# Data Characterization

Sam Finn, Gabriela Gonzalez, Mijan Huq, Soumya Mohanty, Soma Mukherjee

LIGO-G990022-28-M

# Focus

- Artifact identification
  - Resolve data set as sum of instrumental artifacts & residual
- Characterization of Artifacts, Residual
  - (non-) Gaussianity
  - (non-) Stationarity
- Goal
  - Automated tools

# Artifact Identification

- Kalman Filtering
  - Estimate state of dynamical system from noisy observation
  - Operates in the time domain
- Components
  - State: $\vec{x}$
  - State Dynamics:

$$x[k+1] = A \cdot x[k] + N[k]$$

  - Observation:

$$y[k] = C \cdot x[k] + M[k]$$

# Example: Violin Modes

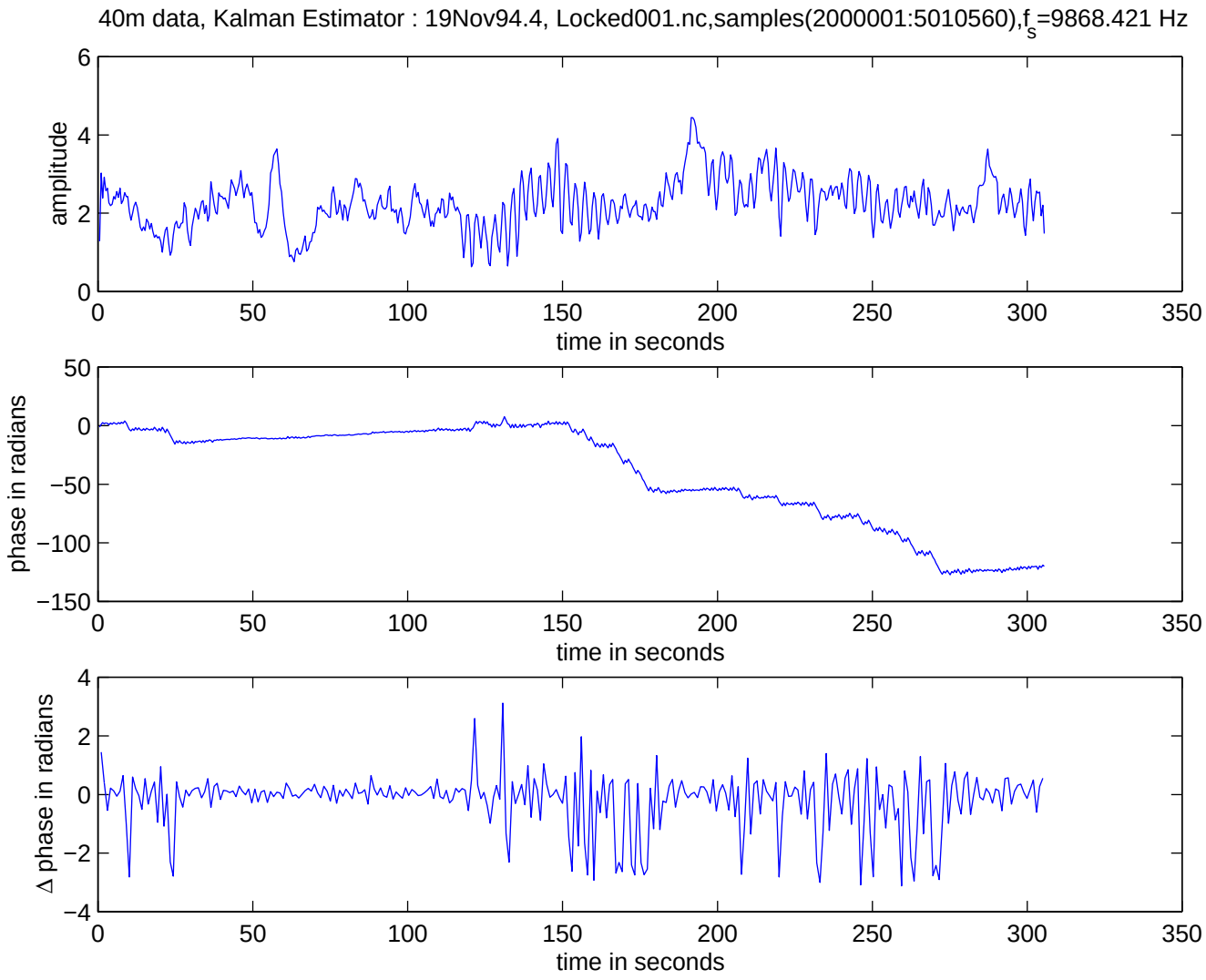
- Observation
  - (Bandwidth limited) noisy detector output
- System
  - Damped oscillator driven by white noise
- State
  - State: x, dx/dt

# Analysis Procedure

- QAM demodulate line to ~ 0 Hz
- Resample to low bandwidth
  - ~ 5 Hz
- Apply filter
  - Estimate line contribution to in-phase, quadrature components
- Resample *estimator* to full bandwidth
- QAM Estimator
- Form residual

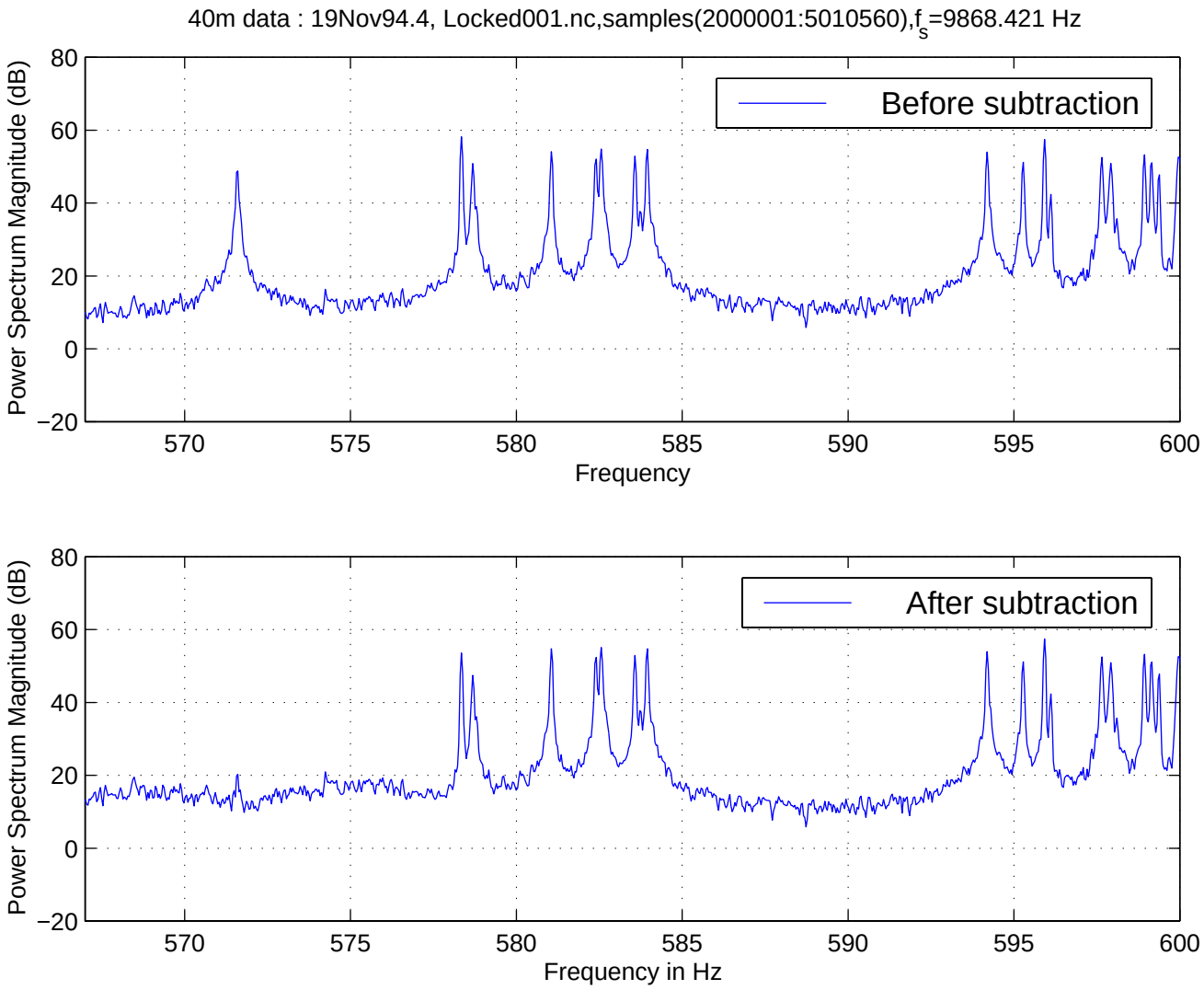
# Line Estimator

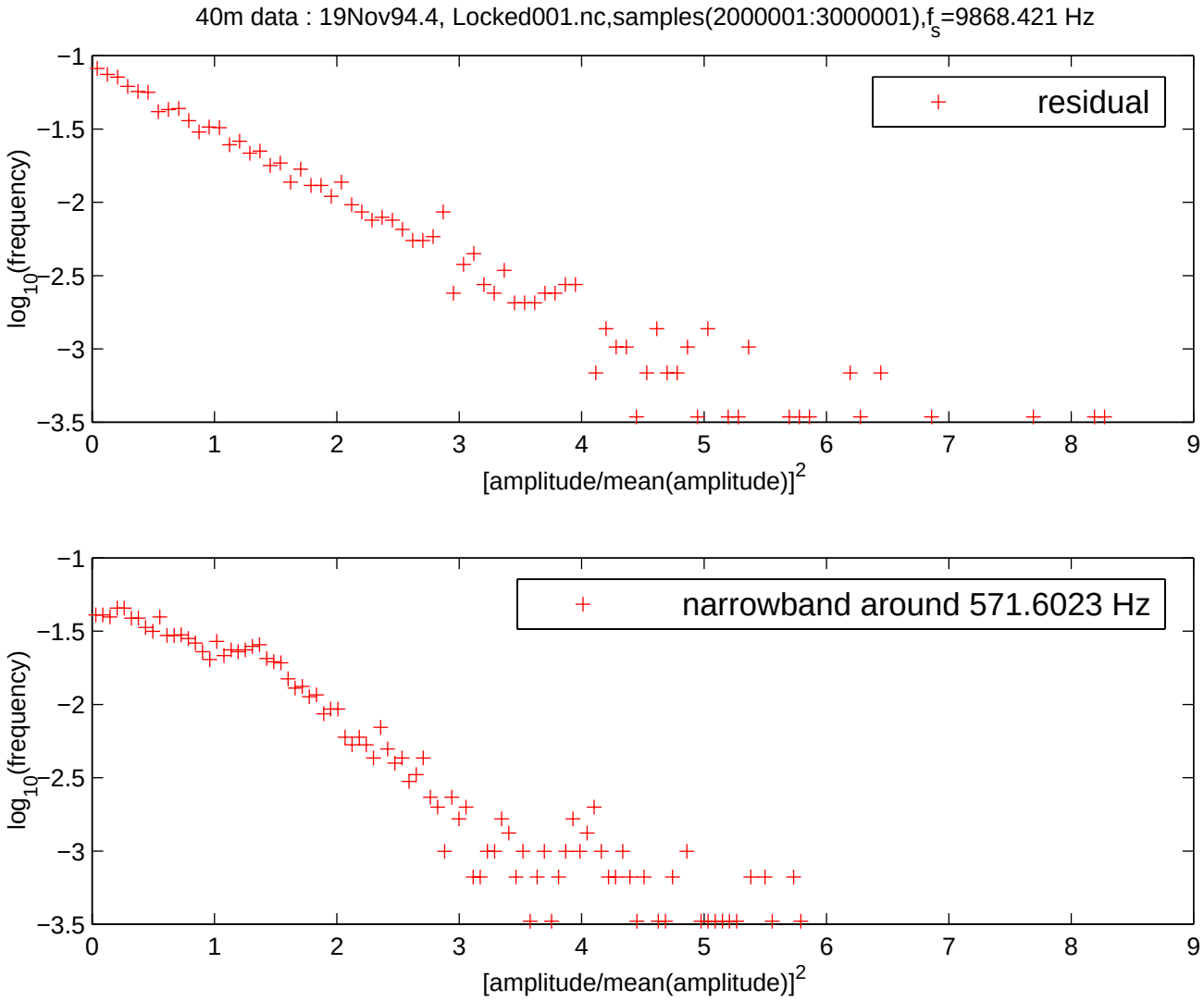
- Amplitude
  - Modulus $x+iy$
- Phase
  - Less $2\pi f_0 t$
- Instantaneous Frequency
  - Less $f_0$

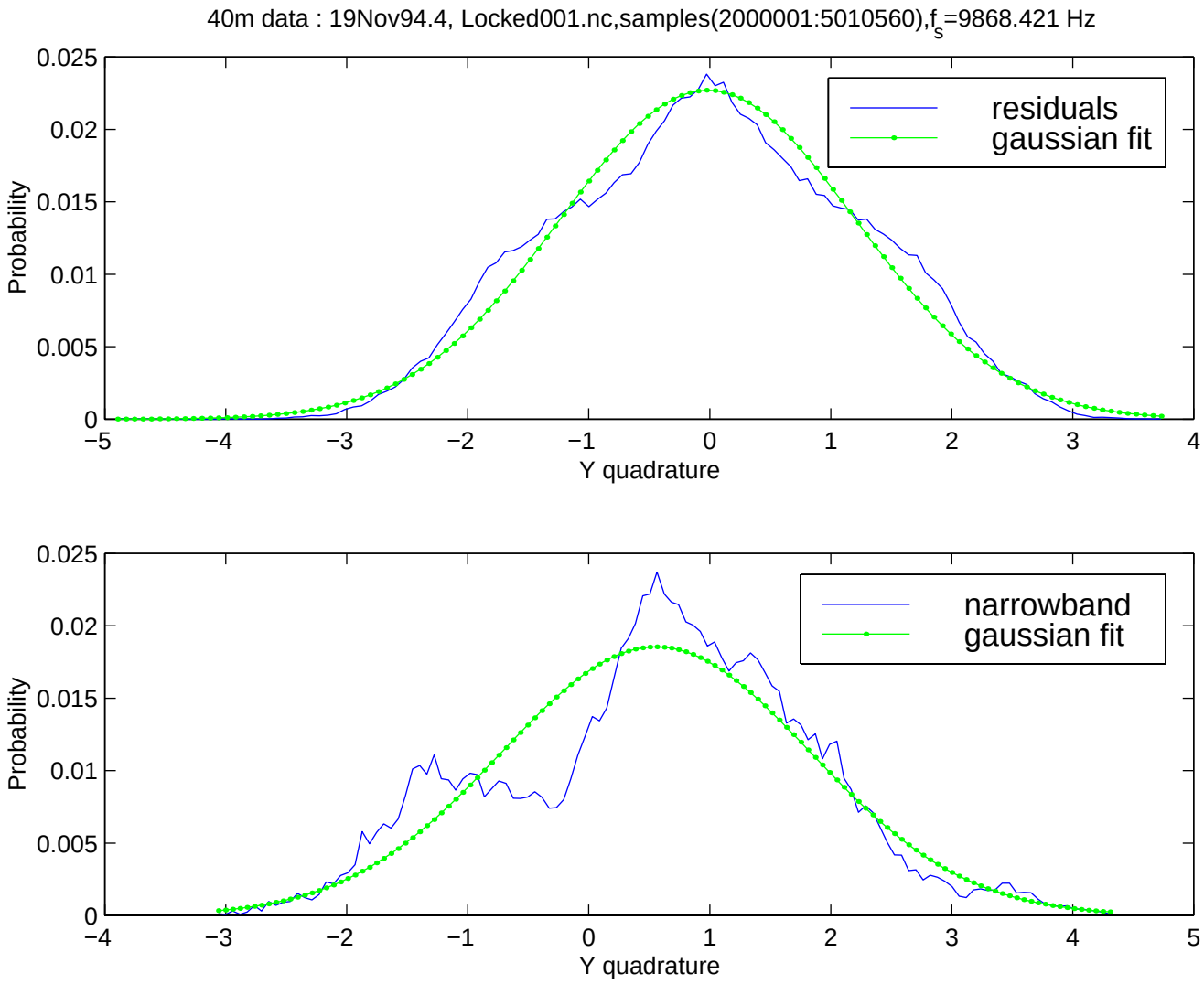

40m data, Kalman Estimator : 19Nov94.4, Locked001.nc,samples(2000001:5010560),f_s=9868.421 Hz
amplitude
time in seconds
phase in radians
time in seconds
Δ phase in radians
time in seconds

# Residual

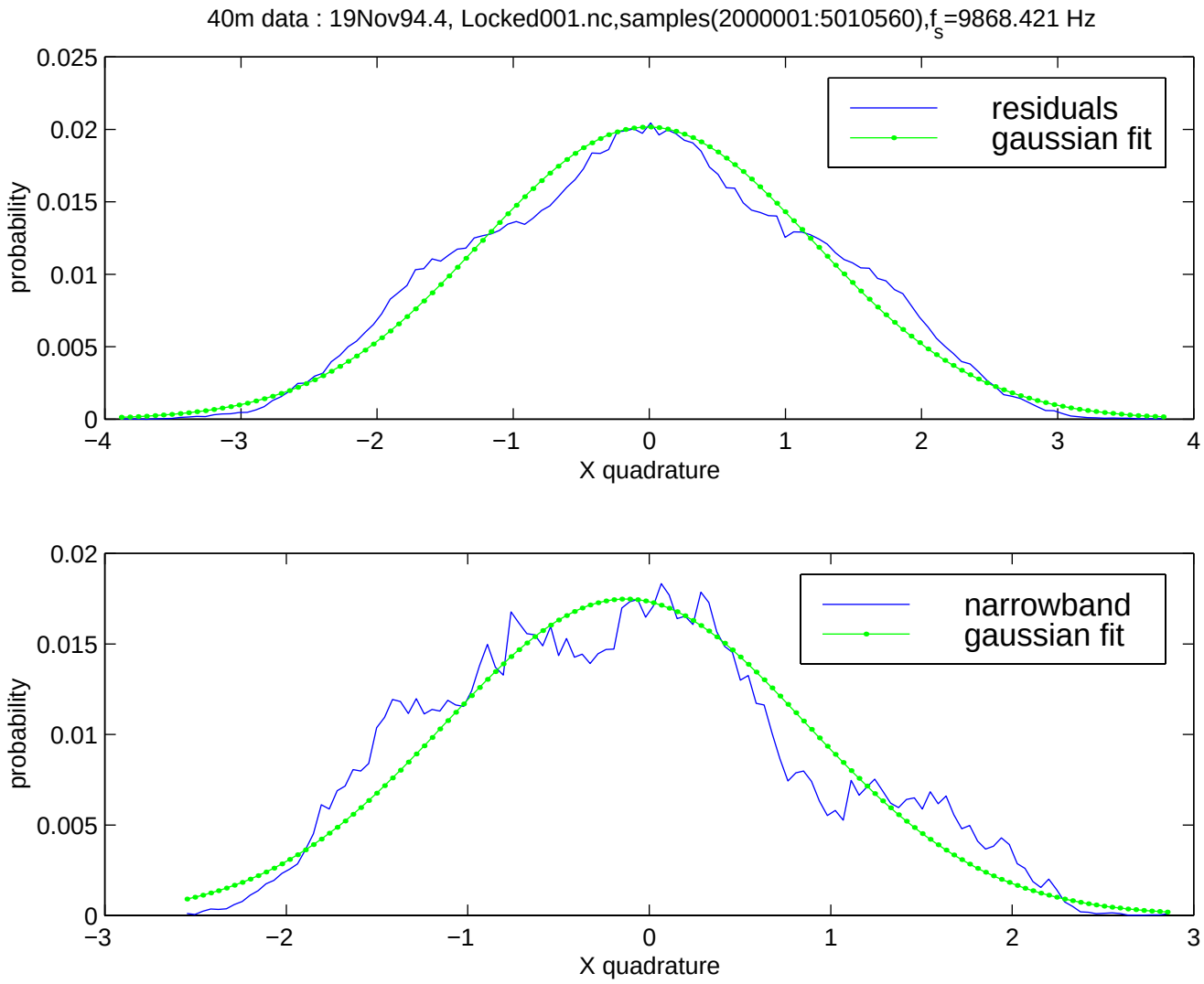
- Spectrum
- Distribution


40m data : 19Nov94.4, Locked001.nc,samples(2000001:5010560),f_s=9868.421 Hz
Before subtraction
Power Spectrum Magnitude (dB)
80
60
40
20
0
−20
570
575
580
585
590
595
600
Frequency
After subtraction
Power Spectrum Magnitude (dB)
80
60
40
20
0
−20
570
575
580
585
590
595
600
Frequency in Hz


40m data : 19Nov94.4, Locked001.nc,samples(2000001:3000001),f_s=9868.421 Hz
residual
log10(frequency)
[amplitude/mean(amplitude)]^2
narrowband around 571.6023 Hz
log10(frequency)
[amplitude/mean(amplitude)]^2


40m data : 19Nov94.4, Locked001.nc,samples(2000001:5010560),f_s=9868.421 Hz
residuals
gaussian fit
Probability
Y quadrature
0.025
0.02
0.015
0.01
0.005
0
−5
−4
−3
−2
−1
0
1
2
3
4


narrowband
gaussian fit
Probability
Y quadrature
0.025
0.02
0.015
0.01
0.005
0
−4
−3
−2
−1
0
1
2
3
4
5


40m data : 19Nov94.4, Locked001.nc,samples(2000001:5010560),f_s=9868.421 Hz
residuals
gaussian fit
probability
X quadrature
narrowband
gaussian fit
probability
X quadrature

# Time-frequency characterization

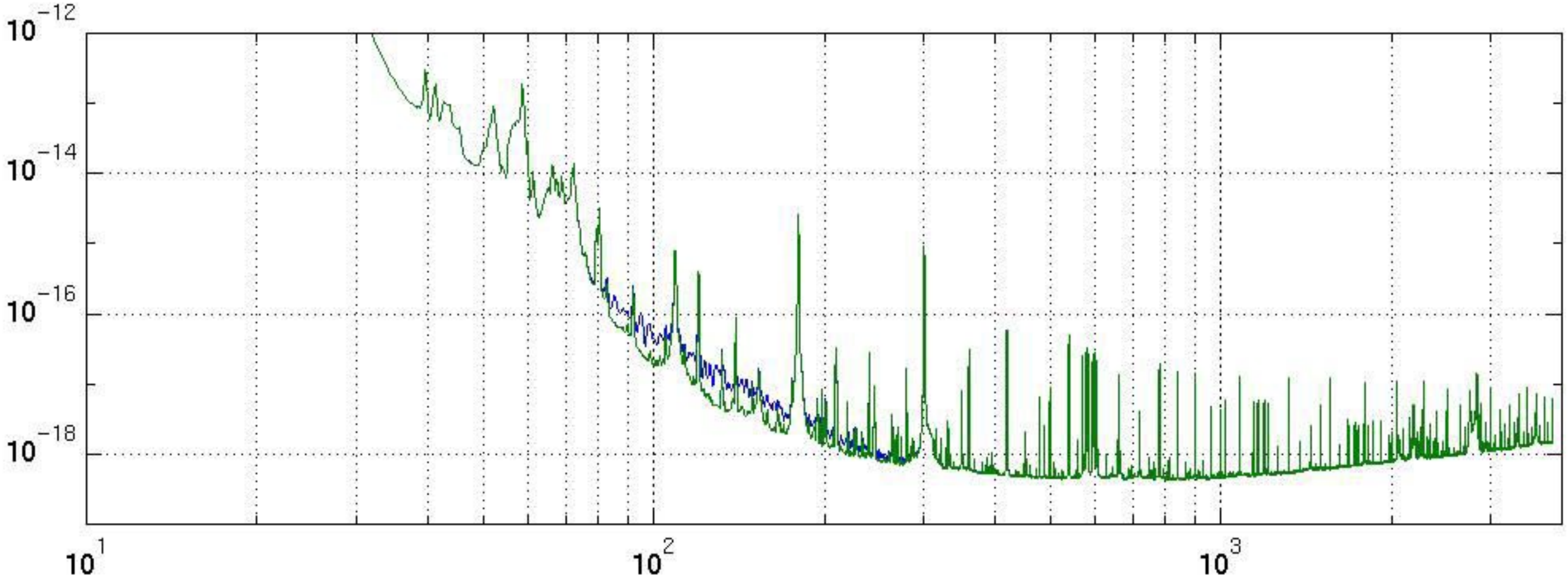
- Rayleigh Statistics

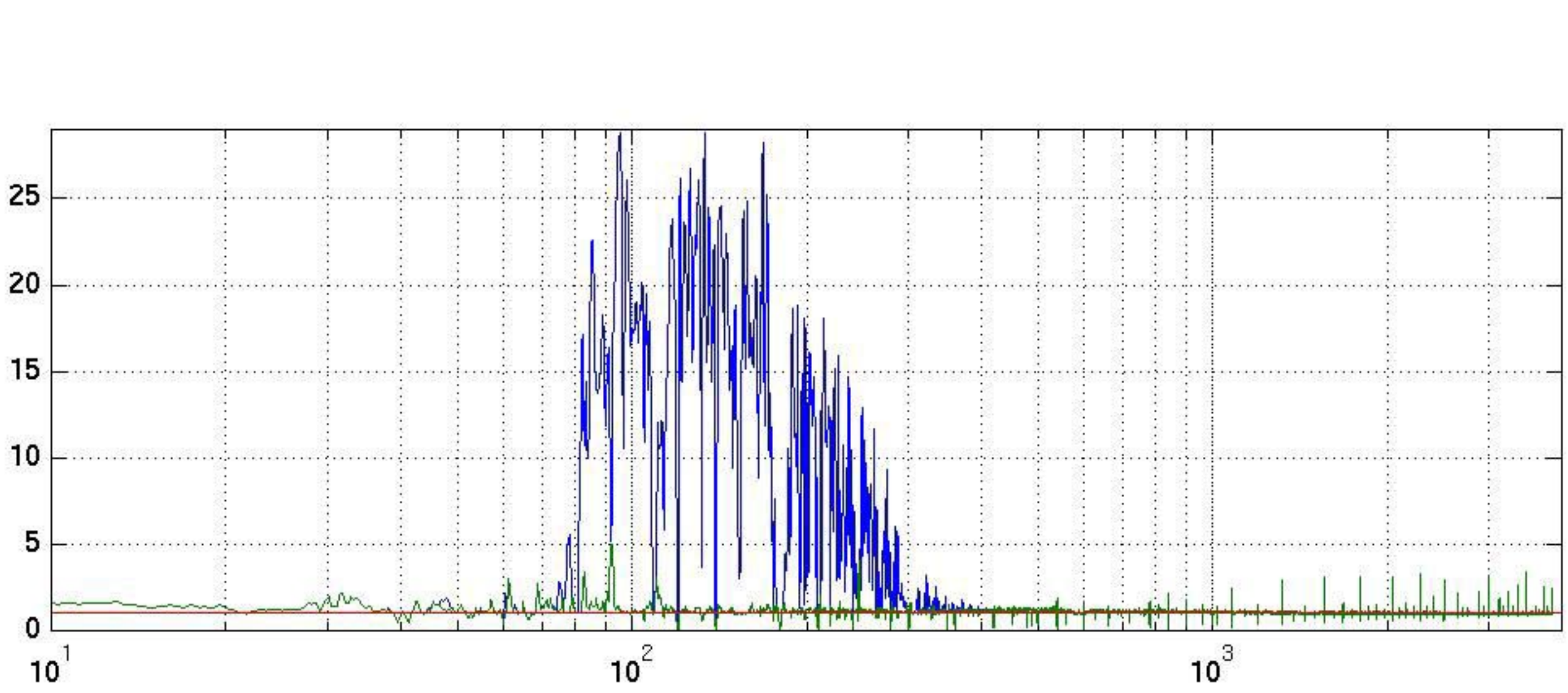
$$P(x) = e^{-x}, \quad \bar{x} = 1, \quad \sigma = 1$$

- Evaluate $r(f) = \overline{\sigma(f)} / \overline{x(f)}$
  - If larger than unity, distribution is too broad
    - E.g., excess noise
  - If smaller than unity, distribution is too peaked
    - E.g., sinusoid
- Focus on $r(f|t)$
  - Short-time Fourier Transforms

# Time-frequency characterization

- Pick frequency with greatest excess noise
  - Time history shows non-stationarity
  - Identify statistically significant excursions and mark as “bad”
  - Re-compute statistics without “bad” data segments


$10^{-12}$
$10^{-14}$
$10^{-16}$
$10^{-18}$
$10^{1}$
$10^{2}$
$10^{3}$


25
20
15
10
5
0
$10^{1}$
$10^{2}$
$10^{3}$

Time history of 133.414 Hz freq bin in 40m Psd

psd($m^2$/Hz)

$10^{0}$ $10^{-2}$ $10^{-4}$ $10^{-6}$ $10^{-8}$

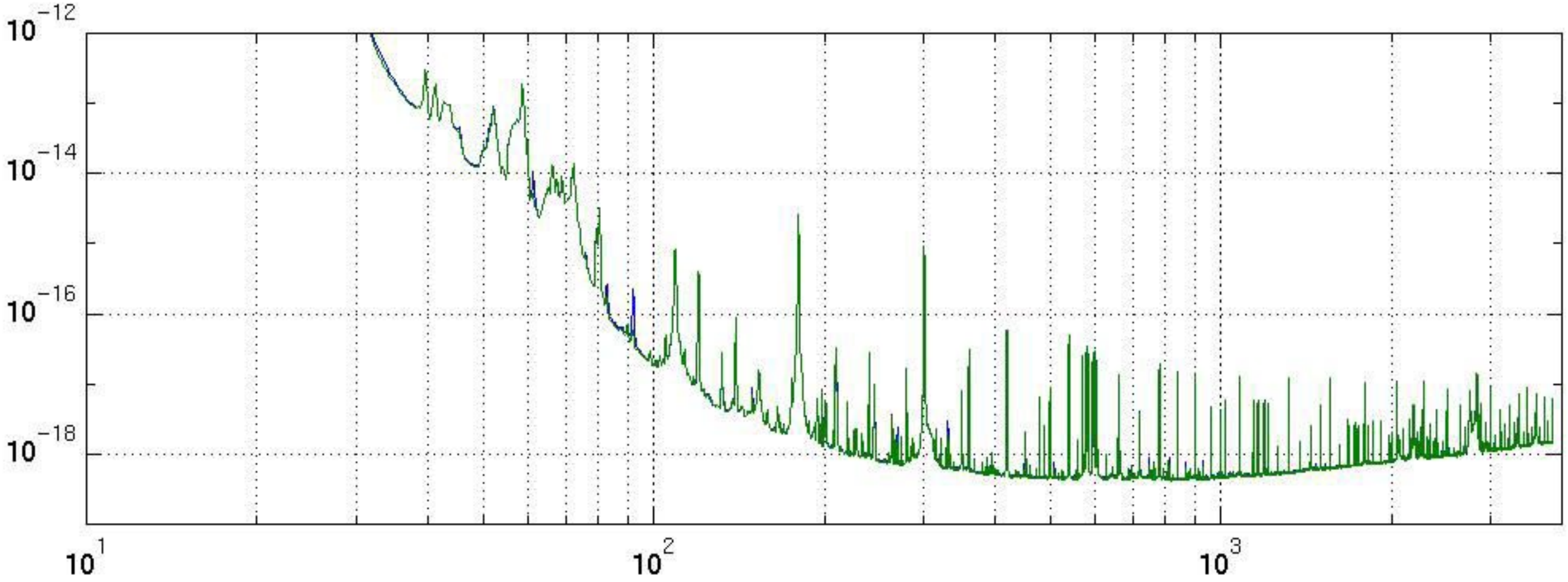
0 500 1000 1500 2000 2500

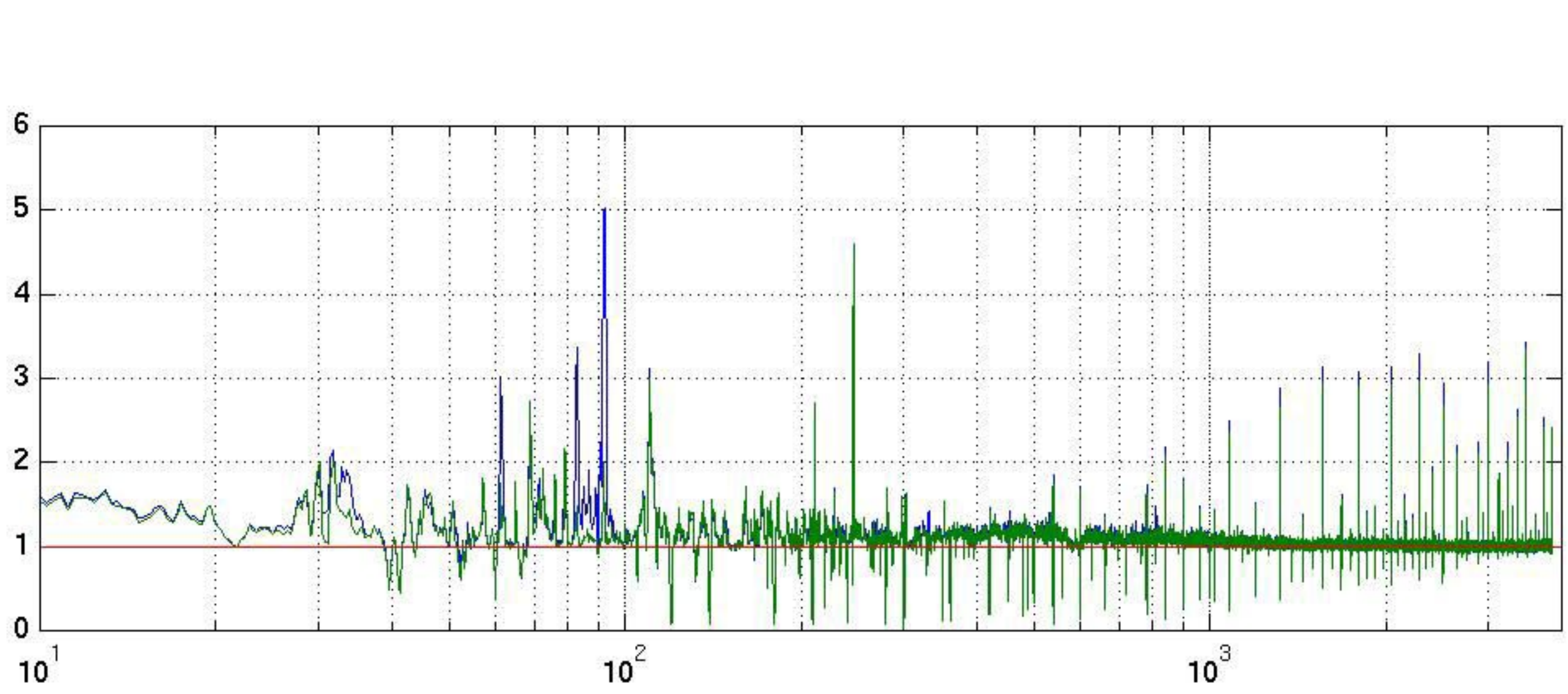
psd($m^2$/Hz)

$10^{0}$ $10^{-2}$ $10^{-4}$ $10^{-6}$ $10^{-8}$

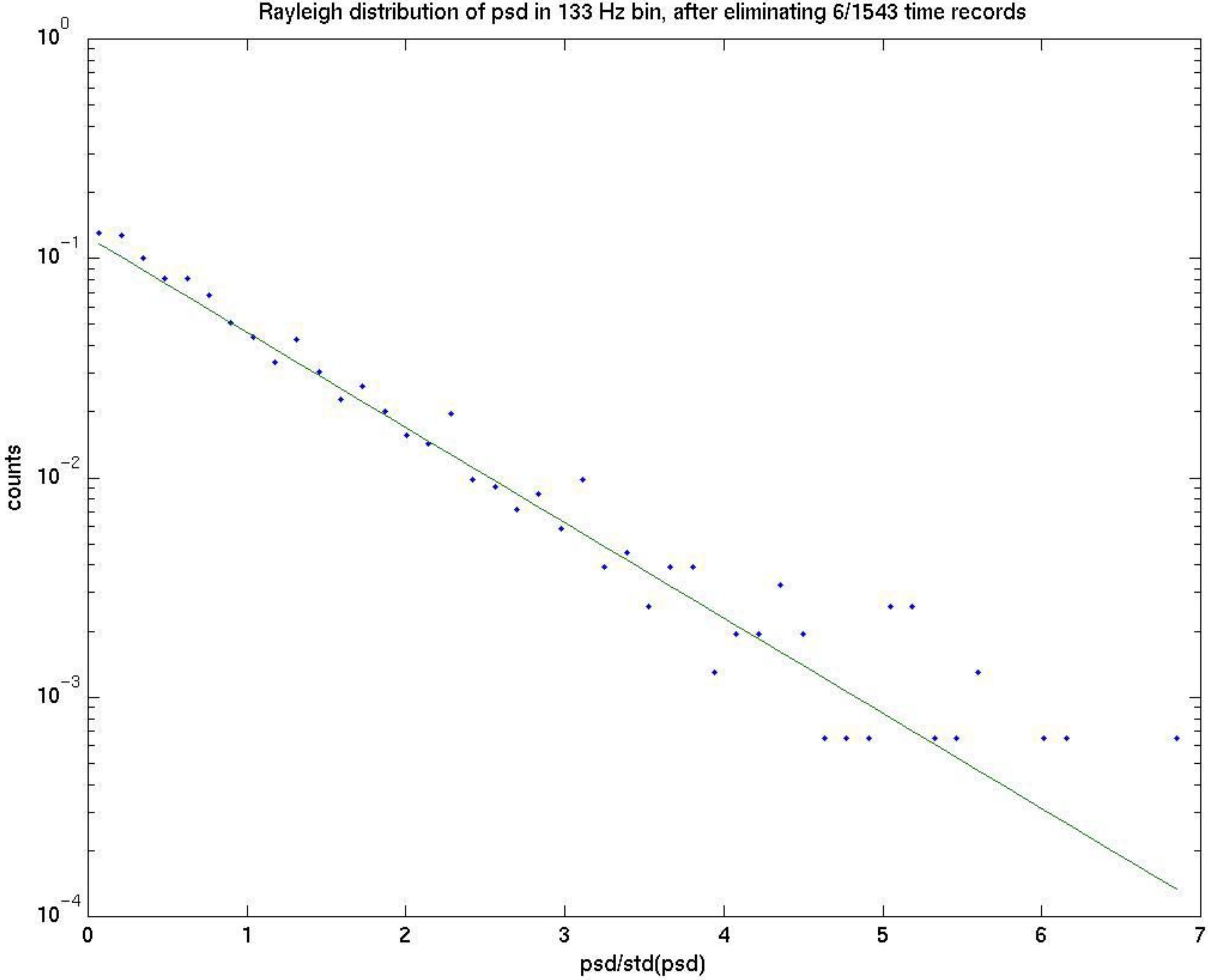
0 500 1000 1500 2000 2500

time(sec)


10^-12
10^-14
10^-16
10^-18
10^1
10^2
10^3


6
5
4
3
2
1
0
10^1
10^2
10^3

Rayleigh distribution of psd in 133 Hz bin, after eliminating 6/1543 time records



# Lessons Learned

- Instrumental artifacts can be identified
  - Resolve data into known artifacts and residuals
  - Artifacts tend to be poorly-behaved, residuals well-behaved
- Bad instrument" behavior is confined to short stretches
  - Which can be identified through character in narrow band
- Search for gravitational-waves in residual

# On-going work

- Automated tools
  - Kalman filter
  - Data quality “bit”
- Resolve data stream
  - Artifacts & residual
- System modeling
  - Better characterizations of Gaussianity, Stationarity
- Addendum
  - Locked segments of Nov’94 locked segments available as netCDF fiiles

*Note 1, LIGO, 03/18/99 08:45:23 AM*
LIGO-G990022-28-M