Analysis of the correlations between the non stationary noise and the auxiliary channels in LIGO

Luca D'Onofrio[1]
Gabriele Vajente[2]

[1]Università di Napoli "Federico II," Complesso Universitario di Monte S.Angelo, Napoli, Italy
[2]LIGO, California Institute of Technology, Pasadena, USA

LIGO-1900398-v1

Overview

- Band-Limited Root Mean Square (BLRMS)
- Auxiliary channels
- Standard Linear Regression - Overfitting problems
- Simulated data - LASSO regression
- Results (O3 data)

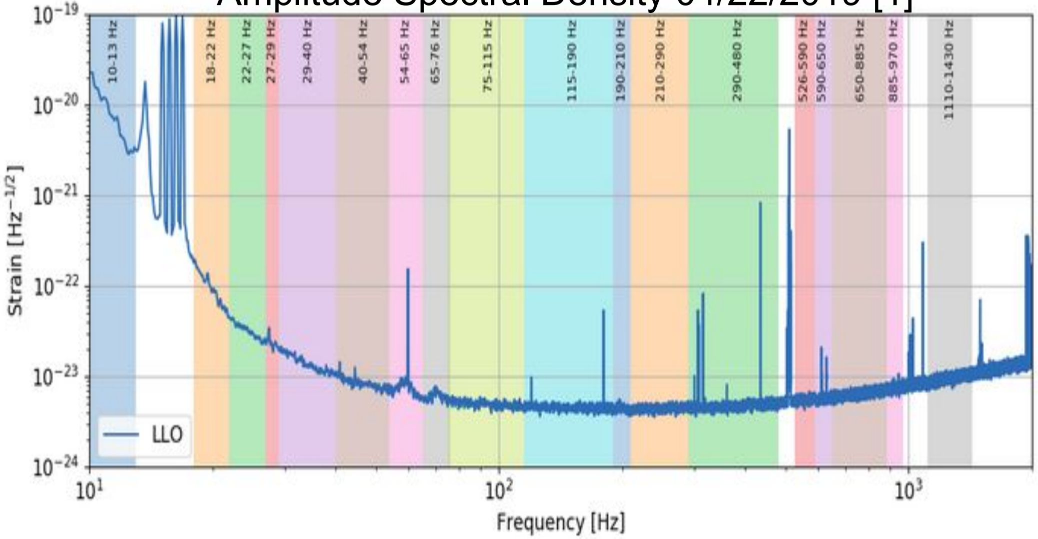
Band-Limited RMS

- Noise is NOT stationary. We need a way to quantify the time evolution of the noise.

- Square root of the integral of the Power Spectral Density (PSD) in a frequency band

$$BLRMS = \sqrt{\int_{\omega_1}^{\omega_2} PSD(\omega) d\omega}.$$

Amplitude Spectral Density 04/22/2019 [1]



- Time evolution of the noise floor: NO lines in PSD, NO glitches in BLRMS evolution

[1] G. Vajente, "Compute band-limited RMS for LIGO O3 data", https://git.ligo.org/gabriele-vajente/blrms

Auxiliary channels

- Thousands of sensors to monitor instrumental and environmental conditions
- Minute trends (without glitches)
- Mean trend and standard deviation trend for each channel

Principal subsystems of LIGO detectors.

ASC	alignment sensing and control
LSC	length sensing and control
PEM	physical environment monitor
PSL	prestabilized laser
SEI	seismic isolation
SUS	suspension
IMC	input mode cleaner
ISI	internal seismic isolation
HPI	hidraulic external pre-isolation
SQZ	squeezed light

Linear Regression

$$y_i = \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k} + \varepsilon_i \,,$$

i-th BLRMS value

k = # channels

i = time index

ε_i = additive noise

$$y = X\beta + \varepsilon \Rightarrow y = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{pmatrix} \quad X = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,1} & x_{i,2} & \cdots & x_{i,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,k} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_k \end{pmatrix}$$

Sum of squared error →

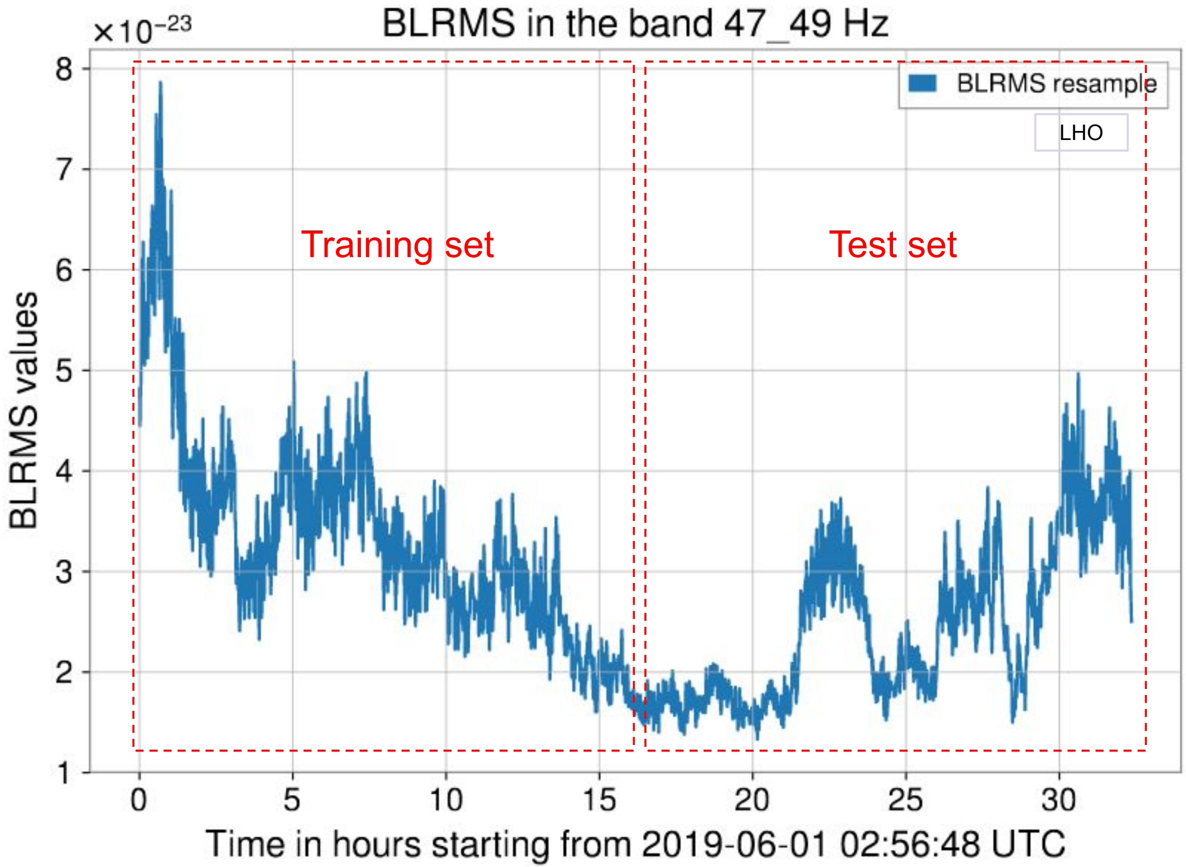
$$S = (y - \beta X)^T (y - \beta X) = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} \beta_j x_{i,j} \right)^2$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

→ Estimate β values on the training set

Compare predictions on the test set

Overfitting Problems



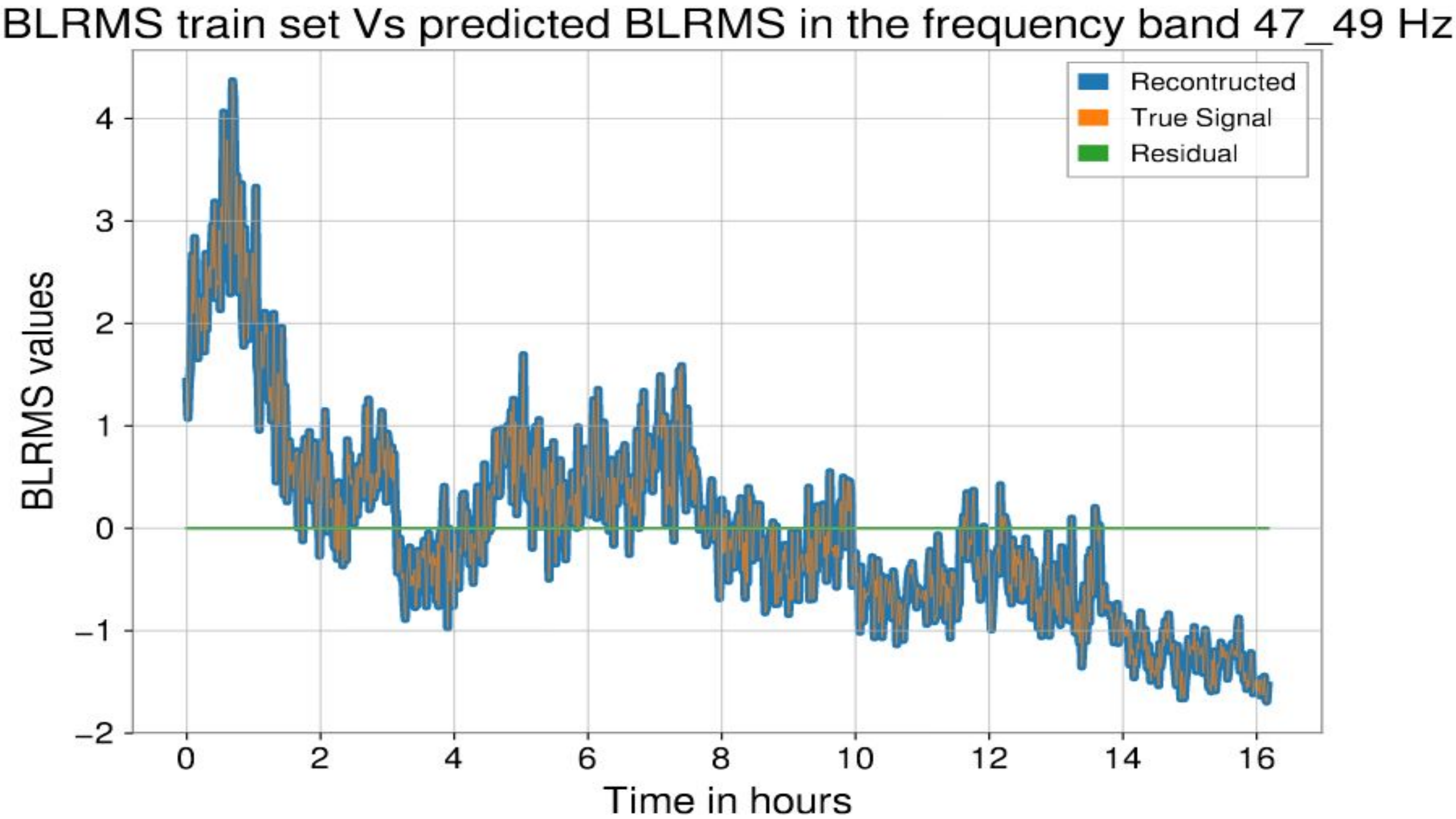
- 47-49 Hz
- 33 hours
- ‘**A**lignment**S**ensing**C**ontrol’ subsystem

↓

- resample
- splitting in training and test set
- scaling
- standard linear regression

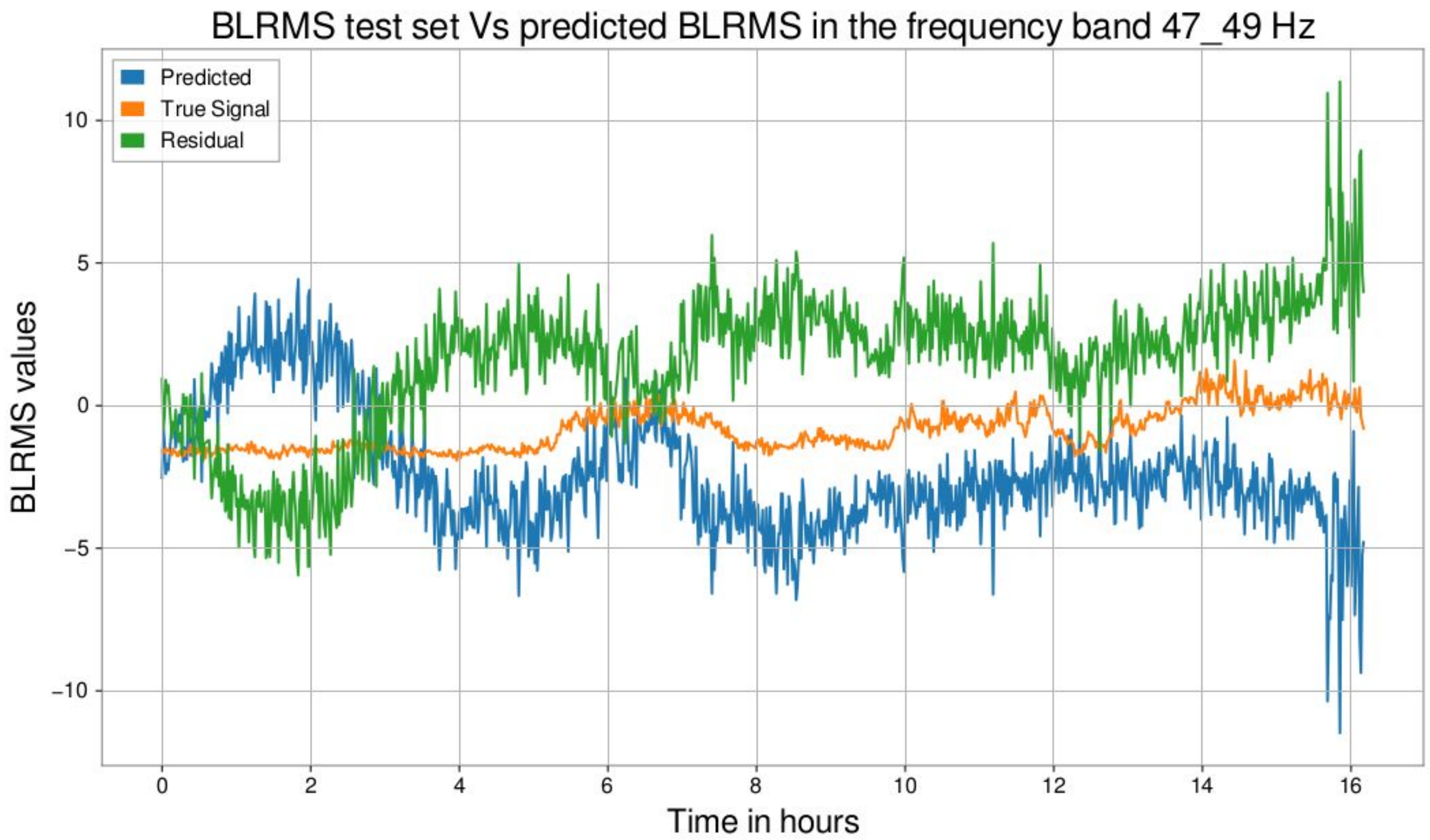
LIGO

Overfitting problems



Perfect prediction for the training set !

Overfitting problems



The model fails to generalize to unseen data.

Simulated data

WHY SIMULATED DATA? → To test possible solutions
→ Known coefficients to estimate

- 20000 random time series (Input signals x_j), 3600 points to simulate auxiliary channels time serie
- Combine first 20 input signals with known coefficients

$$y(t) = \sum_{j=1}^{20} a_j x_j(t) + \varepsilon(t)$$

LASSO Regressor

Least **A**bsolute **S**hrinkage and **S**election **O**perator

$$S_L = \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{k} \beta_j x_{i,j} \right)^2 + \alpha \sum_{j=1}^{k} |\beta_j|$$

The number of selected features is controlled by the regularization parameter α.

This term shrinks many components of the solution to zero.

- α →0 : Standard Linear regression
- Increasing α values, we expect that the number of coefficients different from zero decreases

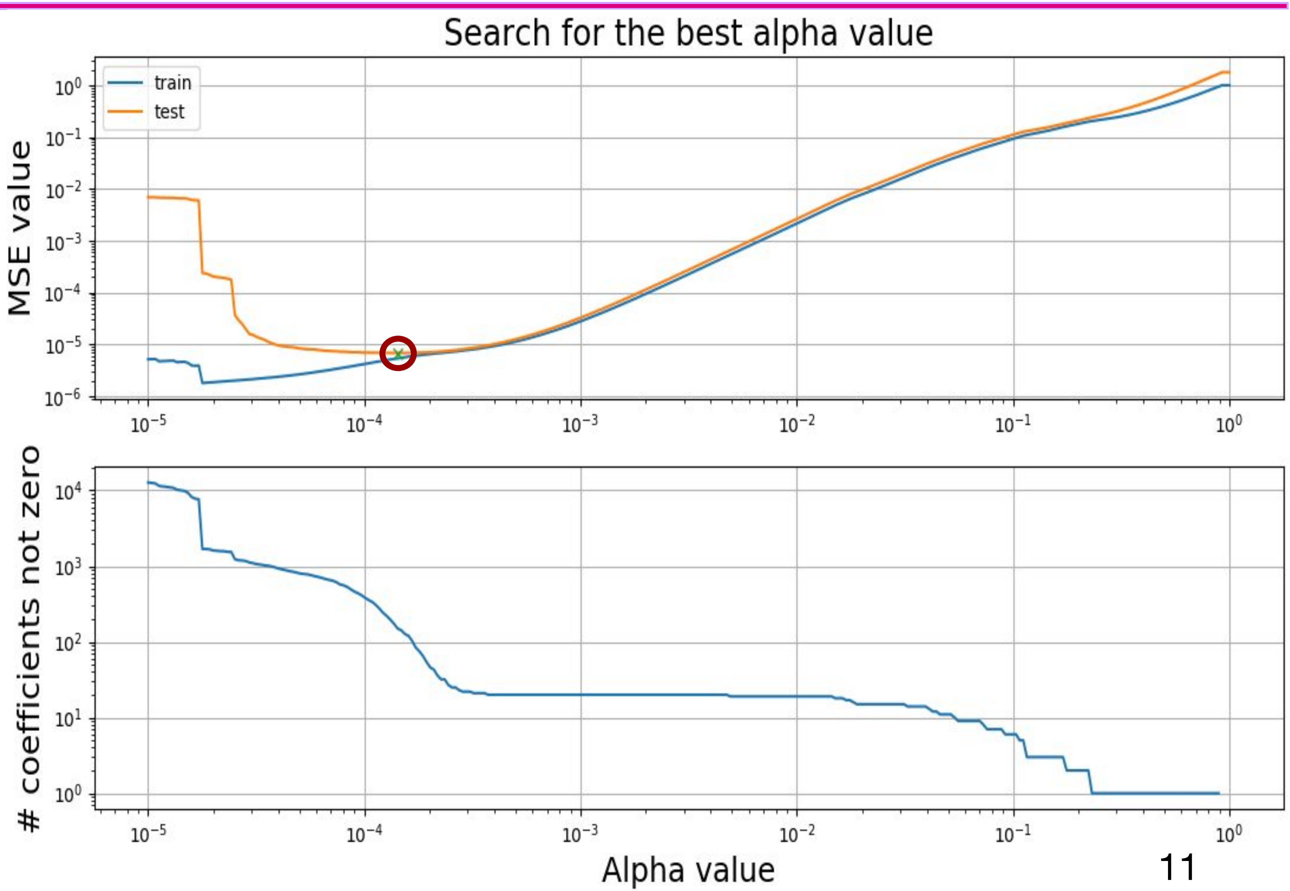
LASSO: results

Mean **S**quared **E**rror:

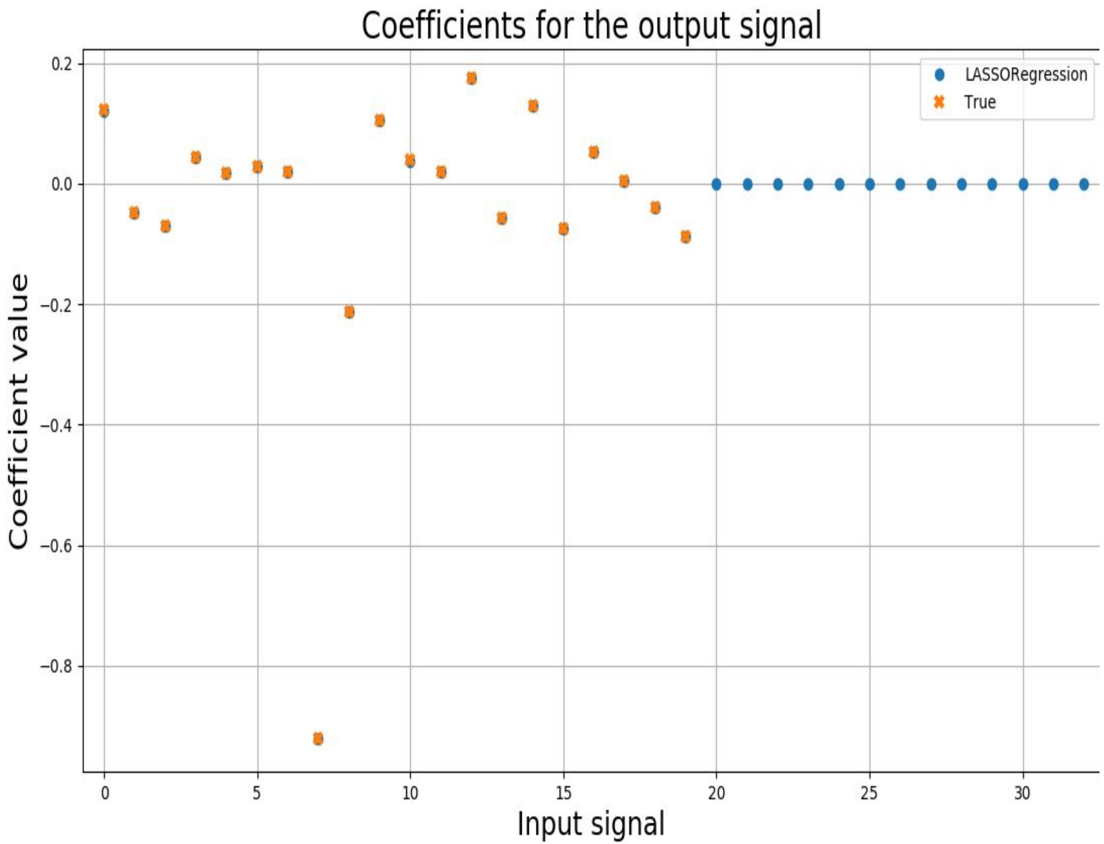
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y^{(i)} - \hat{y}^{(i)})^2$$

Best α : 0.00014

# of coefficients different from zero for Best α : 149



LASSO: results

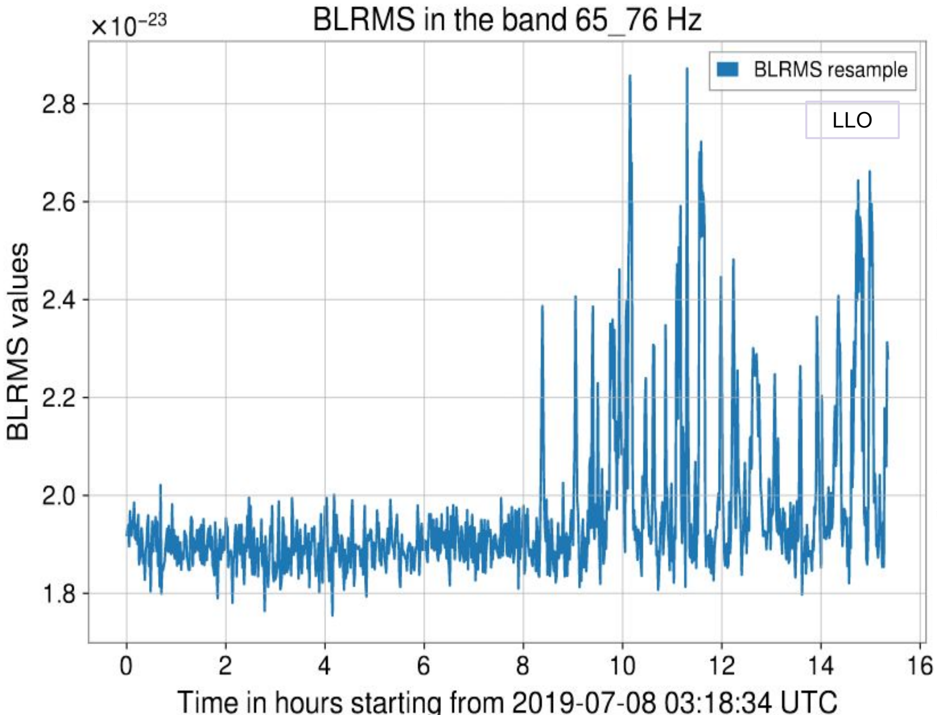


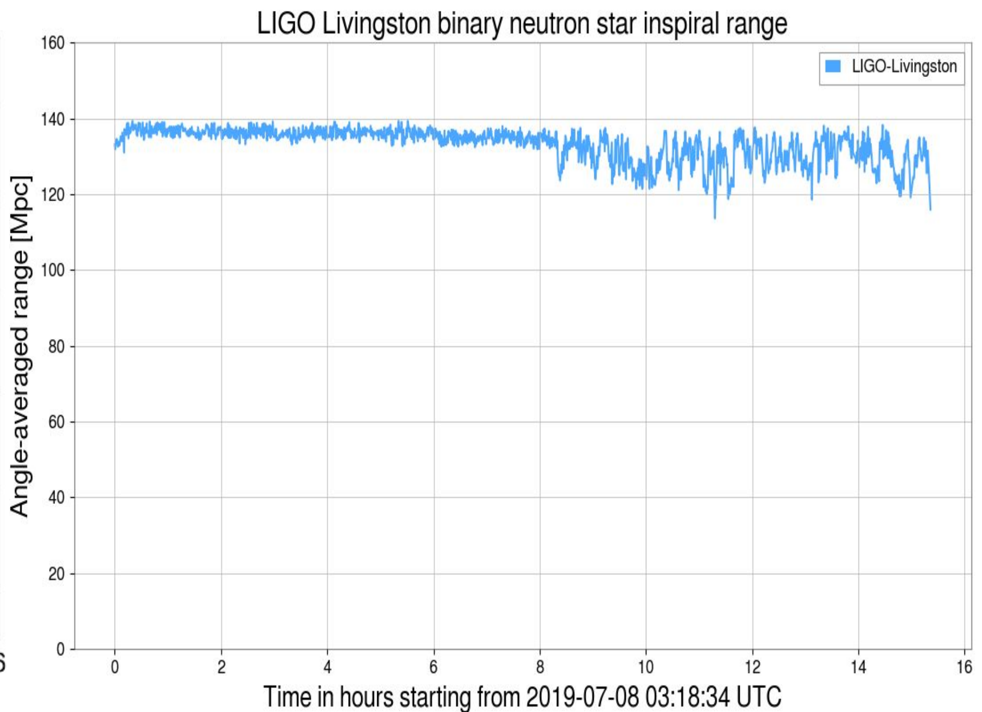
i-th IS	$\varepsilon_i = \frac{\|\|y - \hat{\beta}_i x_i\|\|}{\|\|y\|\|}$
0	0.990
1	0.998
2	0.995
…	…
7	0.078
8	0.980
…	…
14999	1.000
19826	1.000

LIGO

O3 data

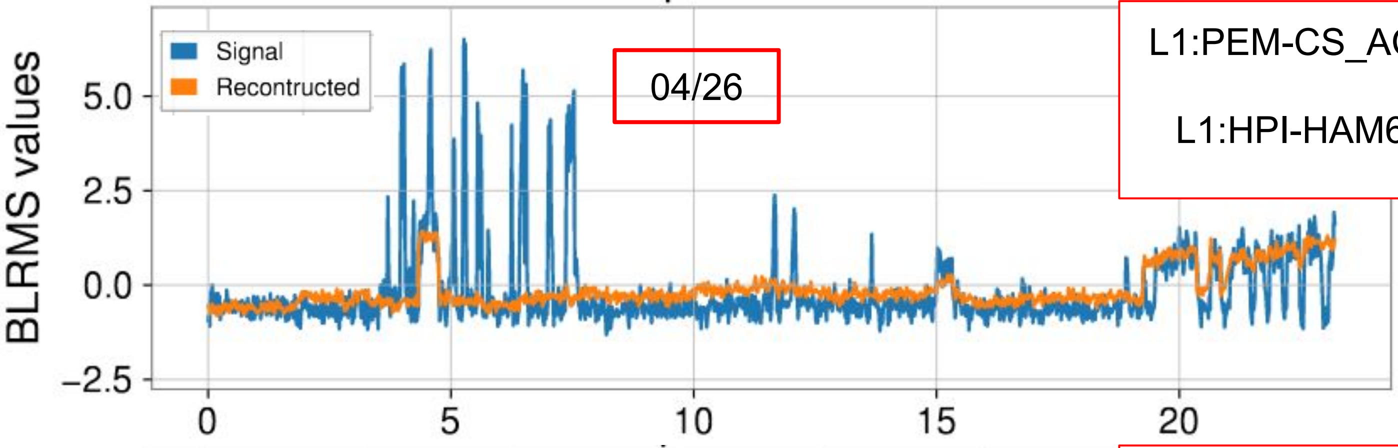
- How to split BLRMS segment in training and test set?



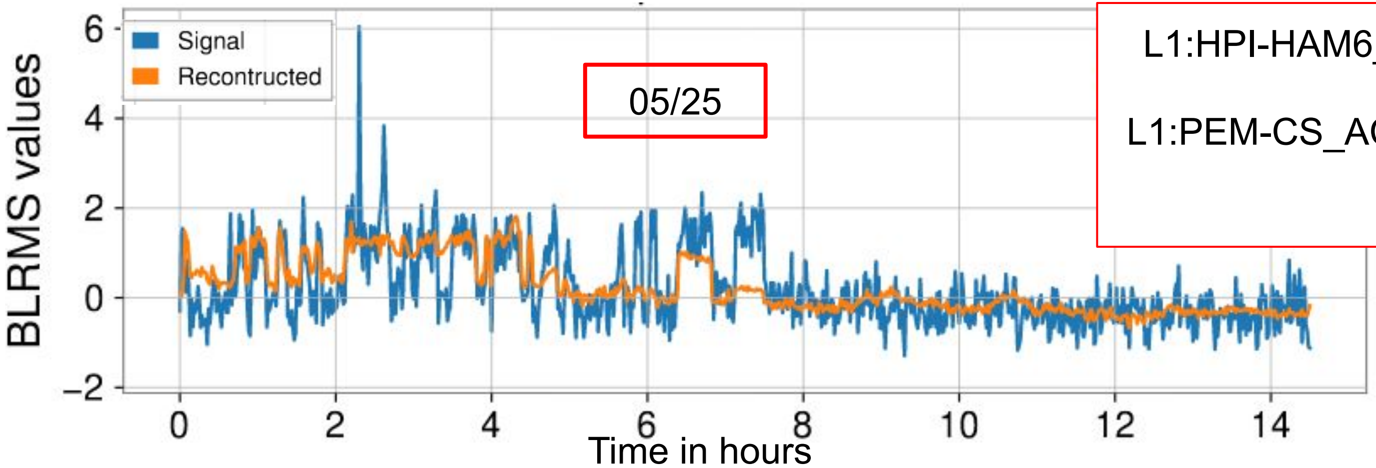


LLO 65-76 Hz

BLRMS test set Vs predicted BLRMS via LASSO

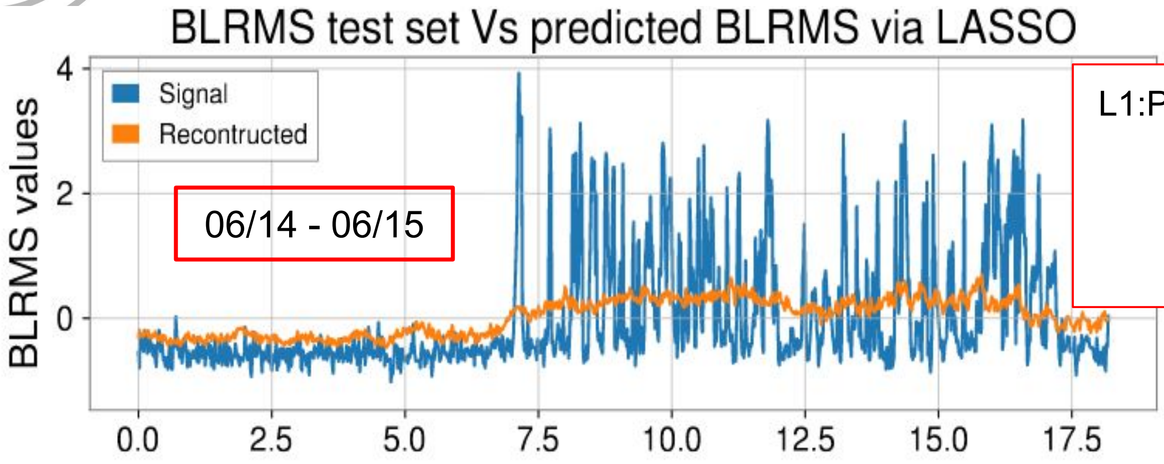


L1:PEM-CS_ACC_HAM6_OMC_X_MON.std
ε score : 0.84
L1:HPI-HAM6_BLND_L4C_VP_INMON.std
ε score: 0.91
+16

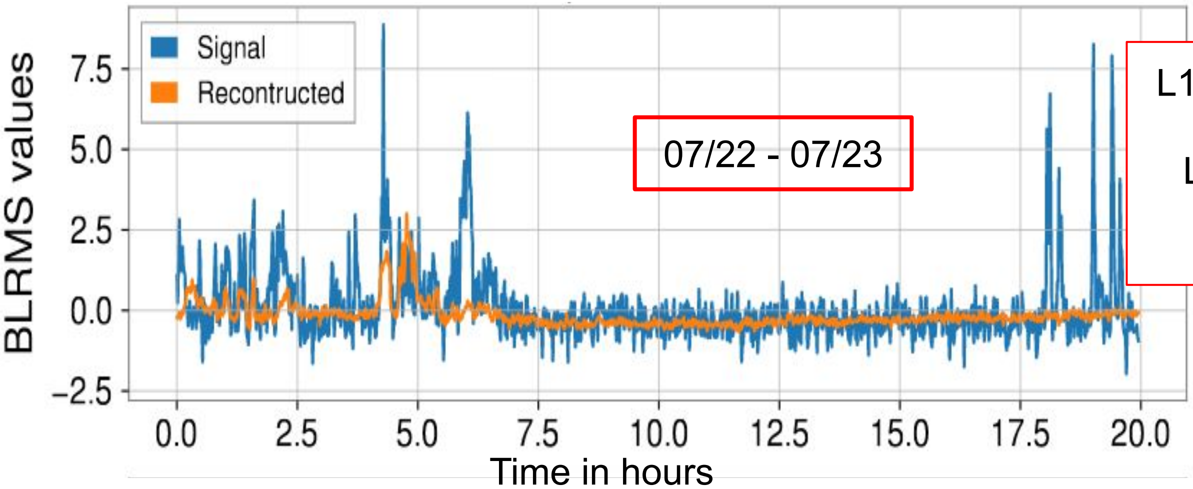


L1:HPI-HAM6_BLND_L4C_VP_INMON.std.
ε score : 0.64
L1:PEM-CS_ACC_HAM6_OMC_X_MON.std
ε score: 0.80
+15

LLO 65-76 Hz



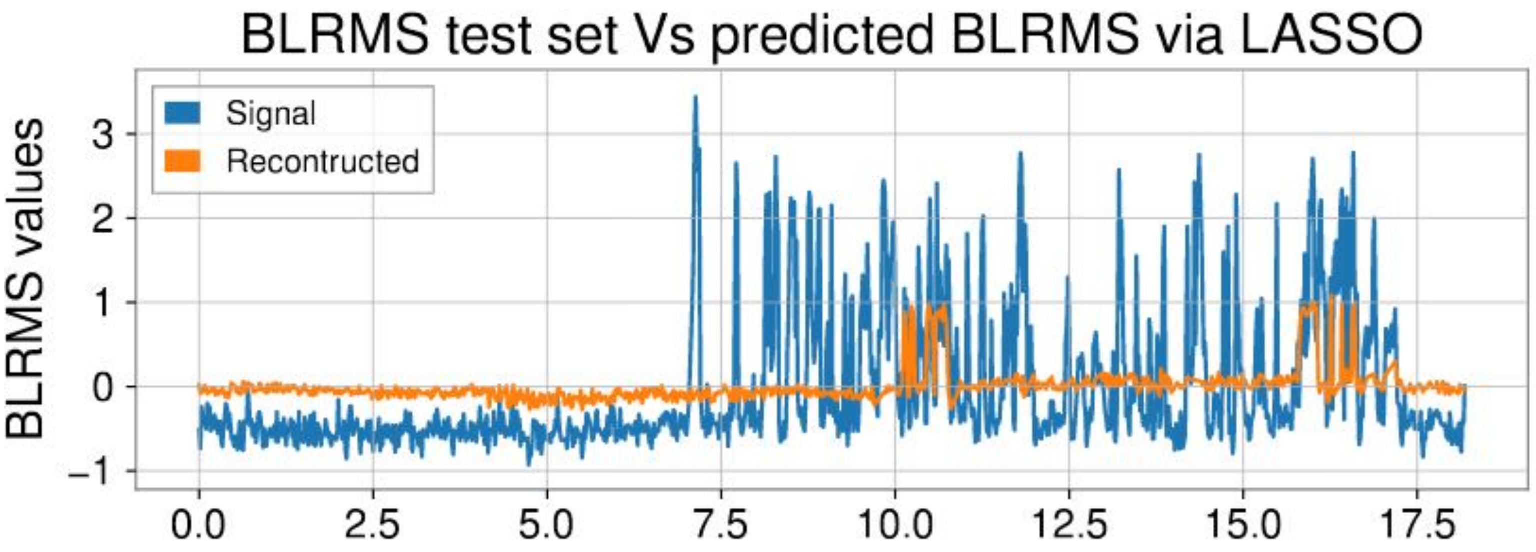
L1:PEM-EX_SEIS_VEA_FLOOR_Z_BLRMS_1HZ3
ε score : 0.90
L1:SUS-ITMY_R0_DAMP_R_INMON.mean
ε score: 0.92
+14



L1:PEM-CS_ACC_HAM6_OMC_X_MON.std
ε score : 0.92
L1:HPI-HAM6_BLND_L4C_VP_INMON.std
ε score: 0.95
+21

LLO 65-76 Hz

06/14 - 06/15



Reconstructed signal : Standard Linear Regression using only 2 channels,
L1:PEM-CS_ACC_HAM6_OMC_X_MON.std
L1:HPI-HAM6_BLND_L4C_VP_INMON.std

Summary

- Using BLRMS time series, we can select particular frequency bands to reconstruct the time evolution of the noise. We used LASSO to solve overfitting problems. We defined a way to quantify the “importance” of each channel (ε-score)

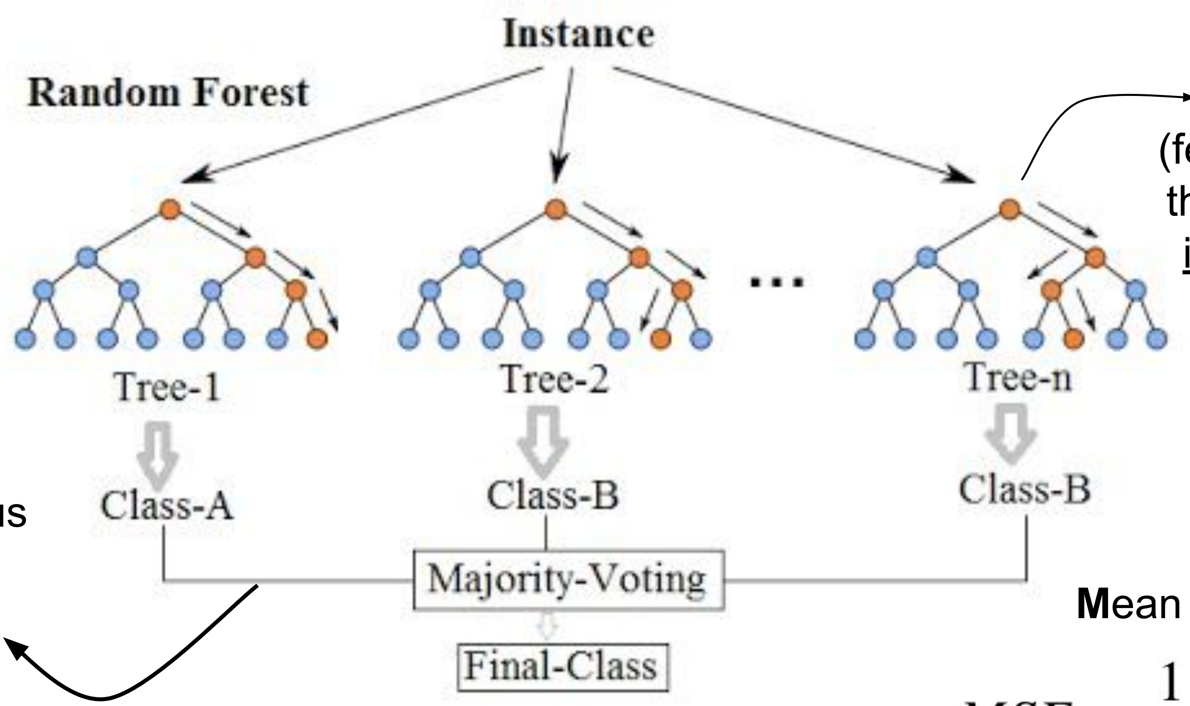
- In the frequency band 65-76 Hz for LLO, we identified 2 channels correlated to time evolution of the BLRMS.
  Other interesting results for LHO in the frequency band 1120-1400 Hz.

- FUTURE WORK :
  - why does not LASSO explain all the bumps in 65 - 76 Hz band?
  - systematic search in all the bands for LLO and LHO

Thank you !

I am grateful to my mentor Gabriele Vajente for the support in these weeks.
I would also like to thank to National Science Foundation and Instituto Nazionale di Fisica Nucleare that supported this research.

Random Forest



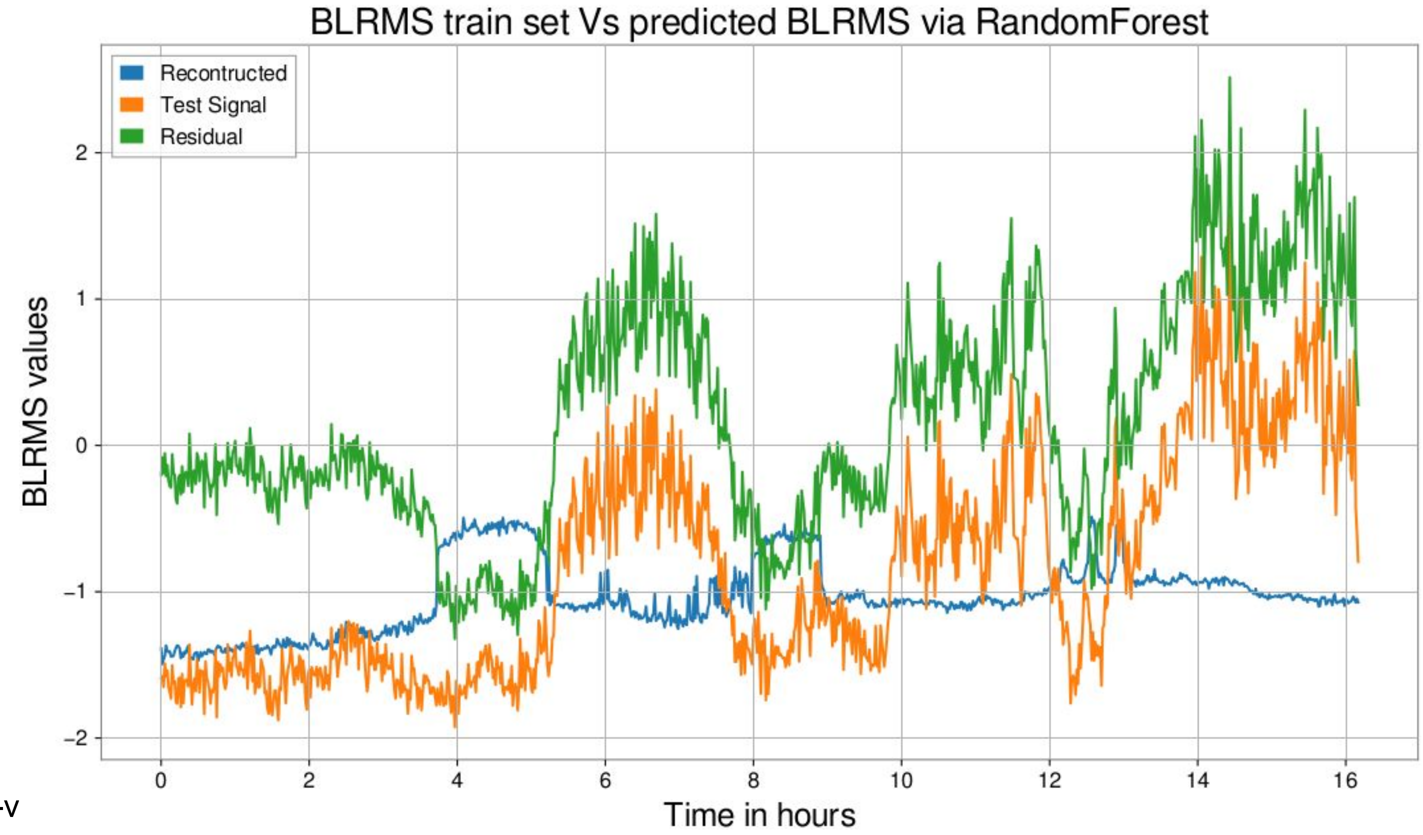
Split :
(feature, threshold)
that minimizes the
<u>impurity function</u>

In the case of regression, trees predict a continuous value and the final prediction is the avarage

Mean **S**quared **E**rror:

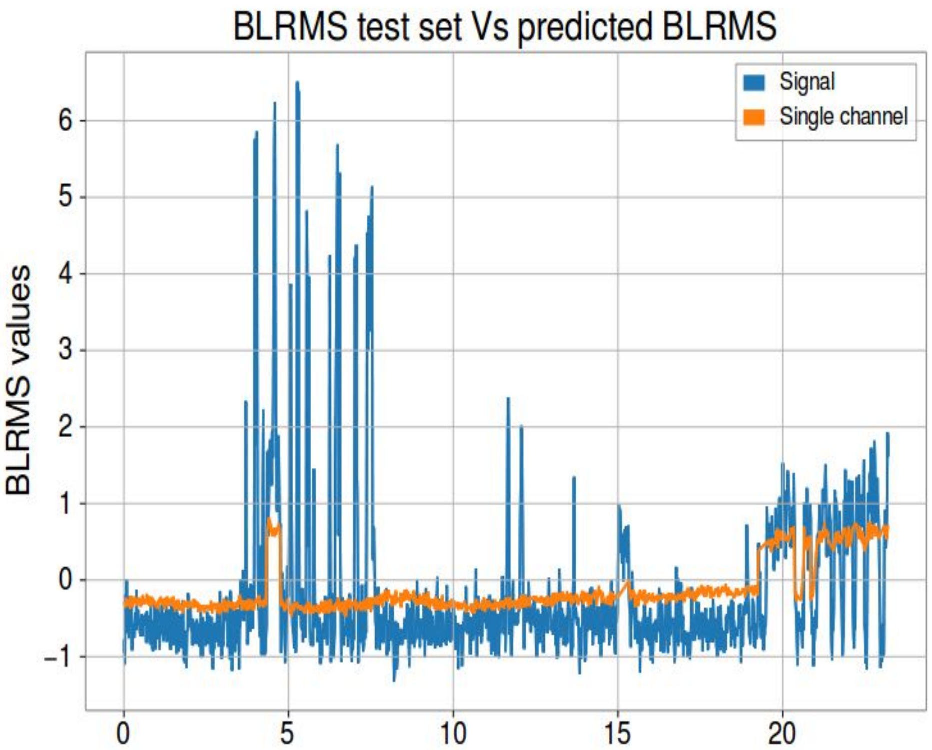
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y^{(i)} - \hat{y}^{(i)})^2$$

Random Forest

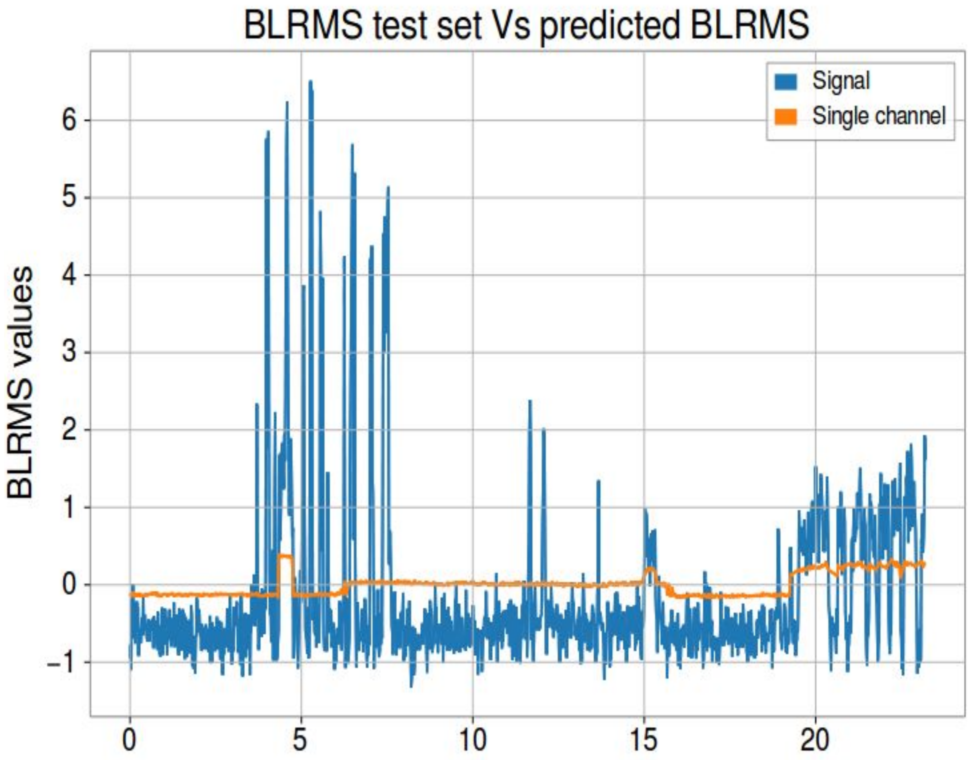


04/26

L1:PEM-CS_ACC_HAM6_OMC_X_MON.std

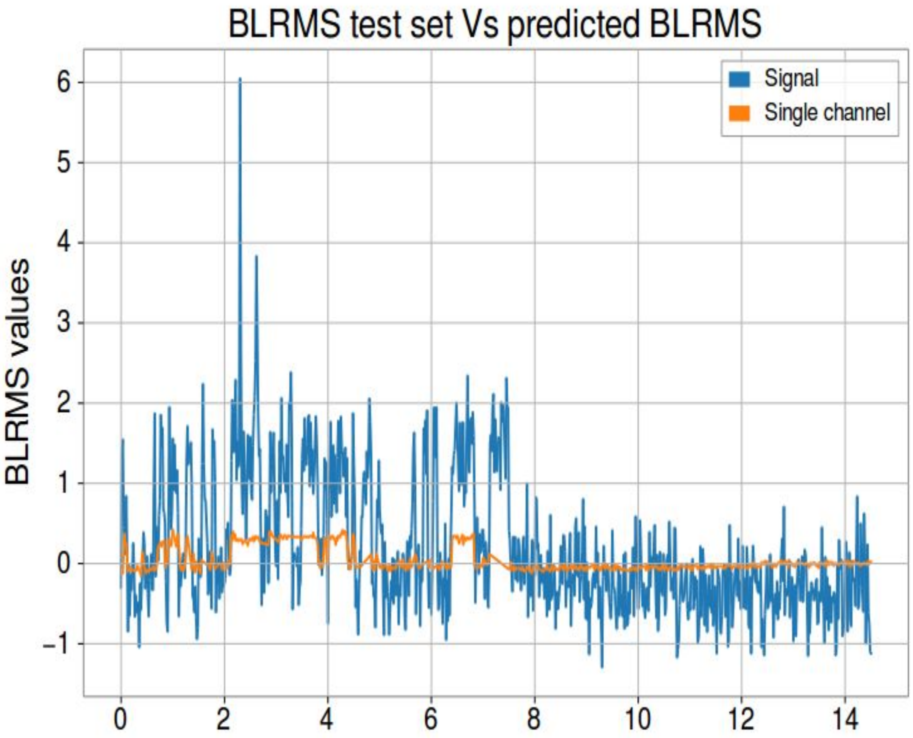


L1:HPI-HAM6_BLND_L4C_VP_INMON.std

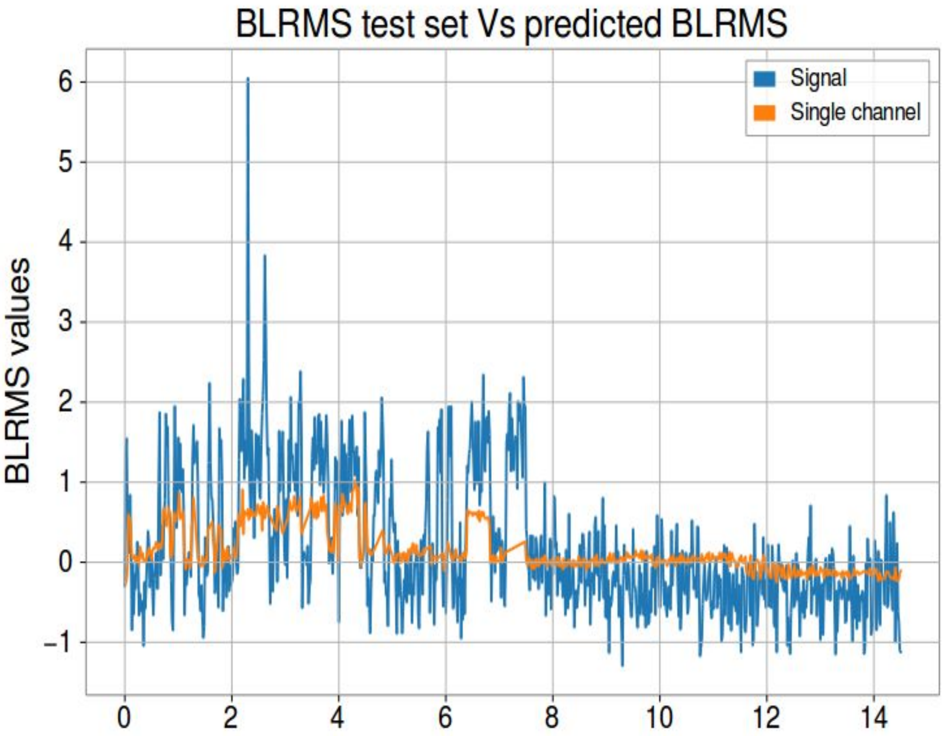


05/25

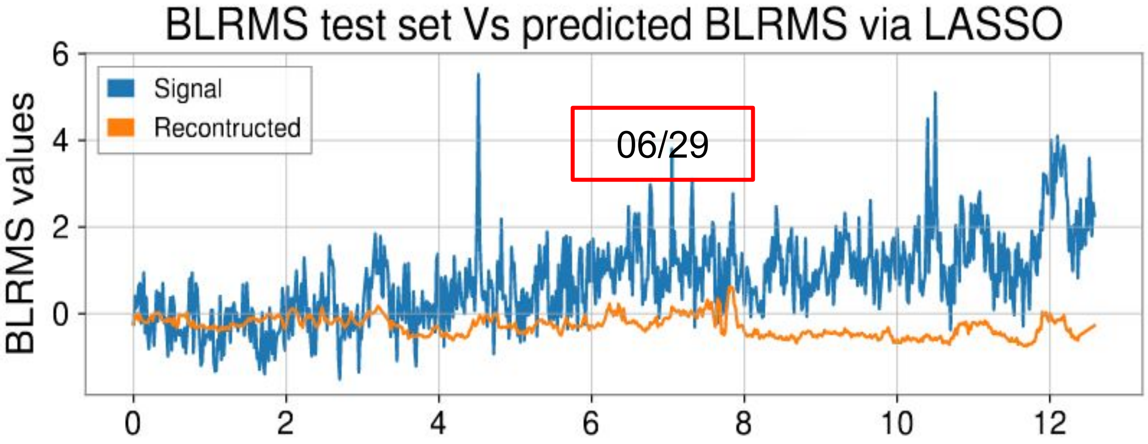
L1:PEM-CS_ACC_HAM6_OMC_X_MON.std



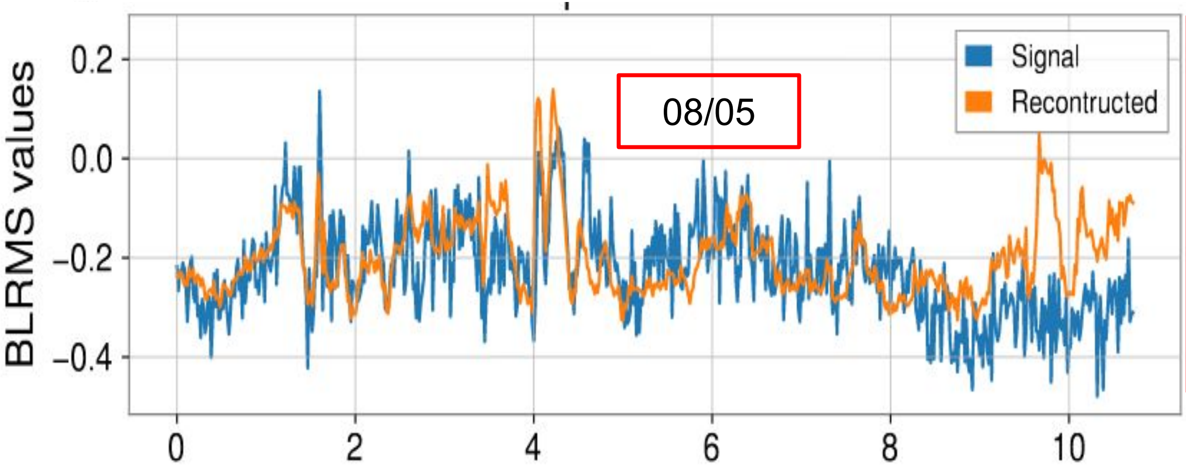
L1:HPI-HAM6_BLND_L4C_VP_INMON.std



LHO 1120-1400 Hz



H1:IMC-REFL_A_DEMOD_RFMON.std
ε score : 0.94
H1:PSL-FSS_PC_MON_OUTPUT.std
ε score: 0.95
+11



H1:PSL-FSS_PC_MON_OUTPUT.mean
ε score : 0.42
H1:PSL-FSS_PC_MON_OUTPUT.std
ε score: 0.58
H1:PSL-FSS_PC_PP.std
ε score: 0.77
H1:PSL-PMC_OSCILLATION_MON.std
ε score: 0.86