SEARCHING FOR GRAVITATIONAL WAVES THROUGH AN AUTOREGRESSIVE APPROACH

DAVID C. Y. HUI

DEPARTMENT OF ASTRONOMY & SPACE SCIENCE
CHUNGNAM NATIONAL UNIVERSITY
SOUTH KOREA

8th KAGRA International Workshop 7-9 July 2021

COLLABORATORS

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

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

- Computationally simple
- Capable to handle various kinds of noise from non-stationary autocorrelated stochastic processes.
- Many applications in diverse fields (e.g. ECG, econometrics), but not many in astronomy until recently (e.g. exoplanet search)

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

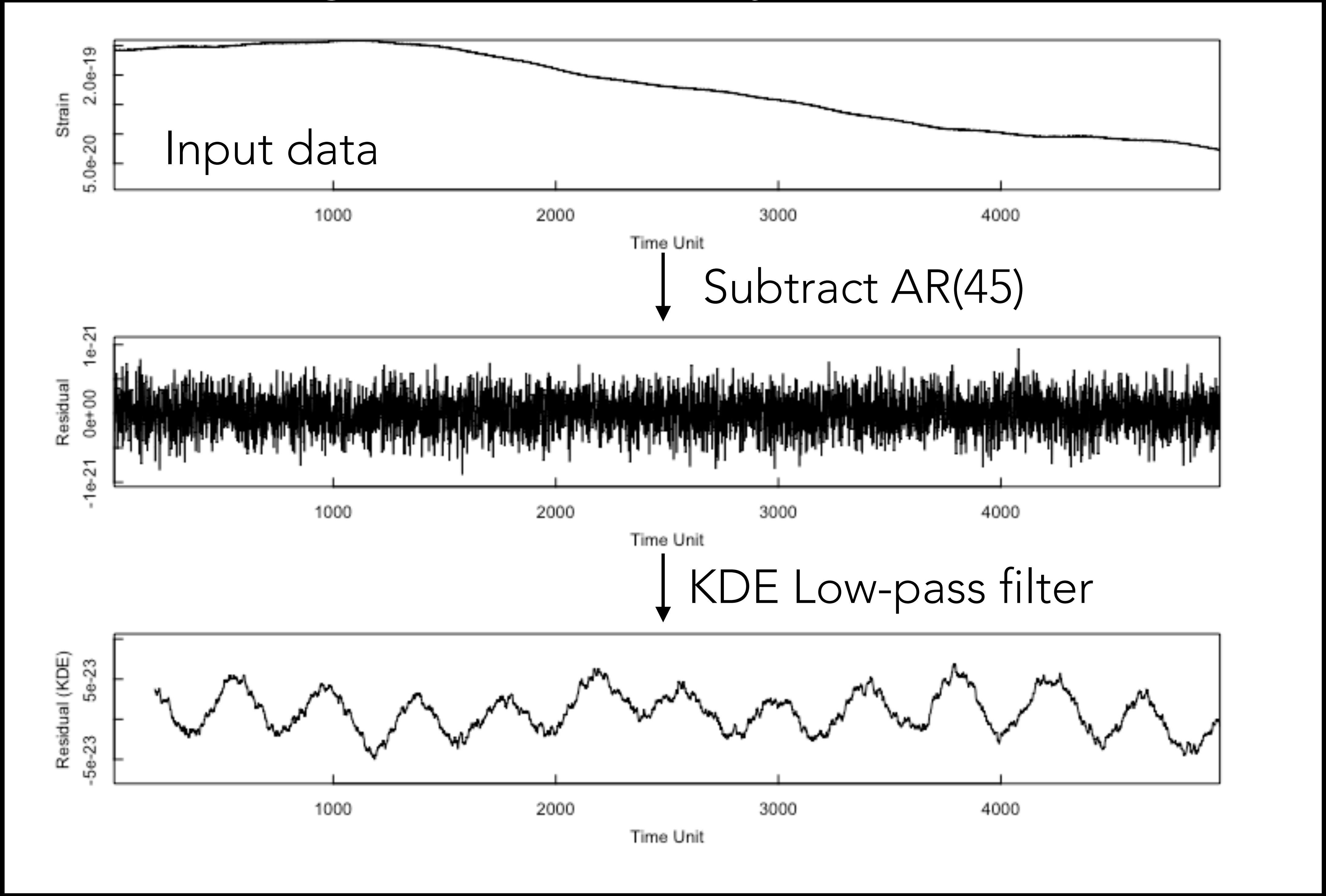
Combining the autoregressive (AR), moving average (MA) and integrated (I) processes together into a single regression procedure, we have ARIMA(*p*,*q*,*d*) model:

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

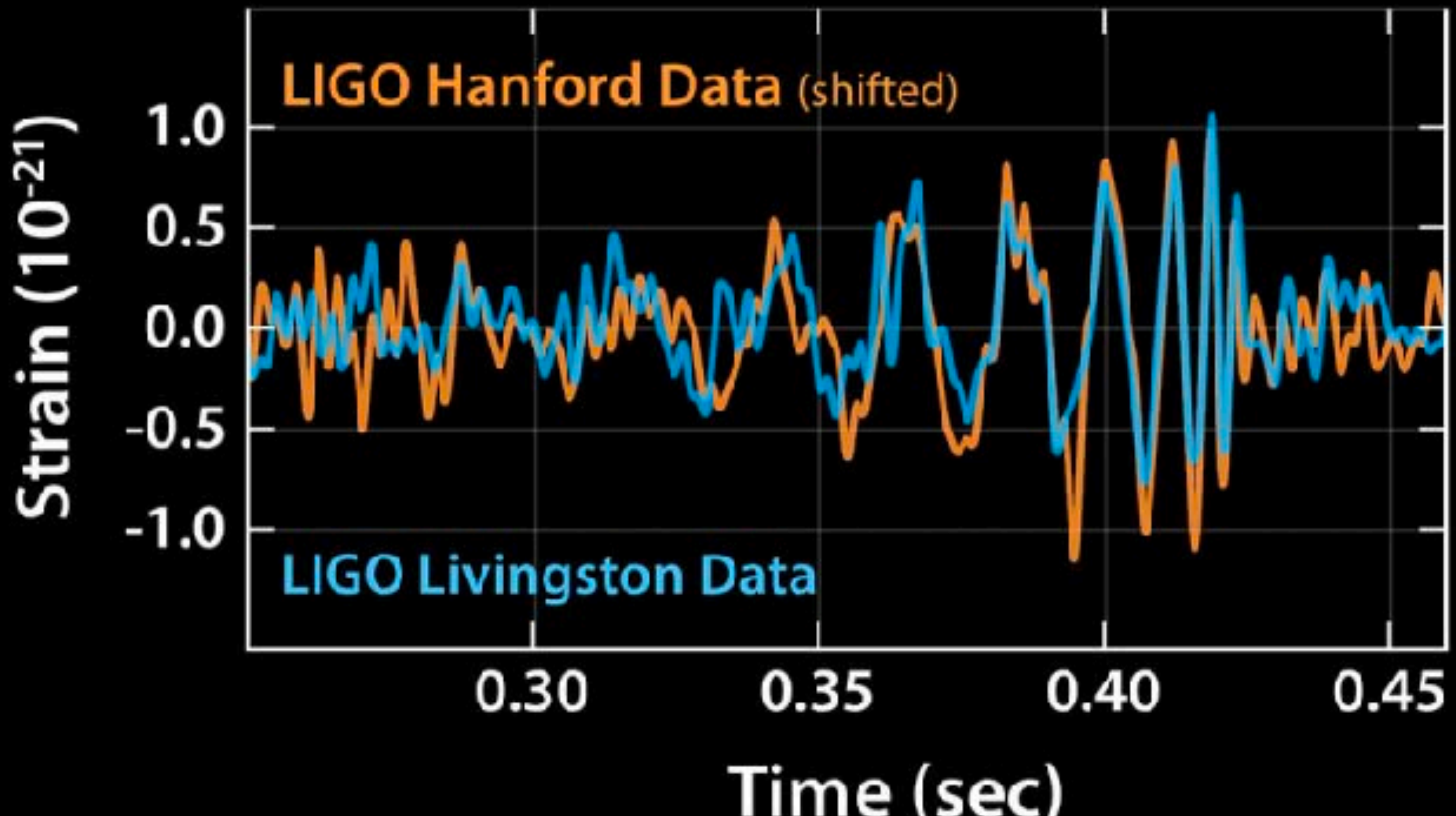
The model parameters can be determined by maximum likelihood estimation with the orders *p* and *q* determined through certain information criterion (e.g. BIC, AIC).

PROOF-OF-CONCEPT

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

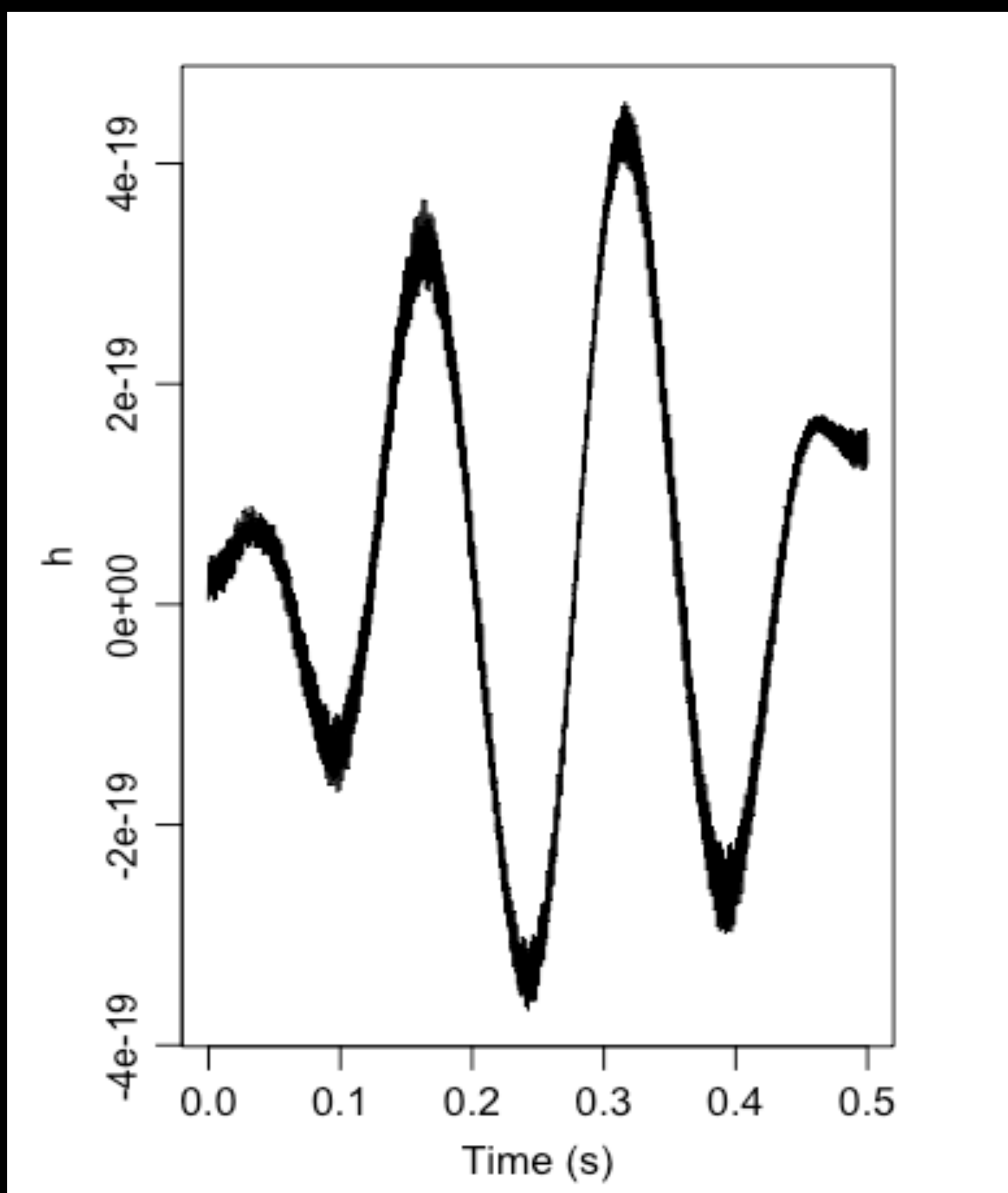


LIGO DATA OF GW150914



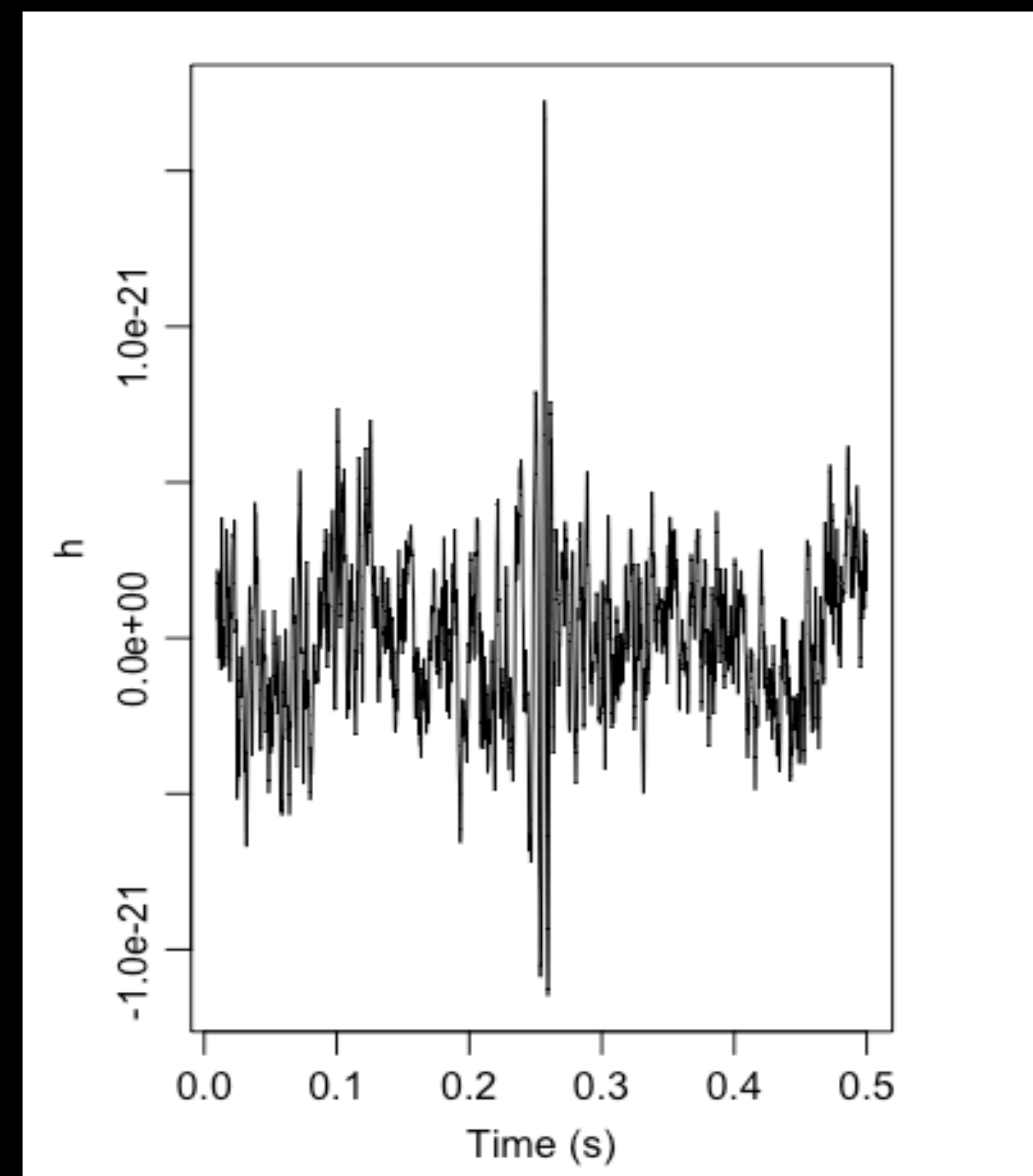
QUICK LOOK OF GW150914

I. LIGO Hanford



Raw data

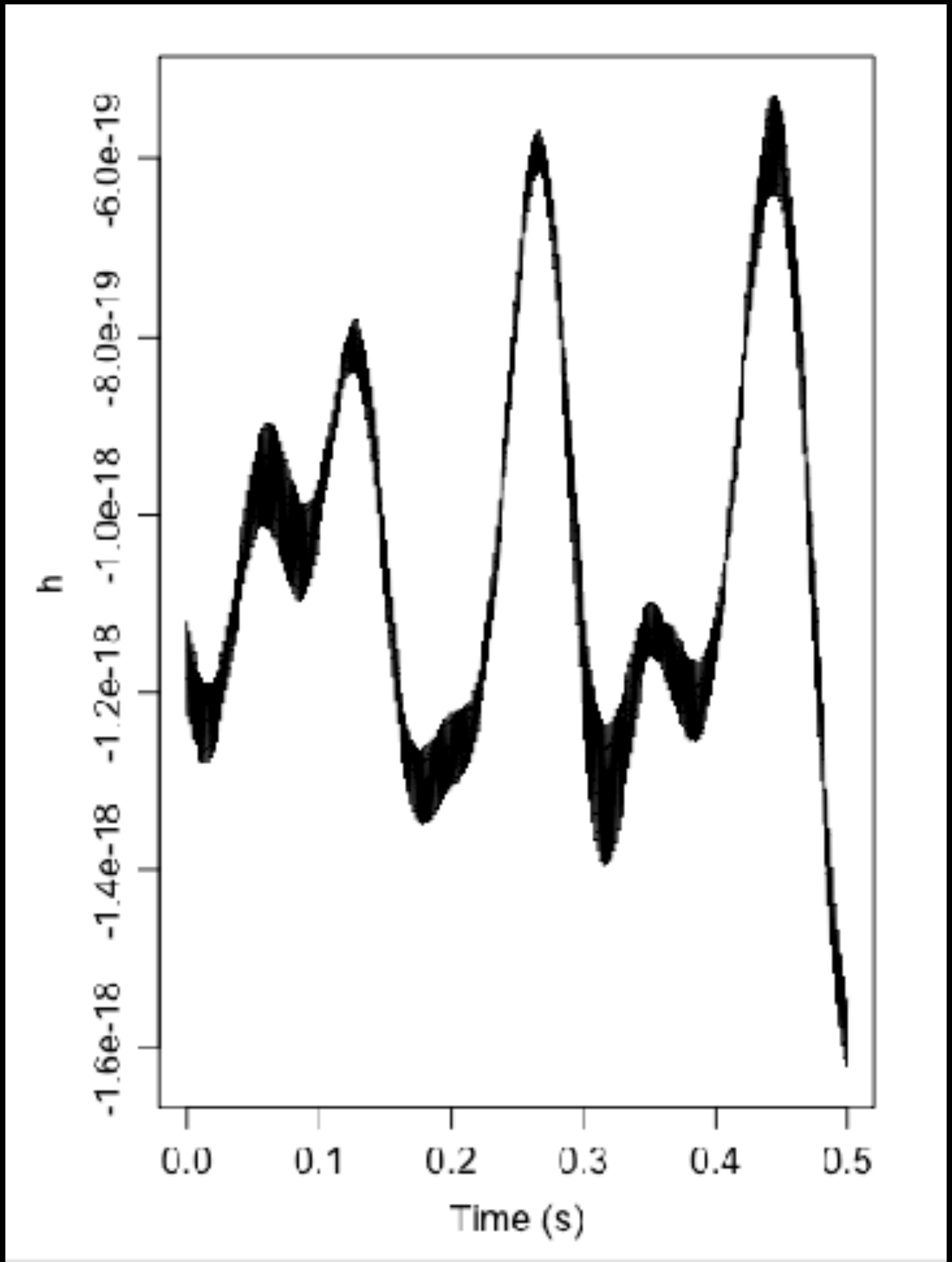
Subtract AR(31) + Low-Pass filter



Residuals

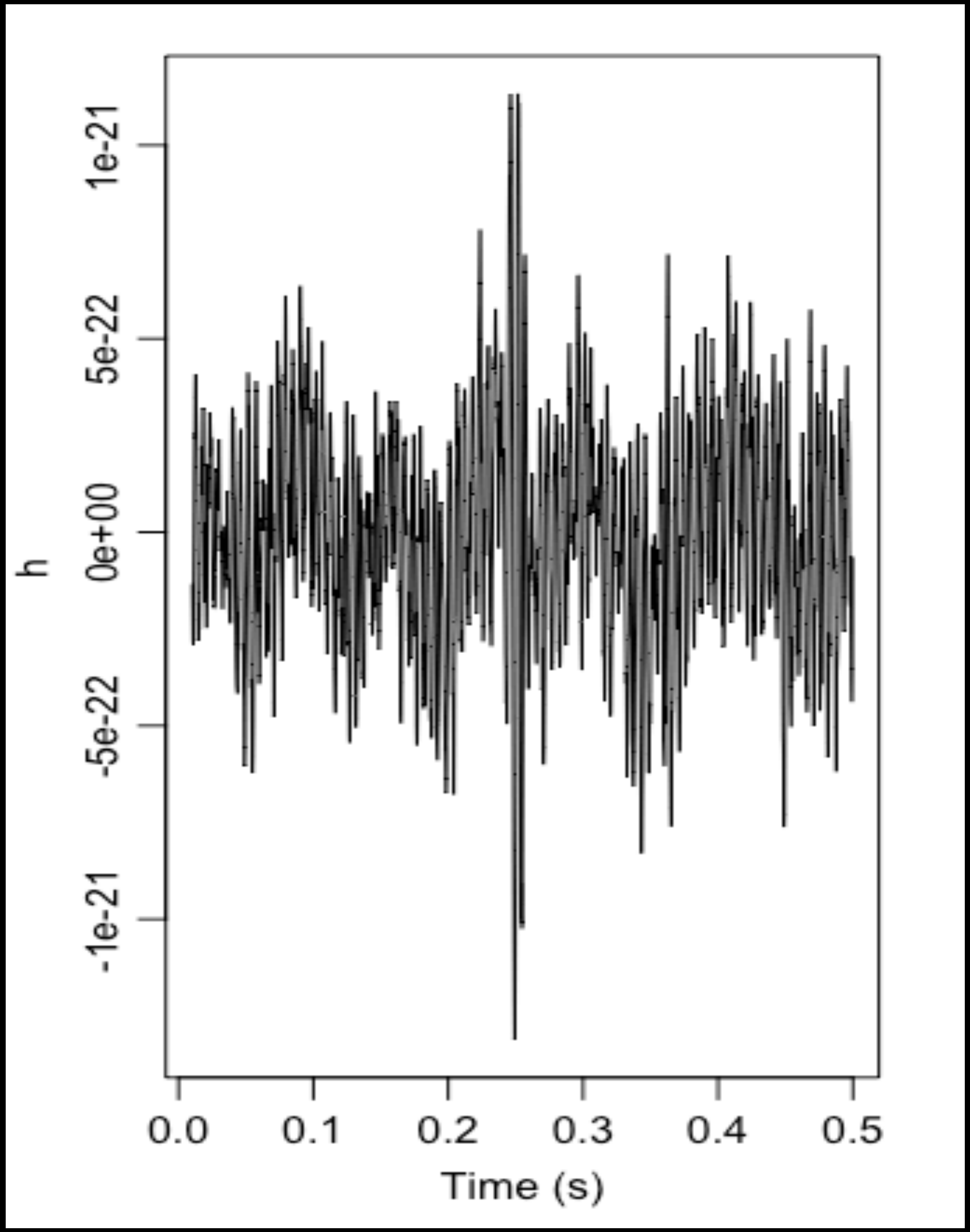
QUICK LOOK OF GW150914

II. LIGO Livingston



Raw data

Subtract
AR(33)
+
Low-Pass
filter



Residuals

EFFECTIVENESS OF AR MODEL

- A diversity of noises (e.g. glitches) are subsumed into AR model without any fine-tuning and a priori knowledge of the noise nature

- A desirable feature of this method is that the transient GW signals were NOT absorbed by AR model.

- AR is a maximum likelihood estimation procedure which weight all data points equally. As the transient signal is a small fraction in a given window, their data points are essentially ignored in the model.

OPTIMIZING S/N RATIO

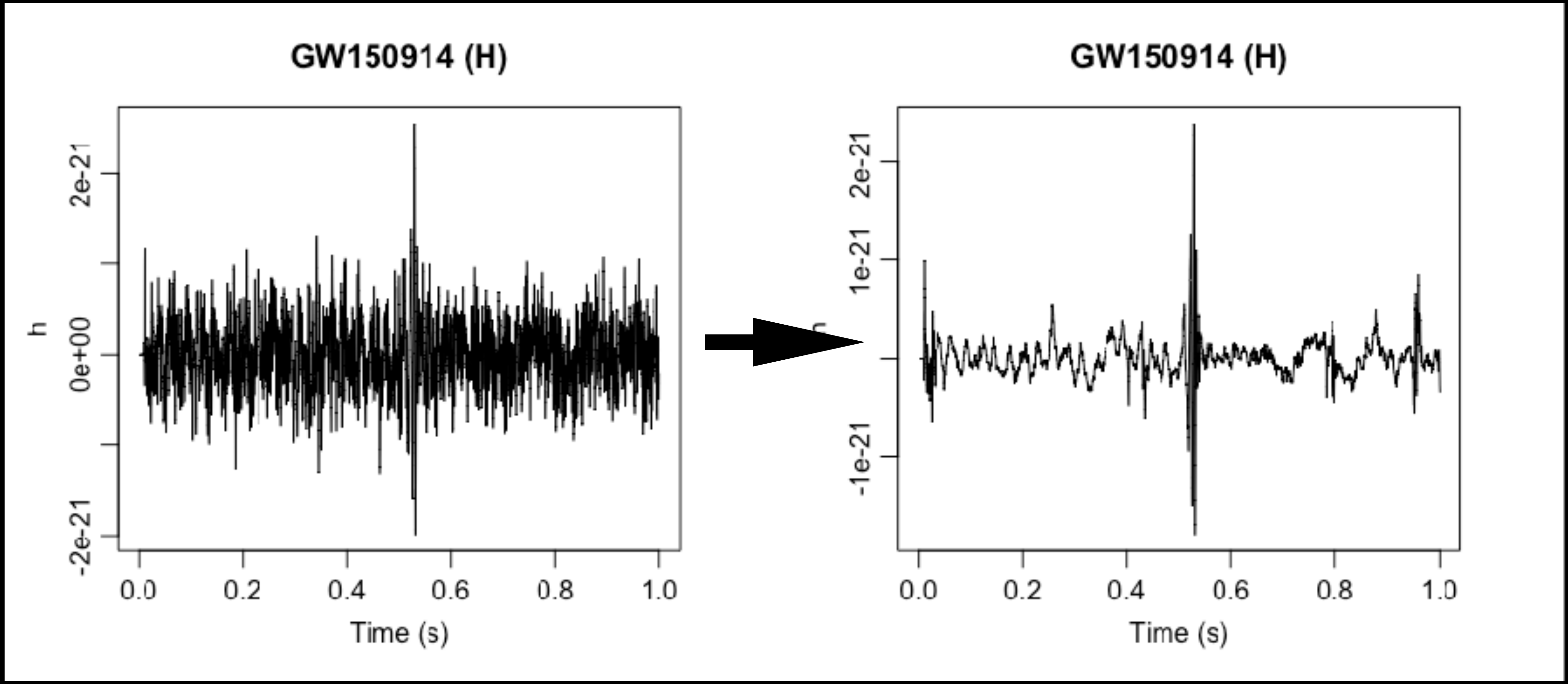
- Kernel size of KDE
- Range of width in an adaptive filter (max./min. Widths)
- Window size

We search for a combination of these hyper-parameters which results in the optimal peak S/N ratio.

ADAPTIVE FILTERING

- Low frequency noise in the residuals can be further removed by using **repeated median regression** (Siegel 1982)

- Low frequency variation is estimated by median of a sliding window with adaptive size (Gather & Fried 2004).



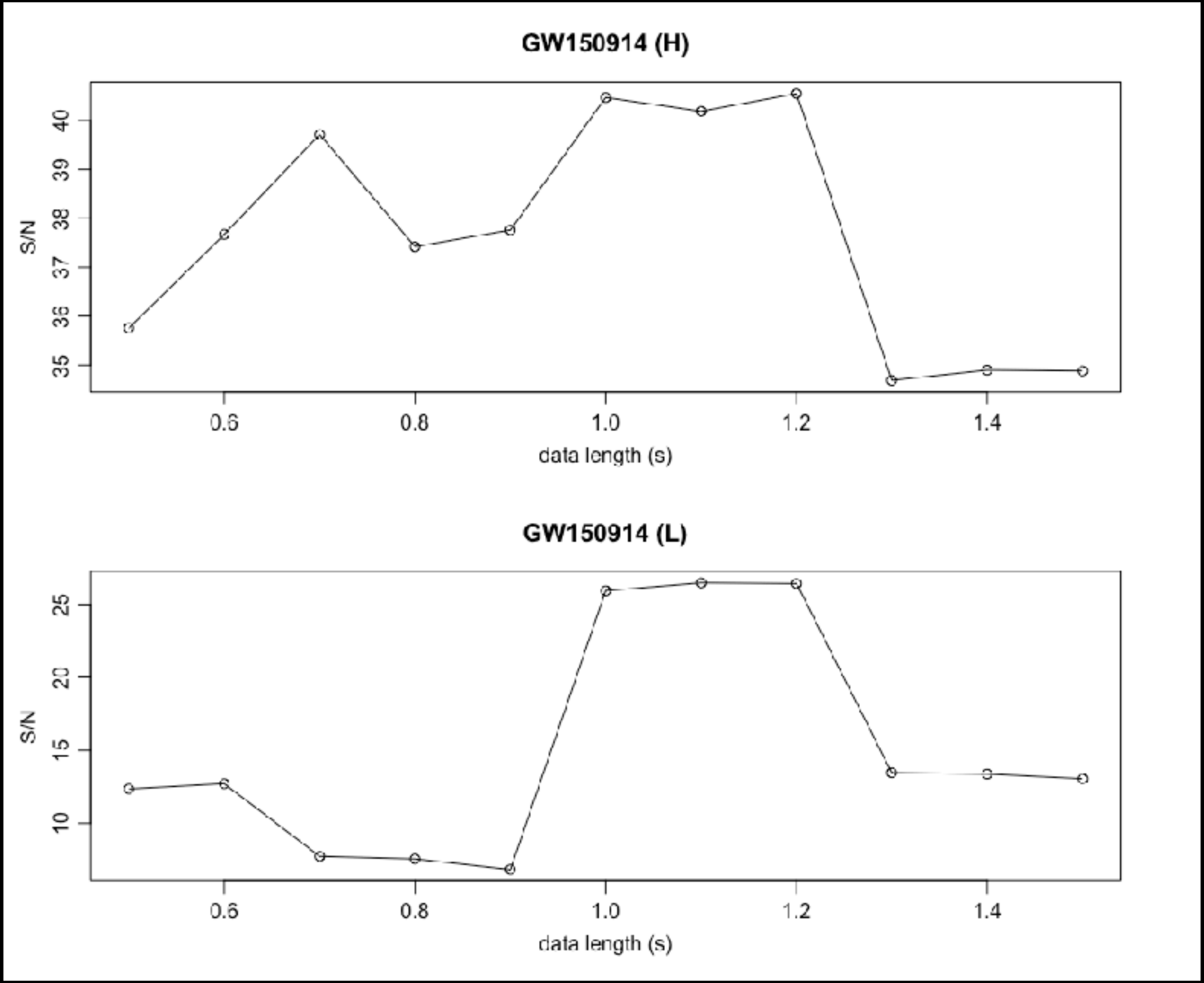
EFFECTS OF WINDOW SIZE

- Depend on the nature of the event one intend to search (BH-BH meger, NS-NS merger)
- Large window results in small variance and smooth model (*May under estimate the noise*)
- Small window results in small bias and adapts quick to changes. (*May absorb the potential signal*)

↓

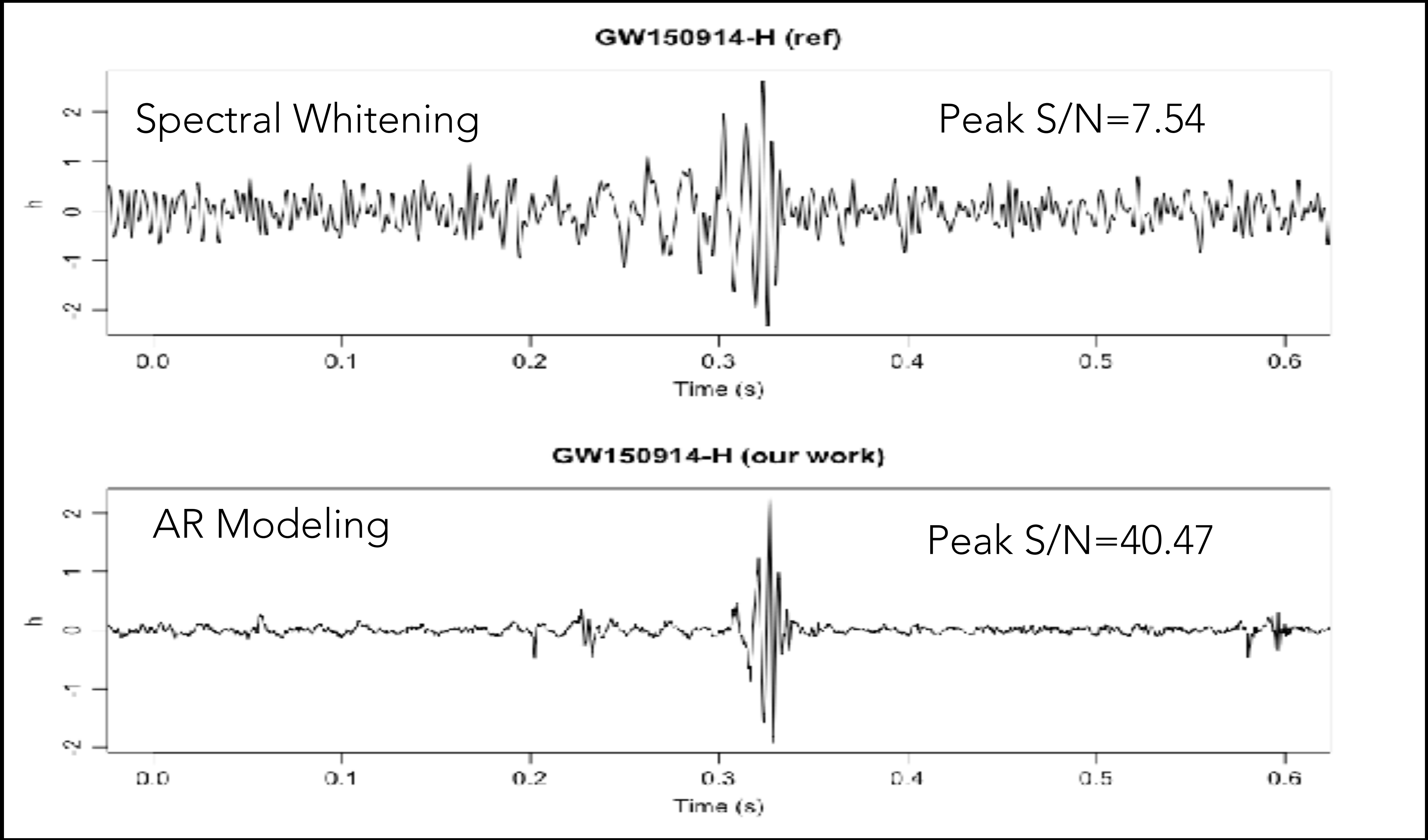
There should be an optimal window size.

EFFECTS OF WINDOW SIZE



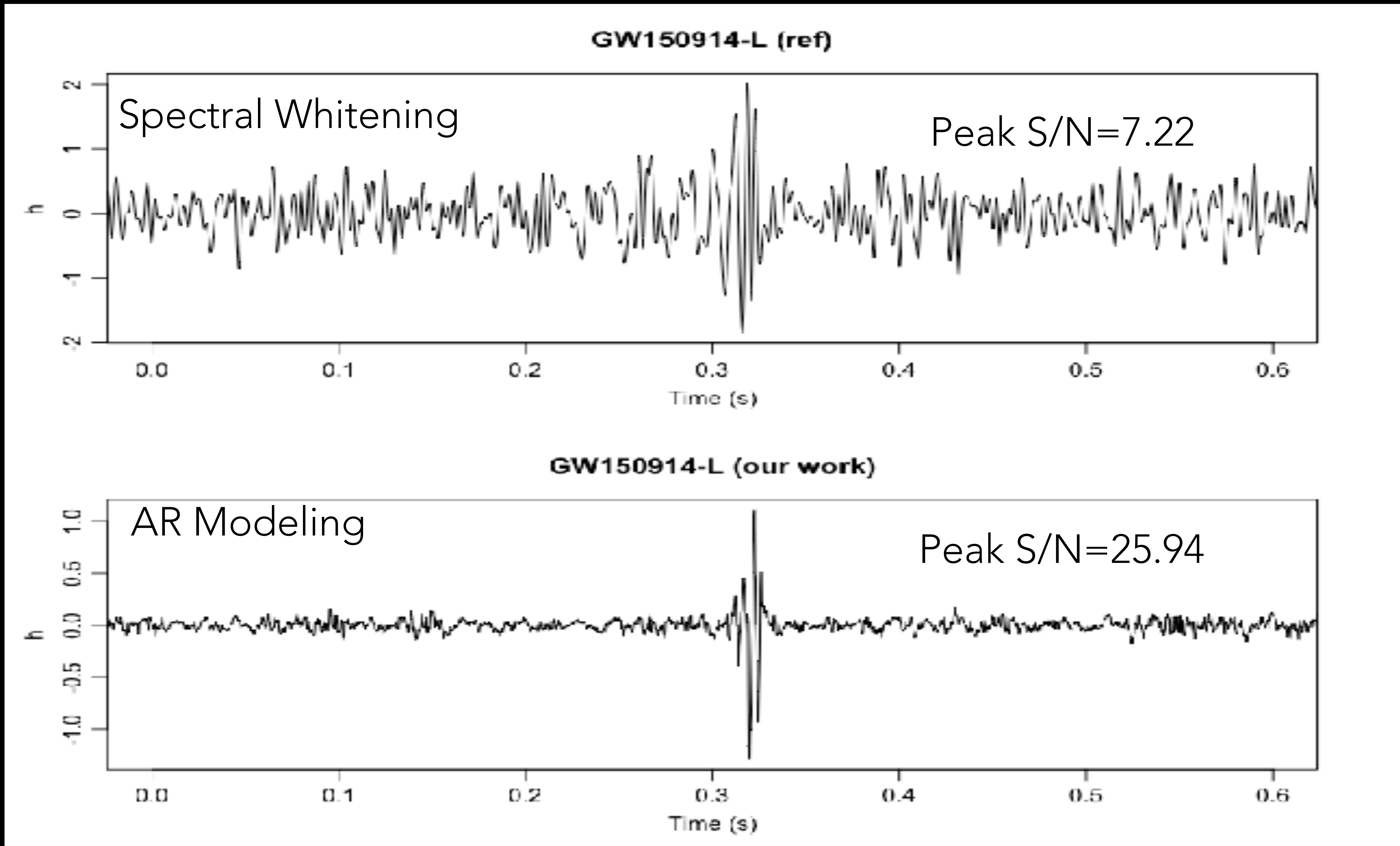
COMPARISON WITH SPECTRAL WHITENING

- Our framework is capable to attain a better S/N in comparison with the conventional spectral whitening

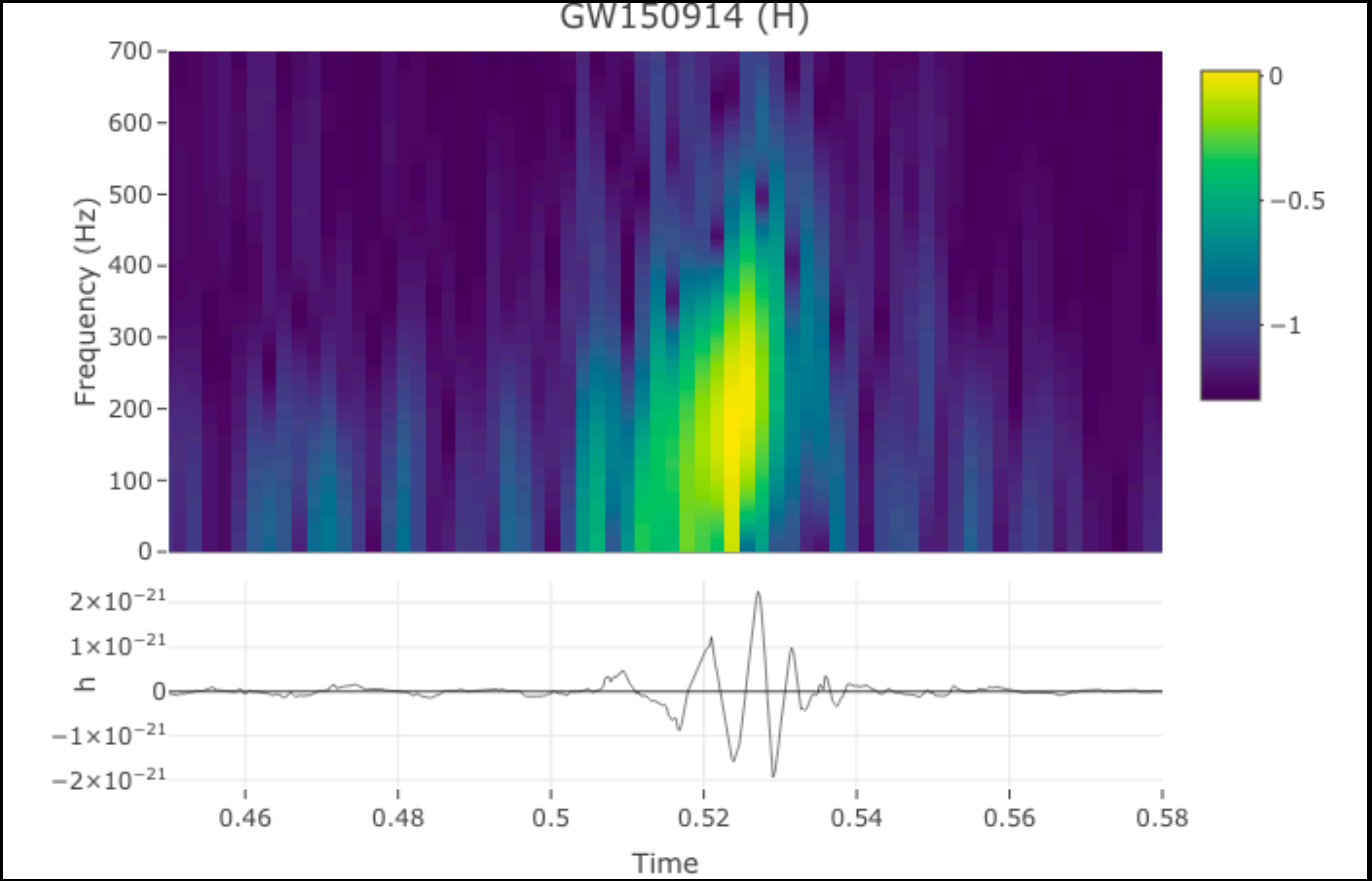


COMPARISON WITH SPECTRAL WHITENING

- Our framework is capable to attain a better S/N in comparison with the conventional spectral whitening

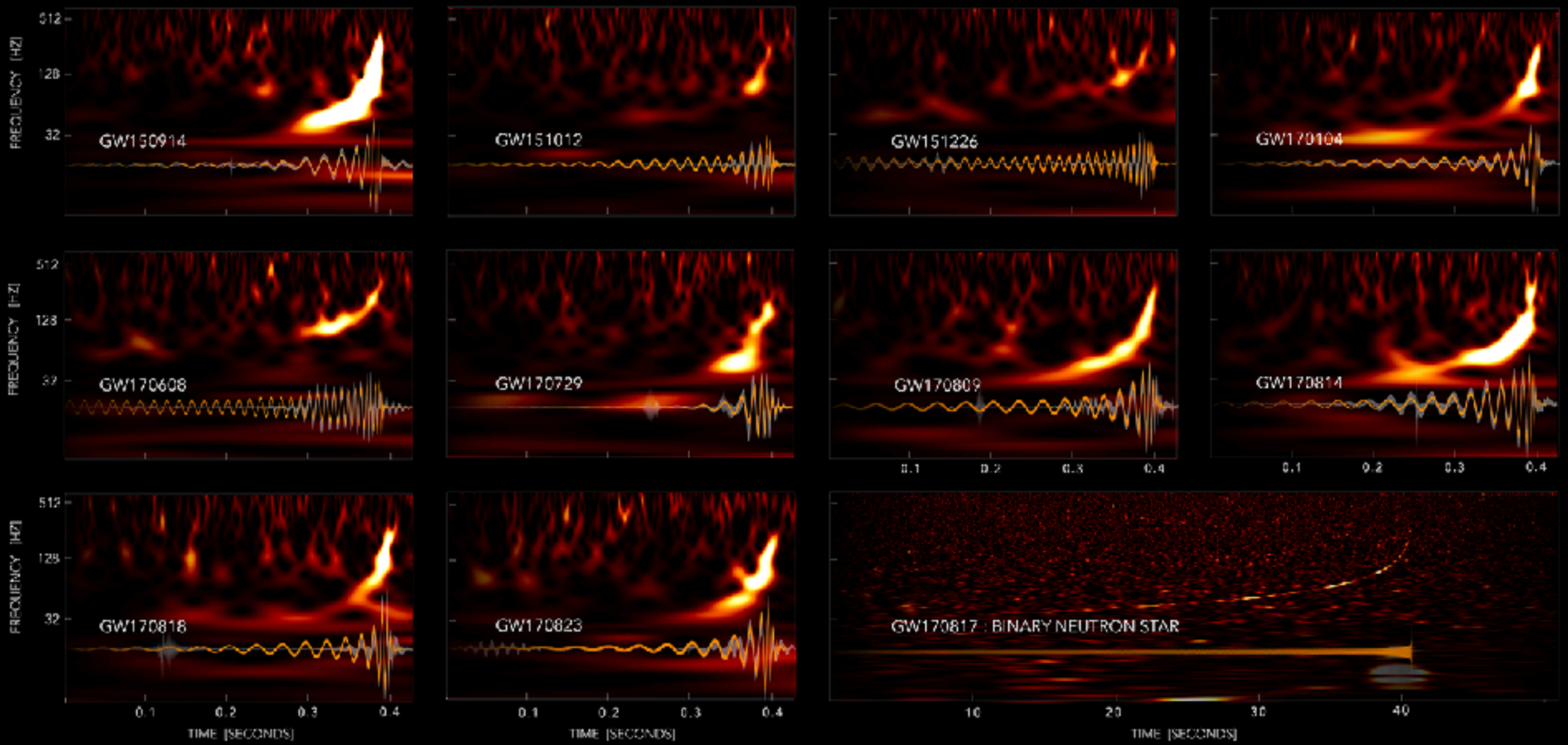


SPECTROGRAM OF OPTIMAL RESULT



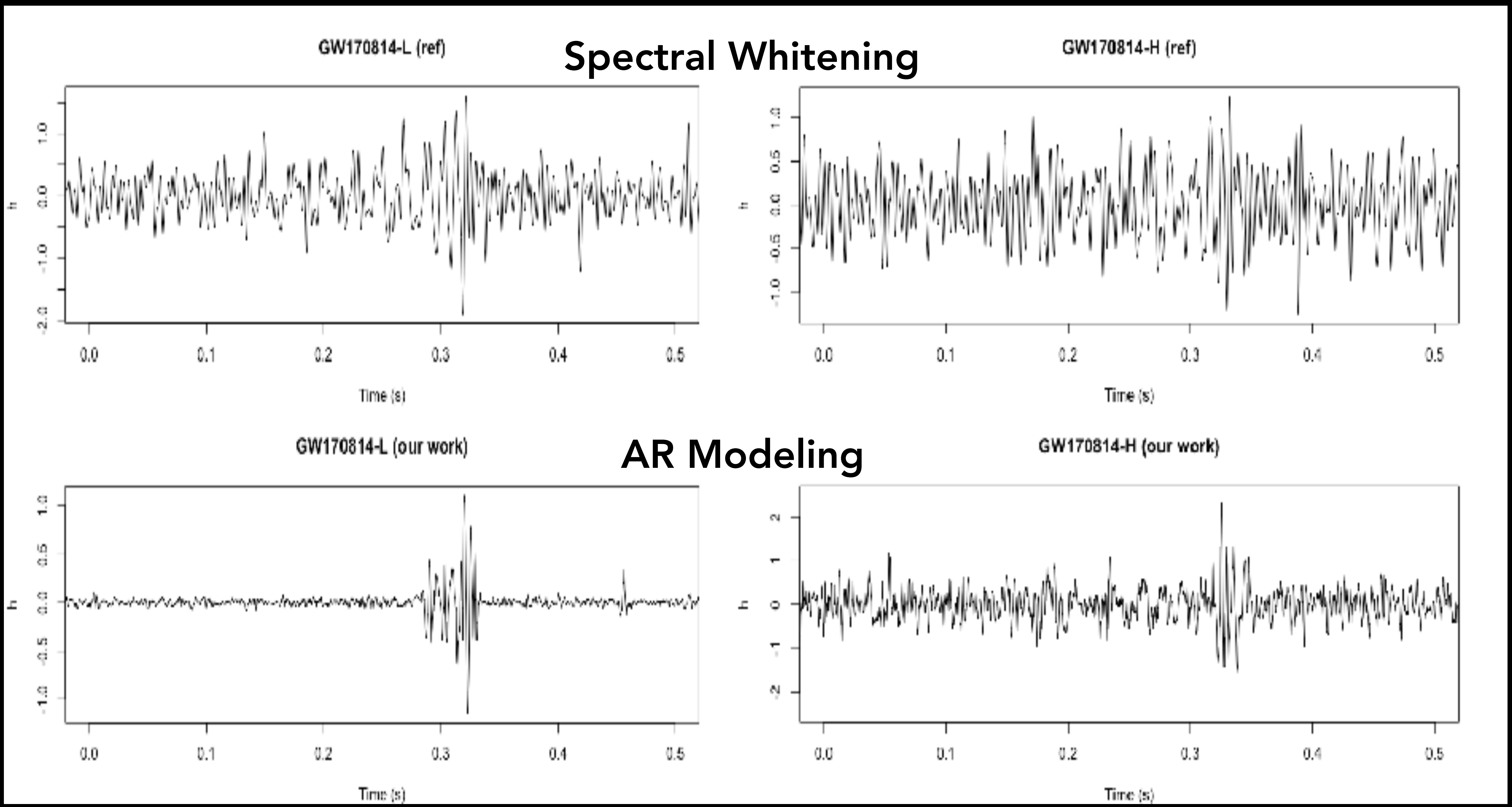
ON-GOING/FUTURE WORKS

1. Recover GWTC-1 events with AR modeling.



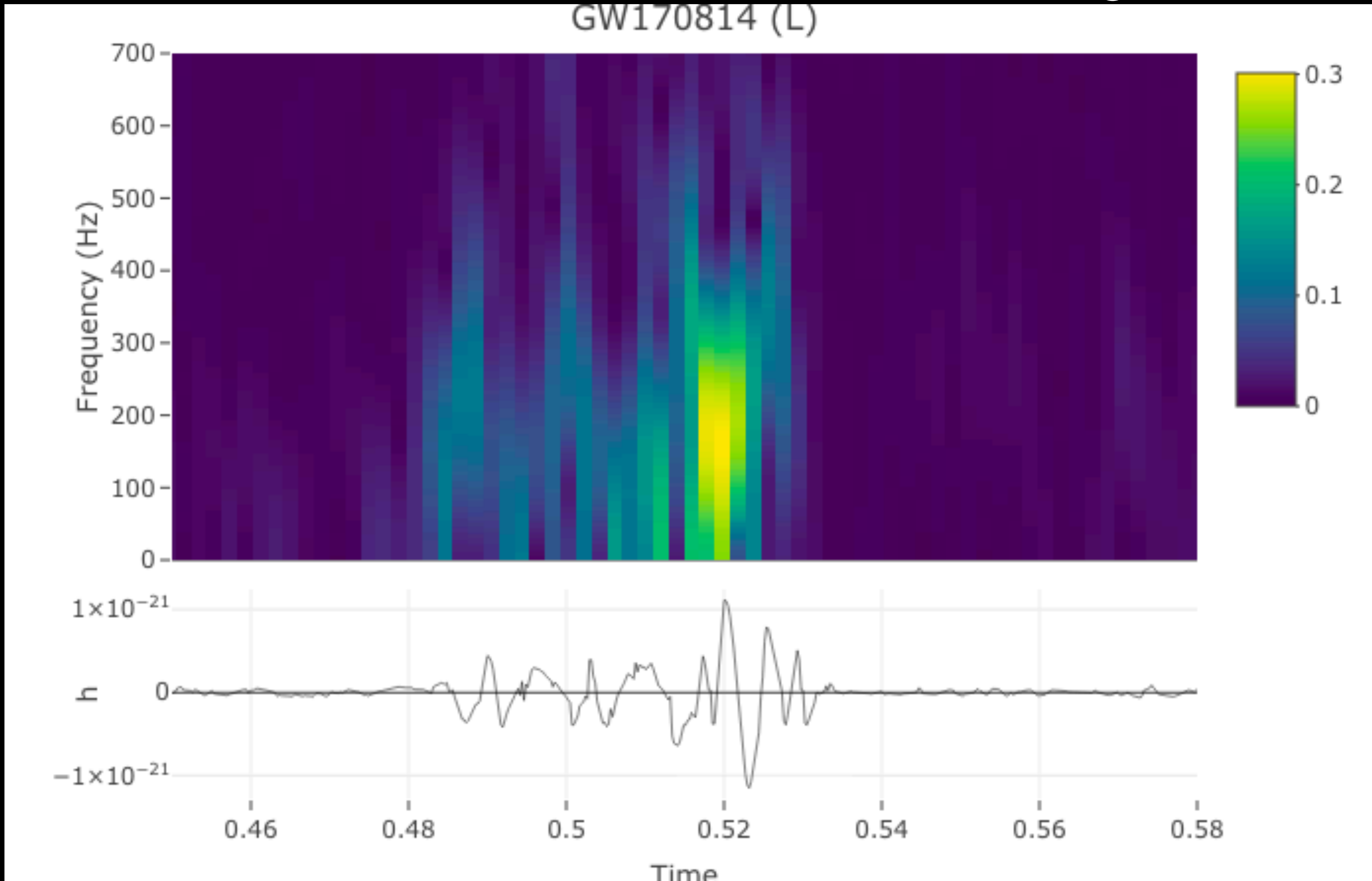
ON-GOING/FUTURE WORKS

1. Recover GWTC-1 events with AR modeling.



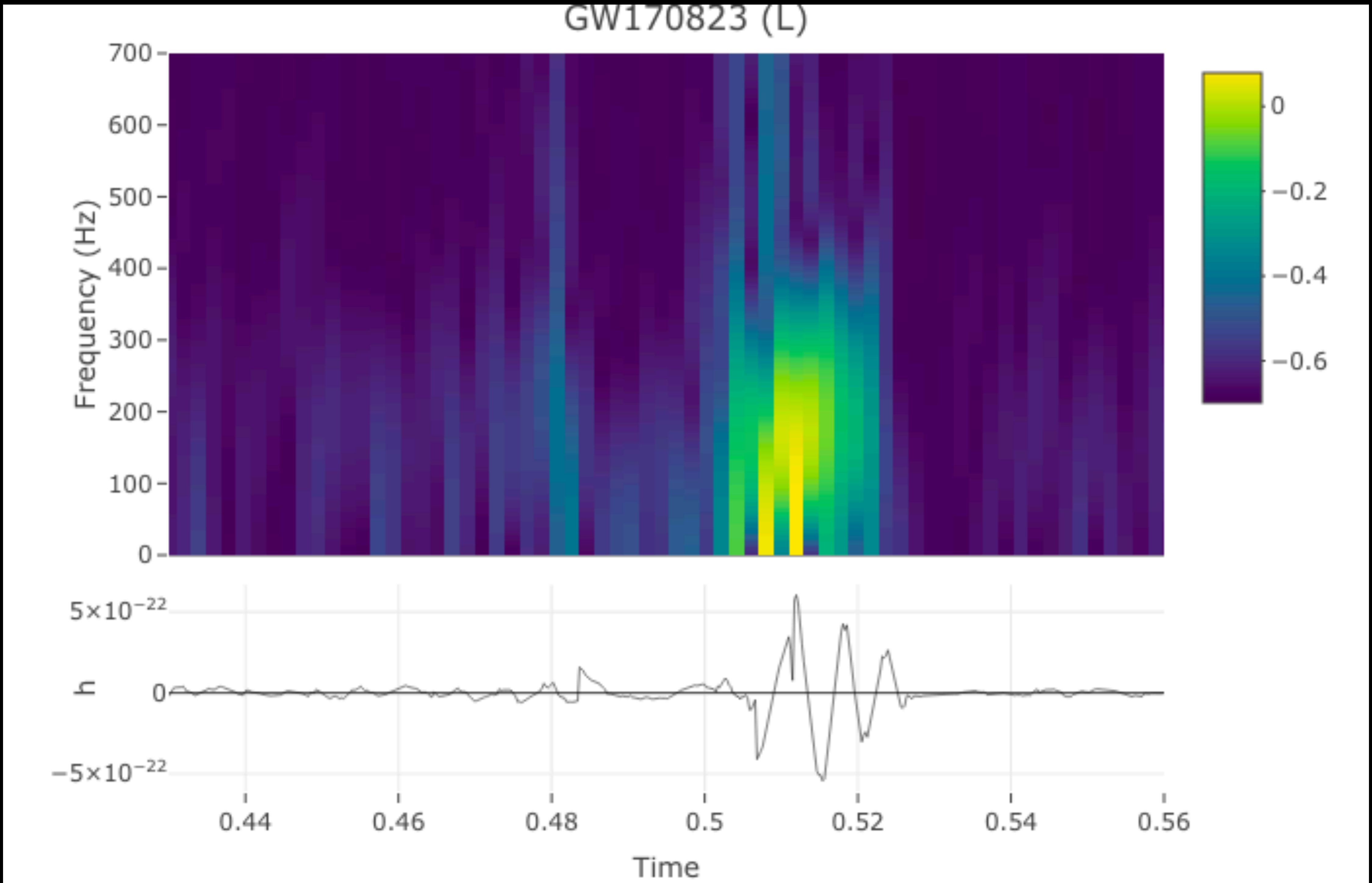
ON-GOING/FUTURE WORKS

1. Recover GWTC-1 events with AR modeling.



ON-GOING/FUTURE WORKS

1. Recover GWTC-1 events with AR modeling.



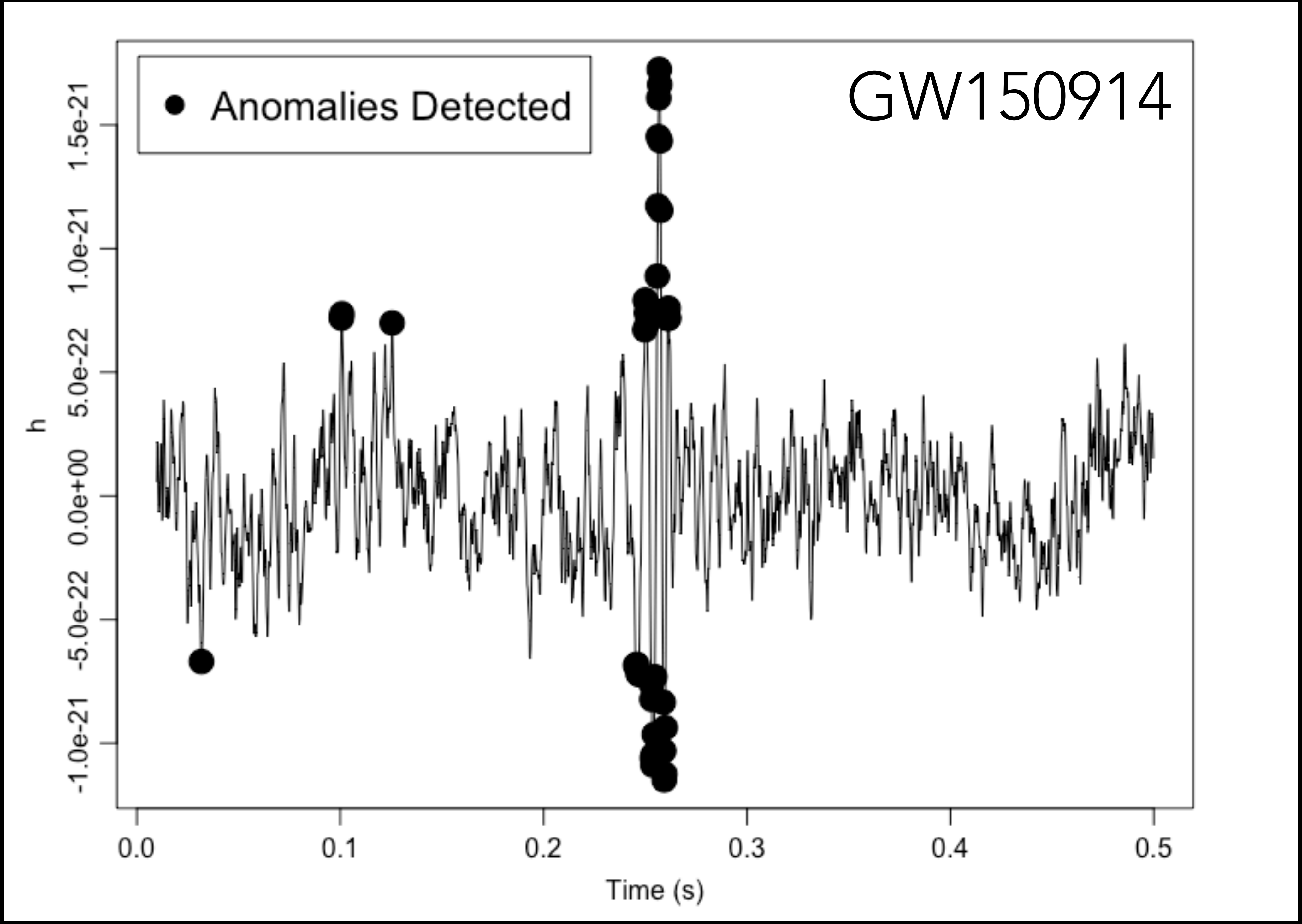
ON-GOING/FUTURE WORKS

2. Anomaly Detections

- After the noise subtraction, events candidates can be identified as anomalies, which differ from normal instances significantly.
- If the duration of the signal is significantly longer than the sampling interval, a cluster of anomalies is expected.
- 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.

ON-GOING/FUTURE WORKS

2. Anomaly Detections

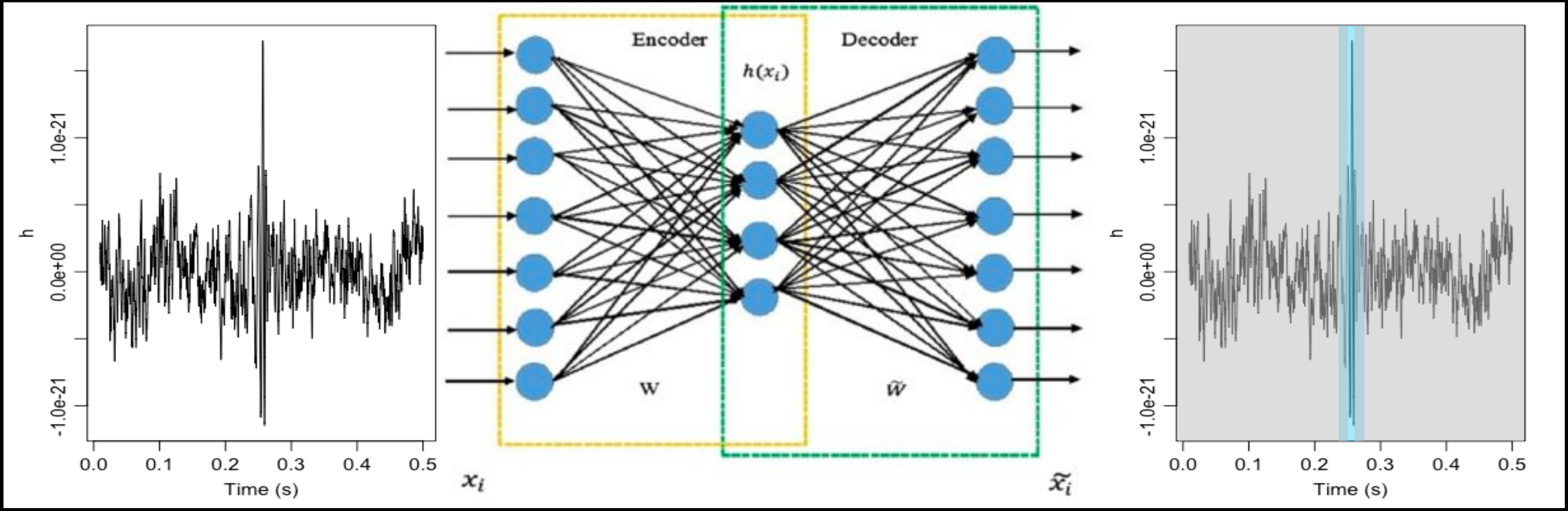


ON-GOING/FUTURE WORKS

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

Input

Output



ON-GOING/FUTURE WORKS

3. Template-free parameter estimation with AR spectral analysis

- Cleaned signal can be fitted with an 2nd stage AR model.
- Signal can be reconstructed from the best-fit model.
- Characteristic equation can be obtained from $\{a_j\}$ and the order p.
- $$F(z) = 1 - \sum_{j=1}^{p} a_j z^j = 0$$
- QNM frequency/Damping can be obtained from the complex roots.

$$z_k = \exp(i2\pi f_k \Delta t)$$

$\mathrm{Re}(f_k) \longrightarrow$ Frequency

$\mathrm{Im}(f_k) \longrightarrow$ Damping

It has been shown that this can extract the ring-down freq./damping timescale from GW150914 (Shinkai 2018,2019).

THANK YOU VERY MUCH