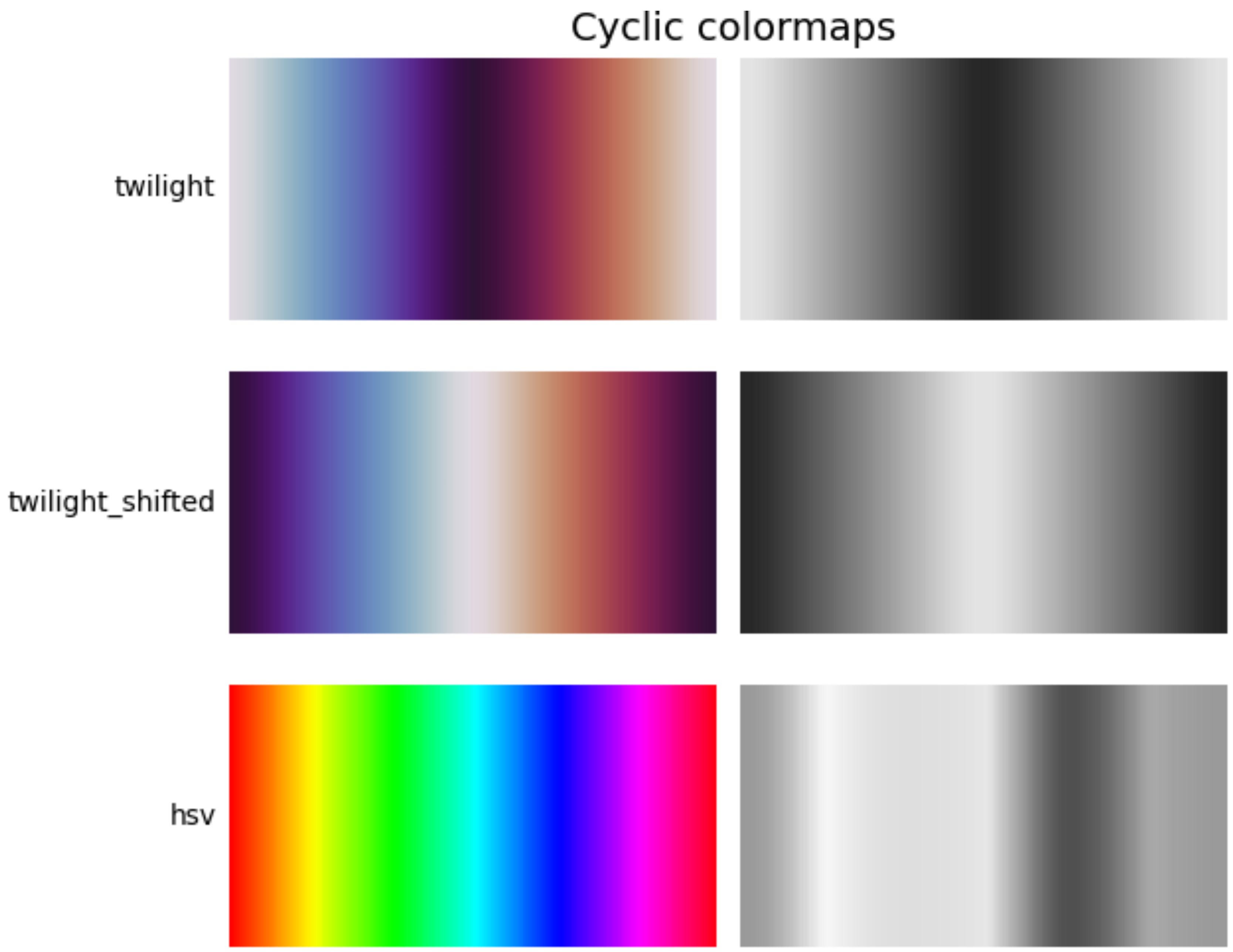
Perceptually Uniform Sequential colormaps
viridis
plasma
inferno
magma
cividis

Colorbars

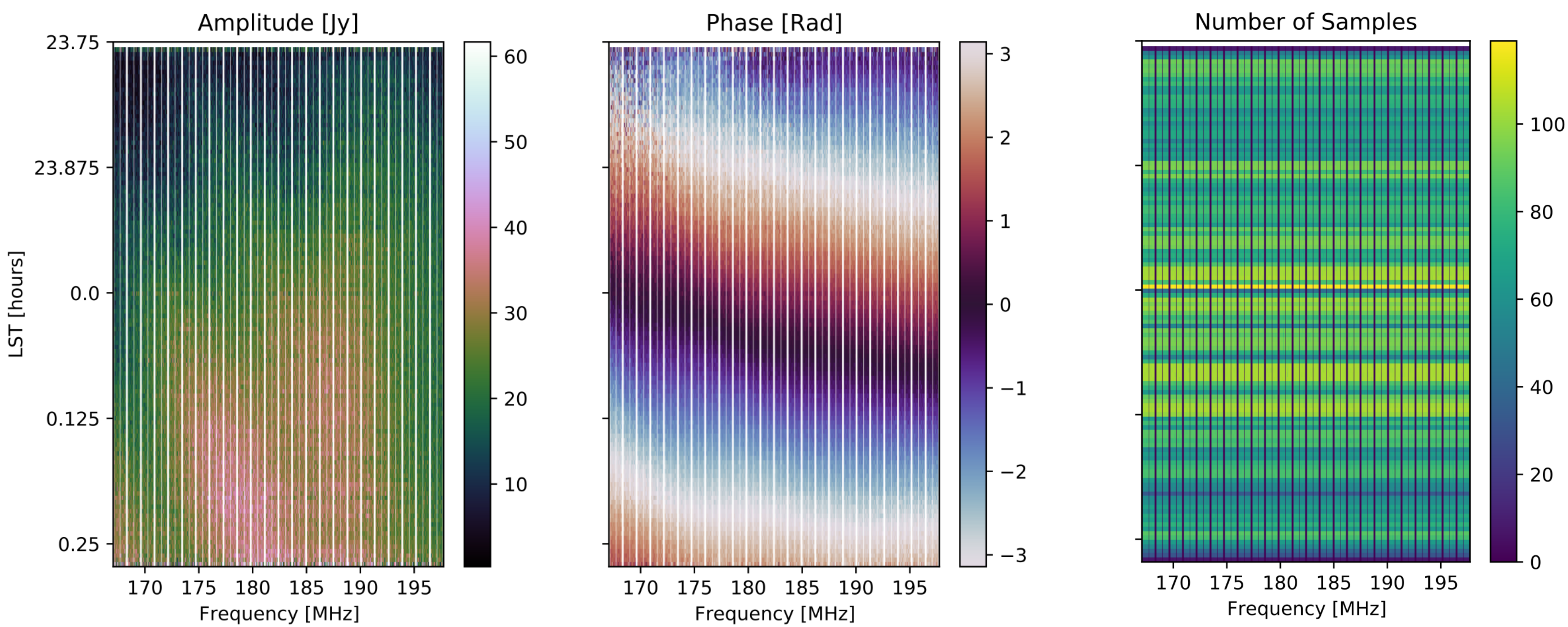


Diverging colormaps
PiYG
PRGn
BrBG
PuOr
RdGy
RdBu
RdYlBu
RdYlGn
Spectral
coolwarm
bwr
seismic

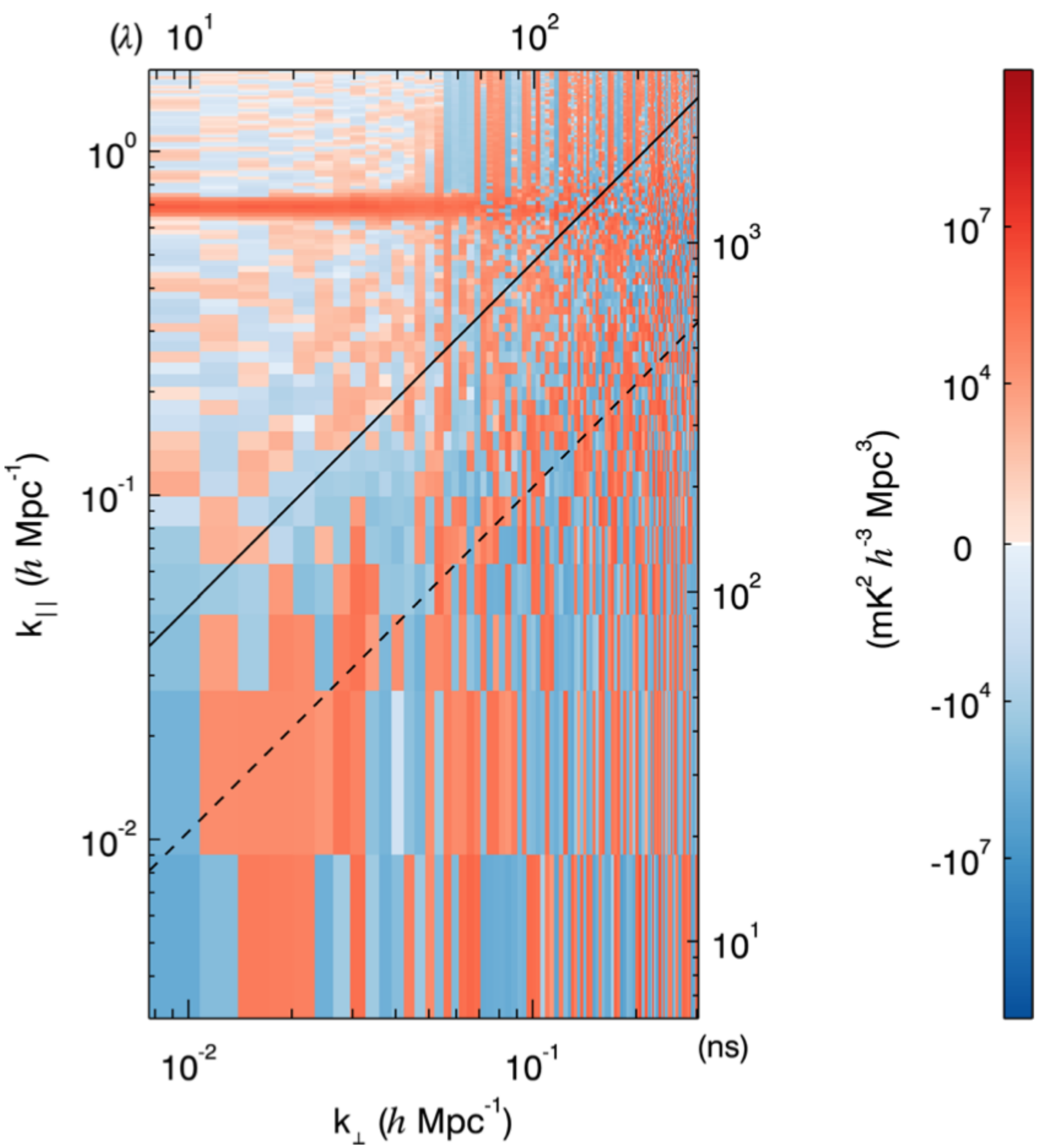
Colorbars



Use of colors



Double log diverging?



Data density examples

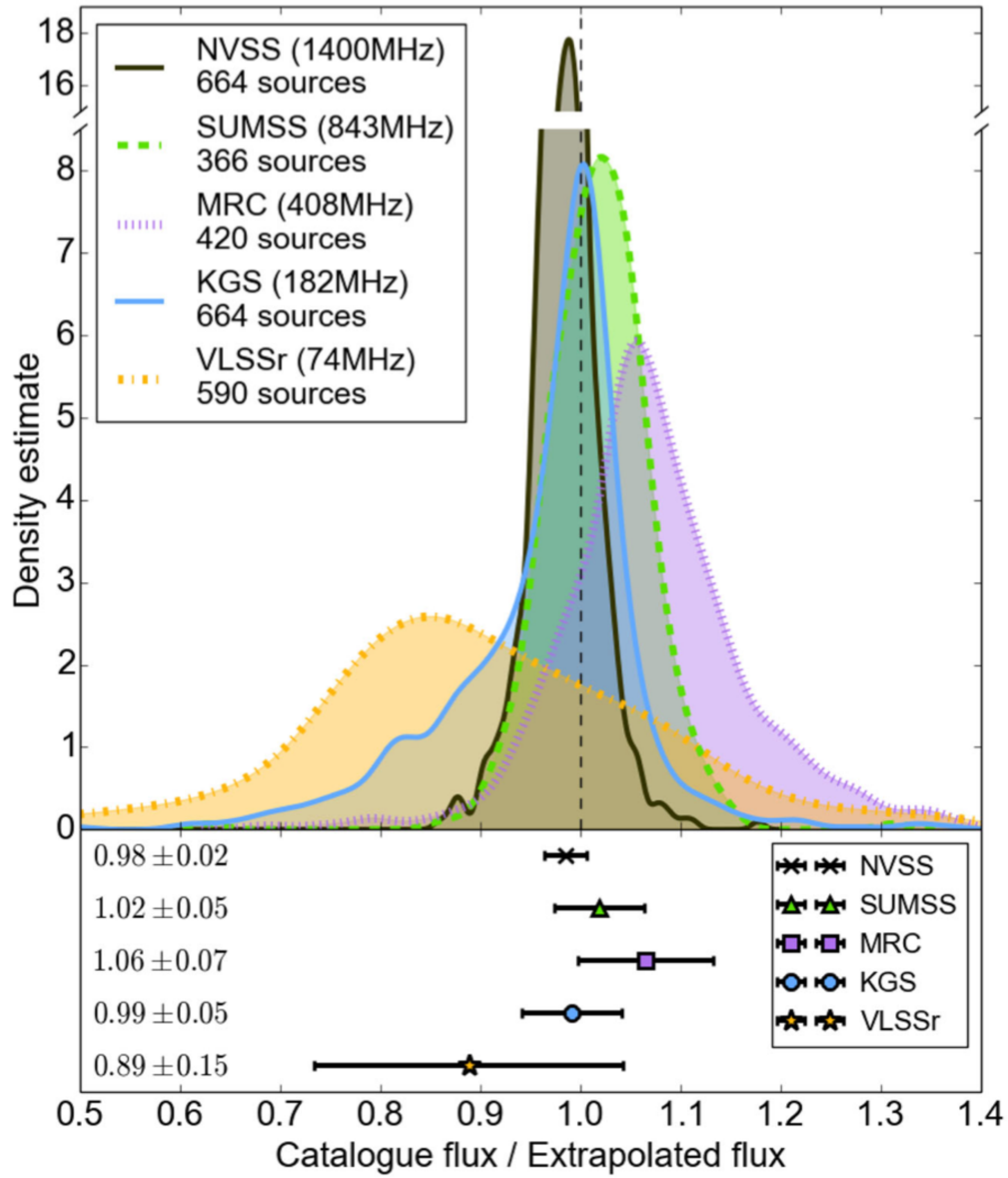
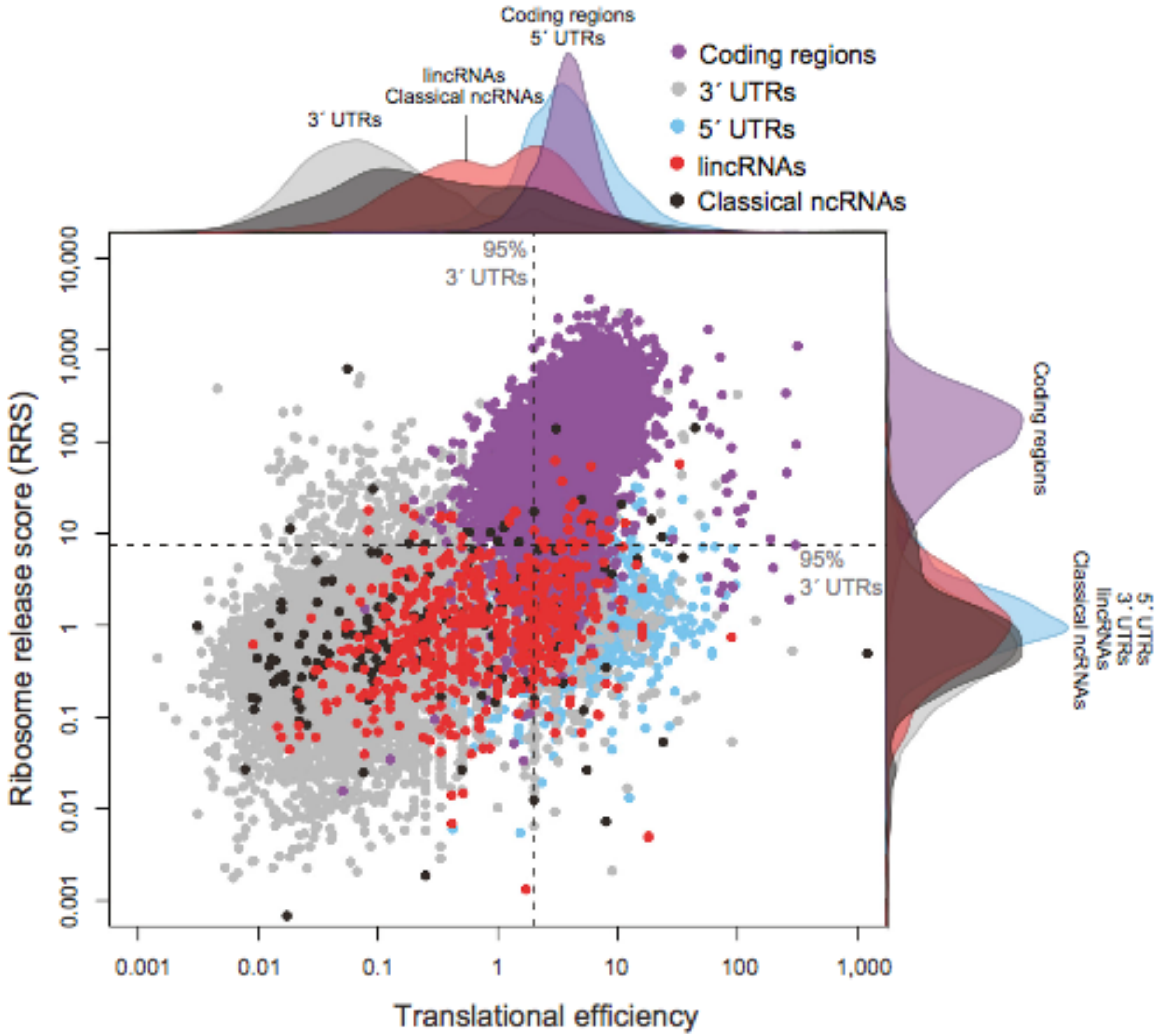
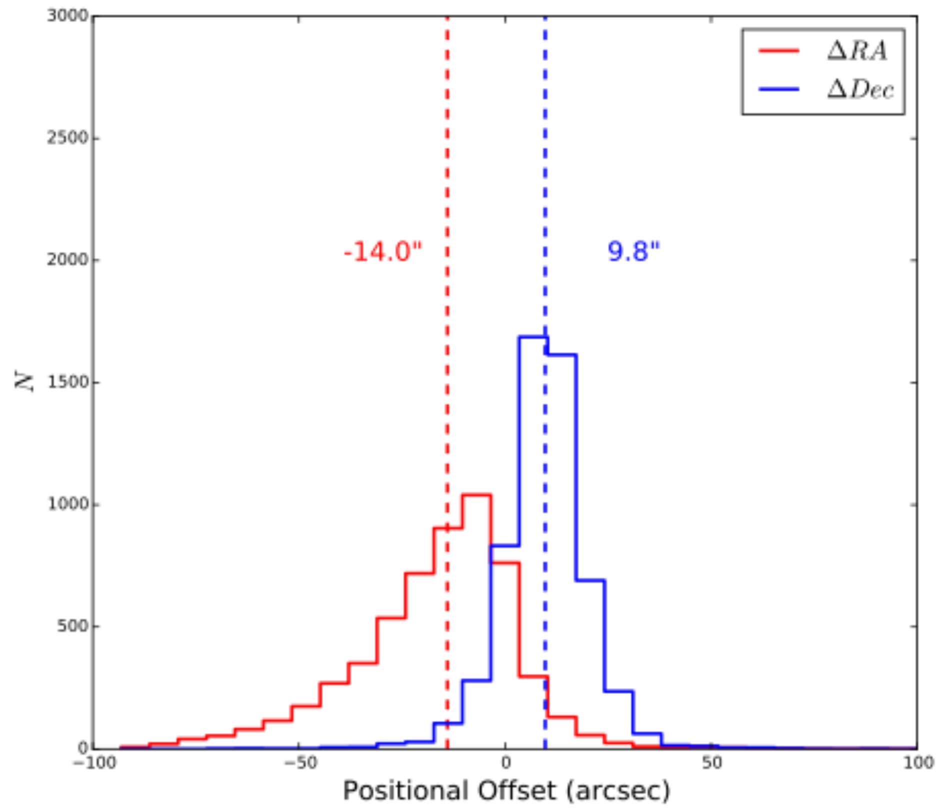


Figure 2. The ratio between observed flux density and extrapolated flux density from a fit to the SED is shown for every time a catalogue appeared in a match with at least two other catalogues for `isolated` sources. The upper panel shows a univariate kernal density estimation of each distribution (note 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.

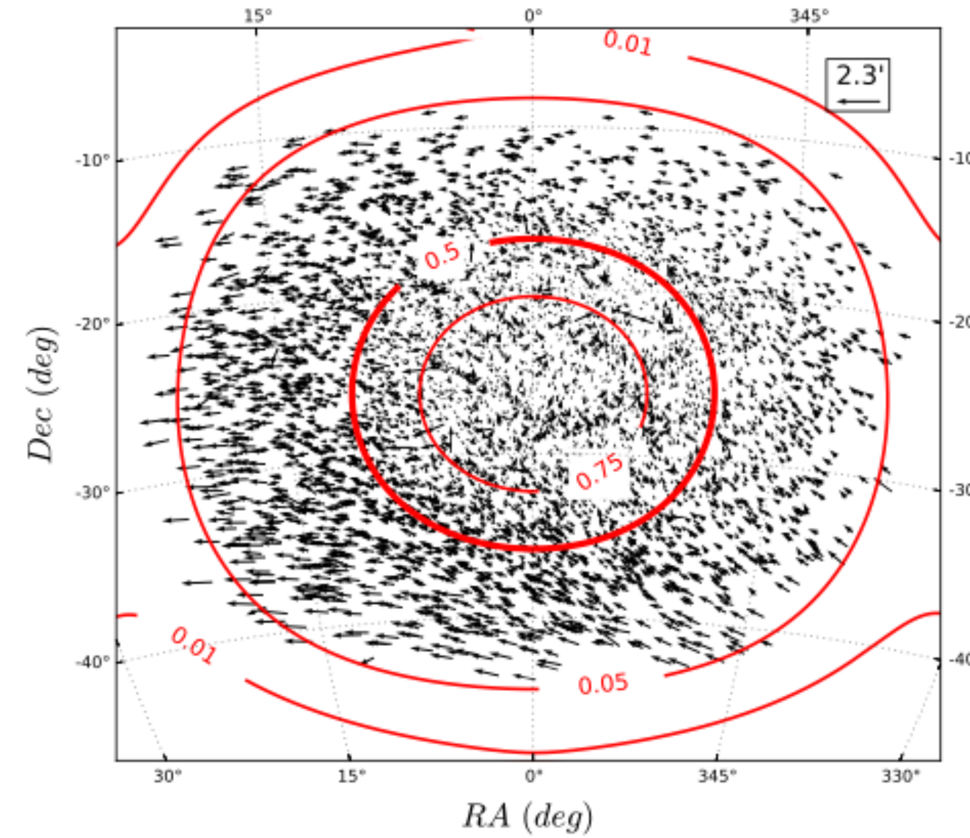


Coding regions
5′ UTRs
lincRNAs
Classical ncRNAs
3′ UTRs
Coding regions
3′ UTRs
5′ UTRs
lincRNAs
Classical ncRNAs
95%
3′ UTRs
95%
3′ UTRs
Coding regions
5′ UTRs
3′ UTRs
lincRNAs
Classical ncRNAs
Ribosome release score (RRS)
10,000
1,000
100
10
1
0.1
0.01
0.001
0.001
0.01
0.1
1
10
100
1,000
Translational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



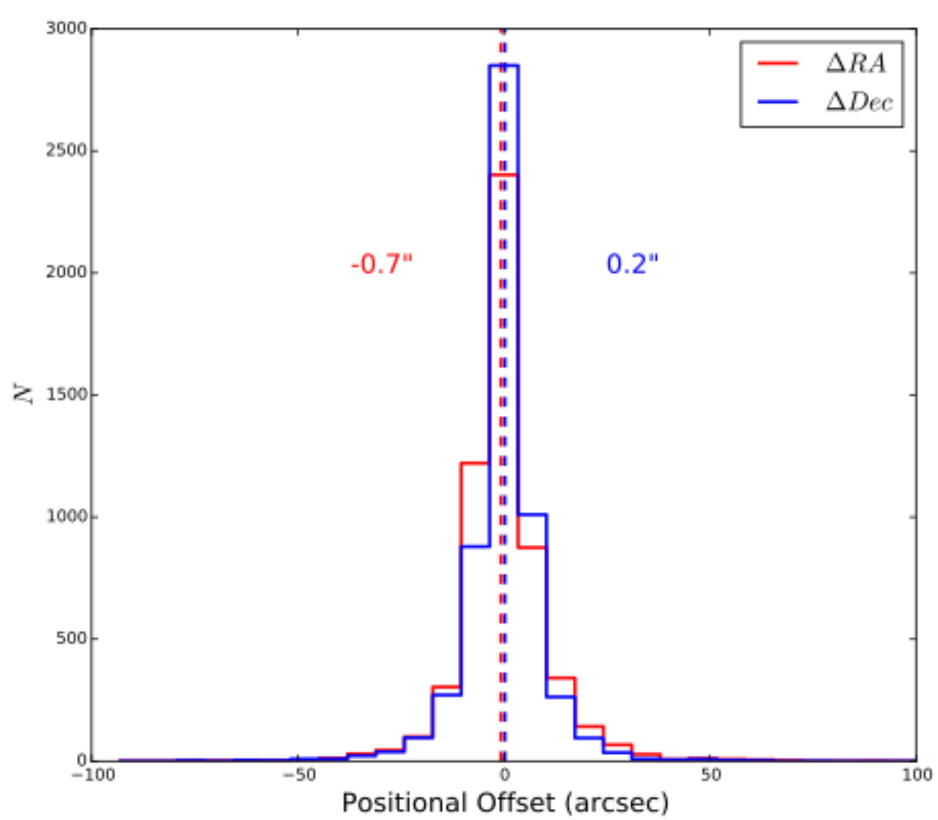
ΔRA
ΔDec
-14.0"
9.8"
N
Positional Offset (arcsec)

(a)



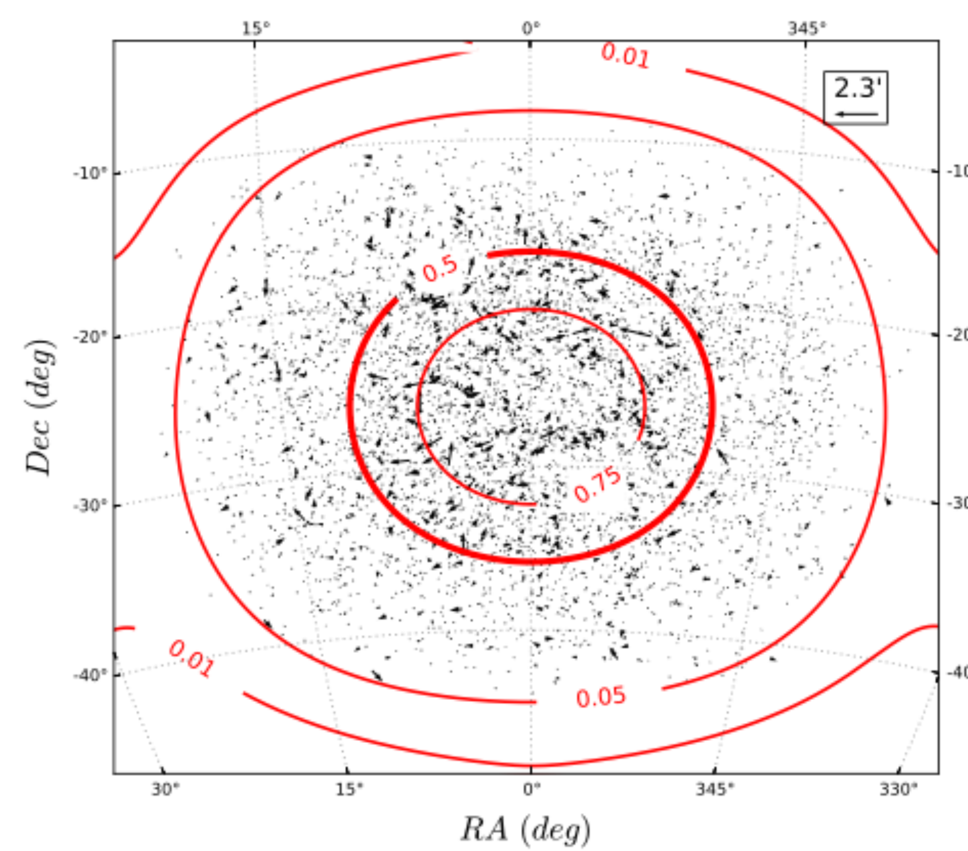
2.3'
0.01
0.5
0.75
0.05
0.01
Dec (deg)
RA (deg)

(b)



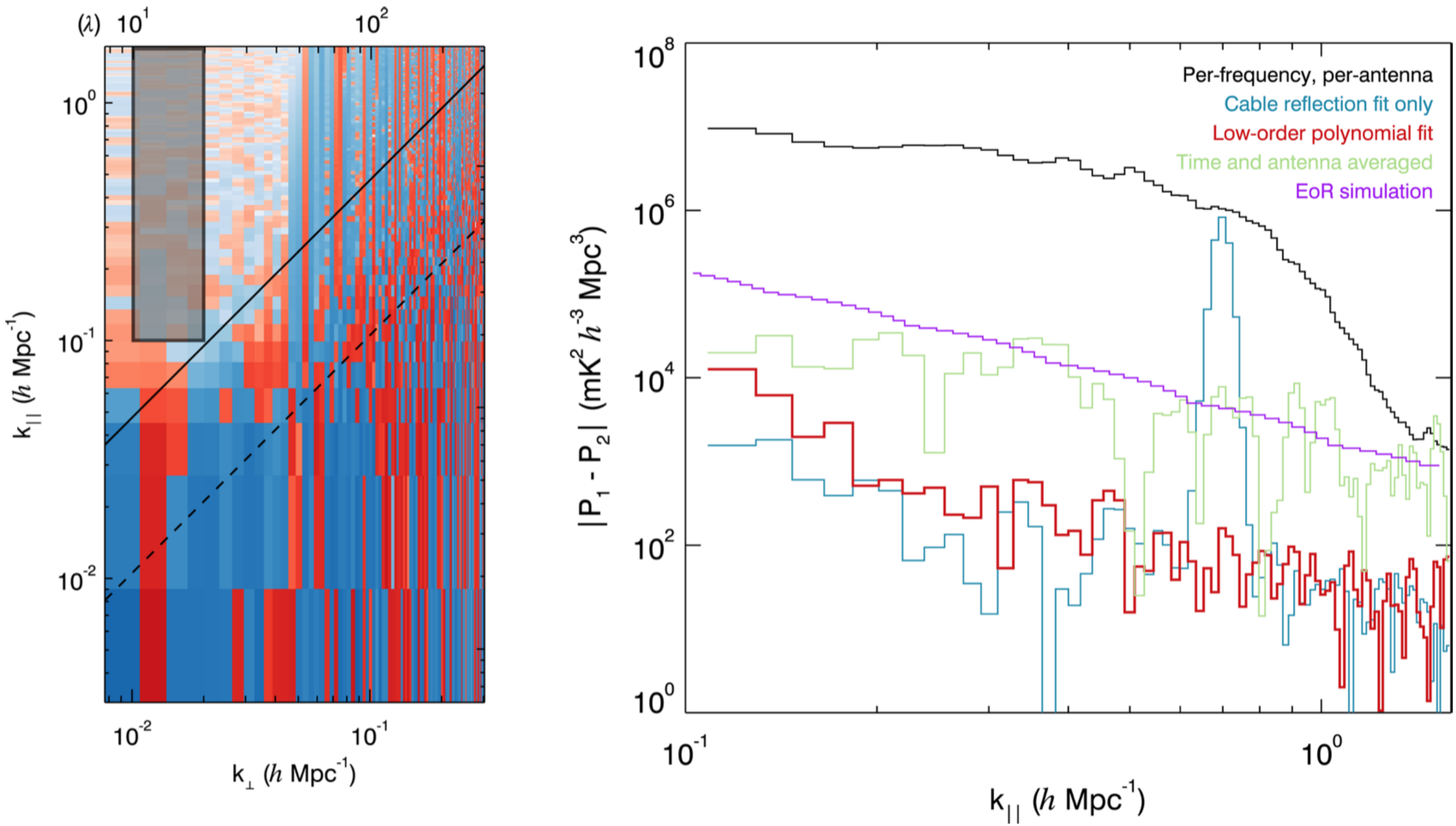
ΔRA
ΔDec
-0.7"
0.2"
N
Positional Offset (arcsec)

(c)



2.3'
0.01
0.5
0.75
0.05
0.01
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)

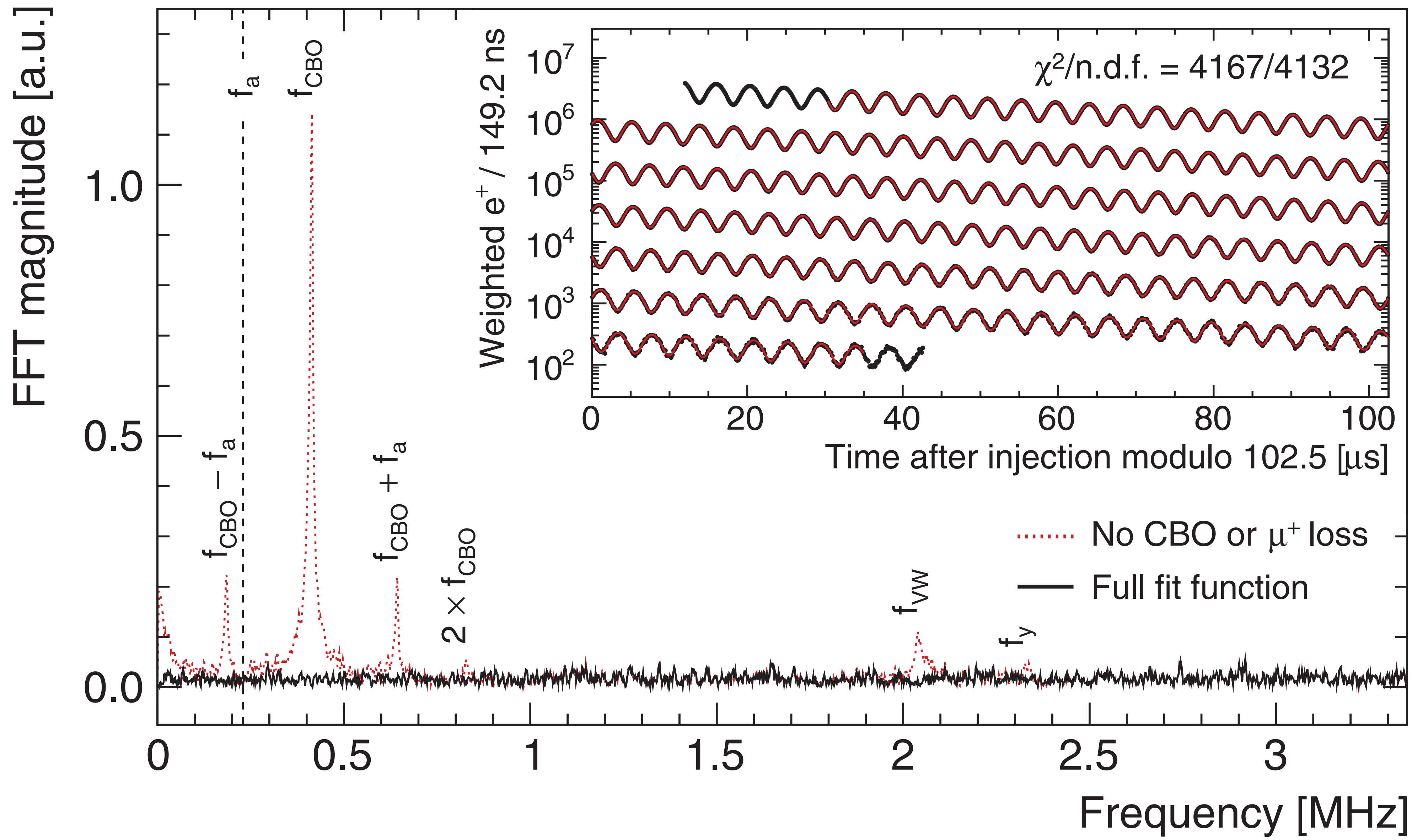
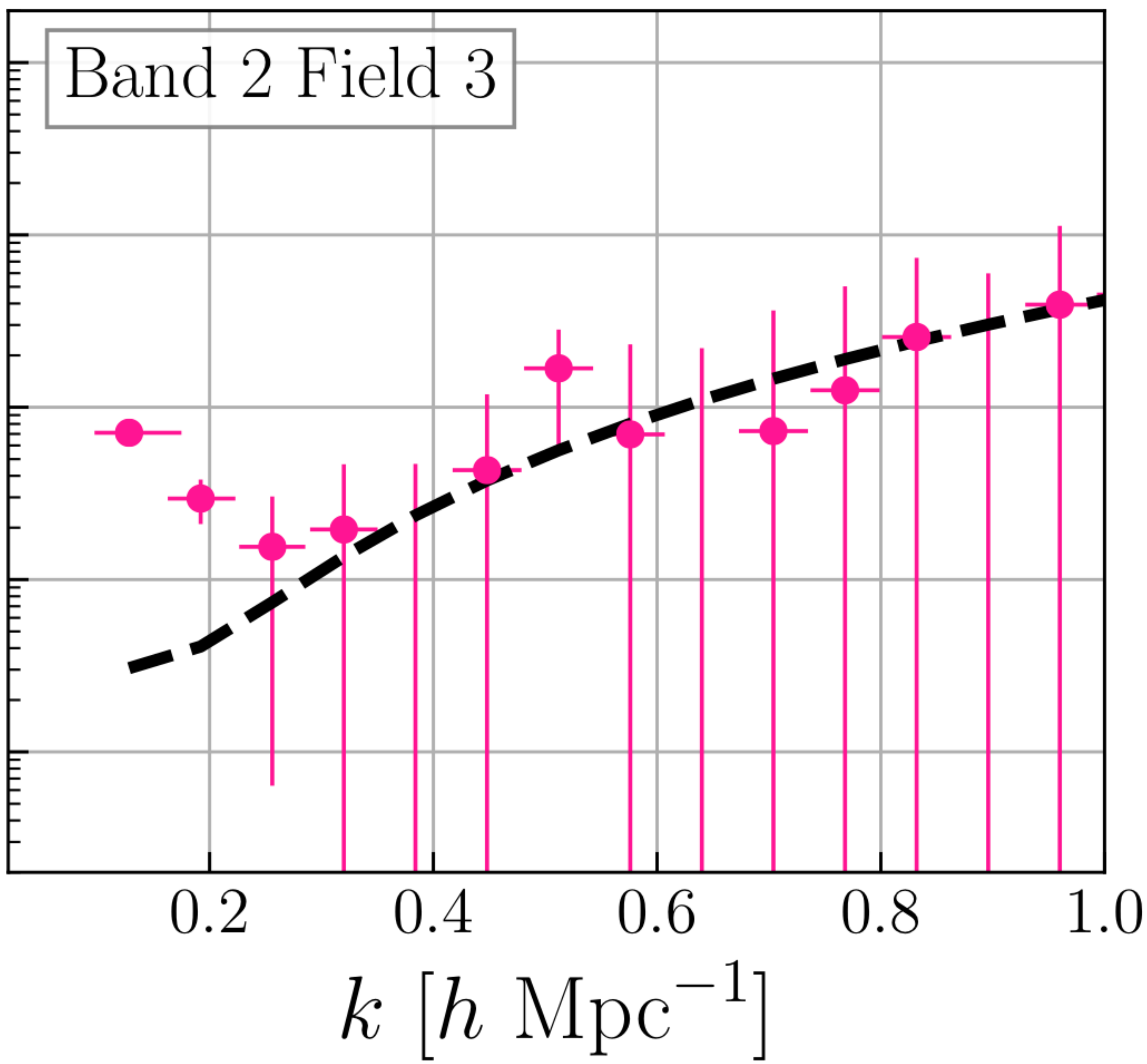


FIG. 2. Fourier transform of the residuals from a time-series fit following Eq. (5) but neglecting betatron motion and muon loss (red dashed), and from the full fit (black). The peaks correspond to the neglected betatron frequencies and muon loss. Inset: asymmetry-weighted e^+ time spectrum (black) from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

Emphasis critiques

Upper limit?



Upper limit

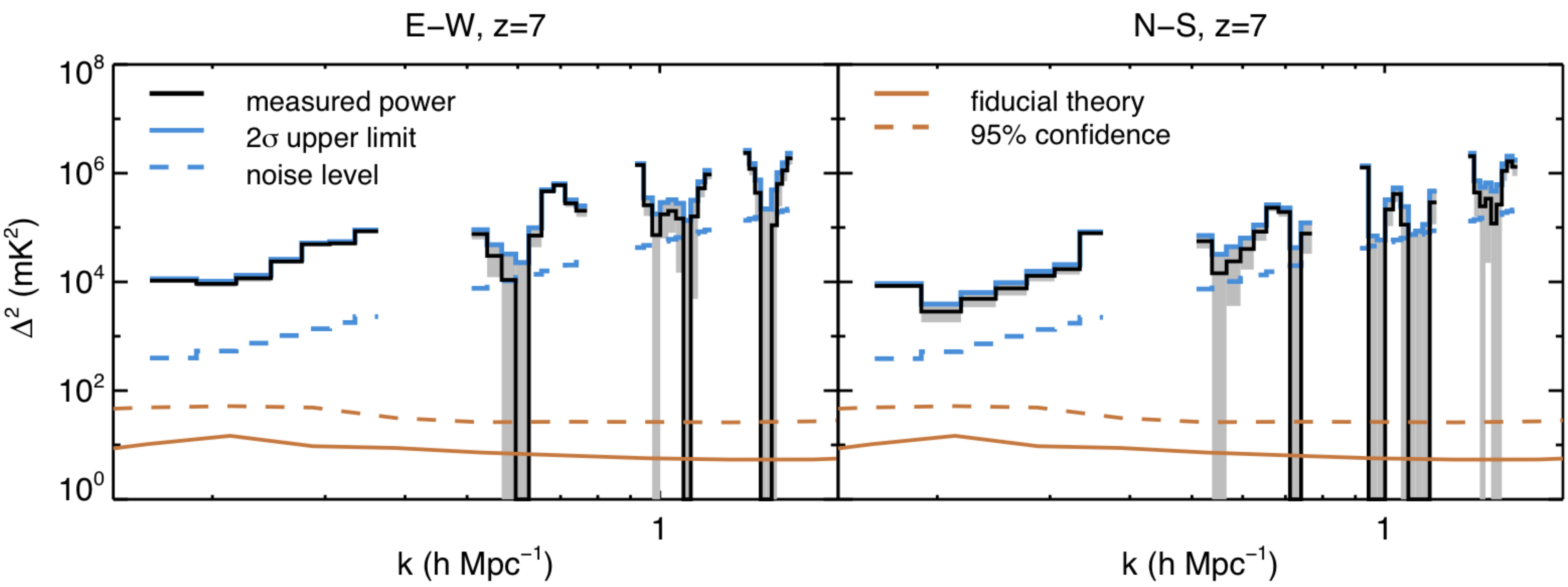
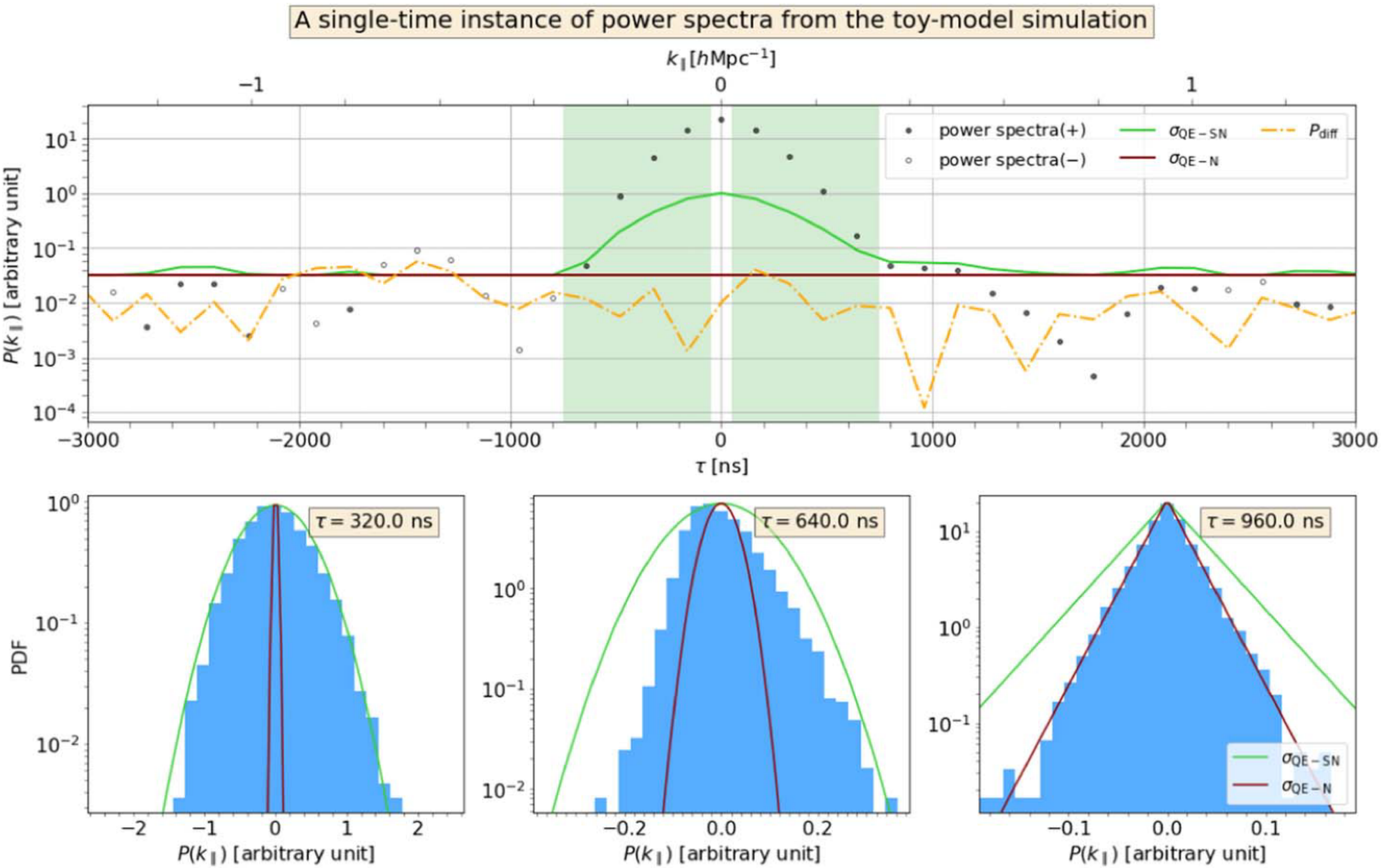


Figure 6. The 1D measured power spectra (black), the 2σ error bars (grey), the 2σ EoR upper limits (solid blue) and the 1σ thermal noise levels (dashed blue) for the E–W and N–S polarizations using 678 observations selected with SSINS. We also present an example fiducial EoR theory power spectrum (solid brown) along with the theoretical 2σ upper limits on the 21 cm power spectrum amplitude (brown dashed) obtained using existing observational constraints (see Appendix A for further details).

These constitute our best EoR upper limits in this work. We are noise dominated for many k-modes, including our lowest EoR upper limit.


A single-time instance of power spectra from the toy-model simulation
$k_\parallel\,[h\mathrm{Mpc}^{-1}]$
power spectra(+)
power spectra(−)
$\sigma_{\mathrm{QE-SN}}$
$\sigma_{\mathrm{QE-N}}$
P_{diff}
$P(k_\parallel)$ [arbitrary unit]
τ [ns]
$\tau = 320.0$ ns
$\tau = 640.0$ ns
$\tau = 960.0$ ns
PDF