Realistic Data Generation : applications and possible implications for burst search.

Soma Mukherjee

Centre for Gravitational Wave Astronomy

Dept. of Physics and Astronomy

University of Texas Brownsville

LIGO-G040362-00-Z

Why do we need a model ?

- Astrophysical searches estimate efficiency from playground data. Data is non-stationary.

- Externally triggered search (Gamma Ray Bursts with GW) – how representative is the ’off-source’ segment of the ‘on-source’ one ?

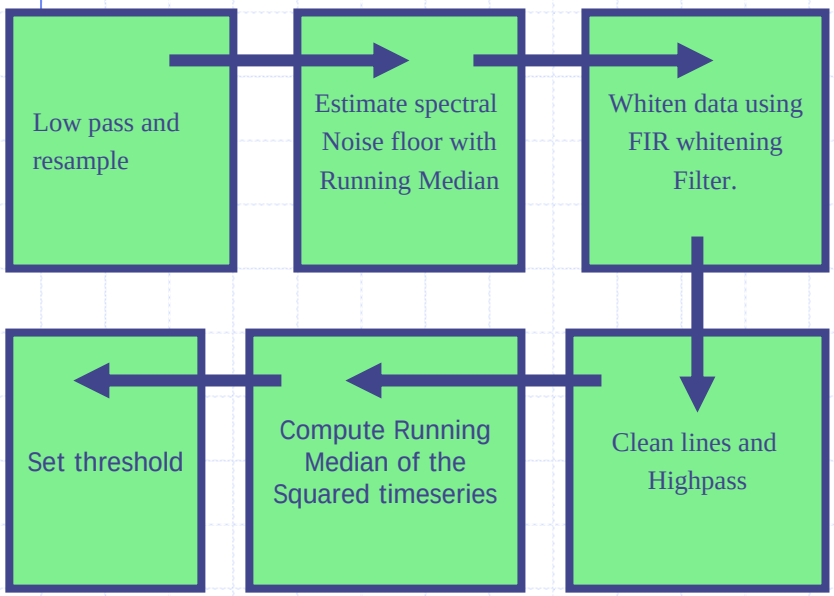
Data Modeling: assumptions

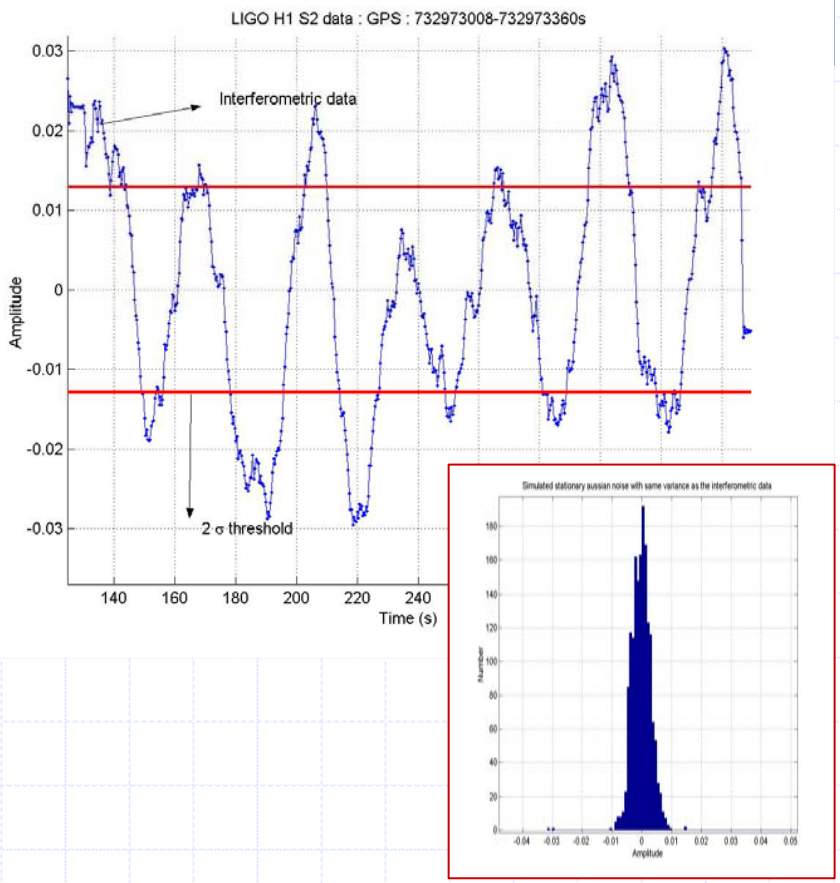
- Interferometric data has three components : lines, transients and noise floor.
- As a first approximation, the three components are independent and appear additively.
  - Physically different sources for each
- Basic idea is to split a channel into these components with mutual exclusion
  - Classify Transients, fit ARMA models to line amplitude and phase modulation, ARMA models for noise floor rms

I :Slowly drifting noise floor

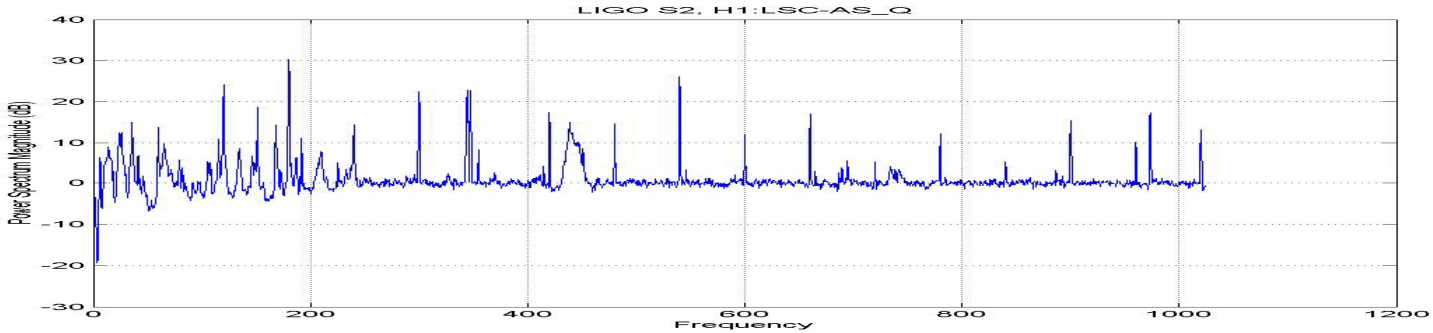
- Data from present generation of interferometers is non-stationary.

Results obtained by running MNFT[1] on a stretch of LIGO S2 data.[1 *Mukherjee, CQG, 2003*]



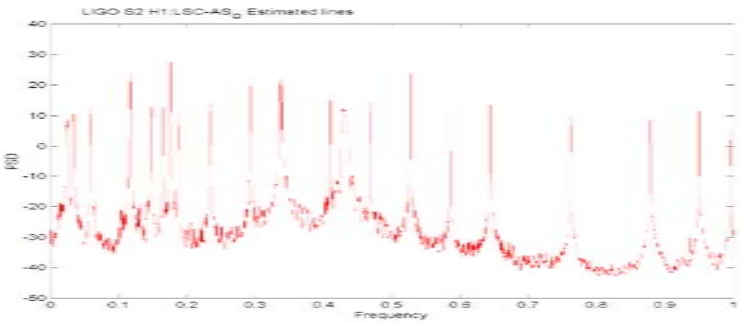


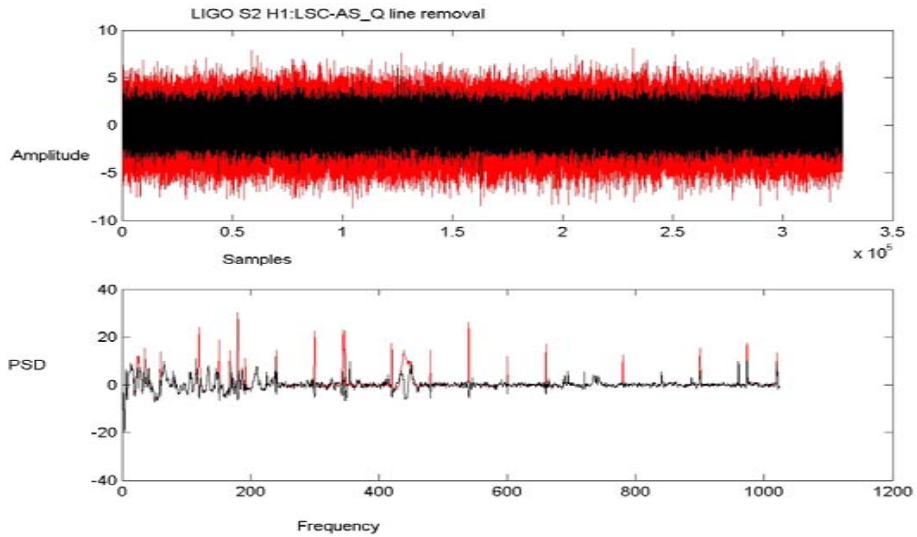
II : Modeling ALL lines



MBLT[1] : Non-parametric Line estimation.

[1 *Mohanty CQG, 2001*]

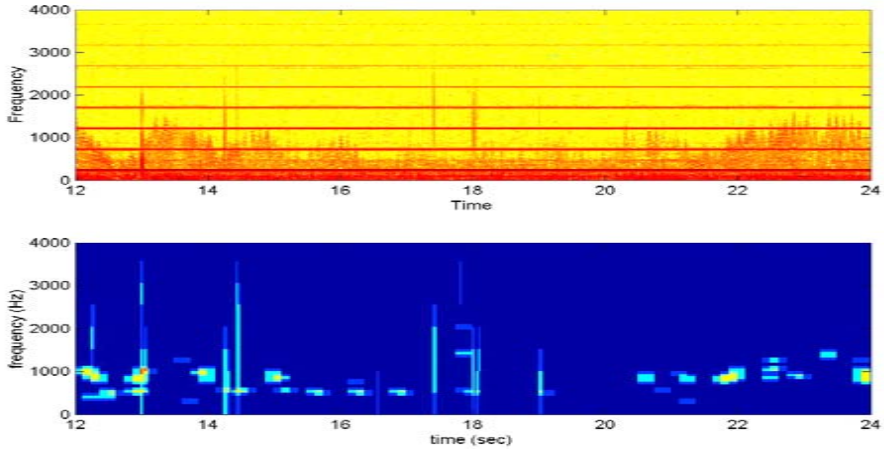


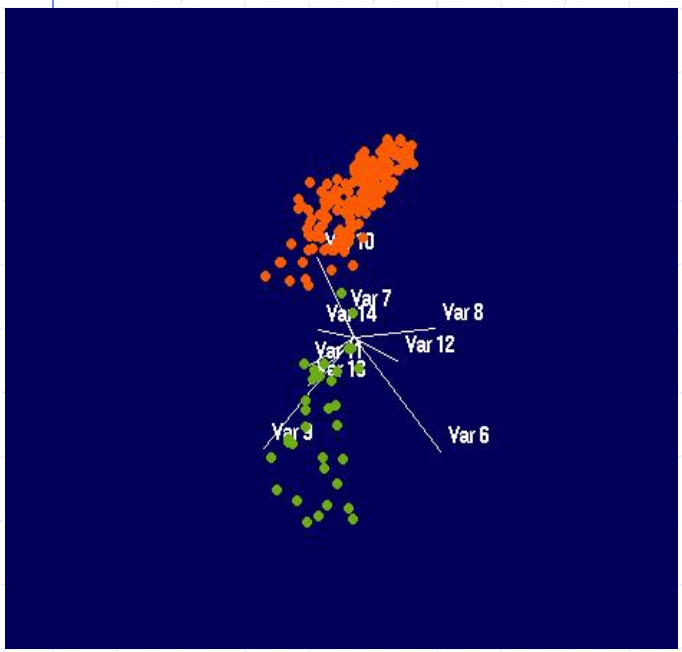


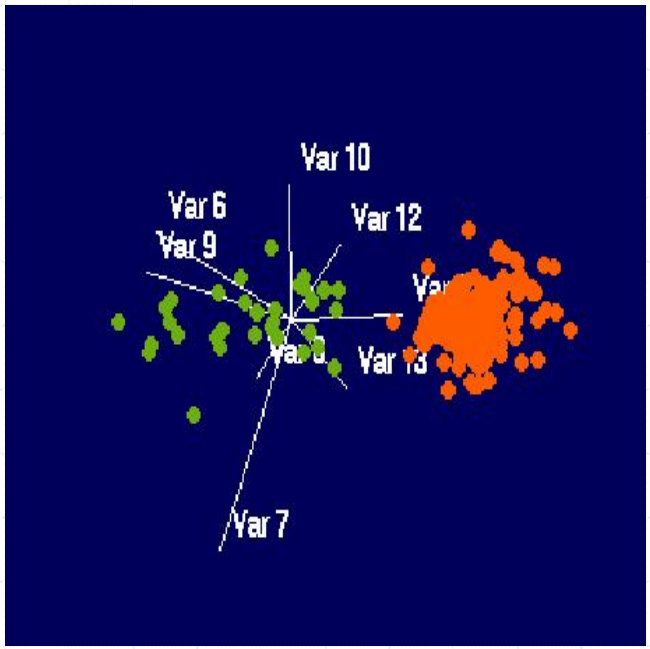
III : Transient Classification

Non-parametric change point detector.

KSCD : Kolmogorv-Smirnov test based Change point Detector *Mohanty, GWDAW (2002); PSDCD, Mohanty, PRD (2000);*







12 top wavelet coefficients of data surrounding each KSCD trigger. *Visualized using GGobi.* (Preliminary).

Mukherjee, 2003, Amaldi, Pisa

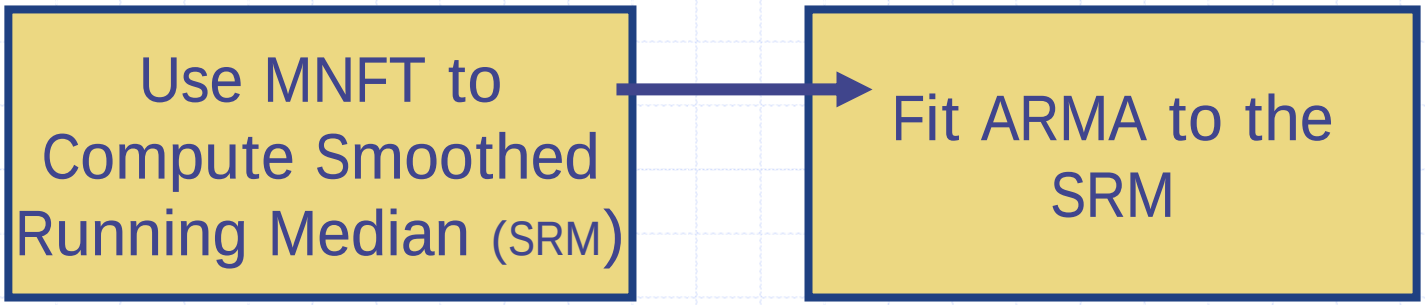
MNFT outline:

Algorithm:

1. Low pass and resample given time series *x(k).*
2. Construct FIR filter that whitens the noise floor. Resulting time series : *w(k)*
3. Remove lines using notch filter. Cleaned time series : *c(k)*
4. Track variation in second moment of *c(k)* using Running Median and apply smoothing (SRM).
5. Obtain significance levels of the sampling distribution via Monte Carlo simulations.

Model Noise Generation

- Model Noise Floor (low order ARMA).



- Estimate lines using MBLT , ARMA model amplitude and phase, add reconstructed lines to synthetic data .
- Add transients.

ARMA (*p,q*)

A(q) y(t) = C(q) e(t)

Y(t) : Output

e(t) : White noise

C(q)/A(q) : Transfer function

q: Time shift operator

A and *C* : Polynomials

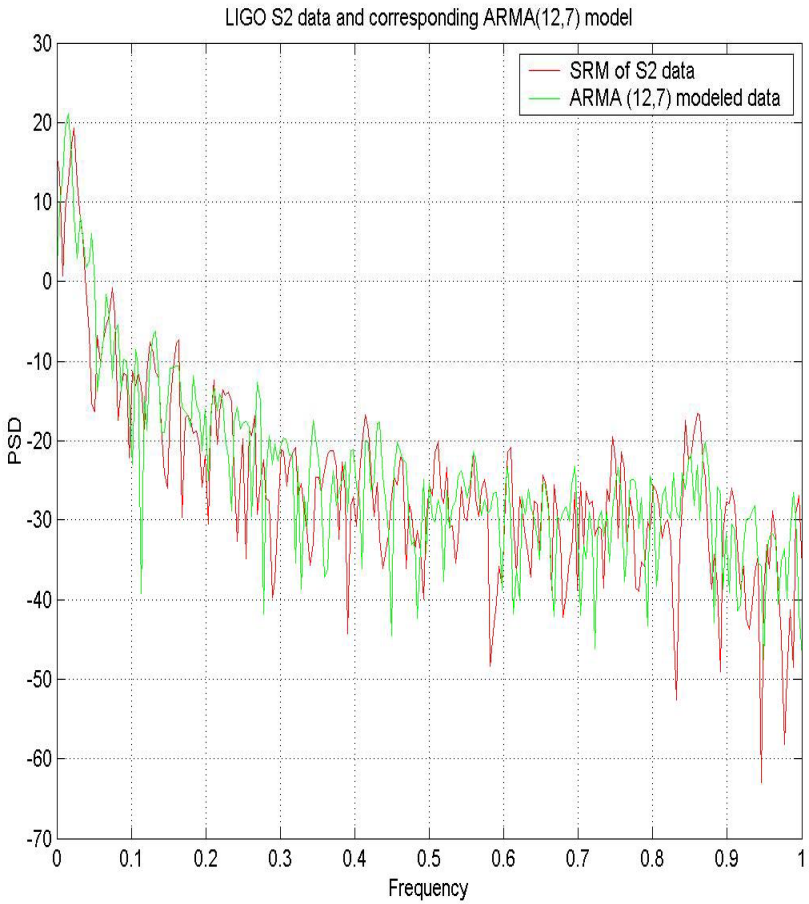
How faithful is the model ?

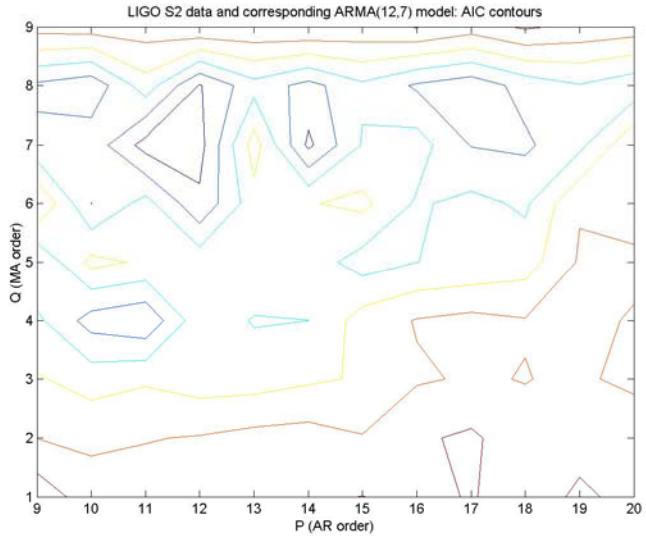
Apply statistical tests of hypothesis

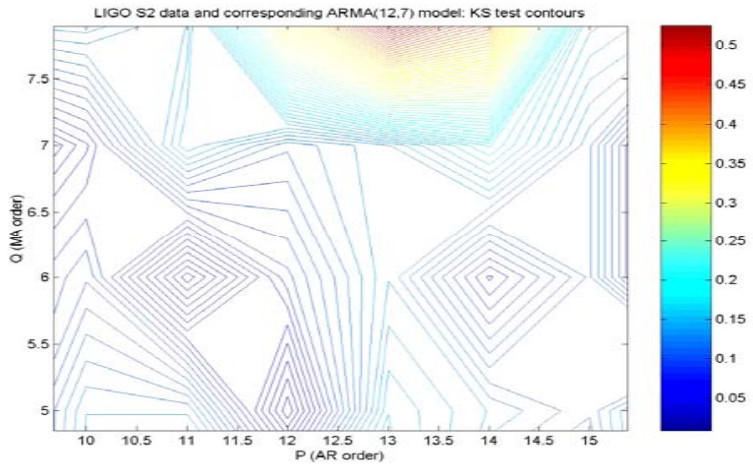
- Kolmogorov-Smirnov
- Akaike Information criterion (AIC)

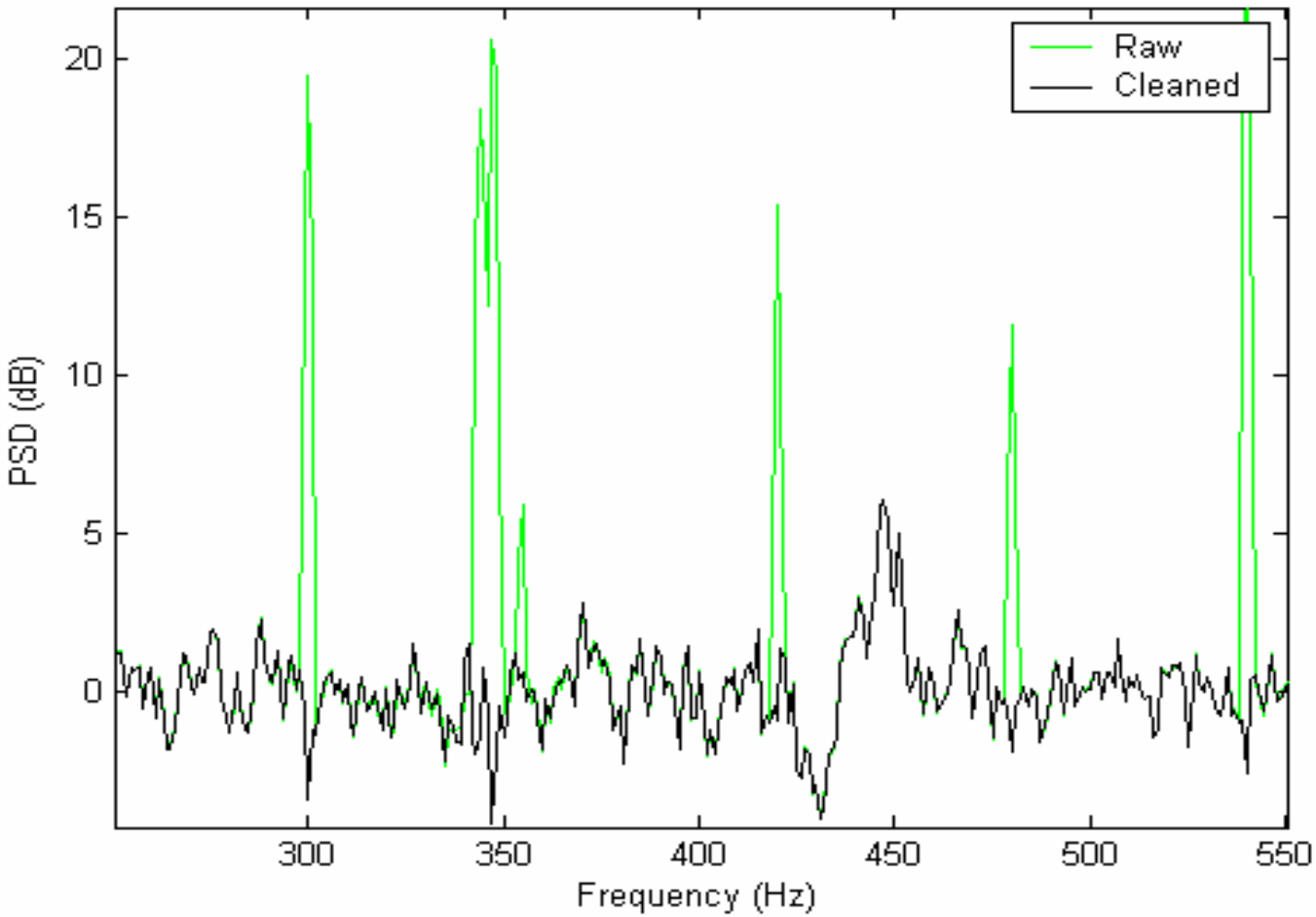
$$I_{akaike}(p,q) = \ln \sigma^{2}_{p,q} + 2(p+q)/N$$

Result I : Noise floor model – ARMA (12,7)







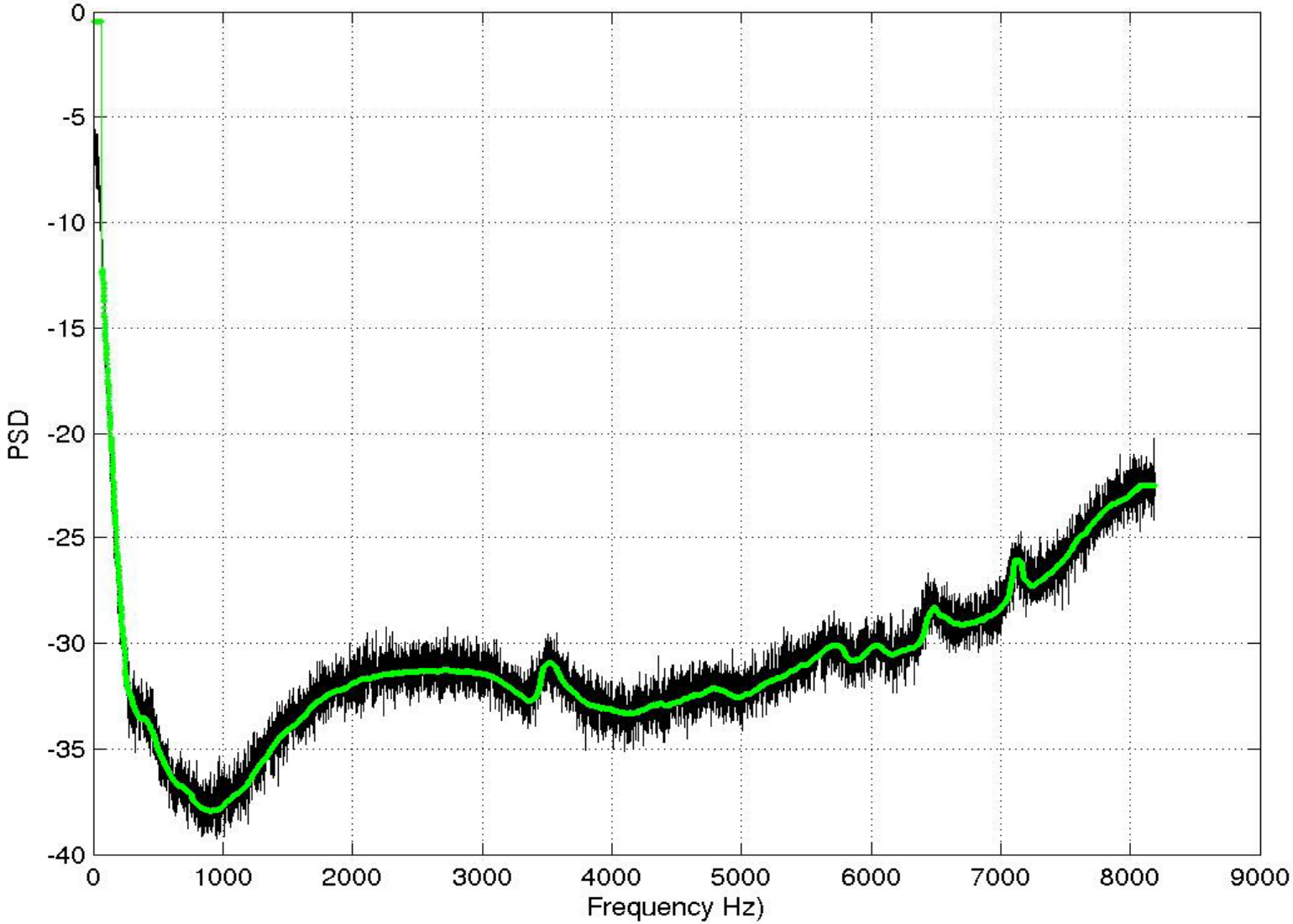

Raw
Cleaned
PSD (dB)
20
15
10
5
0
300
350
400
450
500
550
Frequency (Hz)

Implications for Burst Search

- Use as an ‘infinite’ playground for astrophysical searches.
- Gives a handle on non-stationarity and hence testing the robustness of the search algorithm.
- Allows us to do ‘controlled tests’.
- Signal injection and efficiency estimation

Existing modeled data

- 20,000 s modeled S2 data exists at suhail.ligo.caltech.edu .
- Contains 330 s .mat files of noise floor and lines separately.
- Script exists to allow the user to combine the noise floor and lines in the desired way.


0
-5
-10
-15
-20
-25
-30
-35
-40
PSD
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Frequency Hz)

Further developments ...

- To add transients.
- Generation of band limited noise.
- To write the output in Frames.
- To generate S3 and S4 segments.
- To carry out known waveform injection studies.