SEARCHING GRAVITATIONAL WAVE WITH AUTOREGRESSIVE APPROACH

DAVID C. Y. HUI

DEPARTMENT OF ASTRONOMY & SPACE SCIENCE
CHUNGNAM NATIONAL UNIVERSITY
SOUTH KOREA

7th KAGRA International Workshop 18-20 December 2020

COLLABORATORS

- Albert Kong (NTHU, Taiwan)
- Lupin Lin (UNIST, Korea)
- Ray Kwan Lok Li (NCKU, Taiwan)
- Kwangmin Oh (CNU, Korea)
- Sangin Kim (CNU, Korea)
- Alex Leung (MUST, Macau)
- Jianqi Yan (MUST, Macau)

DATA CHALLENGES IN GRAVITATIONAL WAVE ASTRONOMY

- Presence of non-Gaussian noise (instrumental + environmental) that are far larger than the astrophysical signal.

$$h_{noise} \sim 10^{-19} \quad \text{vs.} \quad h_{signal} \lesssim 10^{-21}$$

- Conventional detection algorithm (e.g. matched filtering) cannot pick the signal without the knowledge of its form.

- Large data volume

WISH LIST FOR DETECTION PIPELINE

- Efficiency in noise reduction
- Scalable
- Robust

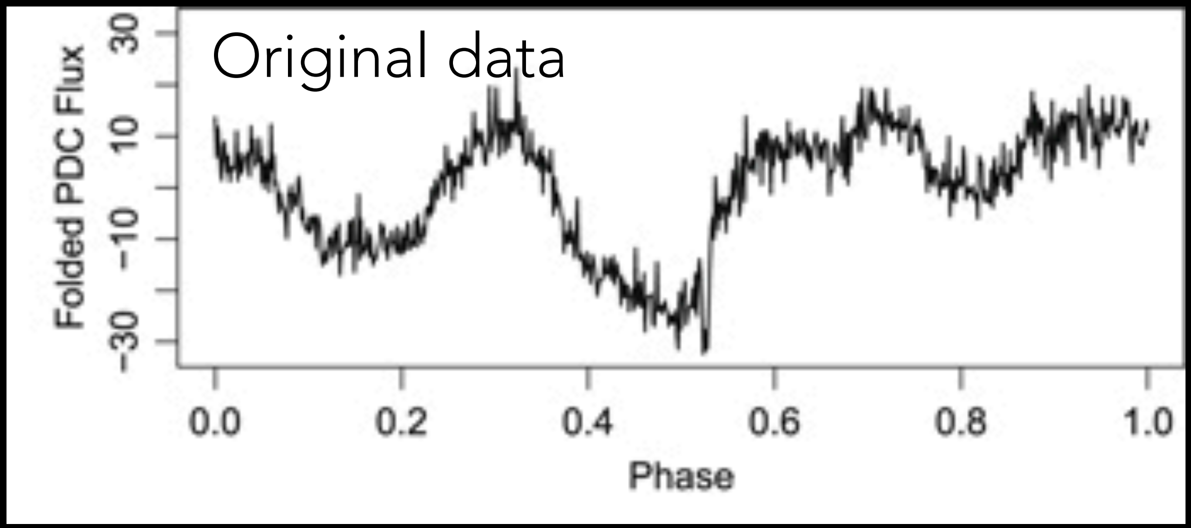
Algorithm Efficiency

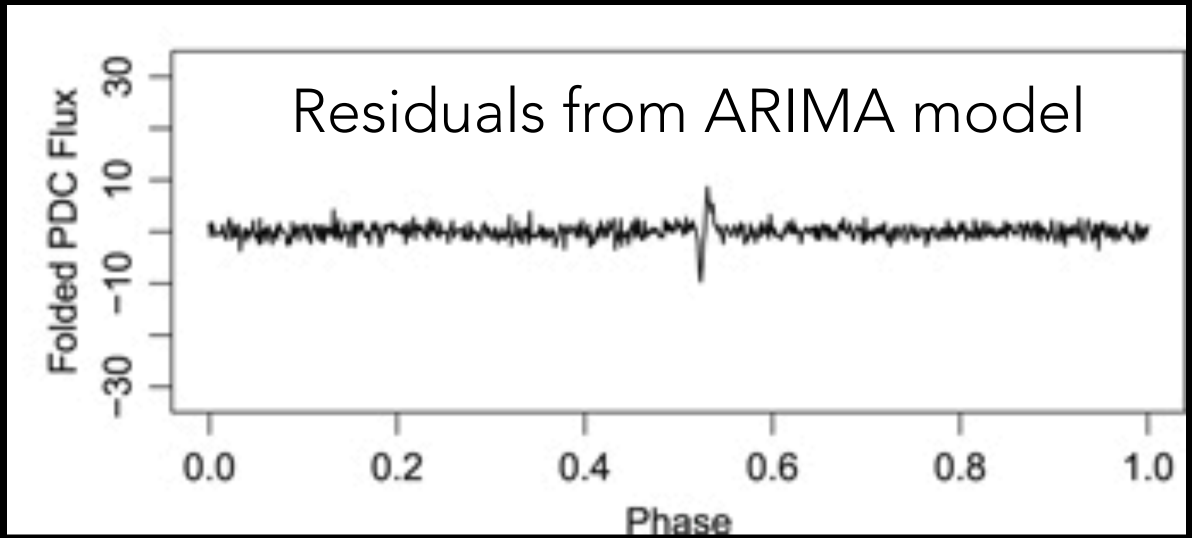
- Capability of picking signals (or candidates) with unknown form

Unsupervised Learning

NOISE REDUCTION WITH STOCHASTIC AUTOREGRESSIVE MODELING

- Simple approach of analysis in time domain
- Capable to handle various kinds of noise from non-stationary autocorrelated stochastic processes.
- Not much application in astronomy until recently (e.g. exoplanet search)





Caceres et al. 2019

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (**ARIMA**) MODEL

- ARIMA model comprises 3 components:

1. If the value of a variable at a given time is influenced by its *past values*, the process is **autoregressive (AR)** which can be regarded as a simple regression problem. An AR process of order *p* is modeled by:

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + \epsilon_t + c$$

ϵ_t is normally distributed random error with zero mean and unknown variance (i.e. white noise). c is a constant.

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (**ARIMA**) MODEL

- ARIMA model comprises 3 components:

2. If the variable has its current value depends on the *noise terms* in the previous *q* time steps, the process is described as a **moving average (MA)** model with order *q*:

$$x_t = b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \cdots + b_q\epsilon_{t-q} + \epsilon_t + c$$

Combining both we have ARMA model which predicts the value at a given time by both lagged values and lagged errors.

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (**ARIMA**) MODEL

- ARIMA model comprises 3 components:

3. For a non-stationary process, an approximately stationary time series can be obtained by *differencing operation*:

$$x_t^{'} = x_t - Bx_t = x_t - x_{t-1}$$

To restore the original light curve, one has to reverse this operation (i.e. *integrate* the series back). An **integrated (I)** process is modeled as:

$$(1 - B)^d x_t = \epsilon_t$$

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (**ARIMA**) MODEL

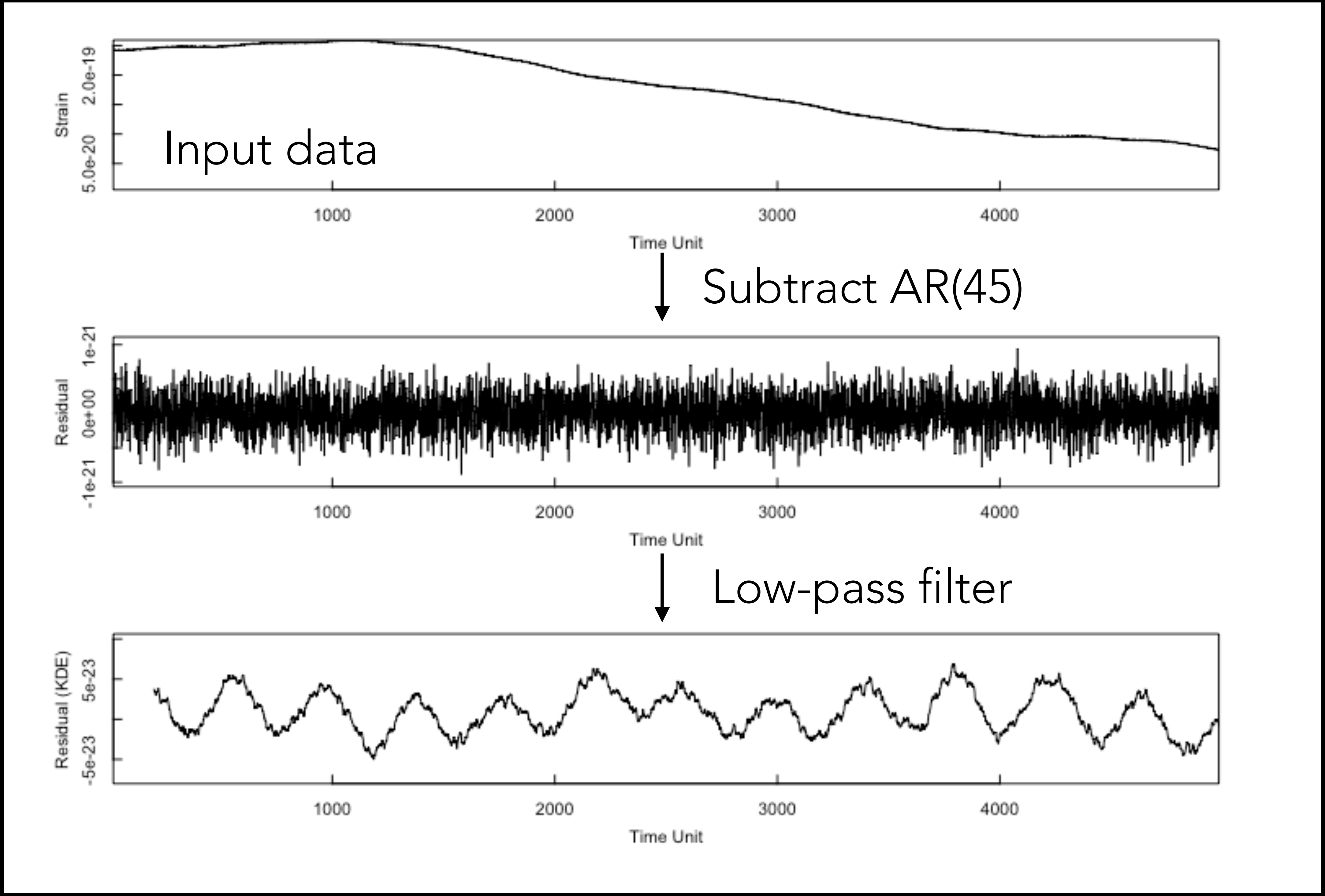
Combing the aforementioned three processes together into a single regression procedure, we have ARIMA(p,q,d) model:

$$(1 - B)^d x_t = \sum_{i=1}^{p} a_i x_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} + \epsilon_t + c$$

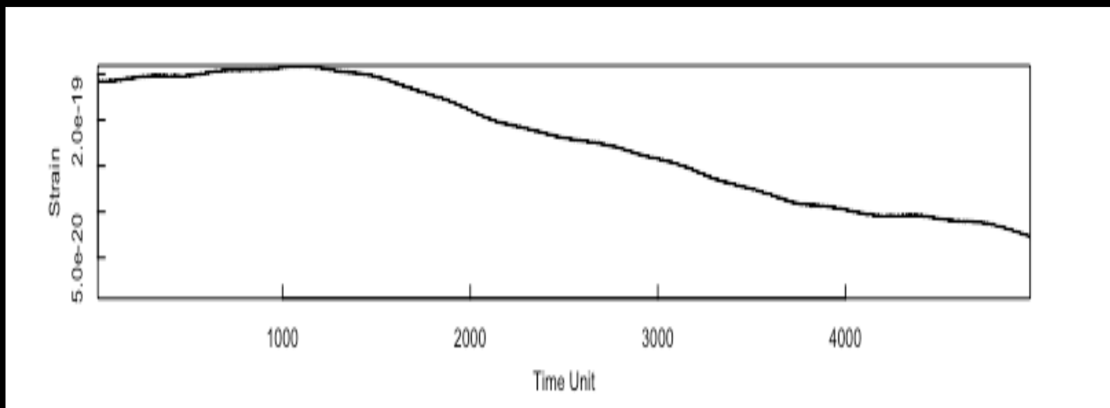
The model parameters can be determined by maximum likelihood estimation with the orders p and q determined through certain information criterion (e.g. BIC, AIC).

TEST CASE#1 SIMULATED DATA

- The simulated LIGO strain series with a constant 10 Hz sinusoidal signal of $h \sim 10^{-21}$ injected.

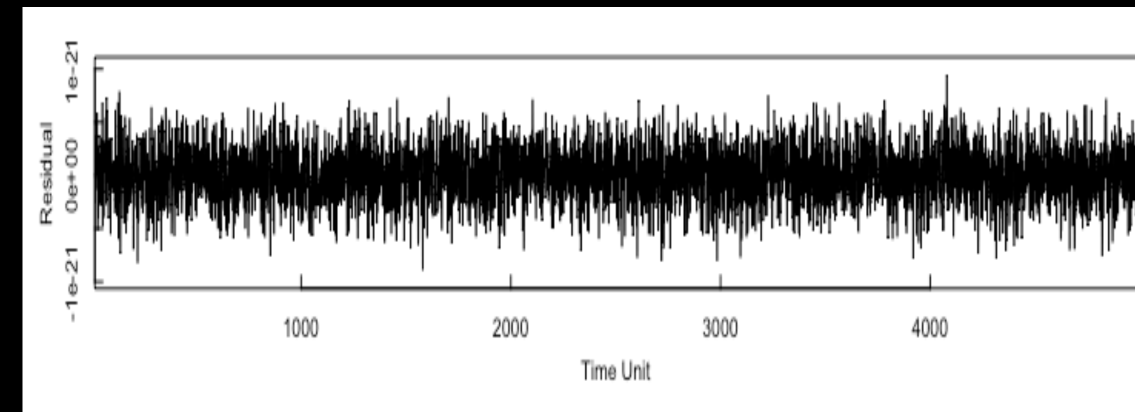


TEST CASE#1 SIMULATED DATA



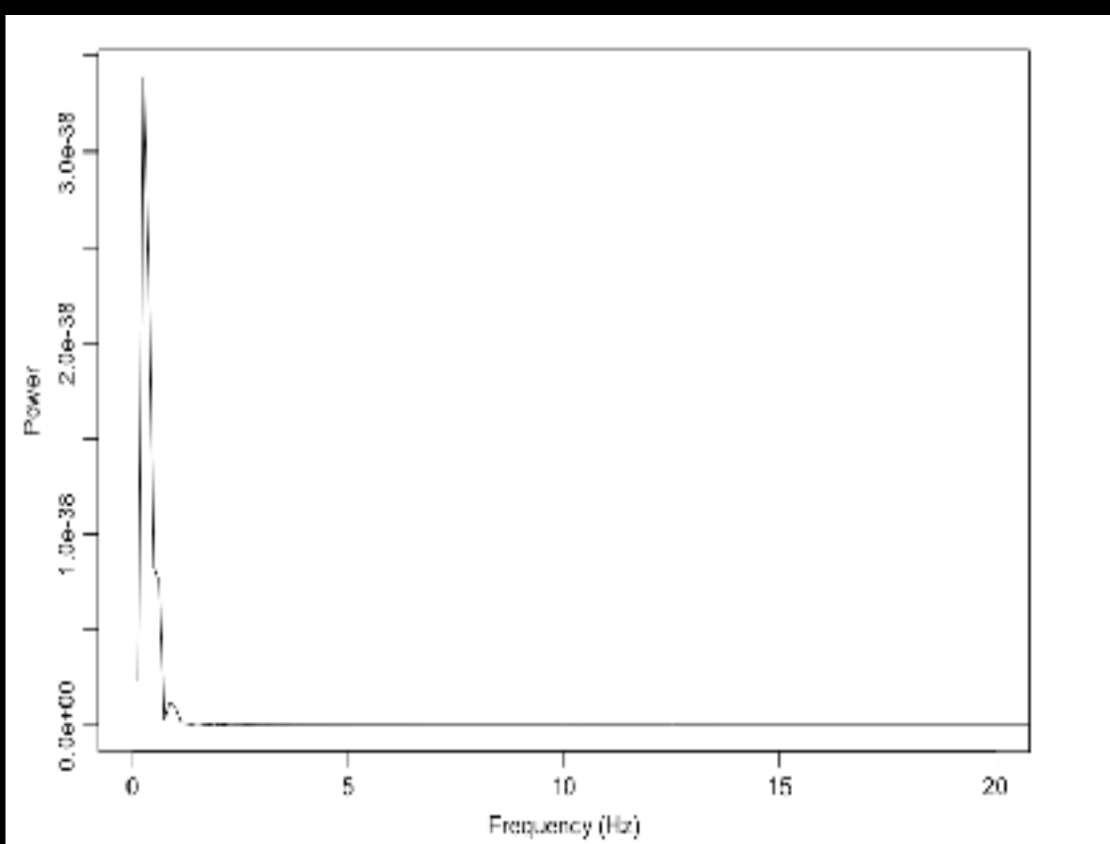
Original data

Subtract AR(45)

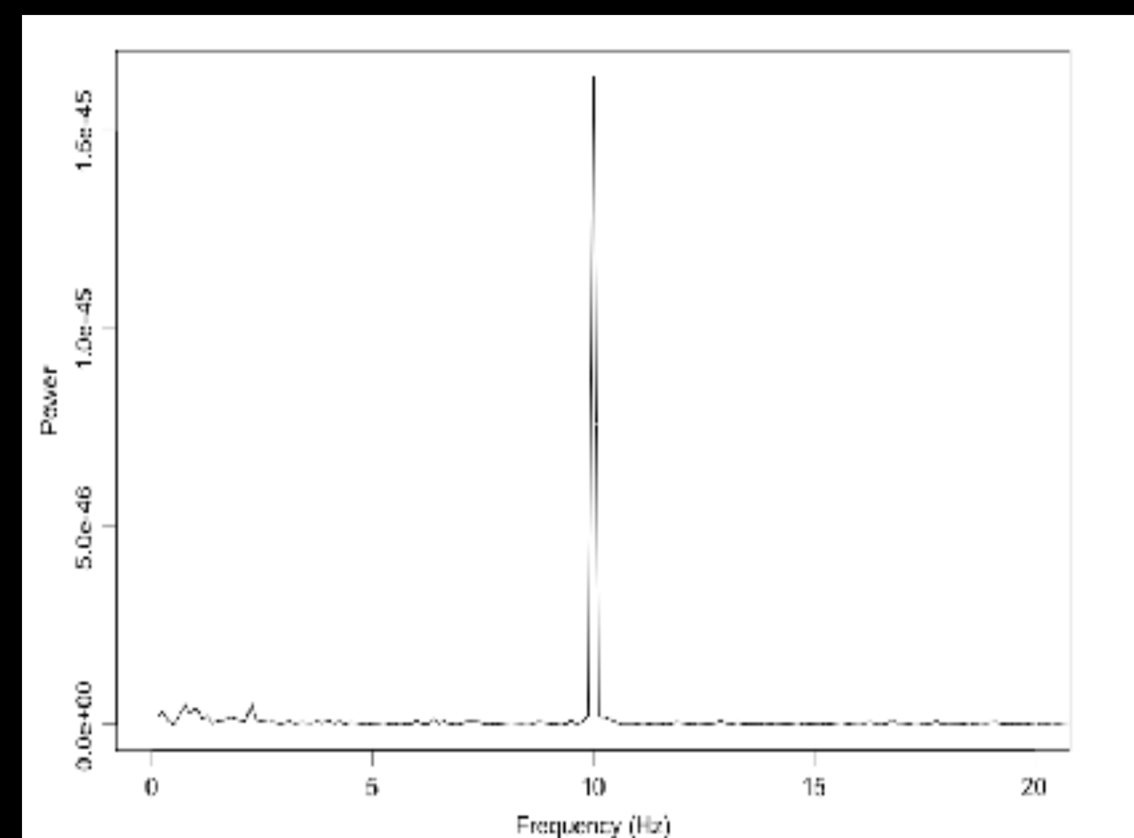


Residual (without low-pass filter)

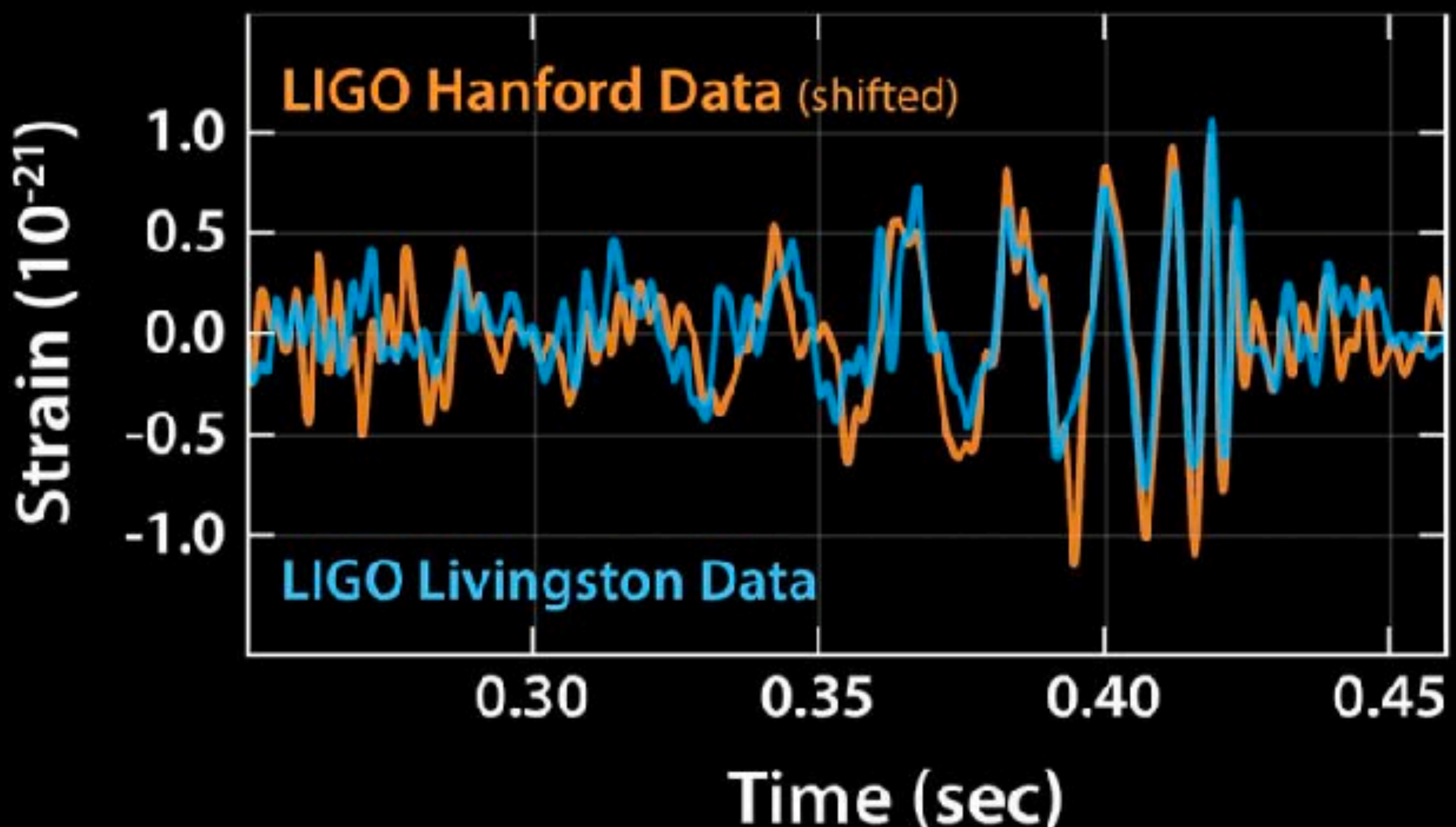
FFT



FFT

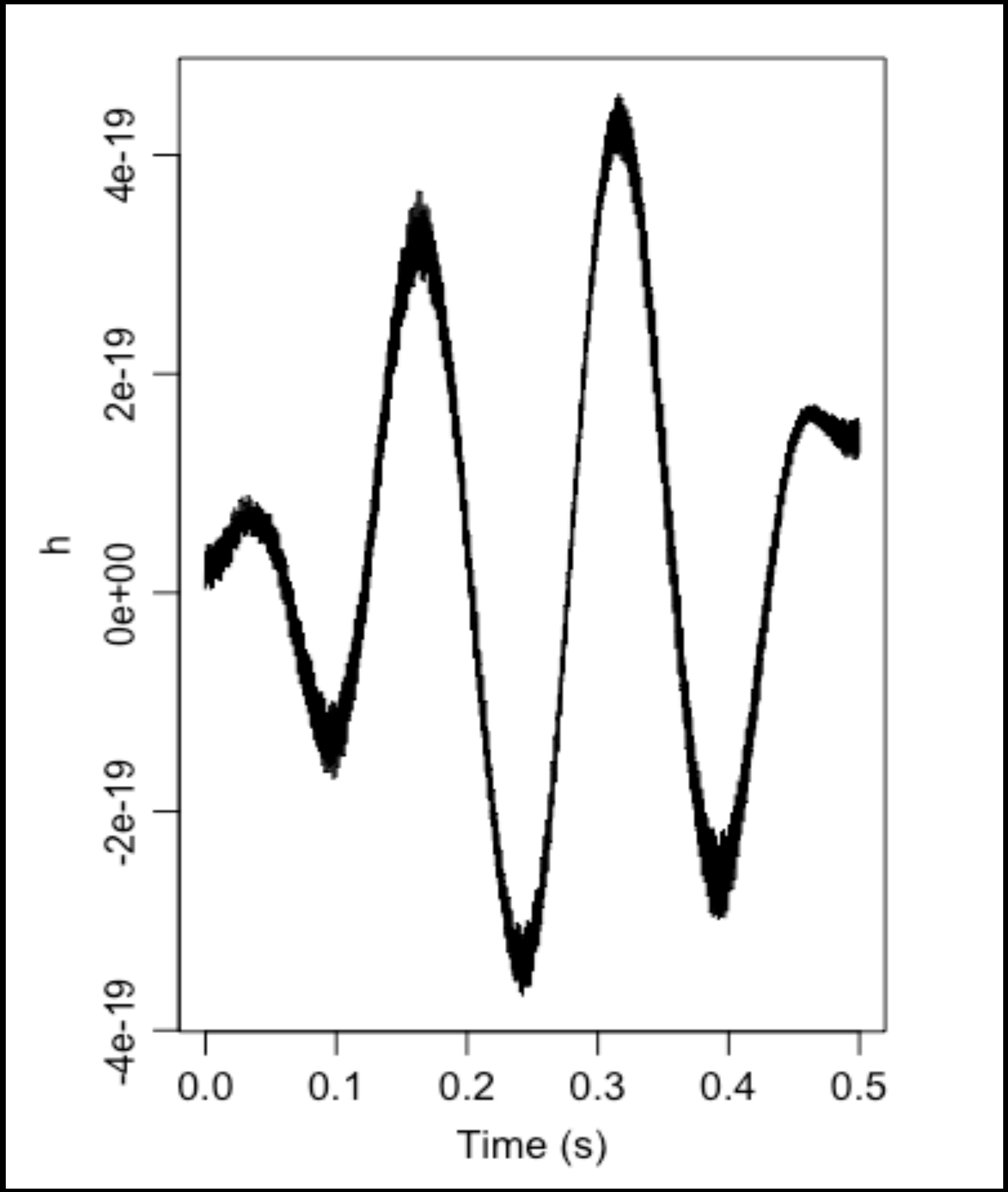


TEST CASE#2: LIGO DATA OF GW150914



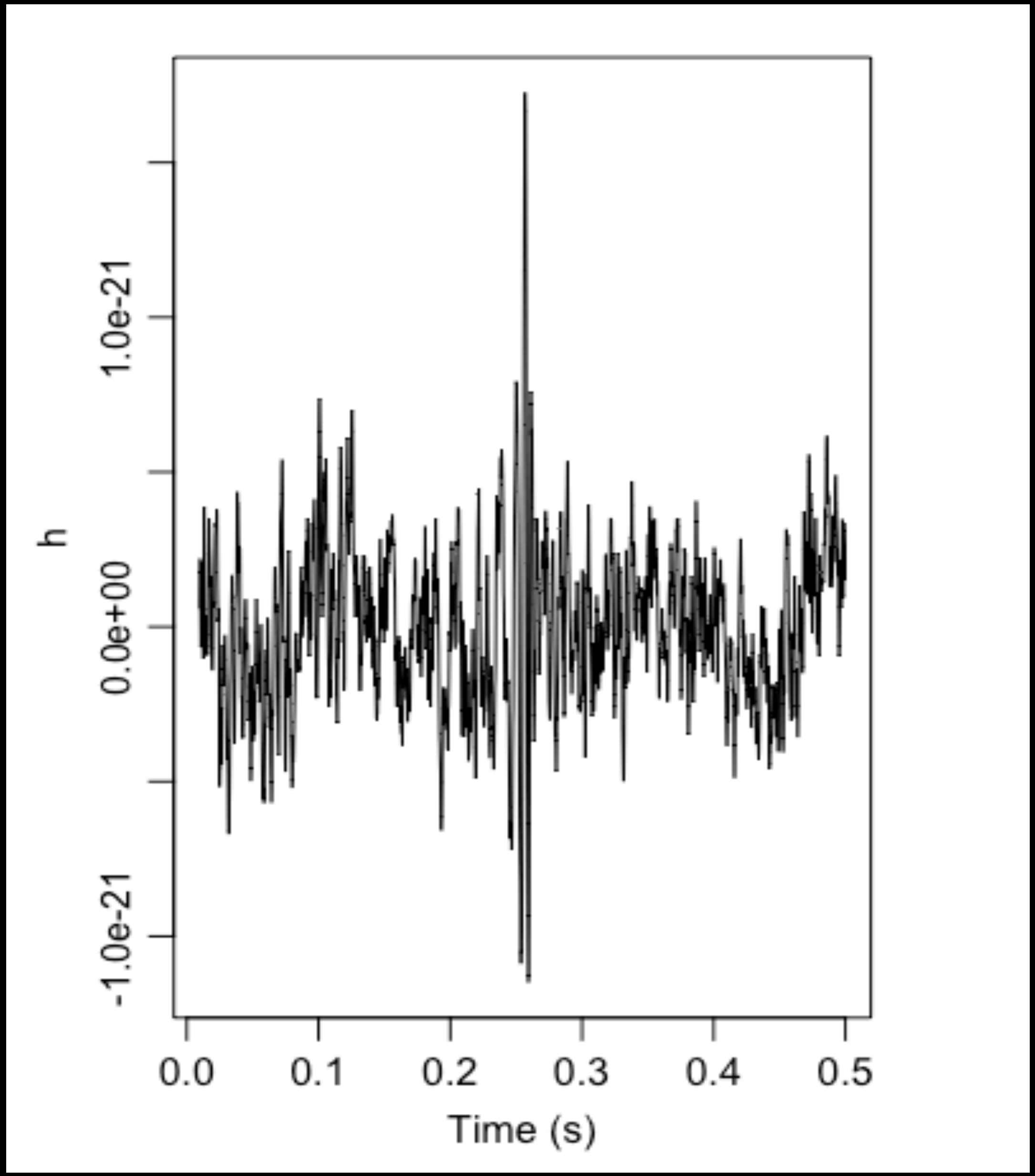
TEST CASE#2: LIGO DATA OF GW150914

I. LIGO Hanford



Raw data

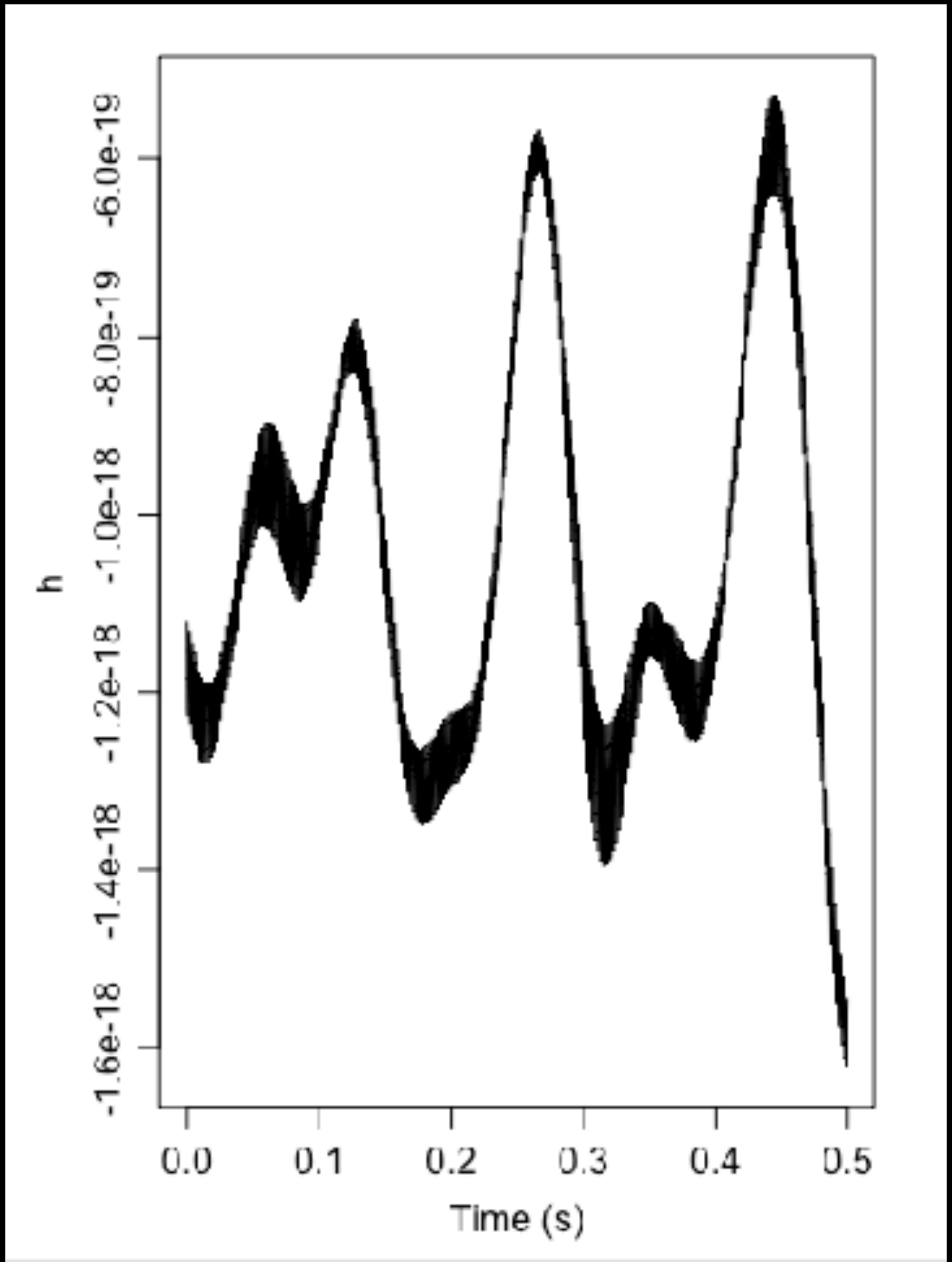
Subtract AR(31) + Low-Pass filter



Residuals

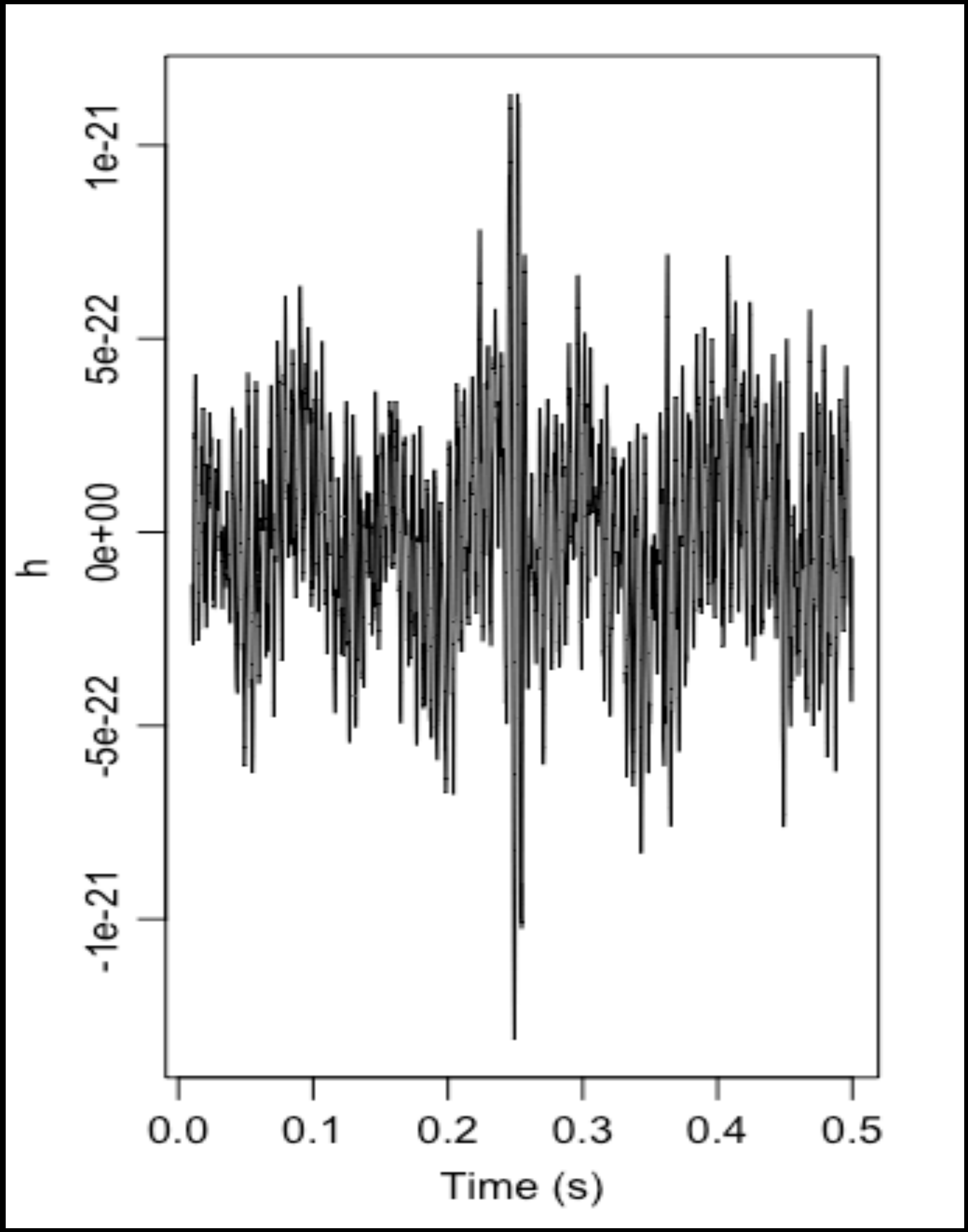
TEST CASE#2: LIGO DATA OF GW150914

II. LIGO Livingston



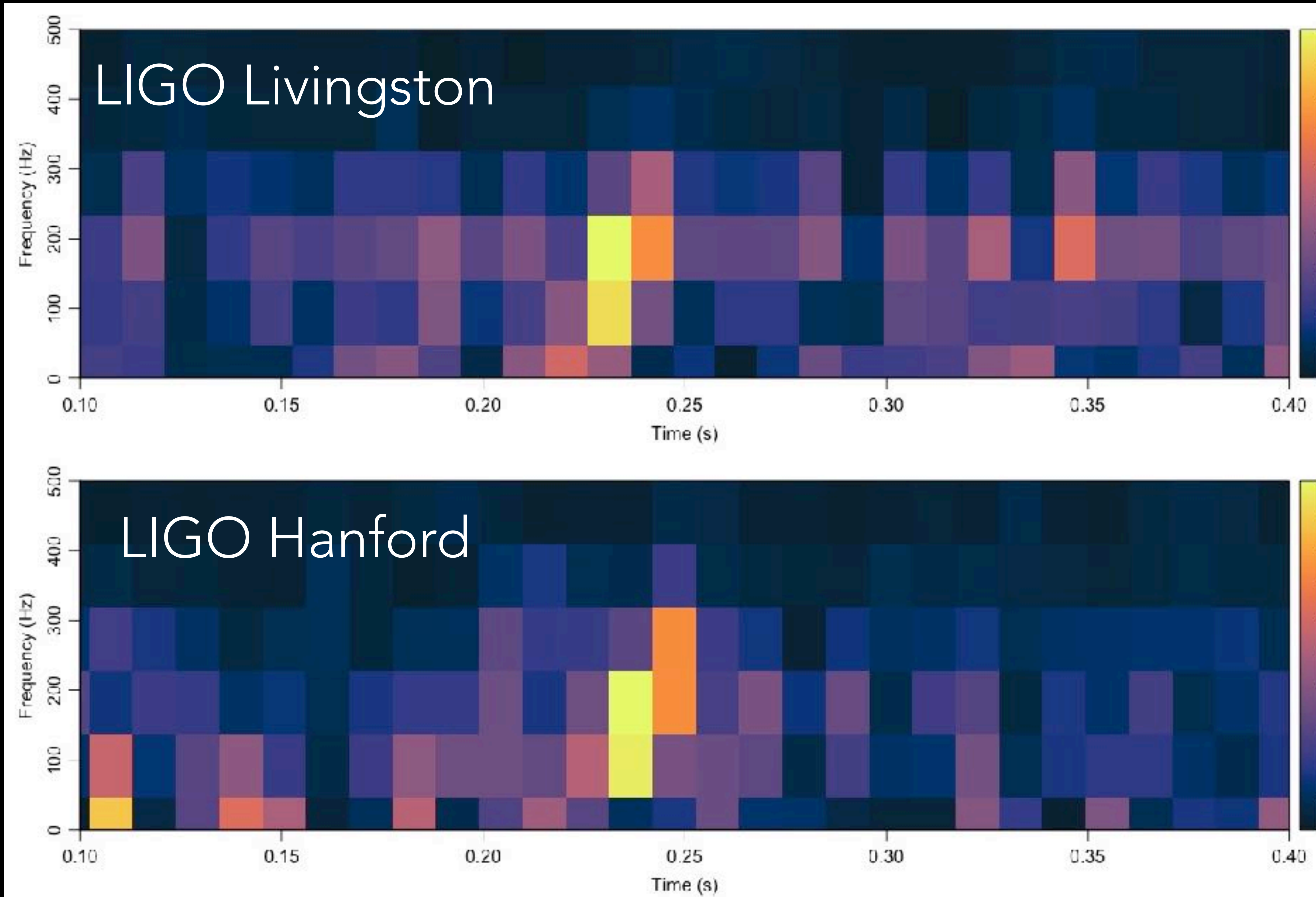
Raw data

Subtract
AR(33)
+
Low-Pass
filter



Residuals

TEST CASE#2: LIGO DATA OF GW150914

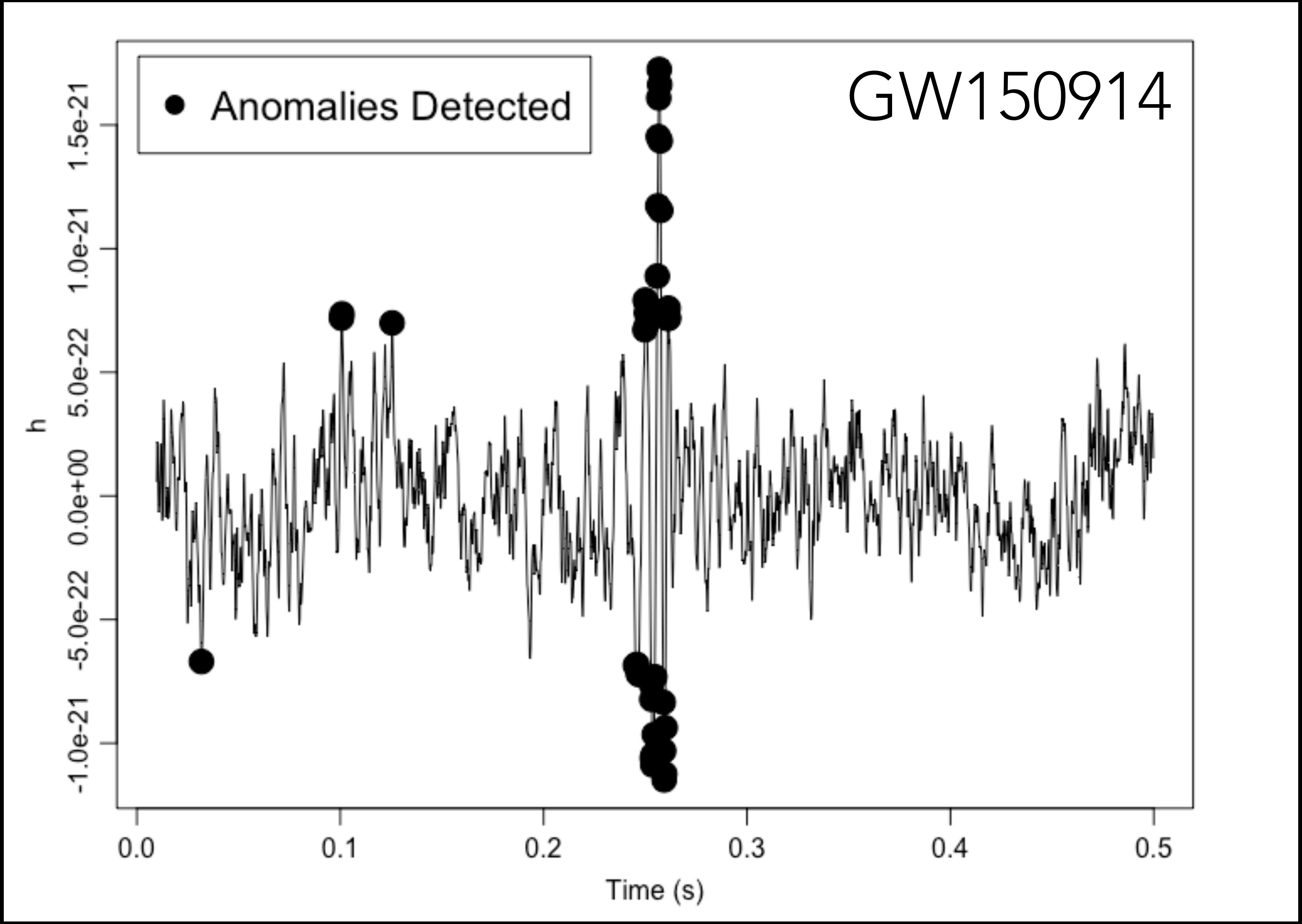


UNSUPERVISED ANOMALY DETECTION

- After the noise subtraction, events candidates can be identified as anomalies, which differ from normal instances significantly.
- Anomalies detected from different detectors (LIGO-H, LIGO-L,KAGRA,VIRGO) can be cross-correlated and analysed with clustering technique.
- The shortlisted anomalies can be taken as event candidates for further analysis.

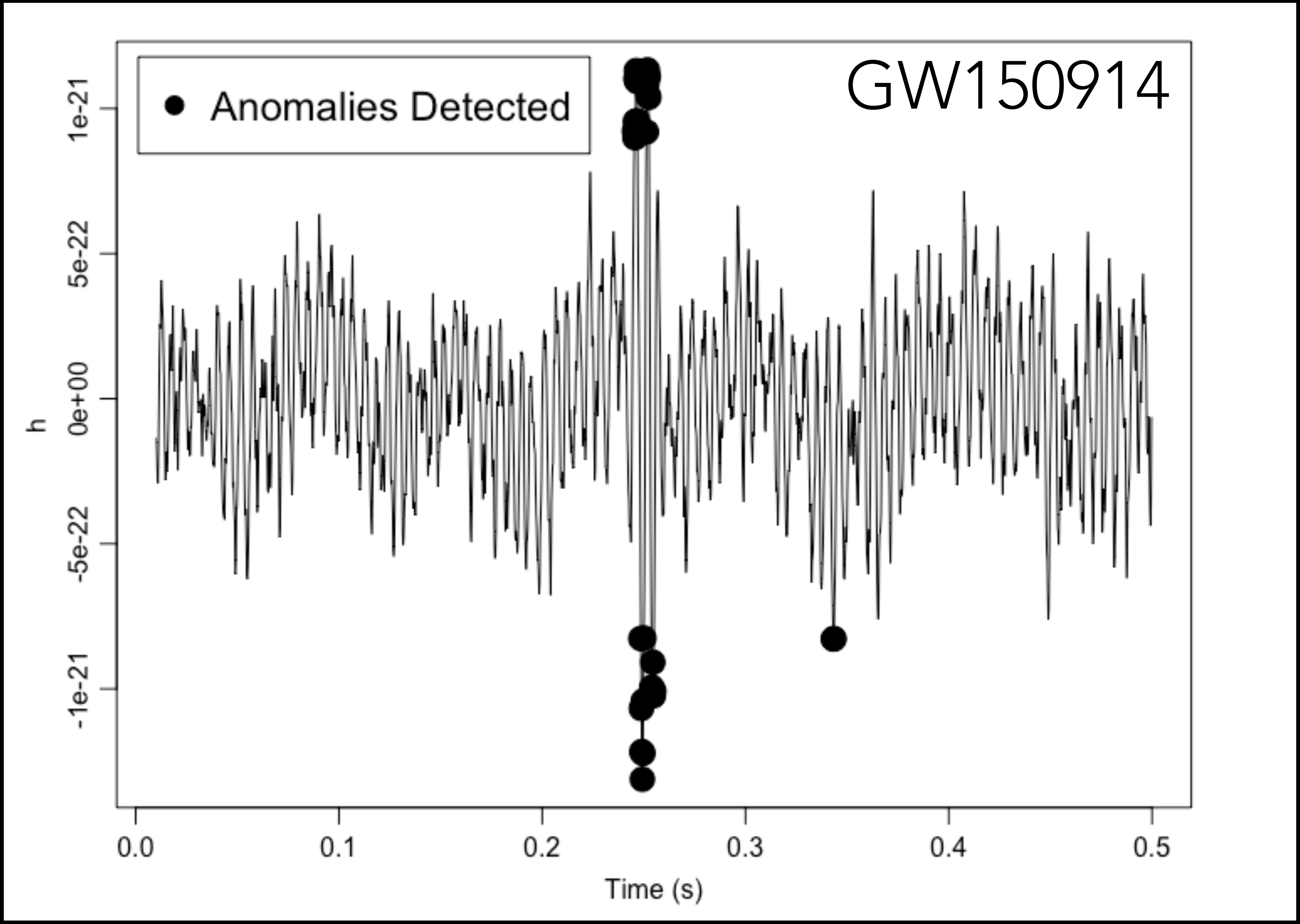
UNSUPERVISED ANOMALY DETECTION

I. LIGO Hanford



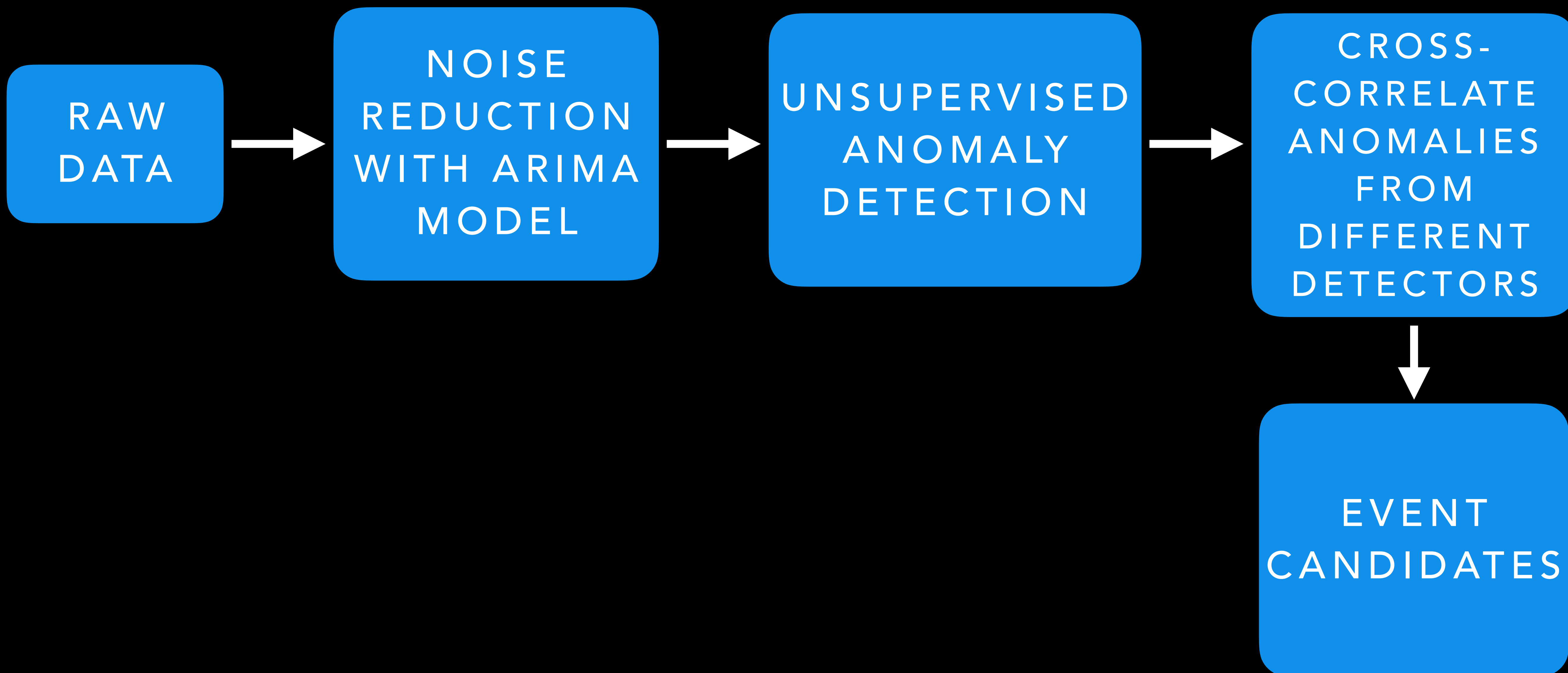
UNSUPERVISED ANOMALY DETECTION

II. LIGO Livingston



SUMMARY & OUTLOOK

We propose a computationally efficient pipeline for handling GW data.

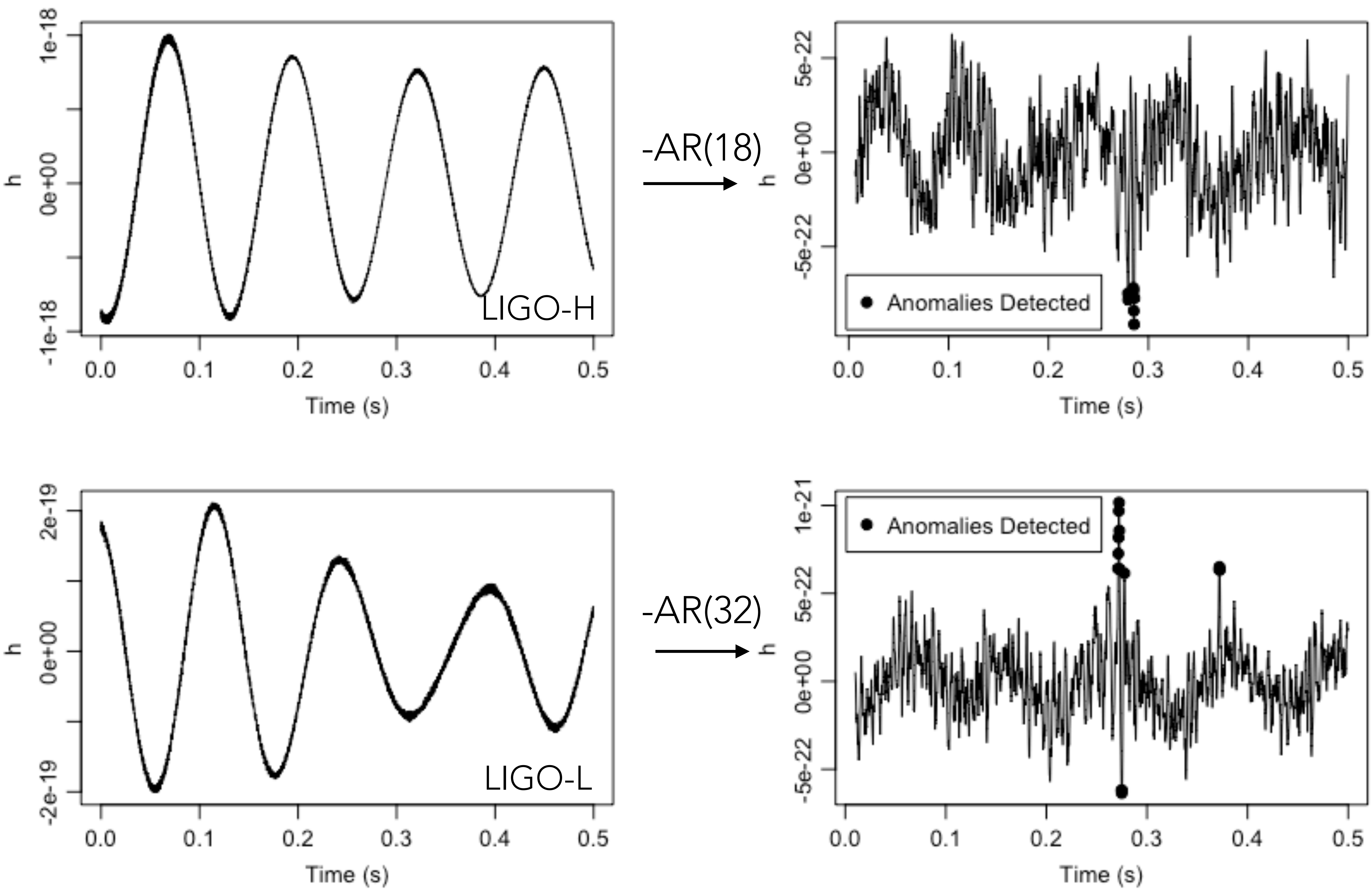


Further work will be devoted to improve the performance of anomaly detection with machine learning techniques (e.g. autoencoder).

THANK YOU VERY MUCH

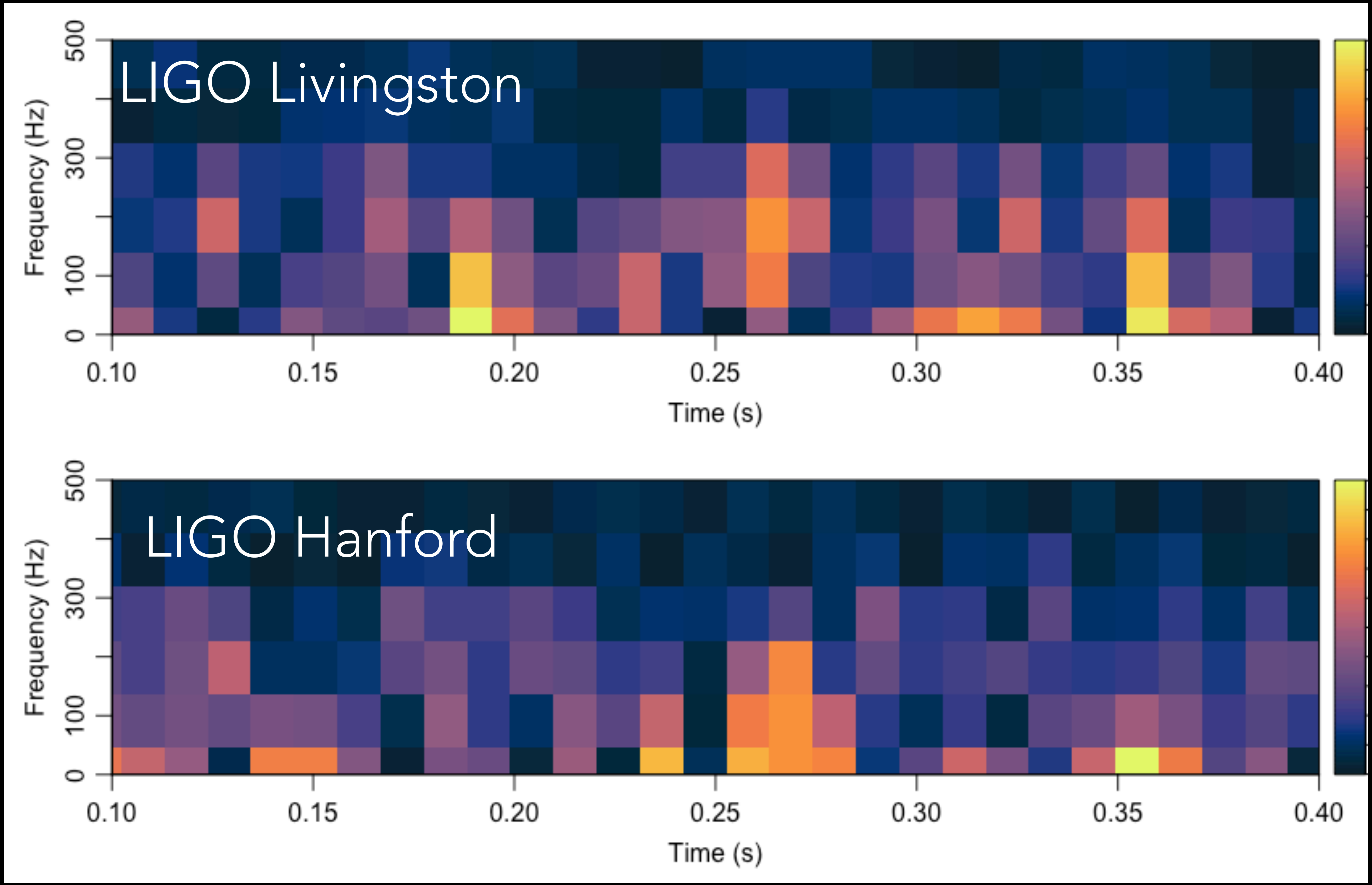
ADDITIONAL SLIDE:
TEST CASE #3

GW170814



h
1e-18
0e+00
-1e-18
0.0
0.1
0.2
0.3
0.4
0.5
Time (s)
LIGO-H
-AR(18)
h
5e-22
0e+00
-5e-22
Anomalies Detected
0.0
0.1
0.2
0.3
0.4
0.5
Time (s)
h
2e-19
0e+00
-2e-19
0.0
0.1
0.2
0.3
0.4
0.5
Time (s)
LIGO-L
-AR(32)
h
1e-21
5e-22
0e+00
-5e-22
Anomalies Detected
0.0
0.1
0.2
0.3
0.4
0.5
Time (s)

GW170814


LIGO Livingston
Frequency (Hz)
500
300
100
0
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Time (s)
LIGO Hanford
Frequency (Hz)
500
300
100
0
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Time (s)