Generation of realistic data segments: production and application.

Soma Mukherjee

Centre for Gravitational Wave Astronomy

D

ept. of Physics and Astronomy

University of Texas Brownsville

LIGO G040361-00-Z

Motivation

An effort to simulate segments of detector noise such that it shares same statistical characteristics as the data itself, i.e. it shows same kind of non-stationarity as present in the data.

We can not predict the *nature* of non–stationarity at a prior time, but we can characterize the variation to construct simulated data.

Astrophysical searches estimate efficiency and tune parameters of search algorithms from playground data. Data is non-stationary. E.g. a 60 s stretch of data may not have same value of variance as another 60 s at a later time. Thus any comparison or any estimation work carried out with sampled real data will have bias introduced in it.

Certain search algorithms require comparison of different data segments. These are different quantities when data is non-stationary.

Outline :

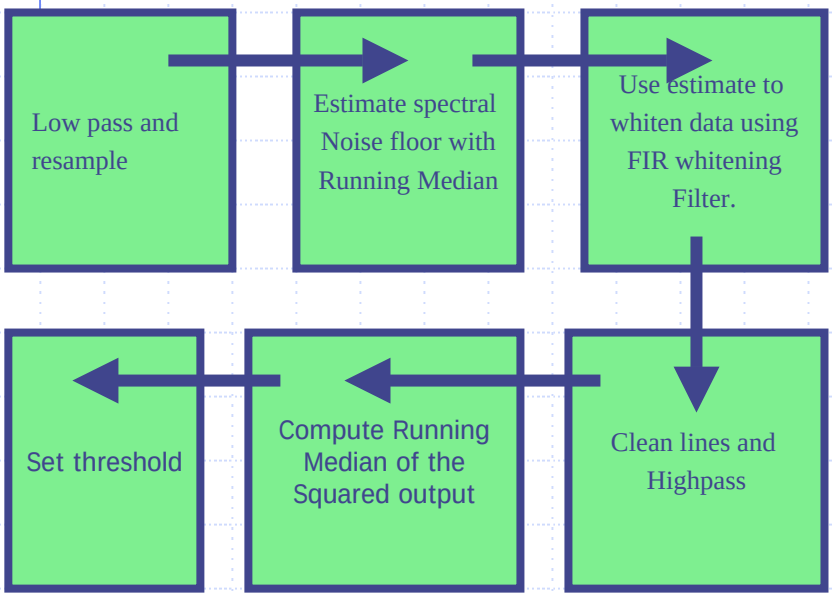
- Description of the model.
- Results.
- Existing modeled data.
- Examples.
- Future.

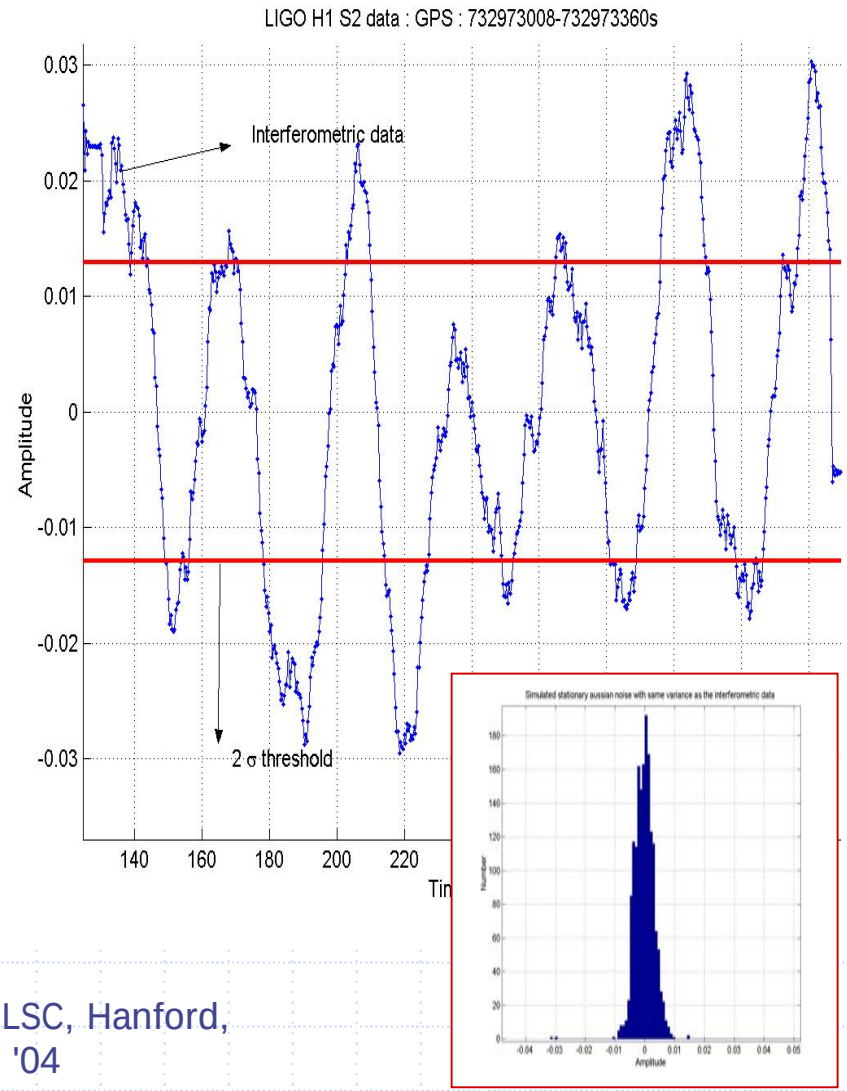
Model description

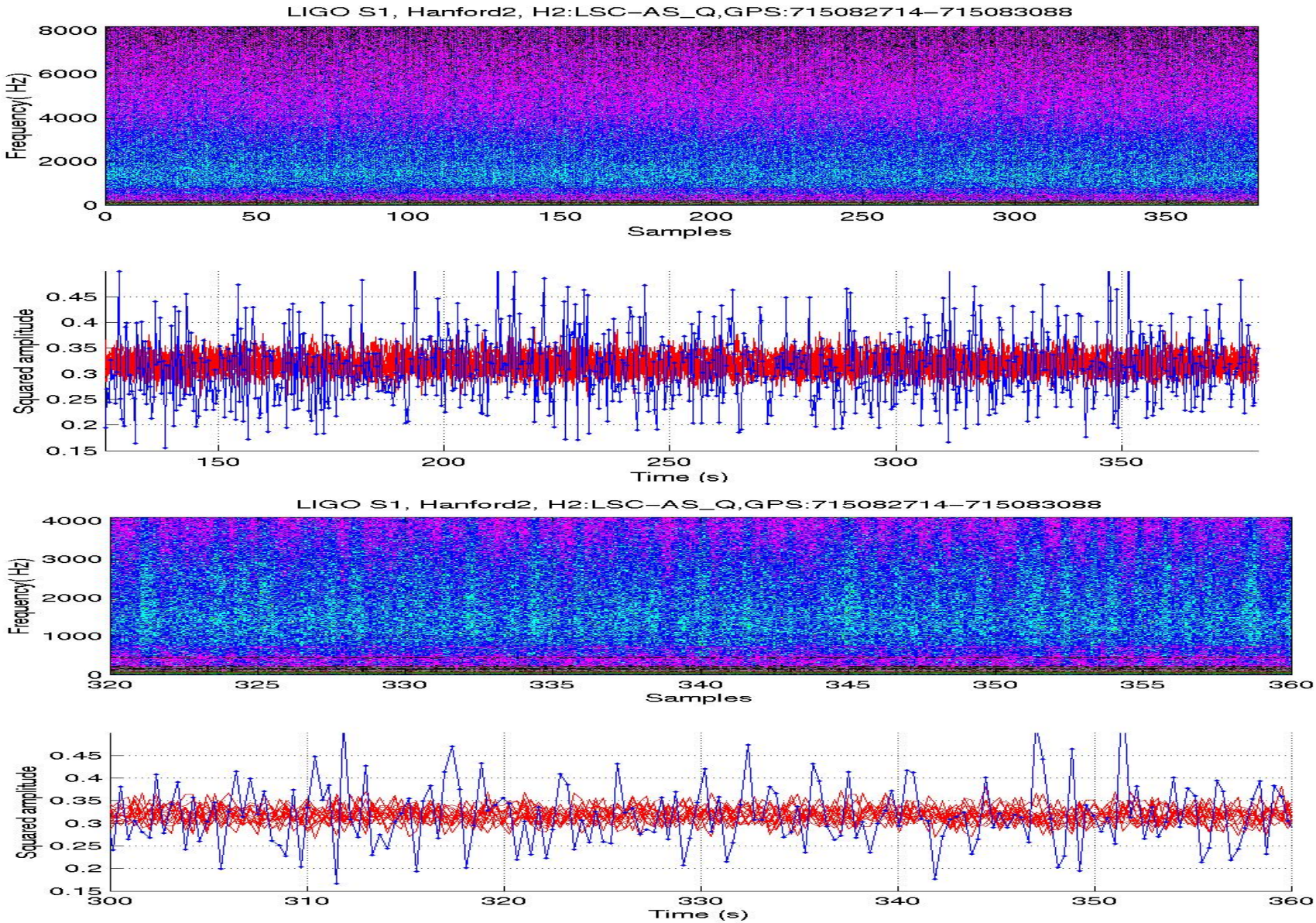
- Each data segment is divided into three independent components viz. lines, transients and noise floor.
- Each component is characterized by low order ARMA model with mutual exclusion in time domain.
- Modeled noise generated from the three independent components are added back together to construct the final product.

Characterizing noise floor variation

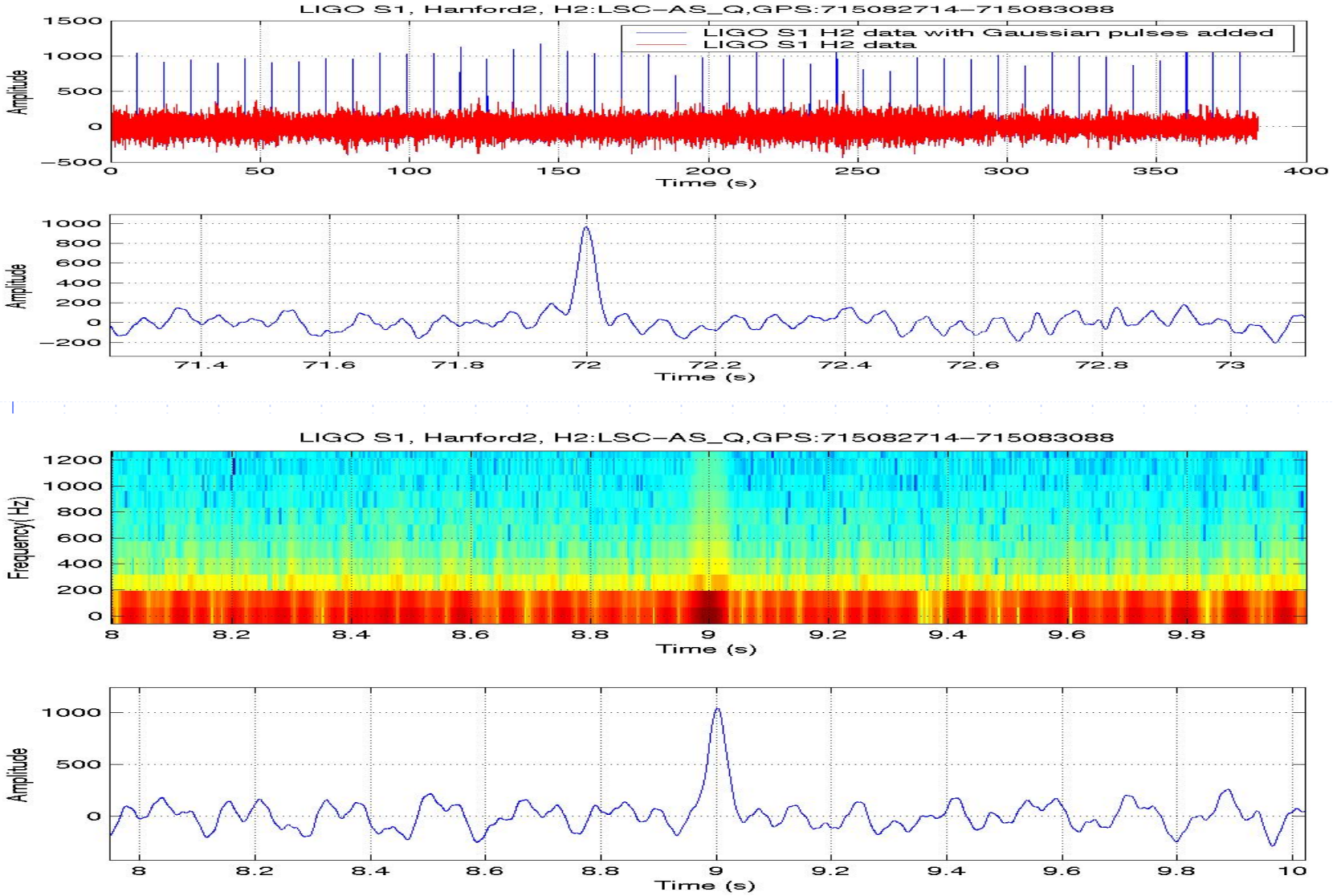
Results obtained by running MNFT[1] on a stretch of LIGO S2 data.[1 *Mukherjee, CQG, 2003*]

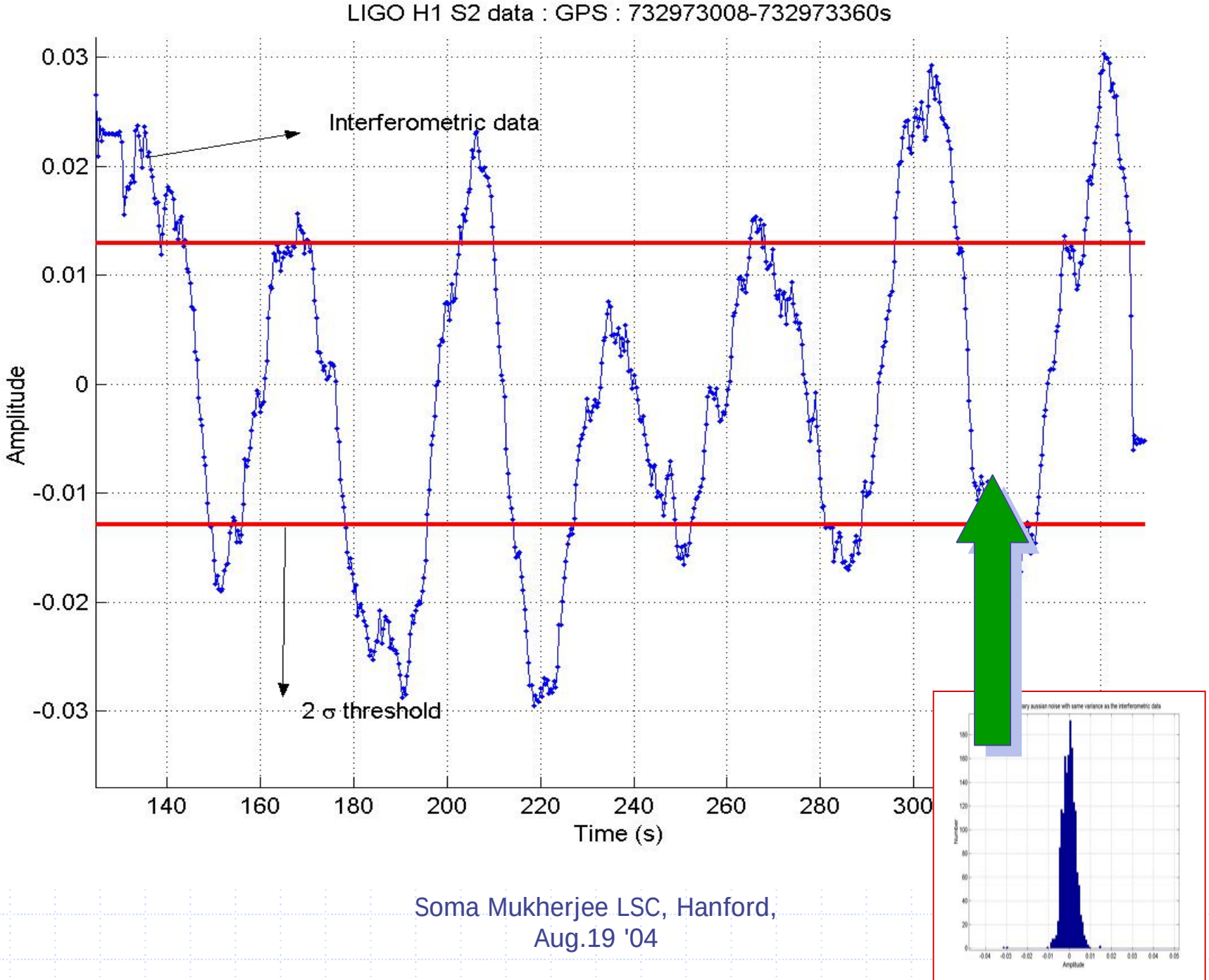




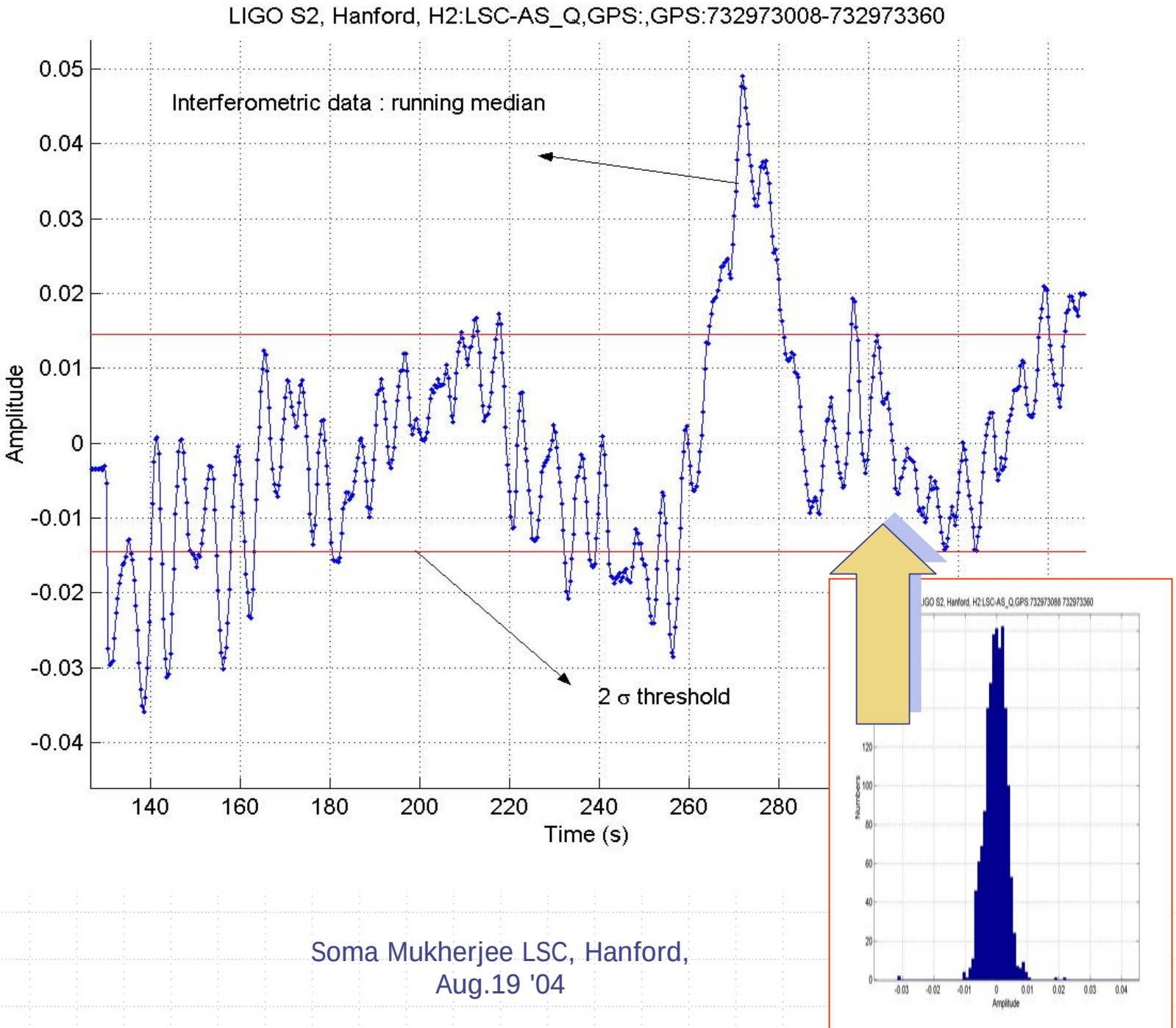


Soma Mukherjee LSC, Hanford,
Aug.19 '04


LIGO S1, Hanford2, H2:LSC−AS_Q,GPS:715082714−715083088
LIGO S1 H2 data with Gaussian pulses added
LIGO S1 H2 data
Amplitude
Time (s)
LIGO S1, Hanford2, H2:LSC−AS_Q,GPS:715082714−715083088
Frequency(Hz)
Amplitude
Time (s)

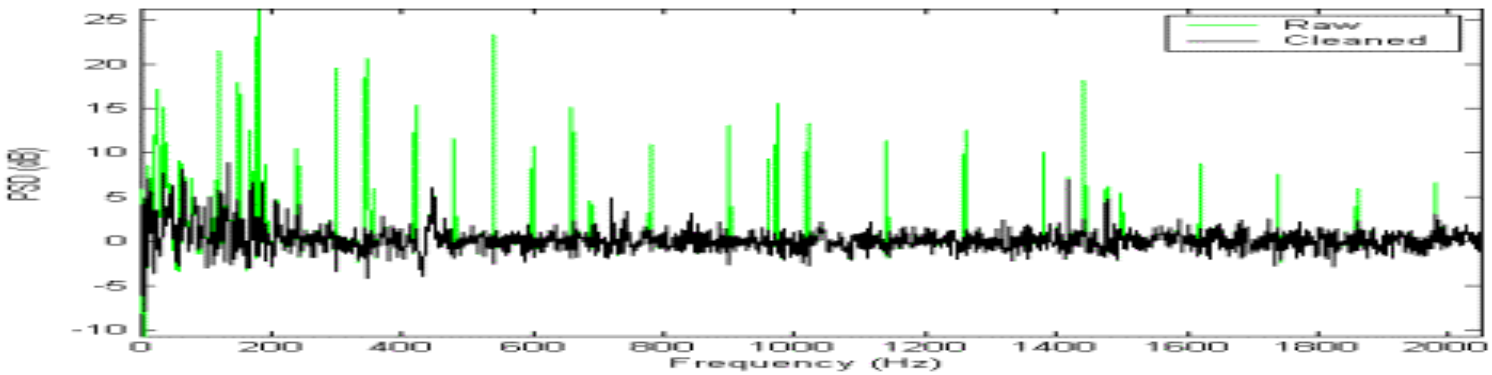


Soma Mukherjee LSC, Hanford,
Aug.19 '04

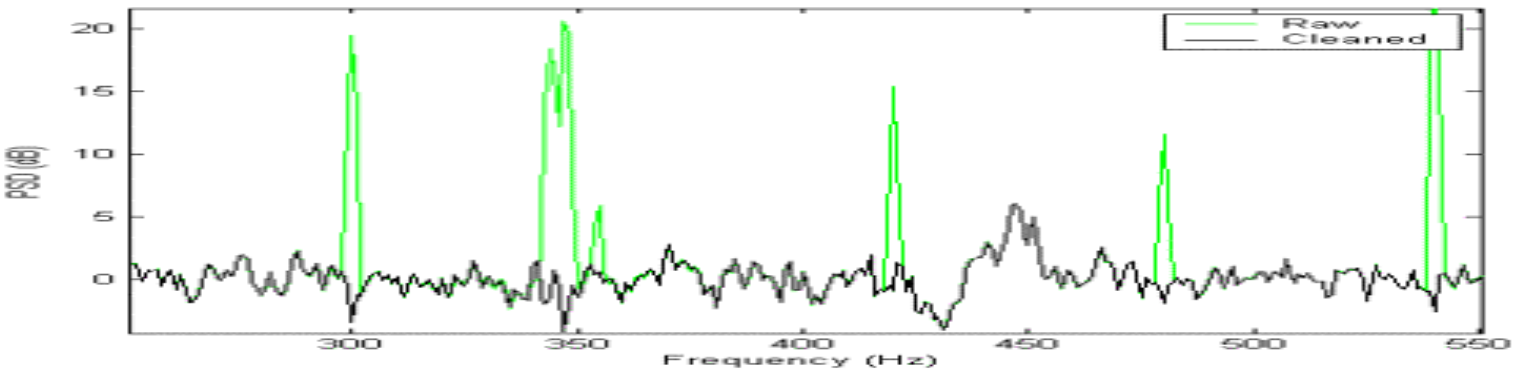


Soma Mukherjee LSC, Hanford,
Aug.19 '04

II : Modeling ALL lines



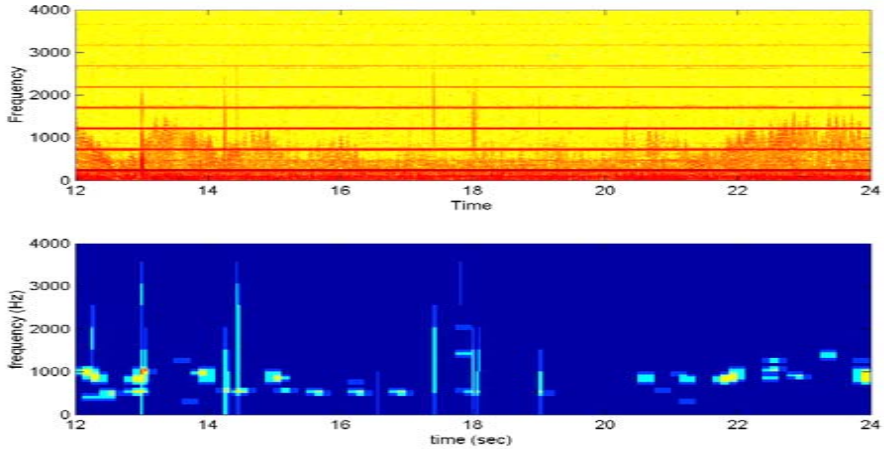
MBLT: Non-parametric Line estimation. [*Mohanty CQG, 2001*]

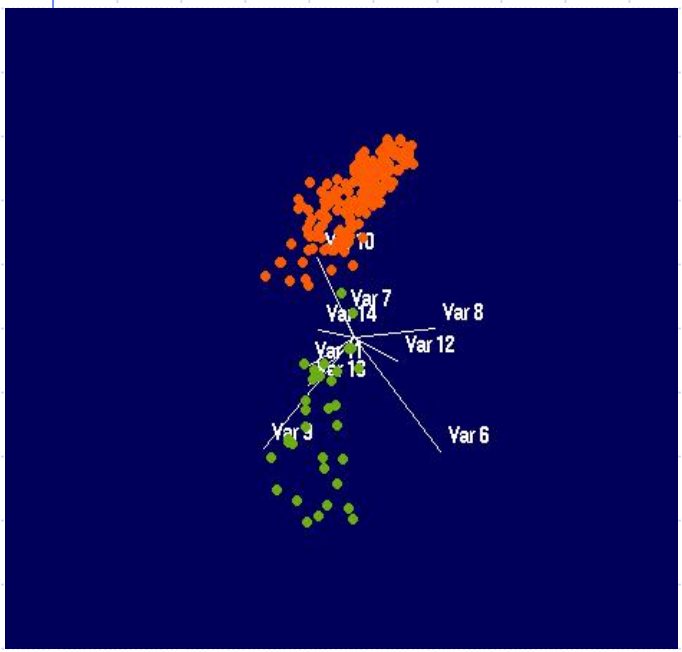


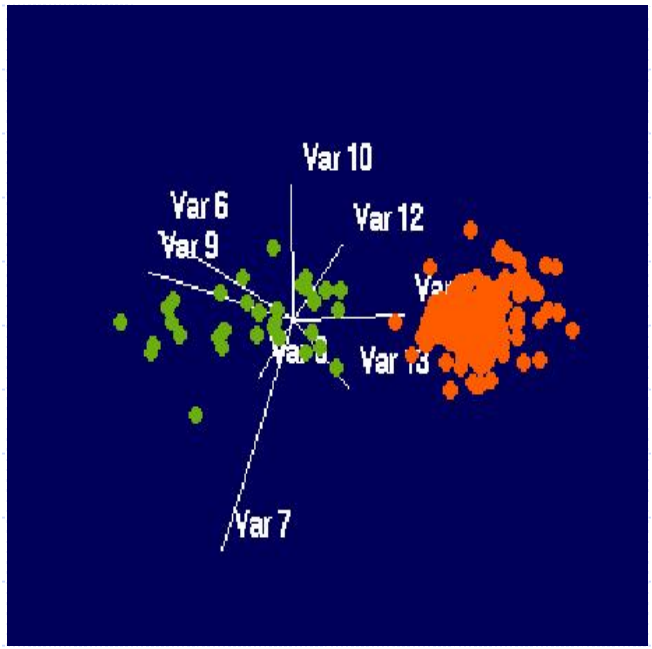
III : Transient Classification

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).

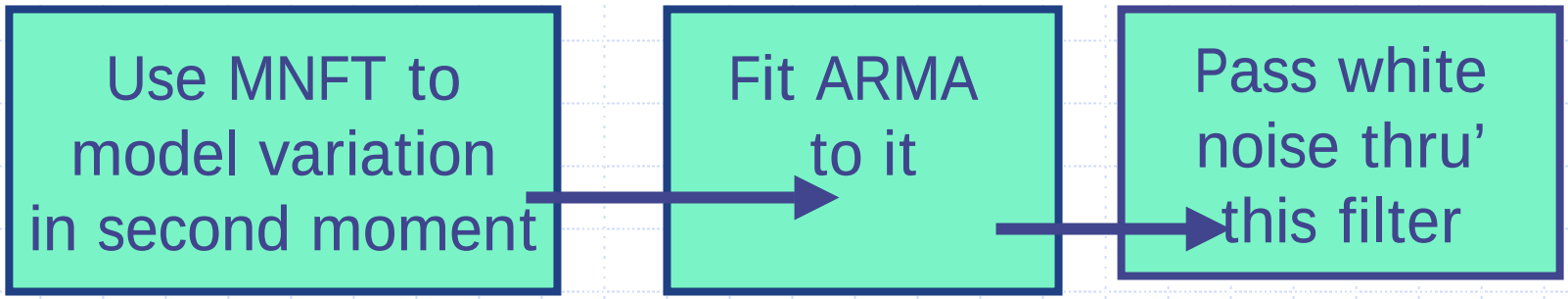
Mukherjee, 2003, Amaldi, Pisa

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. Model variation in second moment of *c(k)* using Running Median and apply smoothing (SRM).

Model Noise Generation

- Characterize MNFT output with low order ARMA.



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

ARMA (*p,q*)

A(T) y[k]= C(T) e[k]

Y[k]: Output

e[k] : White noise

T: Time shift operator , T(x[k])=x[k-1]

A and C : Polynomials of order p and q respectively.

Example ARMA(2,1):

$A_1 y[k-1] + A_0 y[k] = C_0 e[k]$

Testing the model

Apply statistical test of hypothesis

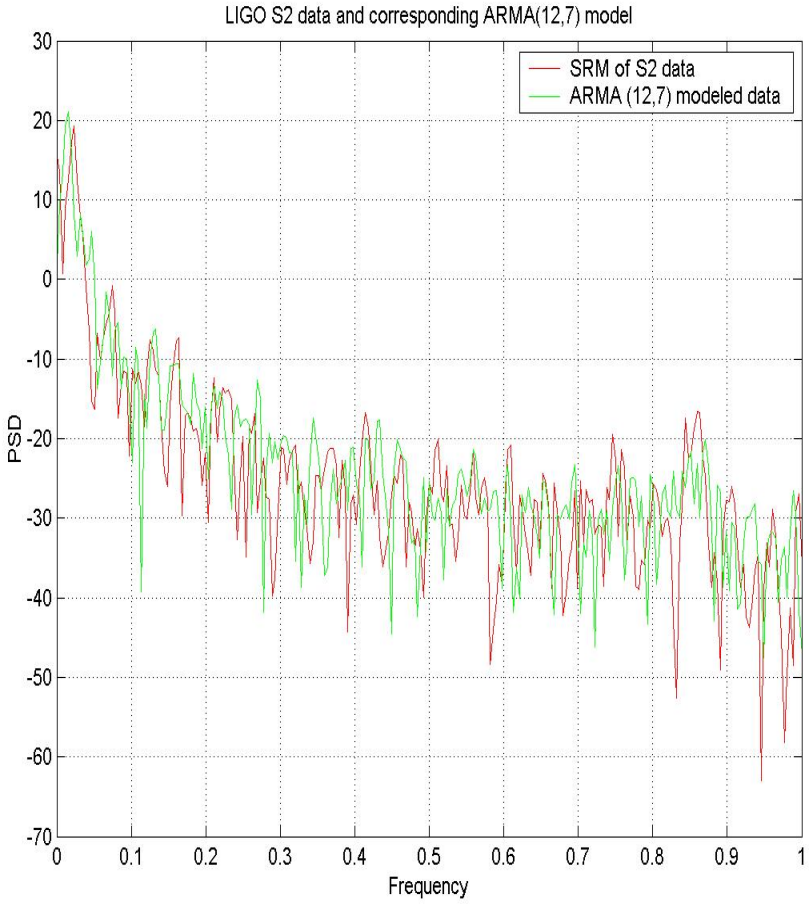
- Kolmogorov-Smirnov

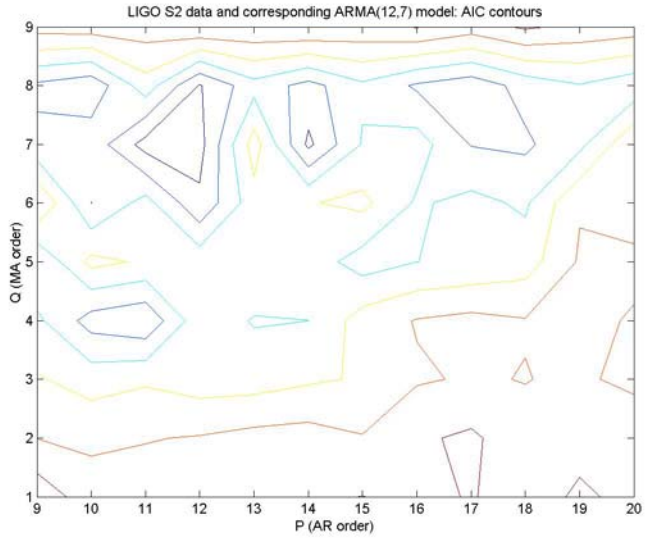
To avoid overfitting :

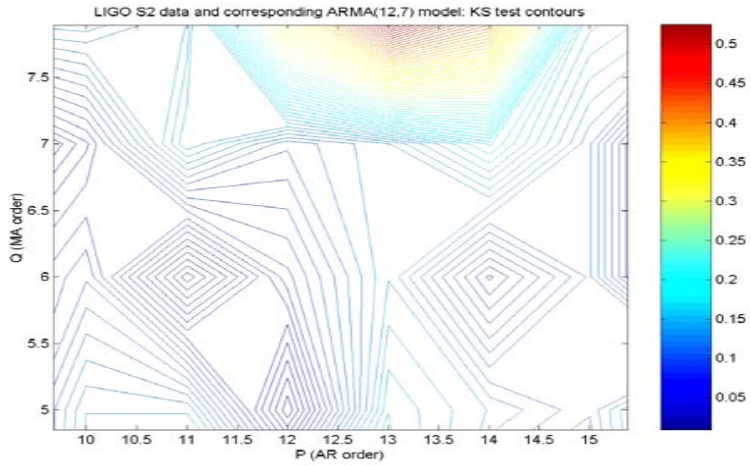
- Akaike Information criterion (AIC)

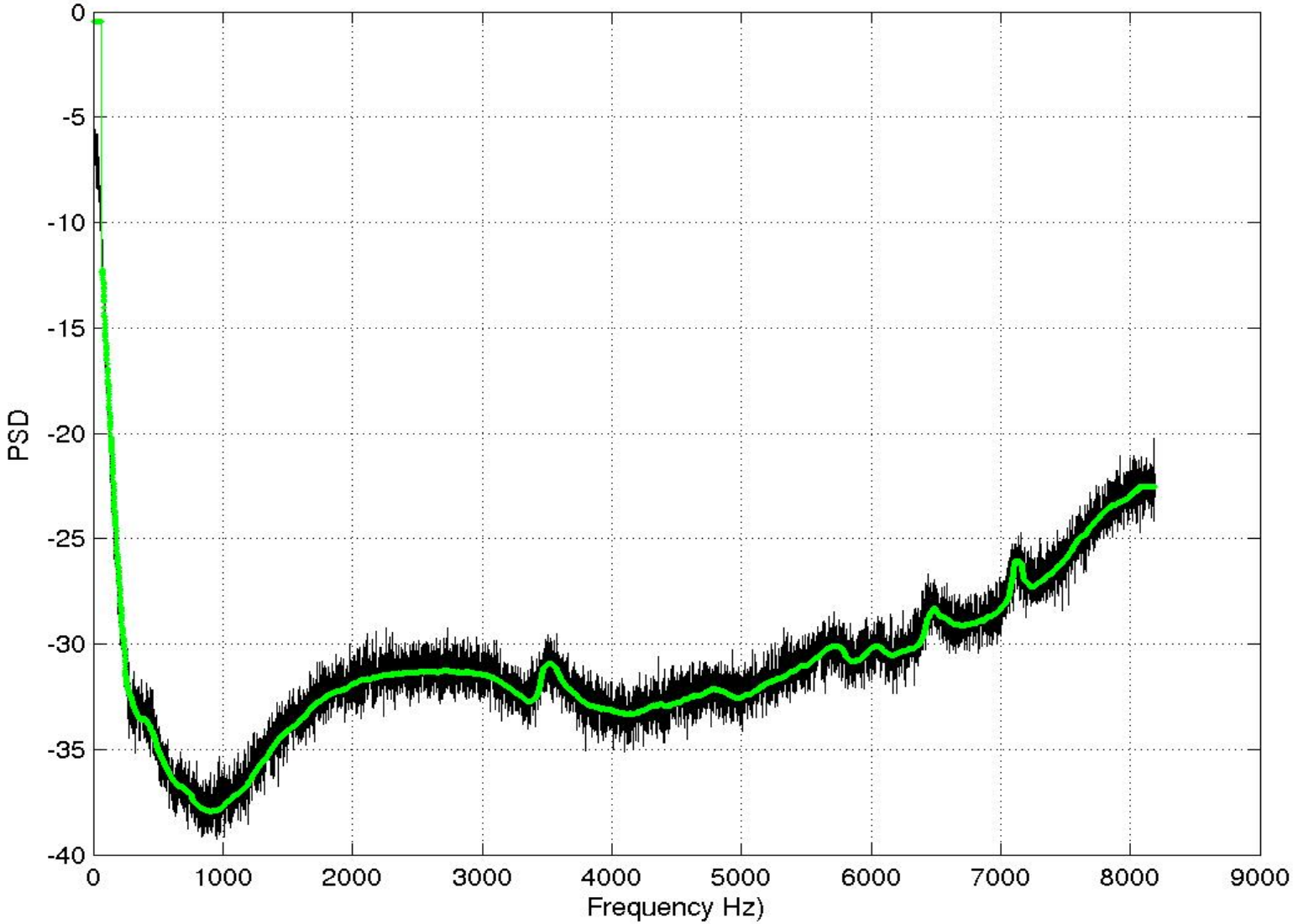
$$I_{akaike}(p,q) = \ln \sigma^2_{p,q} + 2(p+q)/N$$

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

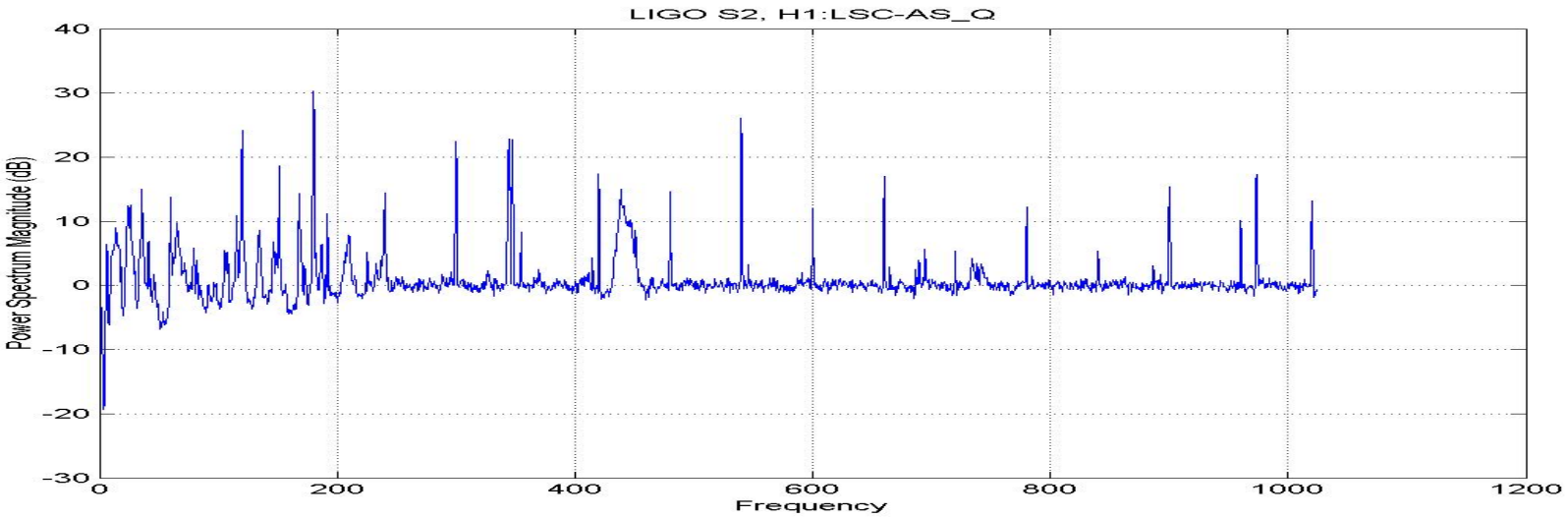


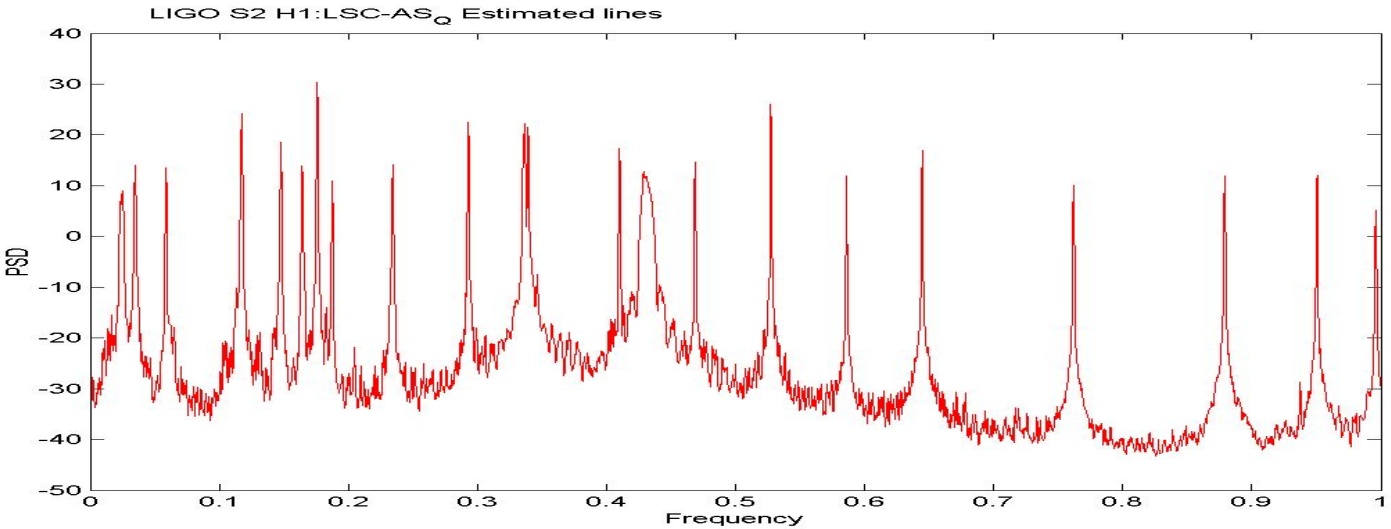






Soma Mukherjee LSC, Hanford,
Aug.19 '04





Soma Mukherjee LSC, Hanford,
Aug.19 '04

More work

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

Existing modeled data

- 20,000 s modeled S2 data generated in April 2004.
- 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.