Class 8: Jackknifes & statistically valid plots

Miguel F. Morales
Bryna Hazelton

Statistically valid plots

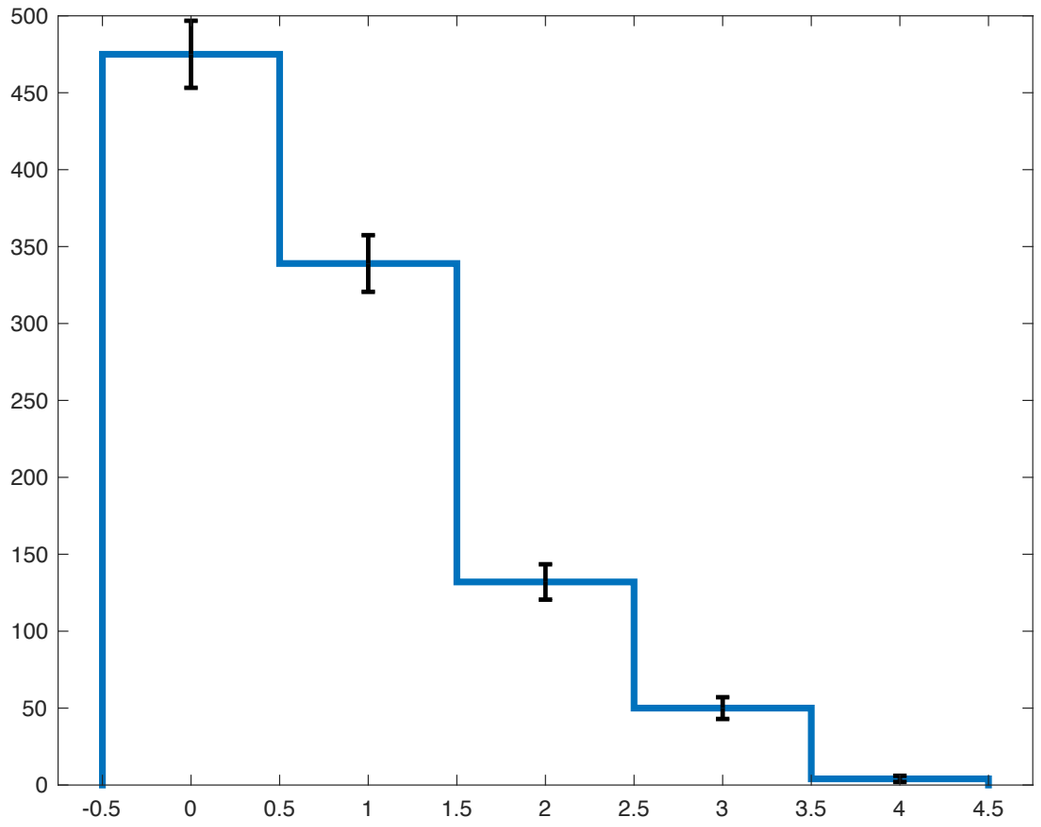
Quantitative visual statistics

Statistical tests & plots

- Requires you know the error (background dist.)
    - for every data value or transformation of data (analysis step or even histograms)
        - Often means propagating your error
        - Allows sophisticated questions
- Statistically normalizing data
- Residual plots (data - model)

Histogram with Error bars

```
figure()
h_all = histogram(data(1:1000),'DisplayStyle',"stairs",'LineWidth',3);
hold on
shift = (h_all.BinEdges(2)-h_all.BinEdges(1))/2;
errorbar(h_all.BinEdges(1:end-1)+shift,h_all.Values, sqrt(h_all.Values), 'k','linestyle','none','LineWidth',2)
```

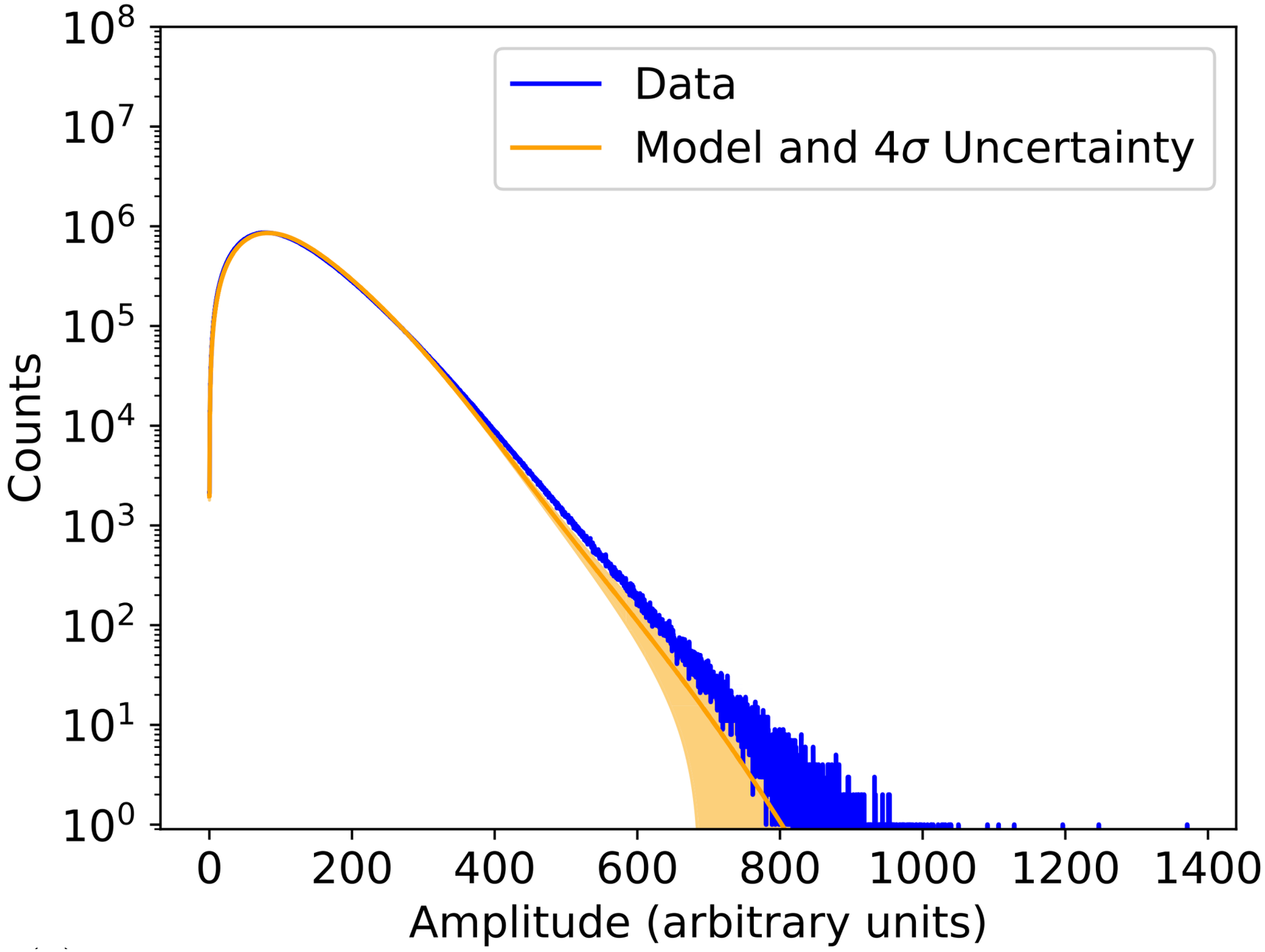


Statistically normalized plots

SSINS

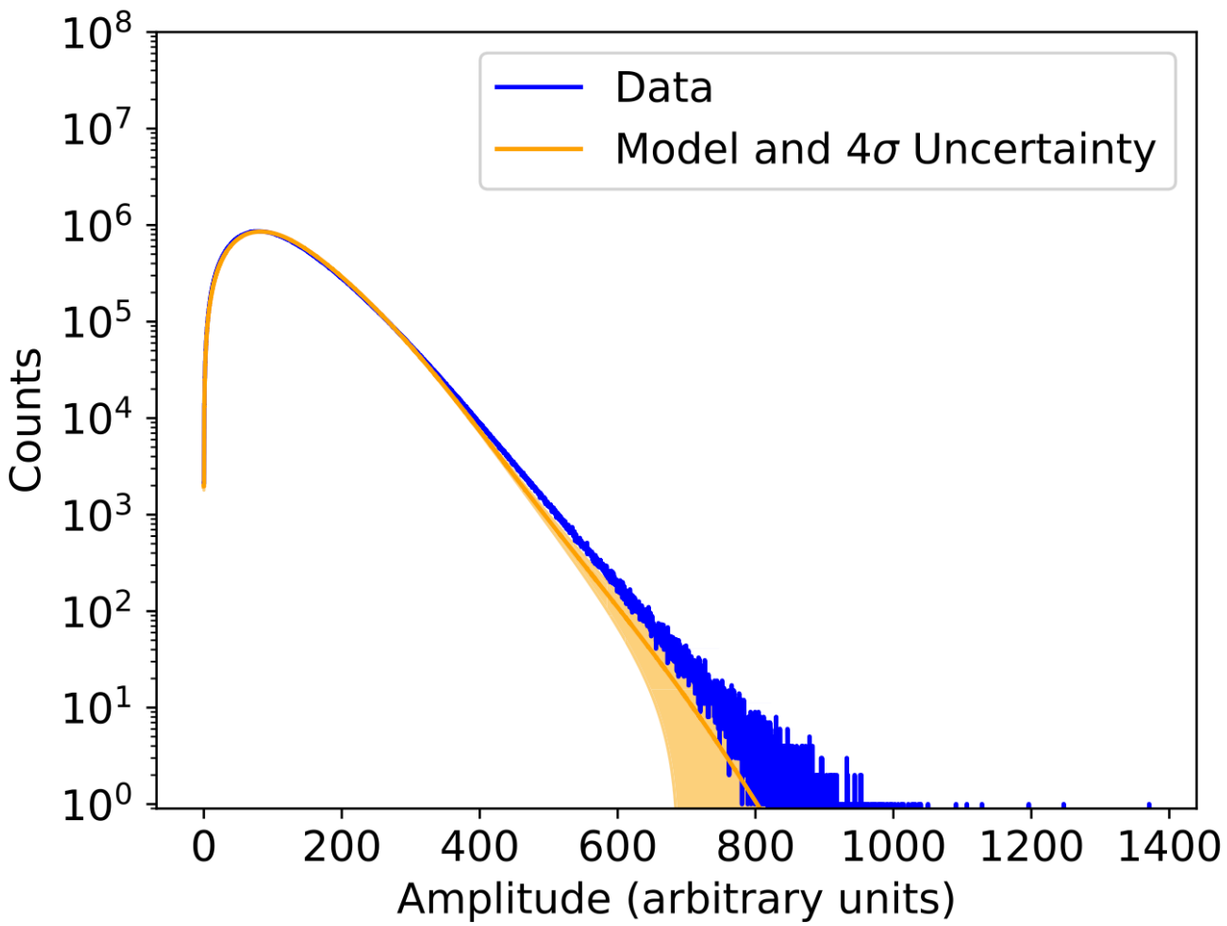
- data from a radio telescope
  - ~8k baselines, ~50 times, ~380 frequencies
    = 1.57 x10^8 measurements
- differenced on short time scales to remove the signal from the sky (short relative to sky rotation)
- result should be noise (background)
  - unless there is contamination that changes on times shorter than the differencing time (RFI)

SSINS

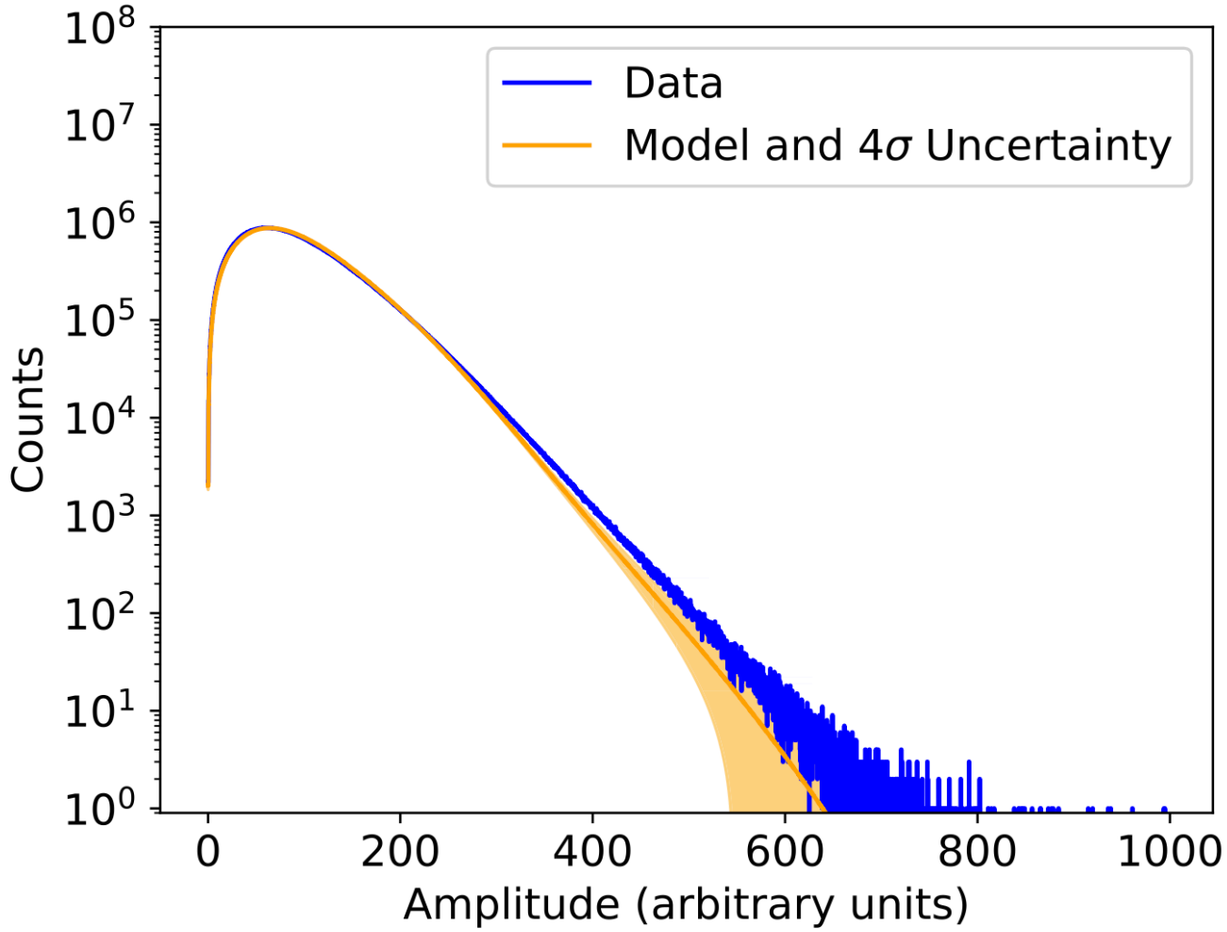

Data
Model and 4σ Uncertainty
Counts
Amplitude (arbitrary units)
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
0
200
400
600
800
1000
1200
1400

SSINS

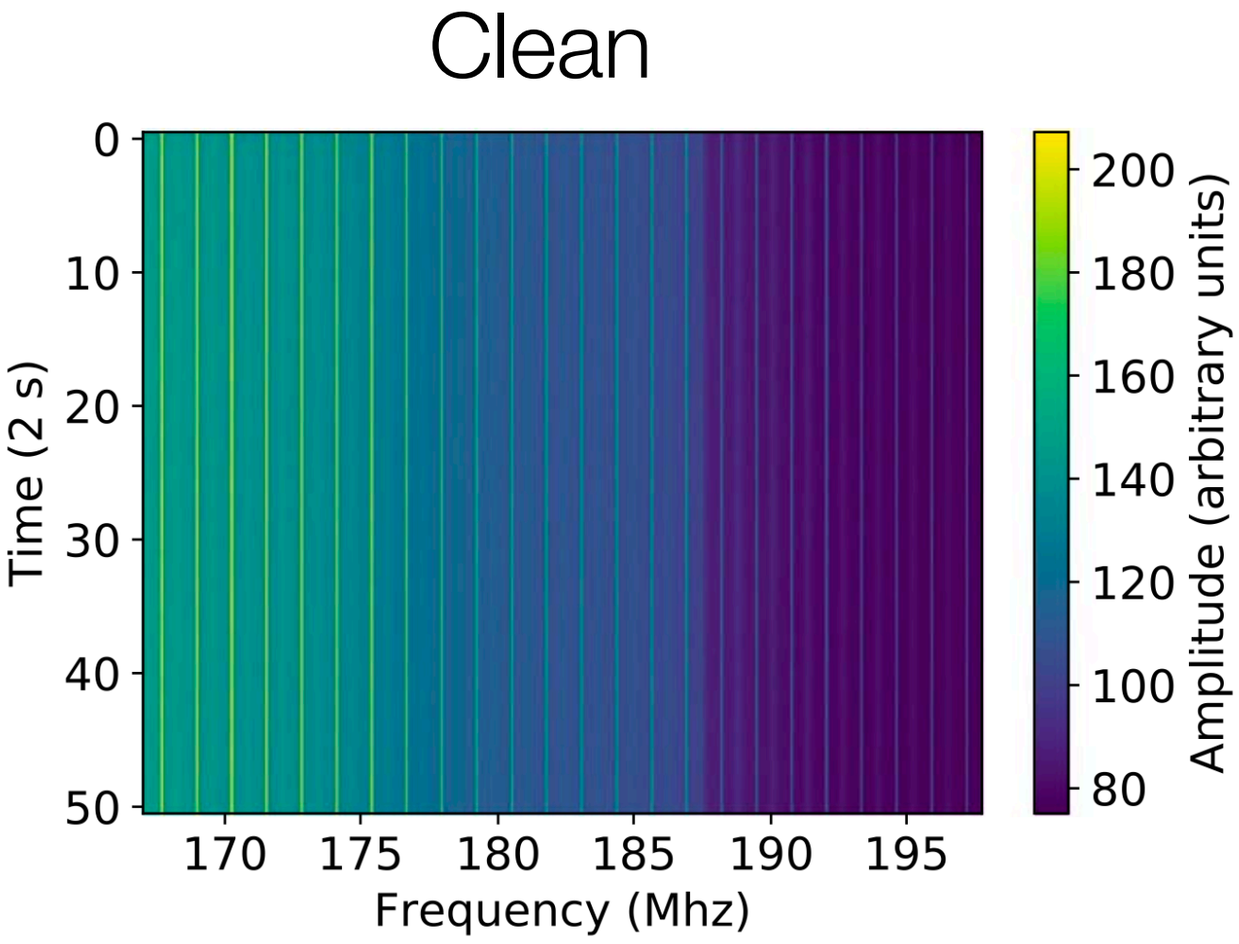
Clean

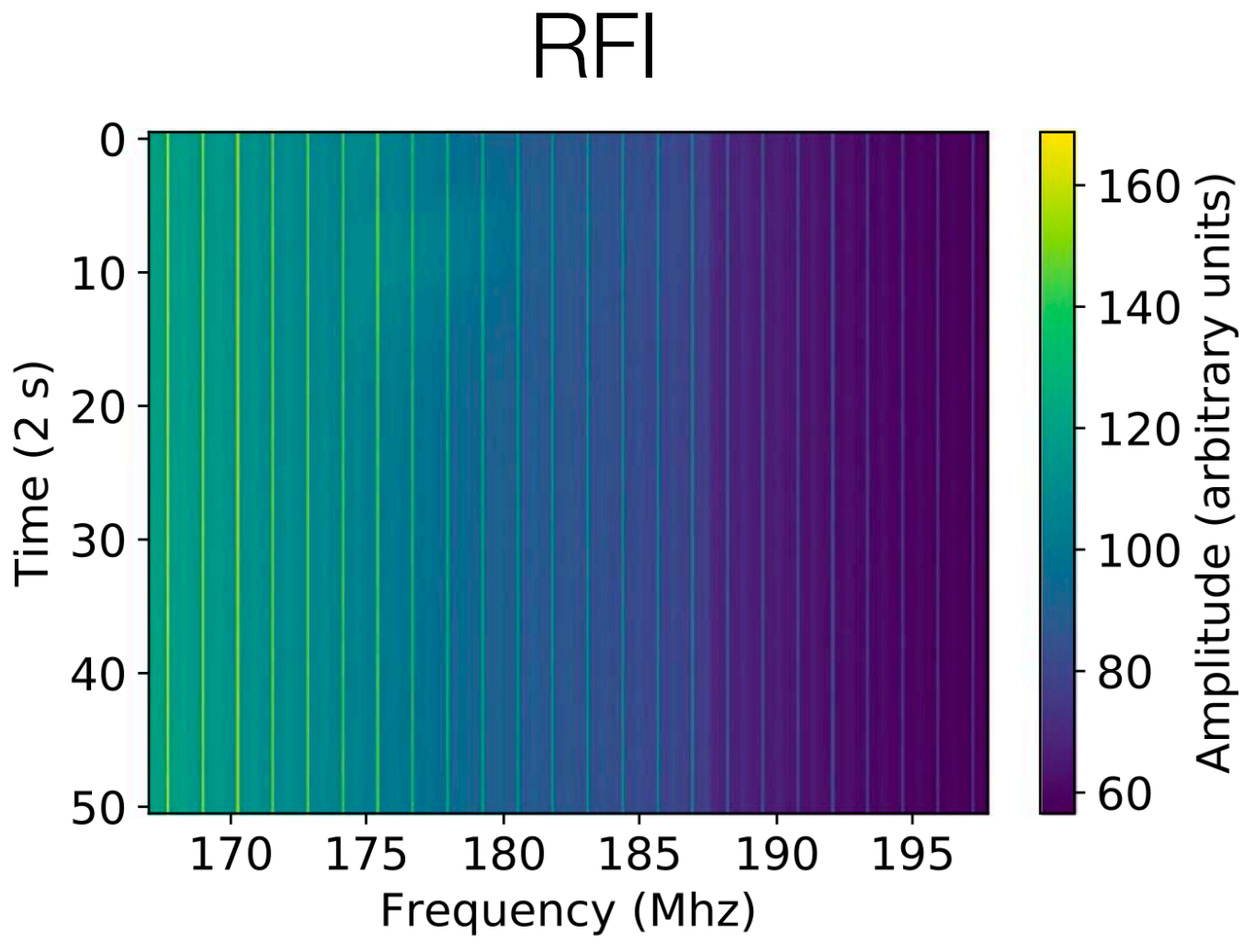


RFI



SSINS


Clean
Time (2 s)
Frequency (Mhz)
Amplitude (arbitrary units)
0
10
20
30
40
50
170
175
180
185
190
195
200
180
160
140
120
100
80


RFI
Time (2 s)
Frequency (Mhz)
Amplitude (arbitrary units)
0
10
20
30
40
50
170
175
180
185
190
195
160
140
120
100
80
60

SSINS

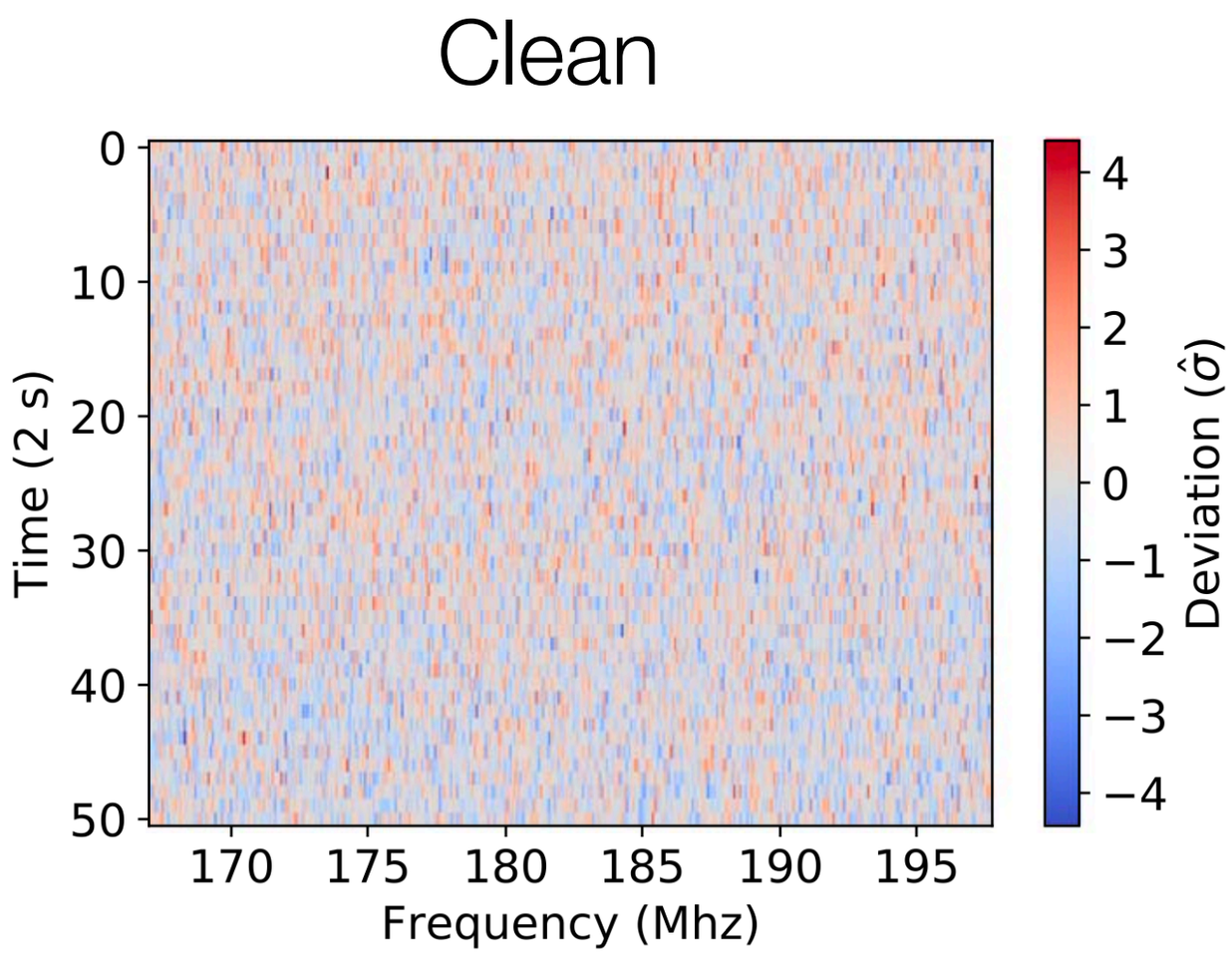
- Background (noise) level varies as a function of frequency but not as a function of time

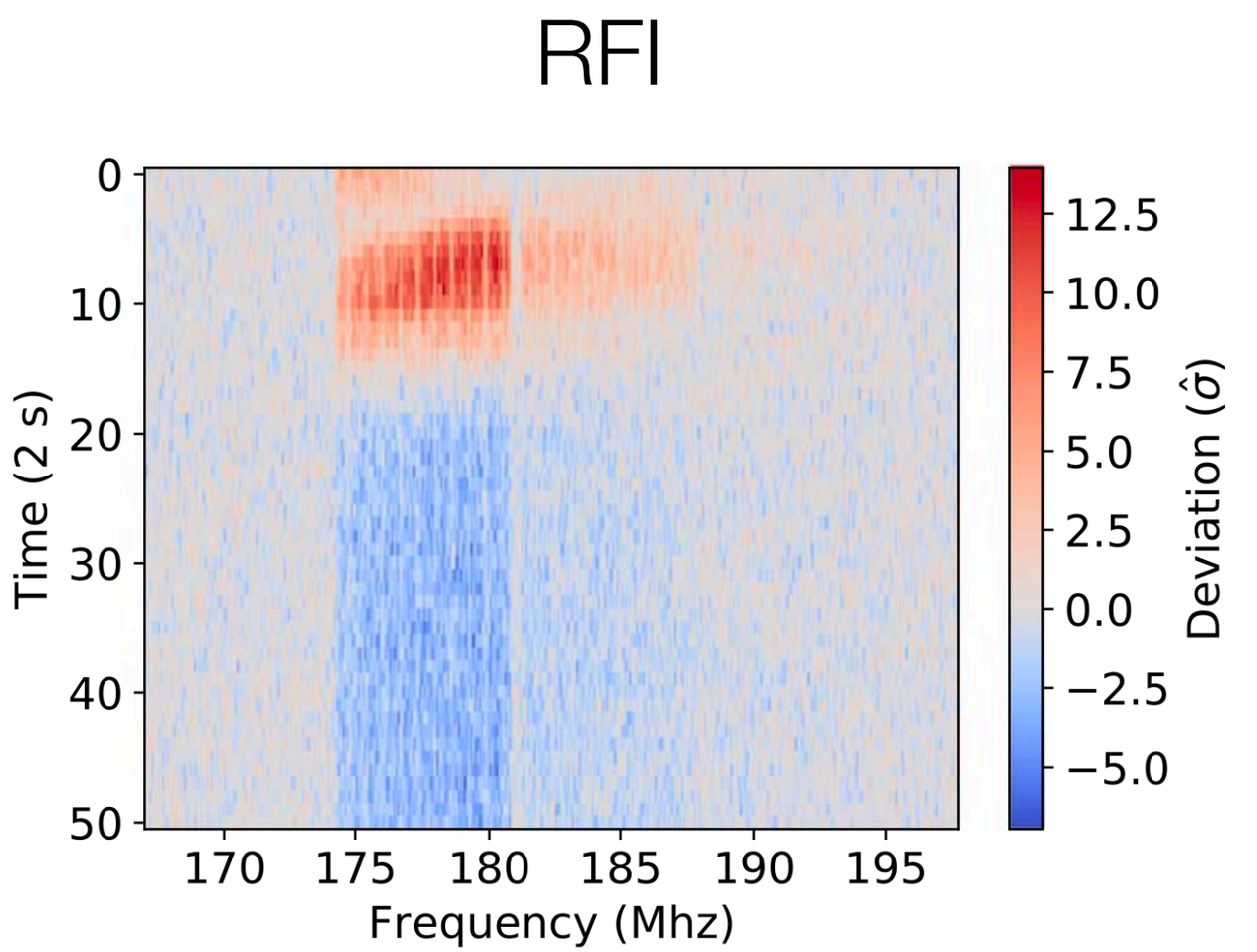
- Because the Rayleigh is a single parameter distribution, the mean predicts the variance — use the mean to normalize the distributions.

- find the mean across time, subtract the mean and normalize by the predicted variance to get a standard gaussian

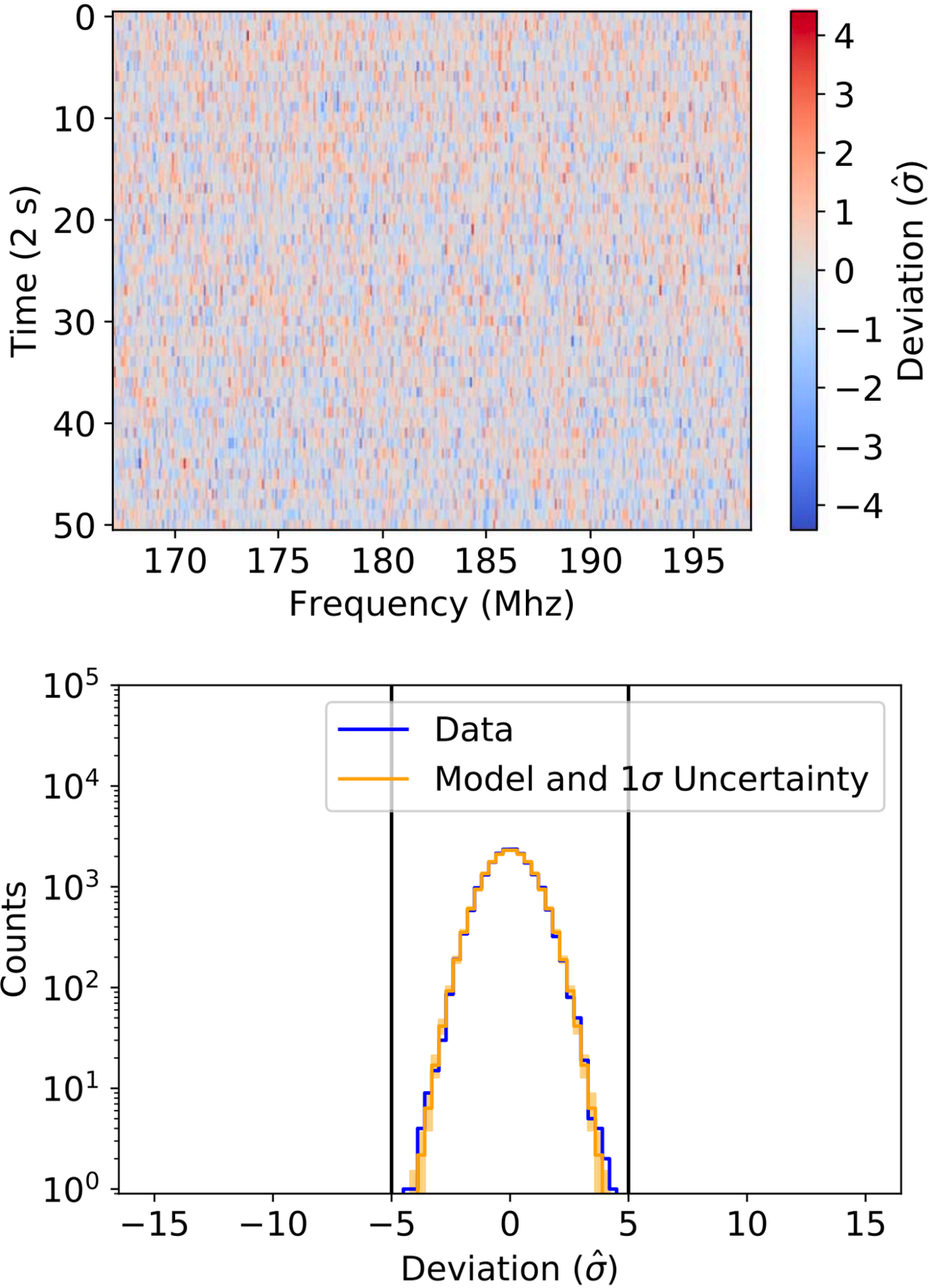
    - After averaging over ~8k baselines, the distribution is gaussian (not Rayleigh), but the variance is predictable because it comes from the Rayleigh distribution

SSINS

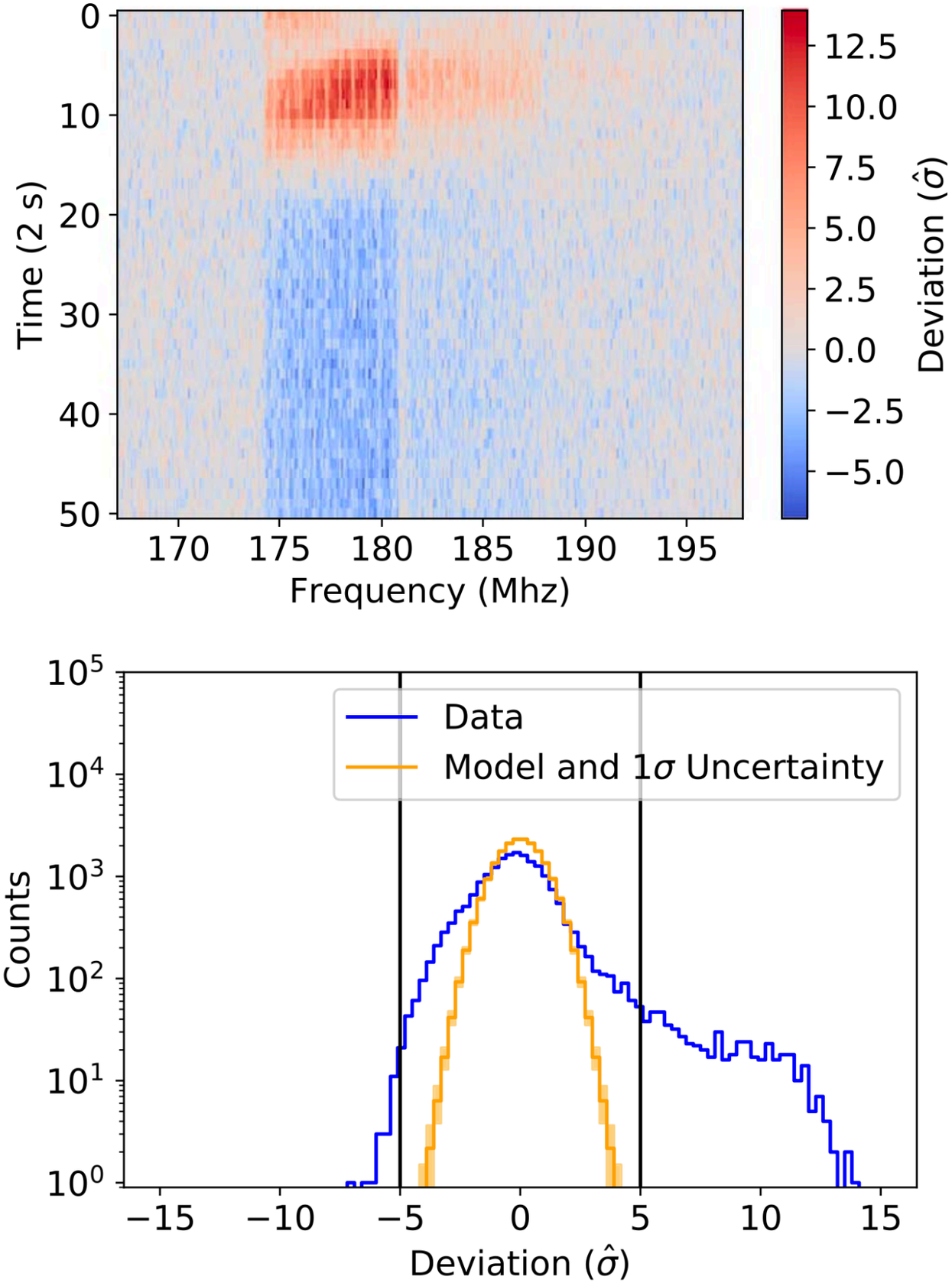

Clean
Time (2 s)
Frequency (Mhz)
Deviation ($\hat{\sigma}$)
0
10
20
30
40
50
170
175
180
185
190
195
4
3
2
1
0
−1
−2
−3
−4


RFI
Time (2 s)
Frequency (Mhz)
Deviation ($\hat{\sigma}$)
0
10
20
30
40
50
170
175
180
185
190
195
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0

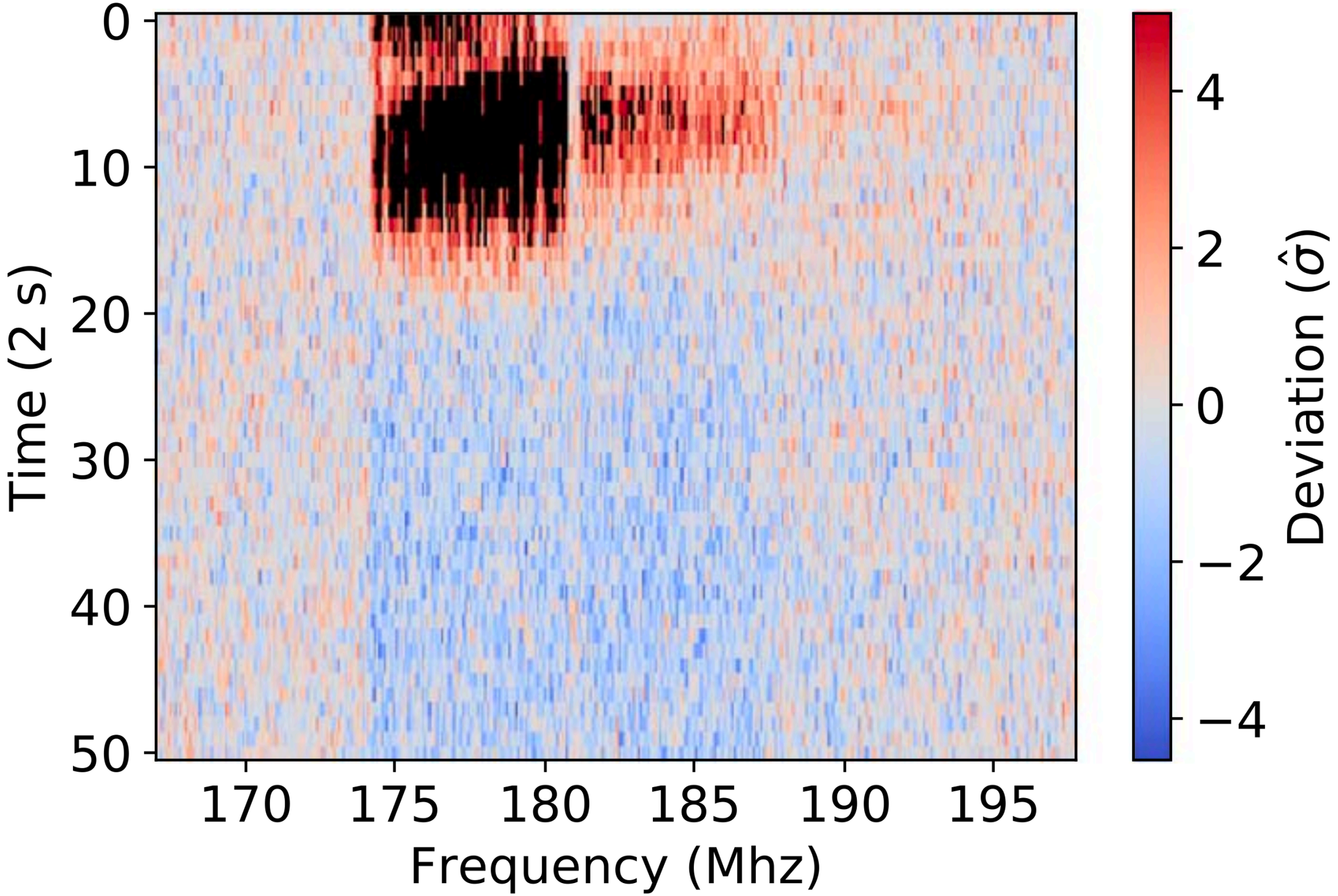
Clean


Time (2 s)
Frequency (Mhz)
Deviation ($\hat{\sigma}$)
Counts
Data
Model and 1σ Uncertainty
Deviation ($\hat{\sigma}$)

RFI


Time (2 s)
Frequency (Mhz)
Deviation ($\hat{\sigma}$)
Counts
Data
Model and 1σ Uncertainty
Deviation ($\hat{\sigma}$)

SSINS


0
10
20
30
40
50
Time (2 s)
170
175
180
185
190
195
Frequency (Mhz)
4
2
0
−2
−4
Deviation (σ)

Australian radiofrequency spectrum allocations chart

The spectrum is a continuous range of electromagnetic radiation extending from the longest radio waves through infrared light, visible light and X-rays to gamma rays.

The radiofrequency spectrum is that part of the total spectrum which is used for transmitting radio waves. The radio frequency spectrum is a national resource that is used but not consumed. It is used by being occupied and the efficiency of its use depends on coordination among users in order to minimise interference to each other.

This chart illustrates how the radio frequency spectrum is allocated among services in Australia. It is derived from the Australian Radiofrequency Spectrum Plan (January 2009), which in turn is based on the International Telecommunication Union (ITU) Radio Regulations.

The radiofrequency spectrum is divided into several broad frequency bands for reference. The ITU conventions for these bands is as follows:

- Very Low Frequency ..... 3–30 kHz
- Low Frequency ..... 30–300 kHz
- Medium Frequency ..... 300–3000 kHz
- High Frequency ..... 3–30 MHz
- Very High Frequency ..... 30–300 MHz
- Ultra High Frequency ..... 300–3000 MHz
- Super High Frequency ..... 3–30 GHz
- Extremely High Frequency ..... 30–300 GHz

Each of these bands is divided into sub-bands which are allocated to particular services such as land mobile radio, broadcasting, aeronautical, maritime or space services^{**} shown in the chart by different colours.

The chart is designed for quick reference. For details of frequency allocations, reference should be made to the Australian Radiofrequency Spectrum Plan (January 2009), since fine details and footnotes cannot be shown on this chart.

Australian Government

acma

LEGEND

VLF

LF

MF

HF

VHF

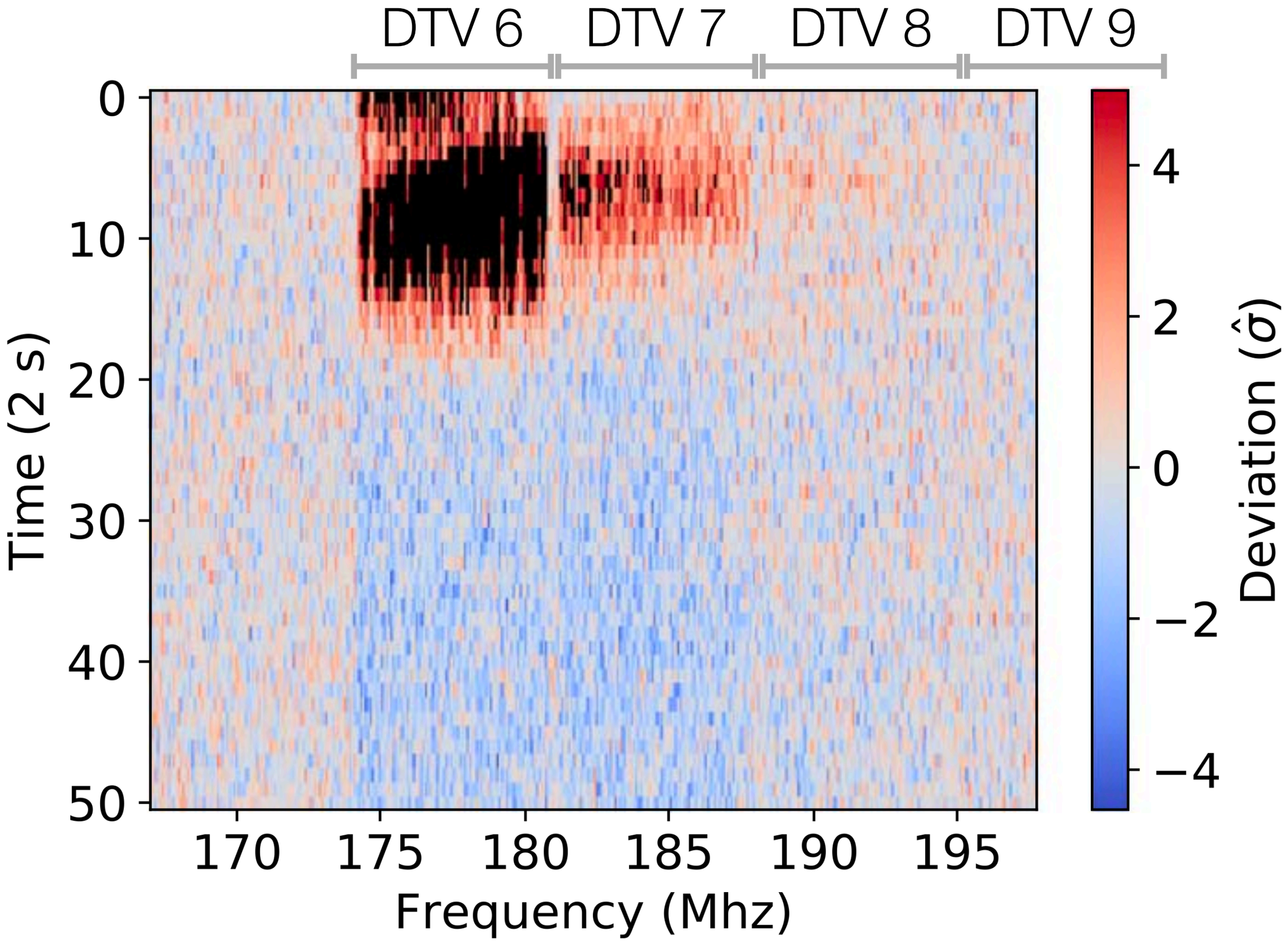
UHF

SHF

EHF

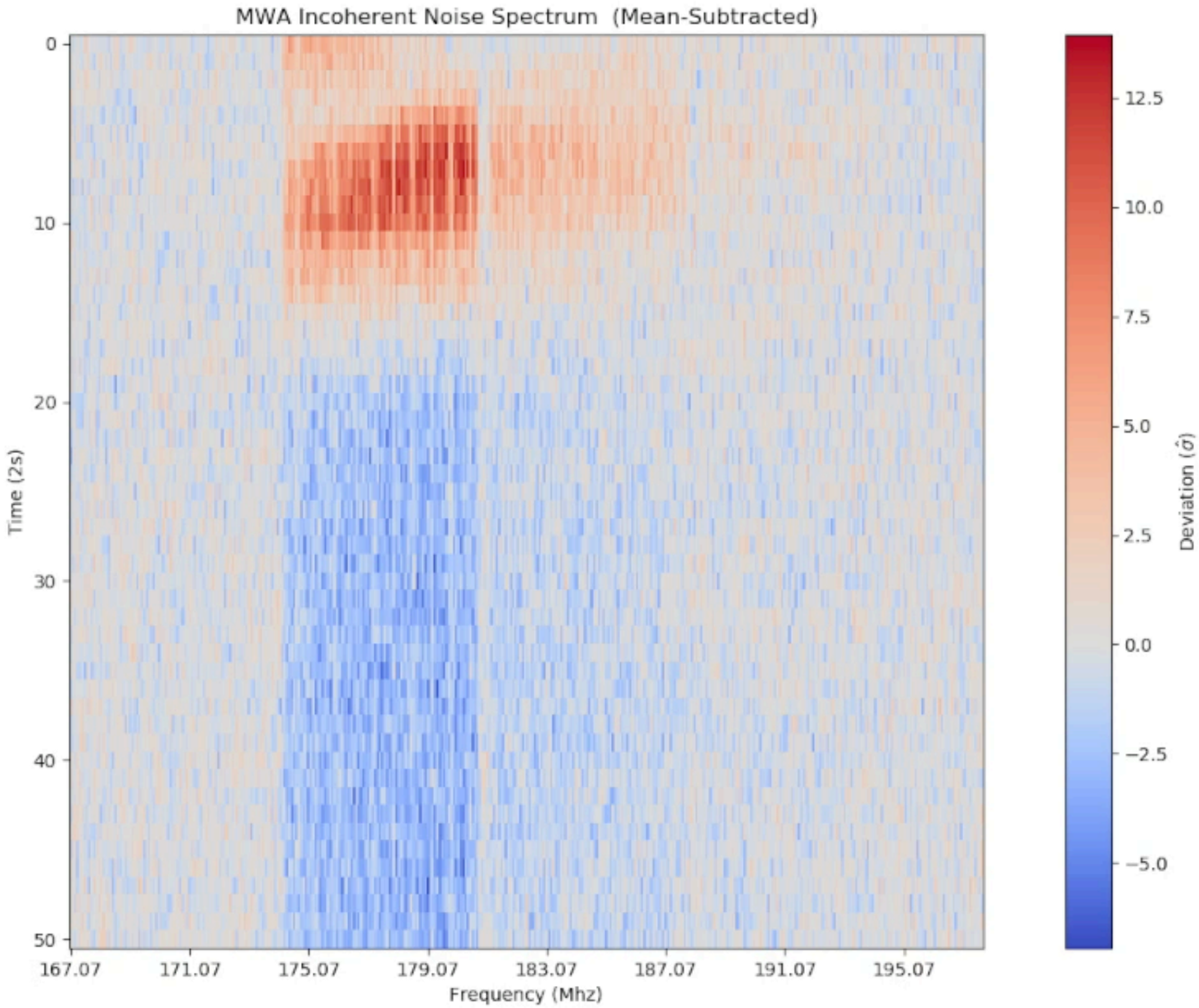
Australian digital TV bands are 7 MHz wide

RF Band	Australia				New Zealand			
	Channel No.	QAM Centre MHz	Vision Carrier MHz	FM Mono Audio Sub-Carrier MHz	Channel No.	QAM Centre MHz	Vision Carrier MHz	FM Mono Audio Sub-Carrier MHz
Band I	0	Reallocated	46.25	51.75	1	Reallocated	45.25	50.75
	1		57.25	62.75	2		55.25	60.75
	2		64.25	69.75	3		62.25	67.75
Band II	3	FM Stereo (1975)	86.25	91.75		FM Stereo		
	4		95.25	100.75				
	5		102.25	107.75				
	5A	Reallocated	138.25	143.75				
Band III	6	177.5	175.25	180.75	4	Reallocated	175.25	180.75
	7	184.5	182.25	187.75	5		182.25	187.75
	8	191.5	189.25	194.75	6		189.25	194.75
	9	198.5	196.25	201.75	7		196.25	201.75
	9A	205.5	203.25	208.75	8		203.25	208.75
	10	212.5	210.25	215.75	9		210.25	215.75
	11[1]	219.5	217.25	222.75	10		217.25	222.75
	12[1]	226.5	224.25	229.75	11		224.25	229.75

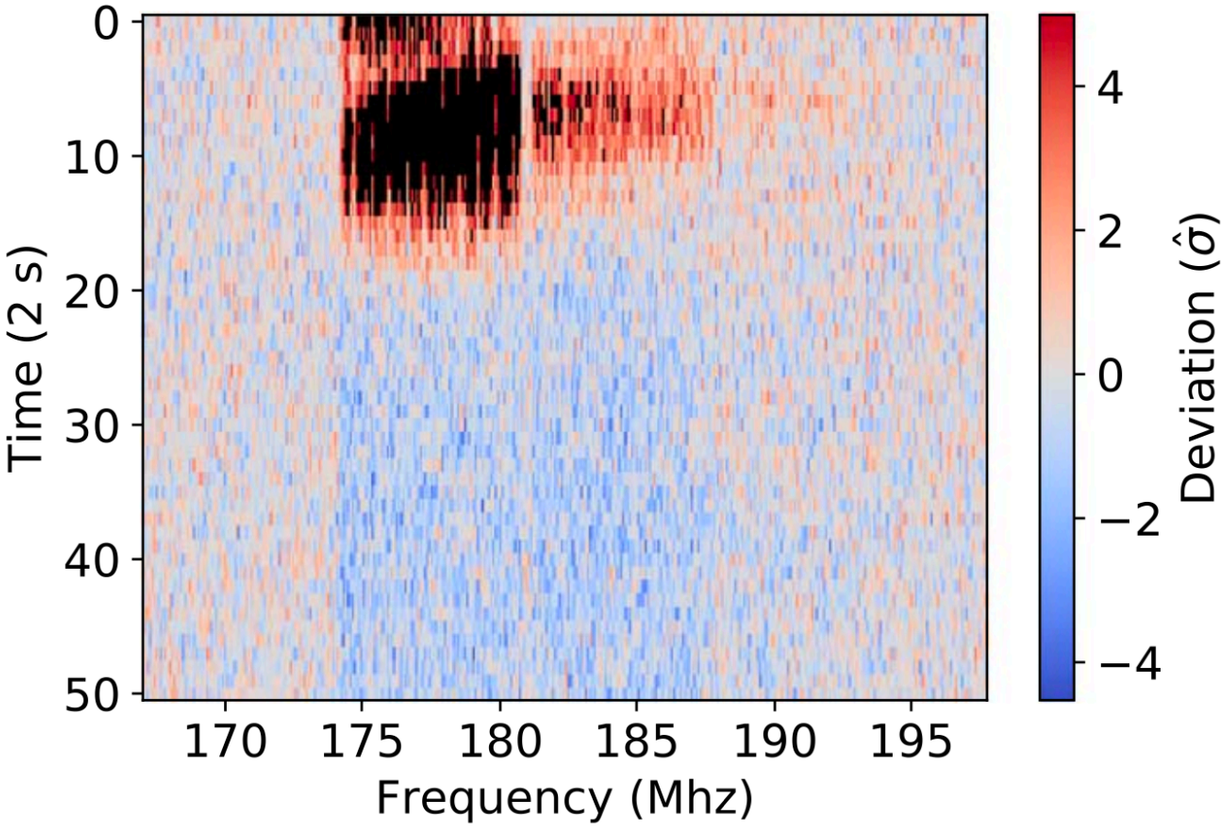


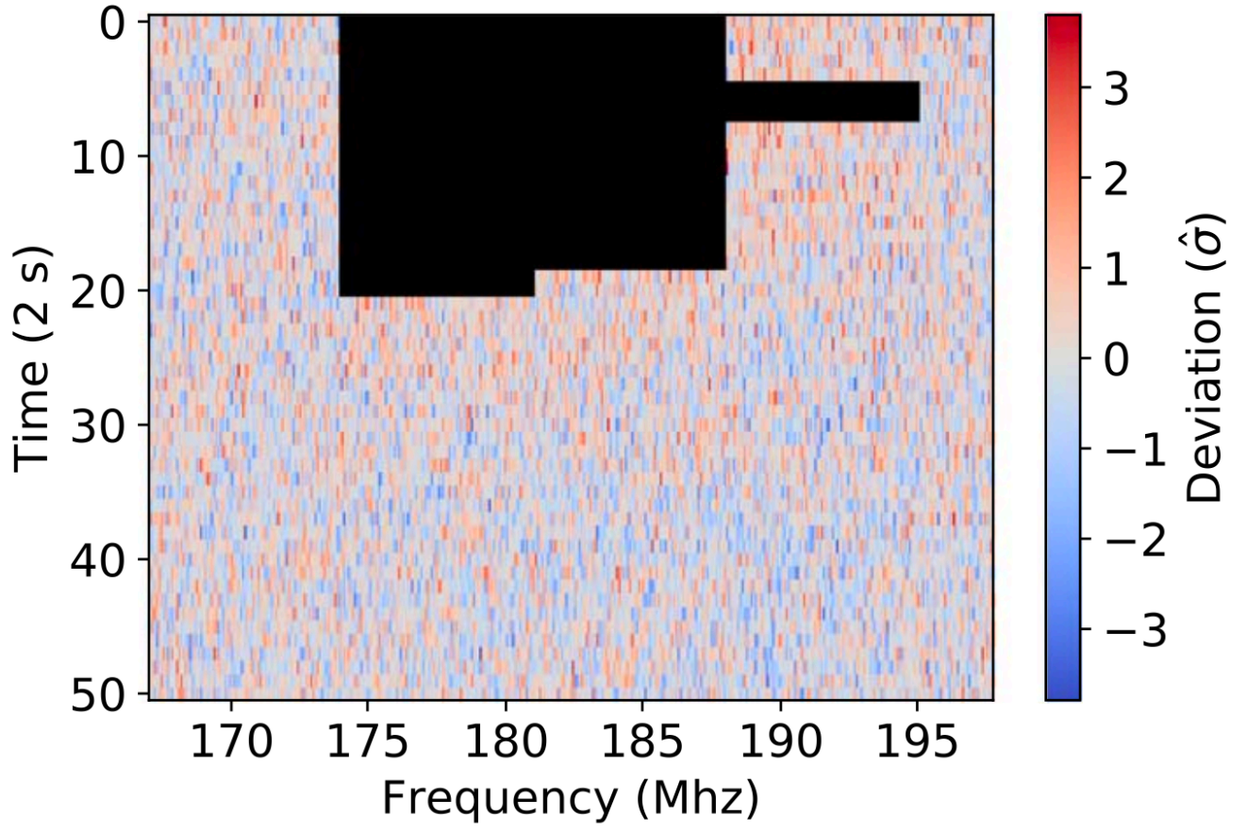
{'TV6': [1.74e8, 1.81e8], 'TV7': [1.81e8, 1.88e8], 'TV8': [1.88e8, 1.95e8], 'TV9': [1.95e8, 2.02e8]}

SSINS

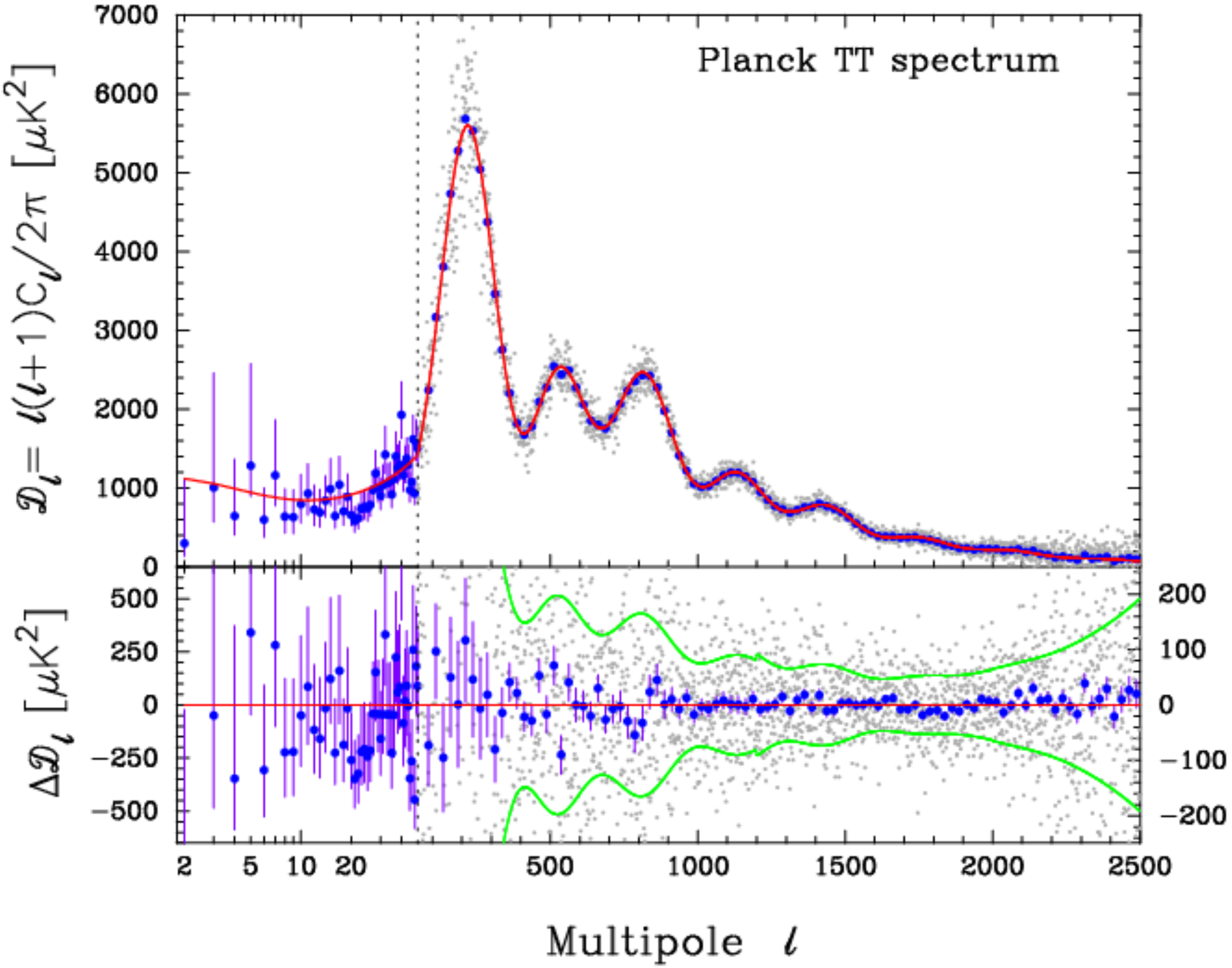

MWA Incoherent Noise Spectrum (Mean-Subtracted)
Time (2s)
0
10
20
30
40
50
167.07
171.07
175.07
179.07
183.07
187.07
191.07
195.07
Frequency (Mhz)
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
Deviation ($\hat{\sigma}$)

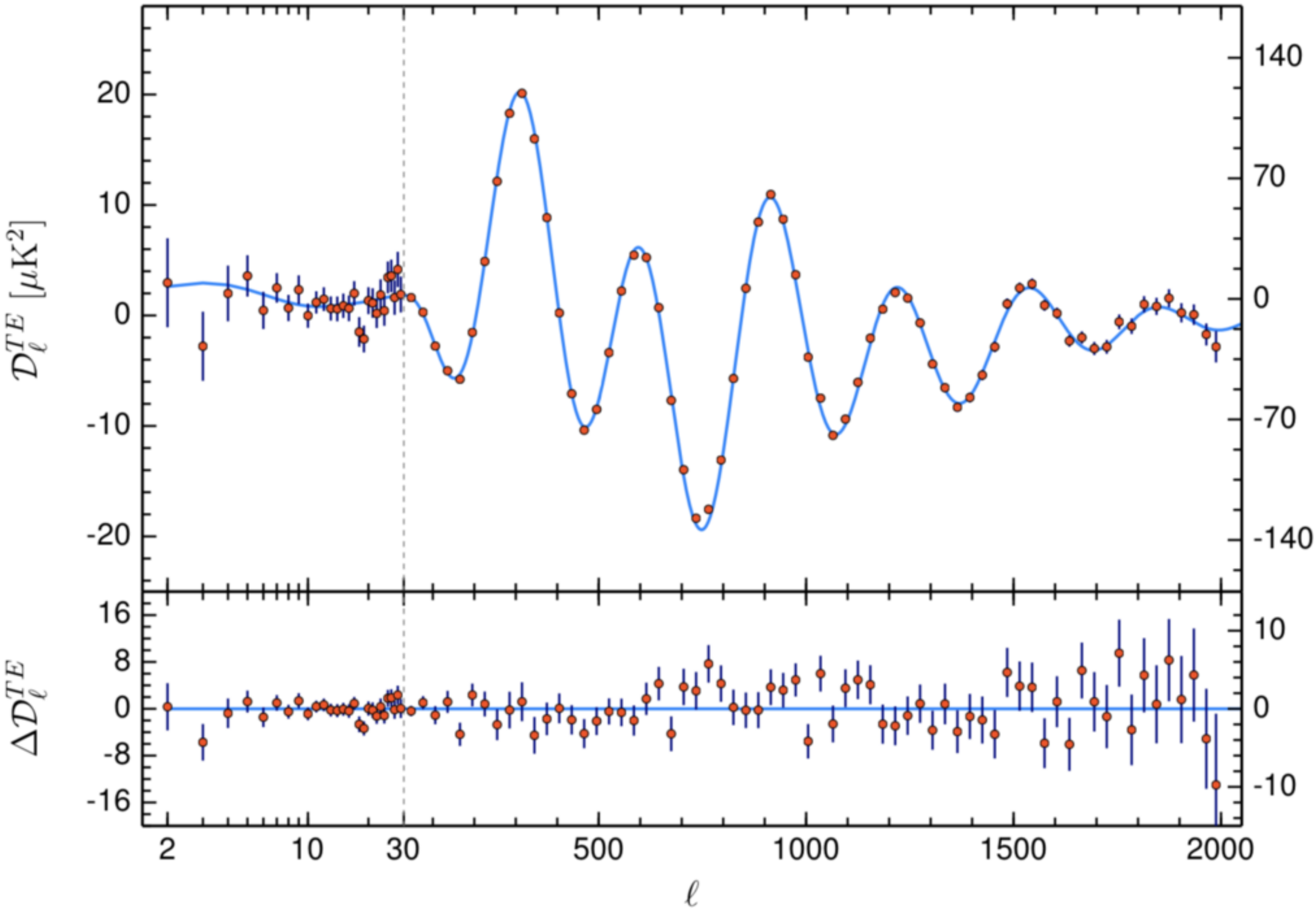
SSINS


0
10
20
30
40
50
Time (2 s)
170
175
180
185
190
195
Frequency (Mhz)
4
2
0
−2
−4
Deviation ($\hat{\sigma}$)


0
10
20
30
40
50
Time (2 s)
170
175
180
185
190
195
Frequency (Mhz)
3
2
1
0
−1
−2
−3
Deviation ($\hat{\sigma}$)

Fit or theory residuals


Planck TT spectrum
$\mathcal{D}_\ell = \ell(\ell+1)C_\ell/2\pi\ [\mu K^2]$
$\Delta\mathcal{D}_\ell\ [\mu K^2]$
Multipole ℓ
7000
6000
5000
4000
3000
2000
1000
0
500
250
0
−250
−500
200
100
0
−100
−200
2
5
10
20
500
1000
1500
2000
2500


$\mathcal{D}_\ell^{TE}$ [μK^2]
20
10
0
-10
-20
140
70
0
-70
-140
$\Delta\mathcal{D}_\ell^{TE}$
16
8
0
-8
-16
10
0
-10
2
10
30
500
1000
1500
2000
ℓ

Jackknife tests

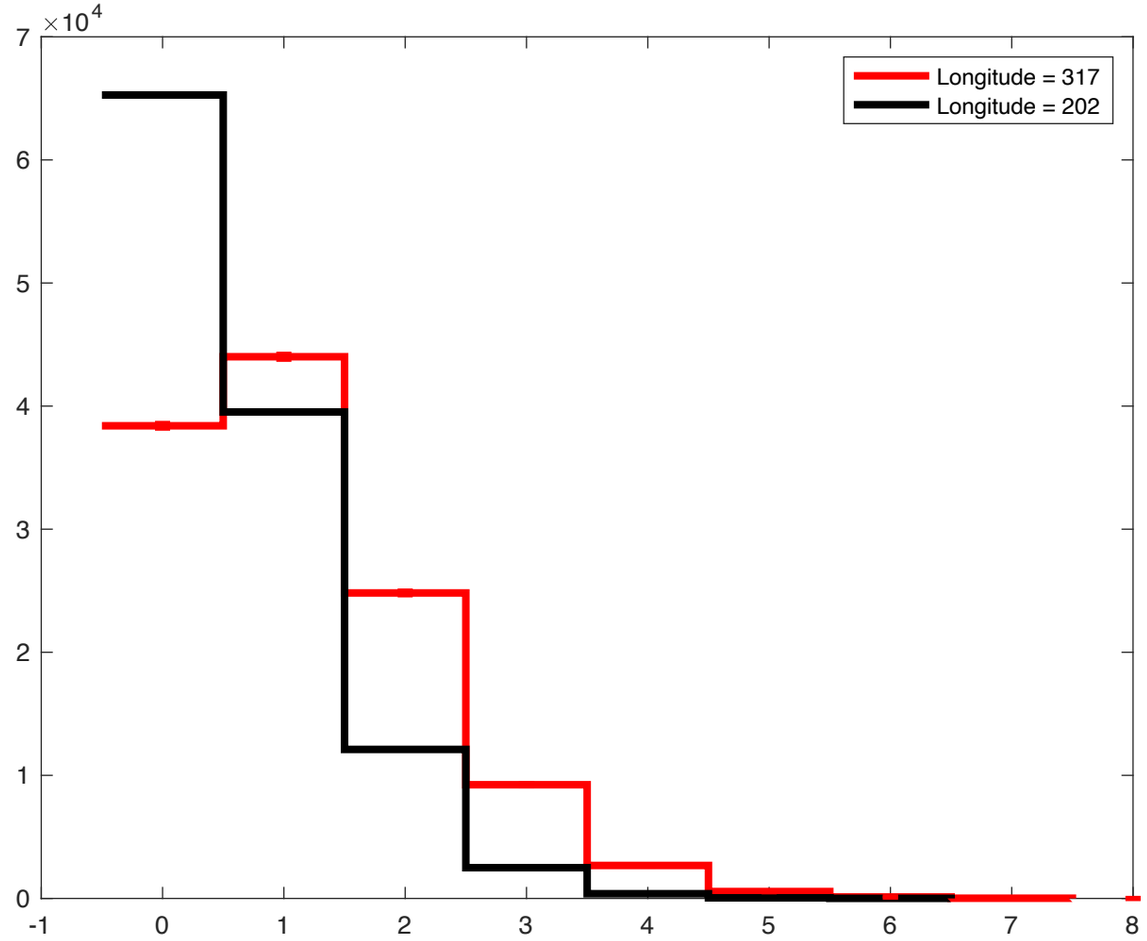
Internal consistency

Classic jackknife tests

- Identify a concern
- Split the data into two based on the concern
- Analyze the data sets separately, compare and difference
- A statistically significant difference means you have a problem, no difference means you can check off the worry
- More sensitive than “scavenger hunts” because matched to the worry and fewer trials factors.

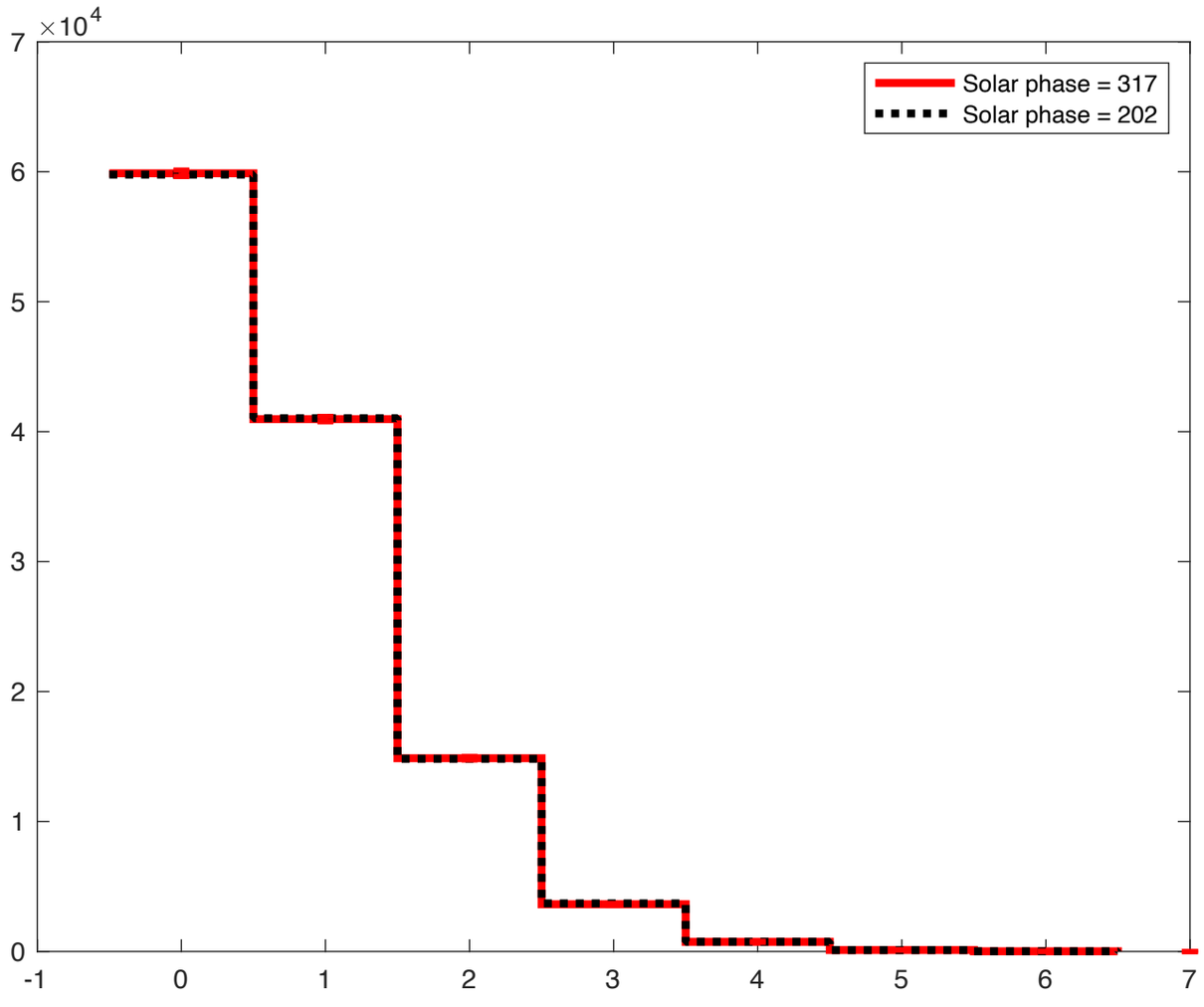
Compare Histograms vs. Longitude

```
figure()
[h1Data h1edges] = histcounts(data(Ephase>315 & Ephase<320));
[h2Data h2edges] = histcounts(data(Ephase>200 & Ephase<205));

stairs(h1edges(1:end-1),h1Data,'LineWidth',3,"Color",'r')
hold on
stairs(h2edges(1:end-1),h2Data,'LineWidth',3,"Color",'k')
shift = (h1edges(2)-h1edges(1))/2;
errorbar(h1edges(1:end-1)+shift,h1Data, sqrt(h1Data), 'r', 'linestyle','none','LineWidth',2)
legend('Longitude = 317', 'Longitude = 202')
```



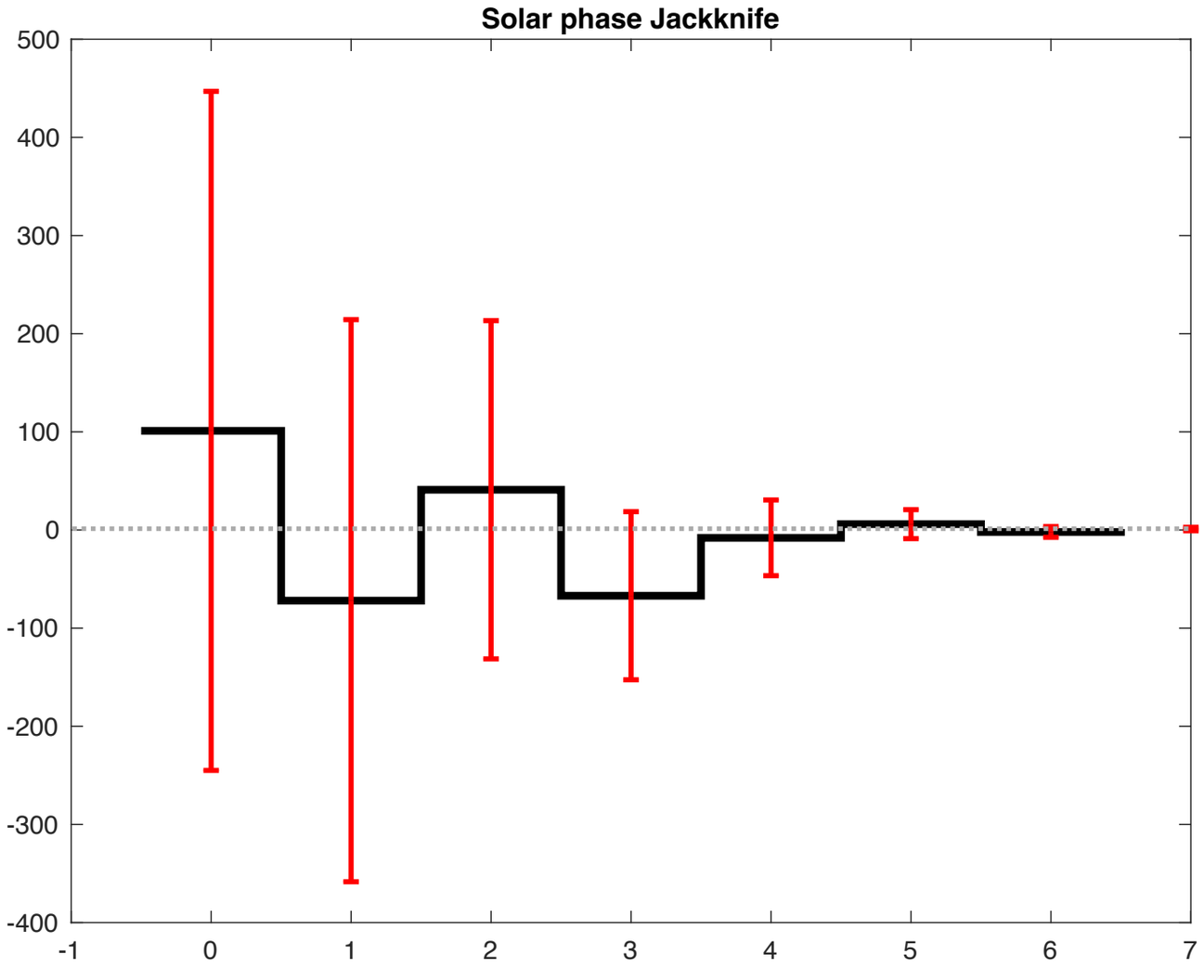
A closer comparison

```
[h1Data h1edges] = histcounts(data(Sphase>315 & Sphase<320));
[h2Data h2edges] = histcounts(data(Sphase>200 & Sphase<205));

stairs(h1edges(1:end-1),h1Data,'LineWidth',3,"Color",'r')
hold on
stairs(h2edges(1:end-1),h2Data,'LineWidth',3,"Color",'k',"LineStyle",":")
shift = (h1edges(2)-h1edges(1))/2;
errorbar(h1edges(1:end-1)+shift,h1Data, sqrt(h1Data), 'r', 'linestyle','none','LineWidth',2)
legend('Solar phase = 317', 'Solar phase = 202')
```



Jackknife test

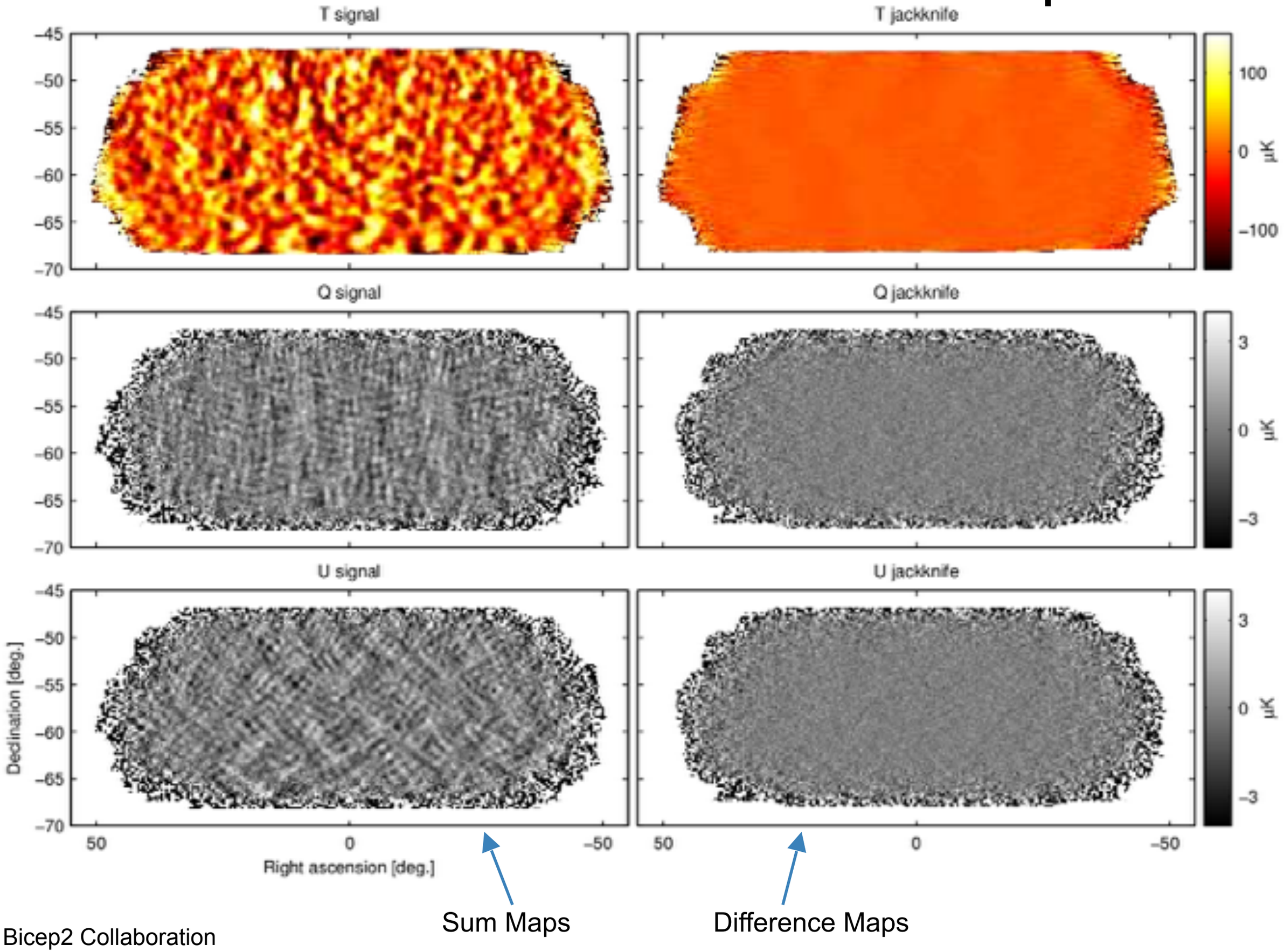
- A - B
- Errors are of the subtraction, not A *or* B



South Pole CMB experiments

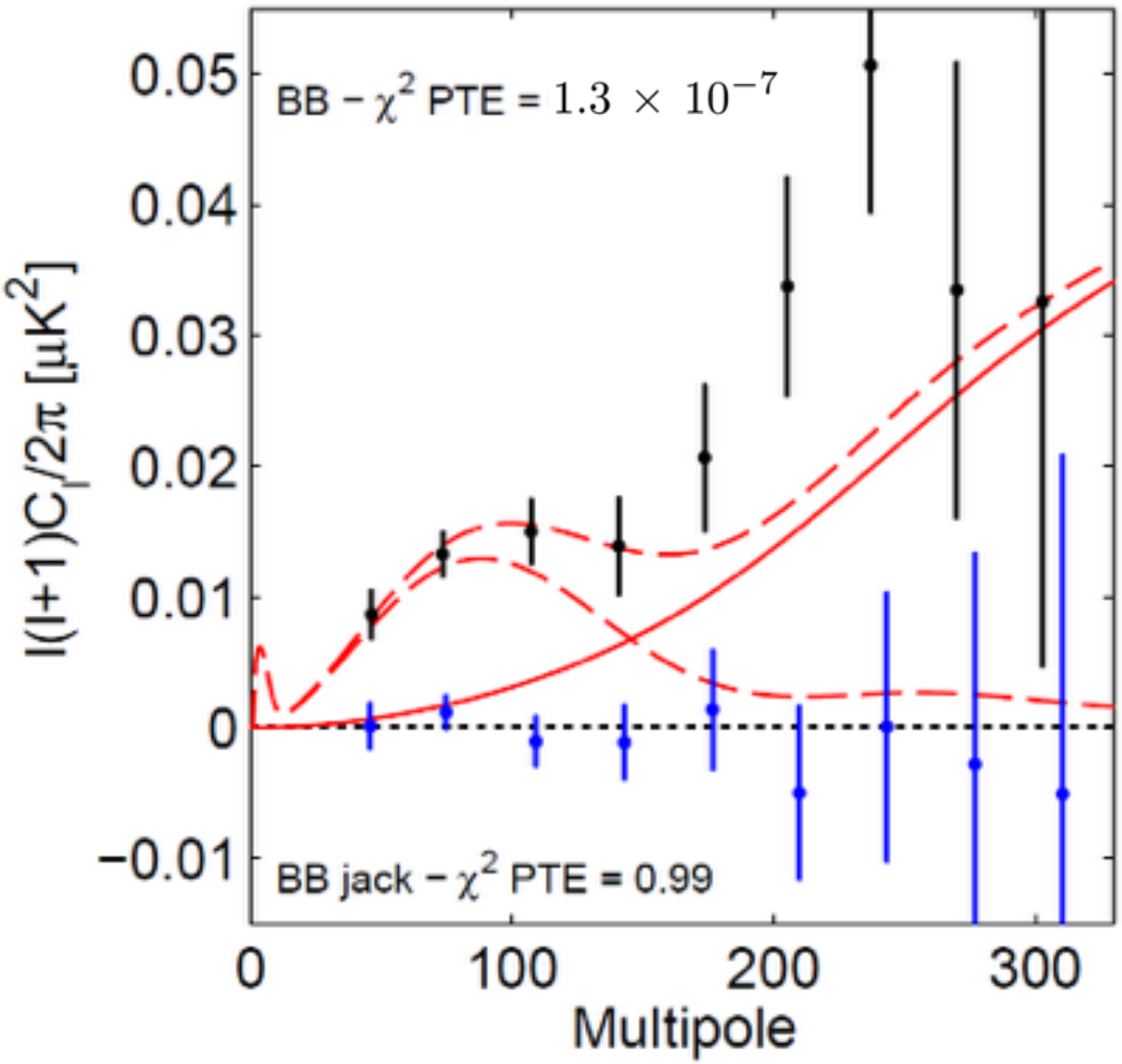
- Worries: differences due to
  - seasonality
  - humidity
  - atmospheric temperature
  - coolant levels
  - how far the sun is below the horizon
  - moon altitude
  - …

BICEP2 T and Stokes Q/U Maps



The Bicep2 Collaboration

BICEP2 B-mode Power Spectrum



B-mode power spectrum
temporal split jackknife
lensed-ΛCDM
r=0.2

B-mode power spectrum estimated directly from Q&U maps, including map based “purification” to avoid E→B mixing

Consistent with lensing expectation at higher l.

At low l excess over lensed-ΛCDM with high signal-to-noise.

For the hypothesis that the measured band powers come from lensed-ΛCDM we find:

χ^2 PTE	1.3×10^{-7}
significance	$5.3\,\sigma$

The Bicep2 Collaboration

South Pole CMB experiments

- Jackknife failure
  - thermal pickup from the south pole station ~100m away
  - compared data when telescope was looking over station vs not. Did not match!
- Response:
  - Throw out data when looking over south pole station
  - Later telescopes built huge shields to protect telescopes from hot south pole station

Kinds of worries

- Some instrument systematic you are hunting for
- Background leakage
- Calibration or selection biases
- Analysis changes

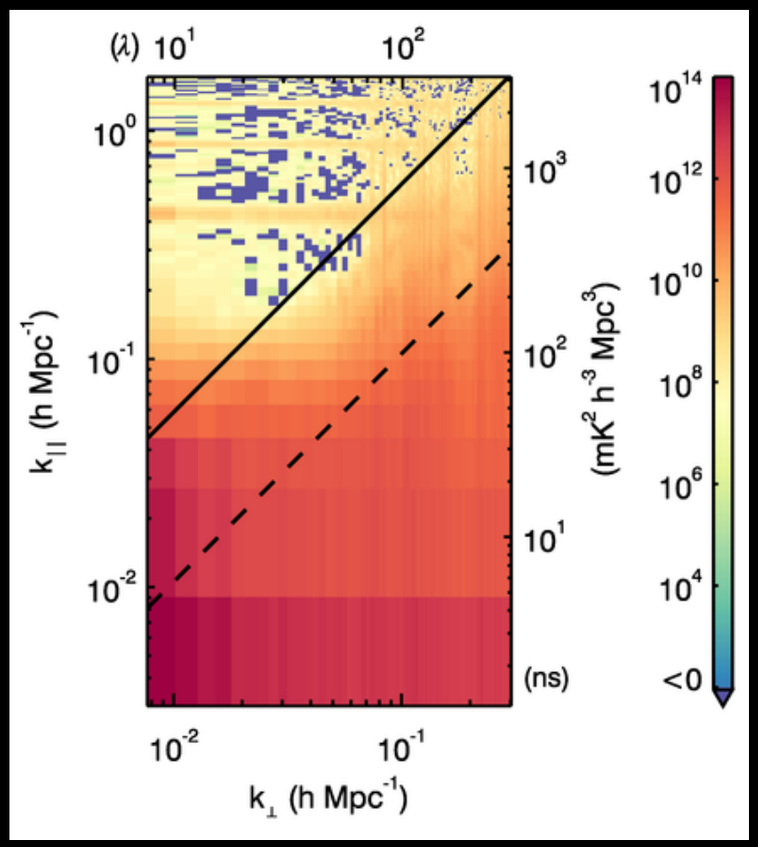
Analysis Jackknifes

Testing below the noise

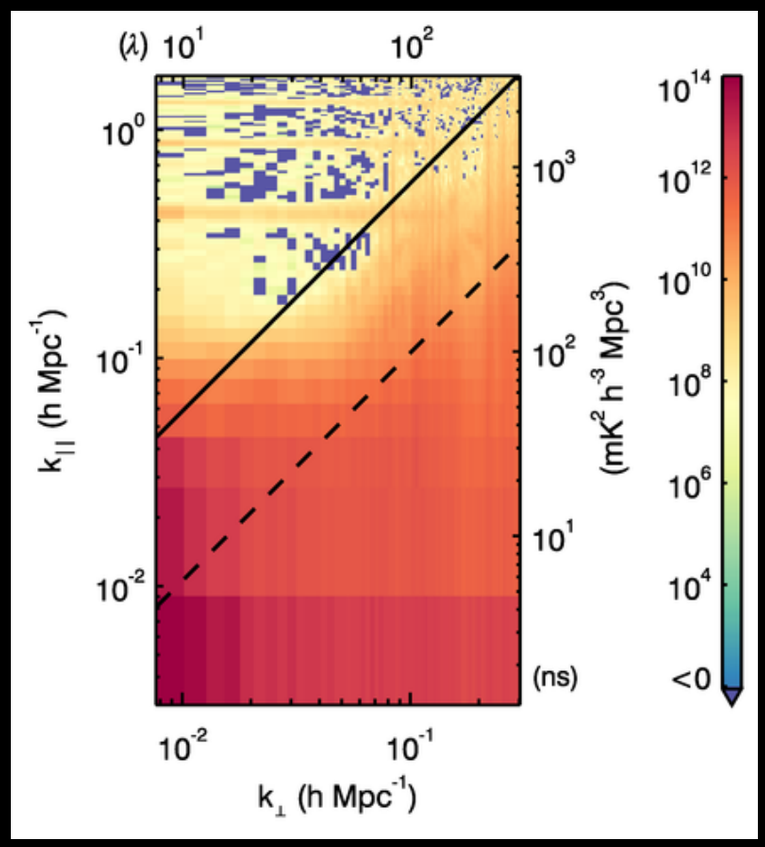
Analysis jackknife tests

- Use the same data run through two different versions of the analysis (or calibration)
- Compare and difference
- The noise realization is the same: can see differences below the level of the noise!

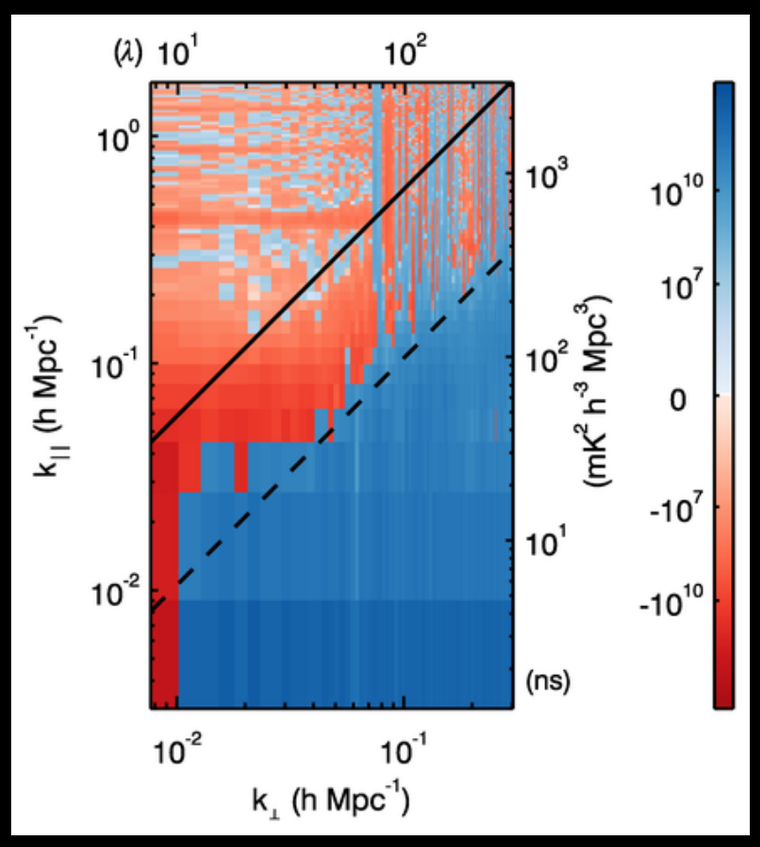
KATALOGSS



−



=

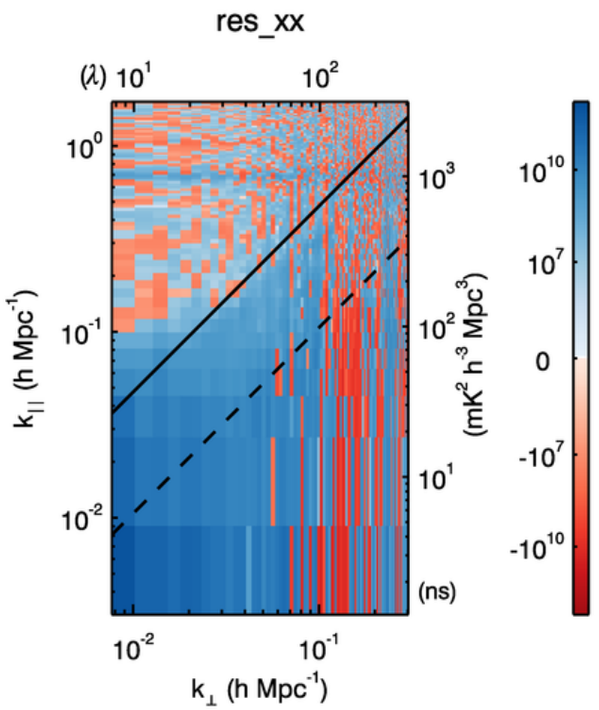


Carroll et al. 2016

Analysis jackknife

2 fhd_core/fhd_struct_init_antenna.pro View

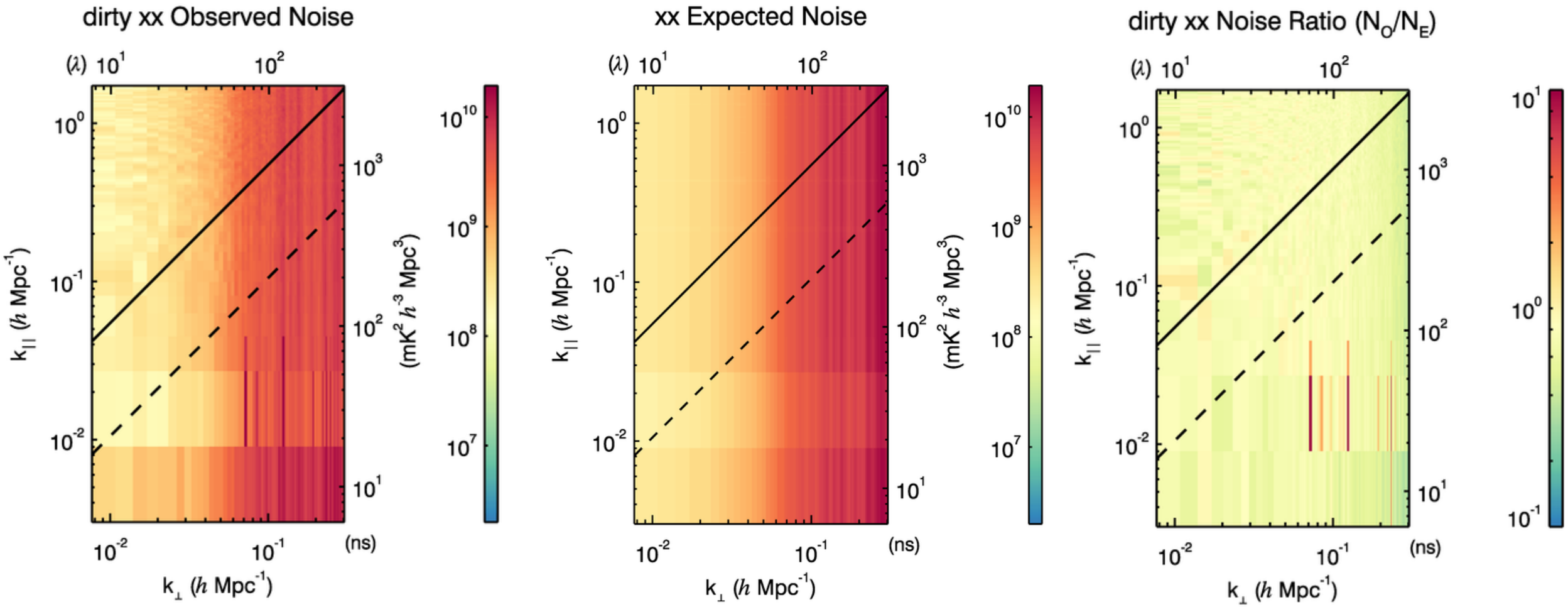
```
@@ -86,7 +86,7 @@ dec_use=dec_arr[valid_i]
 ;NOTE: Eq2Hor REQUIRES Jdate to have the same number of elements as RA and Dec for precession!!
 ;;NOTE: The NEW Eq2Hor REQUIRES Jdate to be a scalar! They created a new bug when they fixed the old one
-Eq2Hor,ra_use,dec_use,Jdate,alt_arr1,az_arr1,lat=obs.lat,lon=obs.lon,alt=obs.alt,precess=1
+Eq2Hor,ra_use,dec_use,Jdate,alt_arr1,az_arr1,lat=obs.lat,lon=obs.lon,alt=obs.alt,precess=1,/nutate
 za_arr=fltarr(psf_image_dim,psf_image_dim)+90. & za_arr[valid_i]=90.-alt_arr1
 az_arr=fltarr(psf_image_dim,psf_image_dim) & az_arr[valid_i]=az_arr1
```



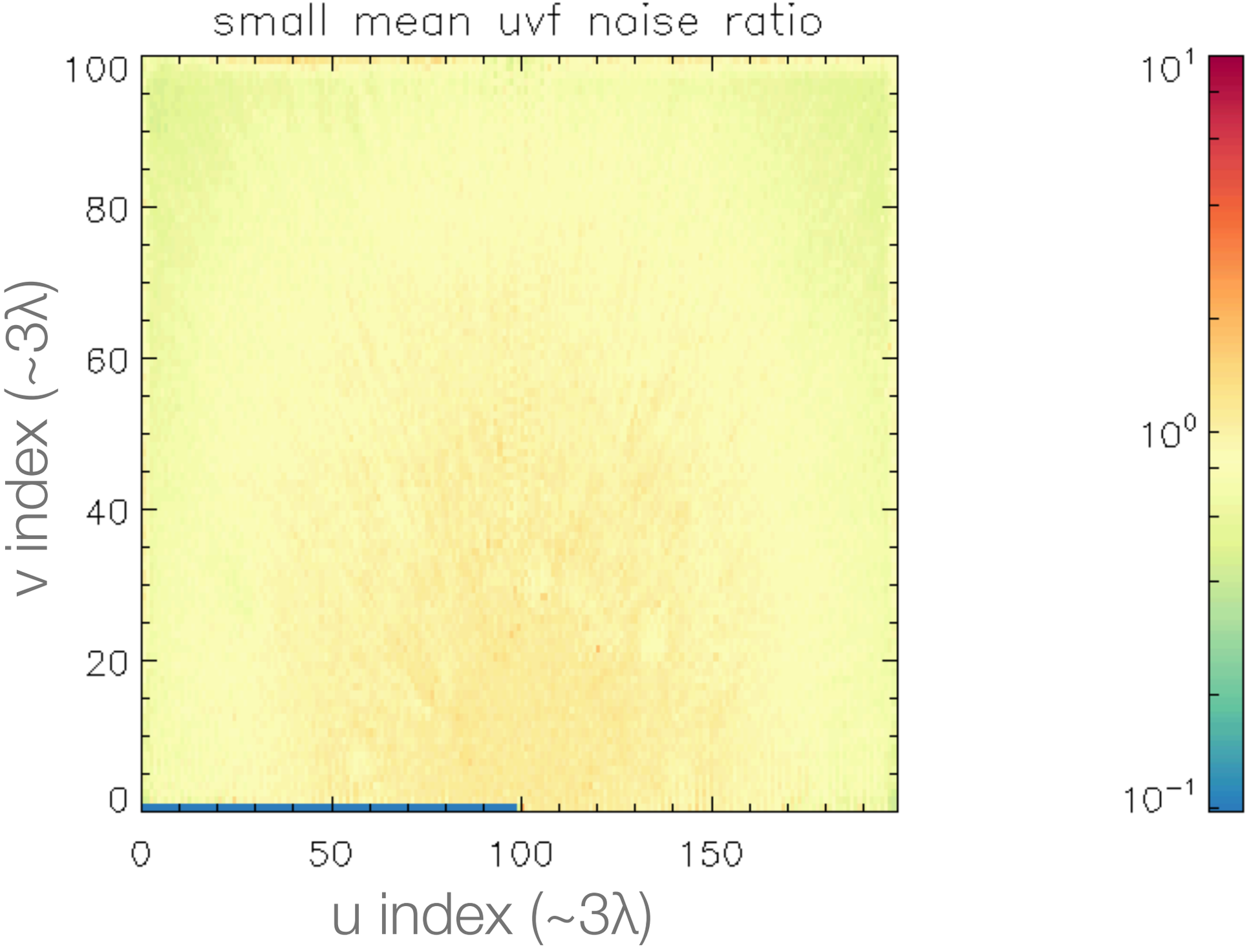
Background ratio plots

Checking your error bars

Ratio of background to expectation



unitless


small mean uvf noise ratio
v index (~3λ)
u index (~3λ)
100
80
60
40
20
0
0
50
100
150
10^{1}
10^{0}
10^{-1}

Statistic plots to answer questions

- If you have error bars, phrase question as
- Jackknife ($data_A$-$data_B$)
- Residual (data - model)
- z-score normalized data
- Ratios (variable backgrounds)
- Analysis jackknife ($analysis_A$ - $analysis_B$)