Strategy for Classification of Noise Transients in Advanced Detectors

Elena Cuoco[1], A.Torres-Forné[3], J.A. Font[3 7], J.Powell[2],R.Lynch[4], D.Trifiró[5], M.Cavaglià[6], Ik S. Heng[2],

[1]EGO and PISA INFN, [2]SUPA and IGR Glasgow, [3]Universitat de València, [4]MIT, [5]Università di Pisa,[6]University of Mississippi, [7]Observatori Astronòmic, Universitat de València

October 7, 2016

Introduction

- A prompt characterization of noise will be critical for improving sensitivity, a fast method for glitch classification was needed.
- The Wavelet Detector Filter (WDF) has been developed for this task.
- WDF has been tested jointly with other 2, PC-LIB and PCAT.
- The methods have been tested succesfully using simulated data.
- We tested our pipelines on LIGO ER7 data.

Principal Components.

- All three methods use at some stage Principal Components (PCs).
- PCs are a set of orthogonal basis vectors, which are ordered so that the first PC represents the most common feature of a set of waveforms.
- Therefore, a few PCs can be used to represent all the common features of the waveforms.
- The signal model consists of a linear combination of PCs

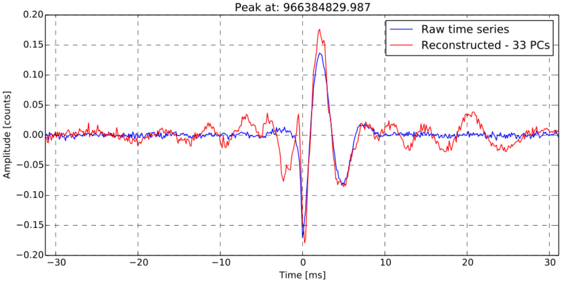


Figure: A glitch reconstructed by PCAT using 33 PCs.

- Results can be strongly effected by the number of Principal Components.
- We use the variance method to choose the ideal number of Principal Components.

Other pipelines

- PC-LIB
  - PC-LIB is an adaptation of the parameter estimation and model selection tool LALInference.
  - Computes Bayes factor for model selection, used to identify the correct glitch type.
- PCAT
  - Principal Component Analysis for Transients (PCAT) is a python-based pipeline based on Principal Component Analysis.
  - PCAT uses the PC coefficients to classify glitches by using a Gaussian Mixture Model (GMM).

WDF-ML: introduction

- WDF-ML consists of an event detection algorithm, Wavelet Detection Filter (WDF), followed by a Machine Learning (ML) classification procedure.
- WDF is part of the Noise Analysis Package (NAP), a C++ library embedded in python, developed by the Virgo Collaboration
- A whitening procedure is applied to the data and is based on a Linear Predictor Filter.
- The parameters are estimated through a parametric Auto Regressive (AR) model fit to the noise PSD.

WDF: denoising

Let us consider a signal x_i which is corrupted by additive Gaussian random noise $n_i \sim N(0, \sigma^2)$ as follow

$$x_i = h_i + n_i \quad i = 0, 1, ...N - 1$$

Let W be an orthogonal wavelet transform. If we apply it to the sequence of data x_i we obtain

$$W(x) = W(h) + W(n)$$

Now let T be a wavelet thresholding function. Then the wavelet thresholding based de-noising scheme can be written

$$\hat{h} = W^{-1}(T(Wx))$$

that is we first take the wavelet transform of our noisy signal and pass it through the thresholding function, then the output is inverted and wavelet transformed.

WDF-trigger

The wavelet coefficients contain the energy of the signal at different scale. After the wavelet thresholding, we selected the highest coefficients of the wavelet transform which are supposed to contain only the signal and not the noise.

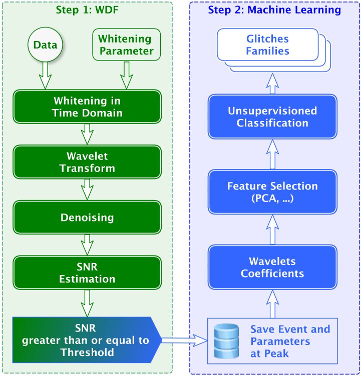
$$E_s = \sqrt{\sum_{k,j} w_{k,j}^2} \tag{1}$$

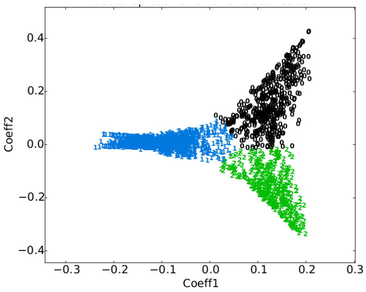
being $w_{k,j}$the wavelet coefficients above the threshold.
In this way $E_s$represent the signal energy content, so we can build our receiver detector which represents the signal to noise ratio, as

$$SNR = \frac{E_s}{\hat{\sigma}} \tag{2}$$

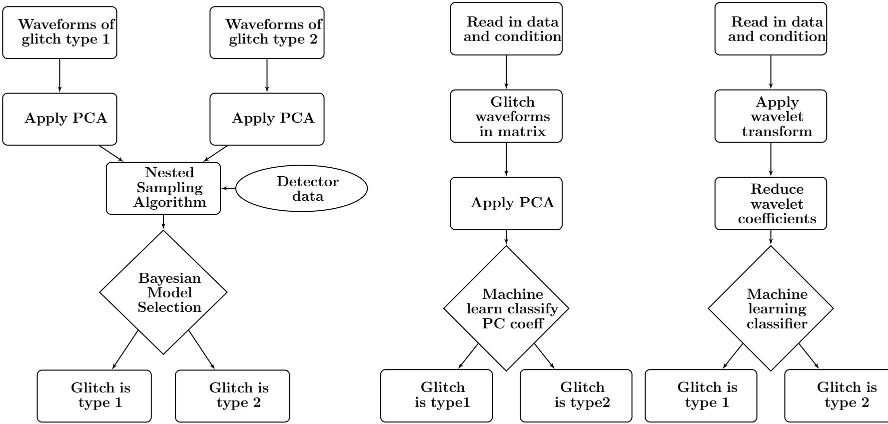
WDF-ML: Machine Learning

- Completely unsupervised algorithms. No target function
- Wavelets coefficients and Meta data (SNR, Freq,Duration) represents our "features"
- Features selection uses PCA transform an Spectral embedding on 2 dimensions
- The Gaussian Mixture Model (GMM) machine learning classifier is then applied to the outputs of WDF for classification.





Pipelines summary: PC-LIB, PCAT,WDF-ML



Simulated data: Datasets

The 3 methods have been tested and compared using 3 simulated data set in aLIGO Gaussian noise.

- Dataset 1:
  - The data set contains 1000 sine Gaussian waveforms and 1000 Gaussian waveforms in simulated Gaussian noise.
  - The sine Gaussian waveforms have a frequency = 400Hz and an SNR between 5 and 30.
- Dataset 2:
  - We use 1000 sine Gaussian waveforms and 1000 Ring-down waveforms.
  - All waveforms have identical frequency 400Hz and a identical duration 2ms.
- Dataset 3:
  - The simulated glitches are Gaussian, sine Gaussian and Ring-down waveforms at five second intervals.
  - The frequencies are distributed linearly between 40-1500 Hz.

Data Set 1 Results

- Table shows the % of detected transients that were classified in each type.
- A few low frequency SG, and low SNR G were in the incorrect classes.
- Overall classification efficiency very good!

	SG	G
PCAT Type 1	99%	0%
PCAT Type 2	1%	100%
LIB Type 1	99.9%	5%
LIB Type 2	0.1%	95%
WDF Type 0	99.5%	2.4%
WDF Type 1	0.3%	46.1%
WDF Type 2	0.2%	51.5%

Data Set 2 Results

- Table shows the % of detected transients that were classified in each type.
- The few transients in the incorrect class are those with the lowest SNR.
- 5PCs PCAT, 7PCs LIB and 10 PCs WDF-ML.
- All methods can classify by waveform morphology alone.

	SG	RD
PCAT Type 1	1.1%	97.4%
PCAT Type 2	98.9%	2.5%
LIB Type 1	97.8%	4.8%
LIB Type 2	2.2%	95.2%
WDF-ML Type 0	8.7%	100%
WDF-ML Type 1	48.0%	0%
WDF-ML Type 2	43.3%	0%

Data Set 3 Results

- PCAT 20PCs, LIB 5PCs, WDF-ML 10PCs.
- All methods have the Gaussians in there own class.
- Cannot distinguish between the sine Gaussian and Ring-down waveforms when the parameter range is so large.

	SG	G	RD
PCAT Type 1	15.5%	0%	13.6%
PCAT Type 2	36.8%	0%	41.4%
PCAT Type 3	14.2%	0%	13.0%
PCAT Type 4	9.1%	0%	13.0%
PCAT Type 5	0.8%	0%	0.3%
PCAT Type 6	21.8%	0%	17.2%
PCAT Type 7	1.8%	100%	1.5%
LIB Type 1	39.5%	4.9%	23.8%
LIB Type 2	17.3%	88.3%	23.2%
LIB Type 3	43.3%	6.8%	53.0%
WDF-ML Type 0	89.5%	9.6%	86.9%
WDF-ML Type 1	5.9%	49.7%	7.0%
WDF-ML Type 2	4.6%	40.7%	6.1%

Powell, J. et al. Classification methods for noise transients in advanced gravitational-wave detectors Class. Quant. Grav., 32 (21), pp. 215012, 2015.

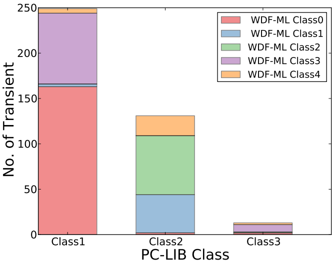
Real Data: ER7 LIGO data

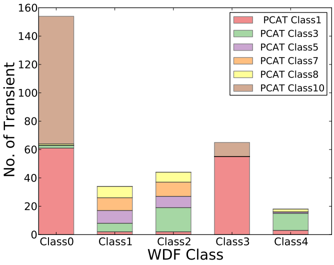
- Data from the 7th aLIGO engineering run (ER7), which began on the 3rd of June 2015 and finished on the 14th of June 2015. The average binary neutron star inspiral range for both Hanford and Livingston detectors in data analysis mode during ER7 was $50 - 60$ Mpc.
- The total length of Livingston data analysed is ~ 87 hours.
- The total length of Hanford data analysed is ~ 141 hours.

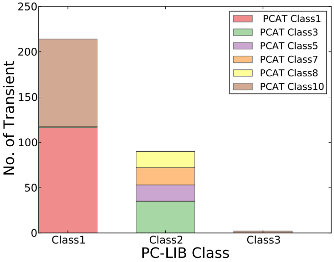
WDF for Real Data

- Downsampling producedure. Allows to work with different sampling frequency.
- The whitening is applied at the beginning of each locked segment. Adaptative whitenining also implemented.
- Good computer performance.
- We run the code in Valencia cluster (ARC1)

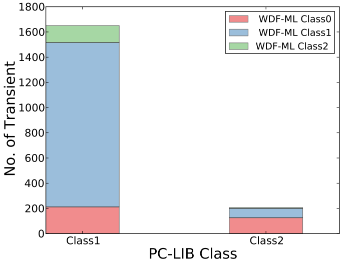
ER7 L1 results

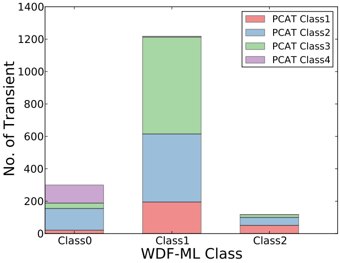


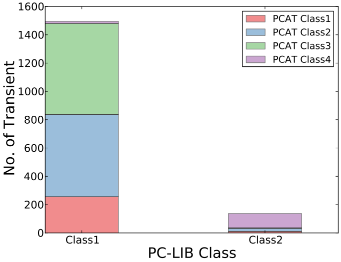




ER7 H1 results







Conclusion

- In the ER7 data from aLIGO Livingston PCAT missed 90 transients and classified 95% of the remaining transients correctly.
- PC-LIB missed 33 transients and classified 98% of the remaining transients correctly.
- WDF-ML classified all transients and 97% of them were correct.
- In aLIGO Hanford PCAT missed 120 transients and classified 99% of the remaining transients correctly.
- PC-LIB missed 6 transients and classified 95% of the remaining transients correctly.
- WDF-ML classified all transients and 92% of them were correct.
- We conclude that our methods have a high efficiency in real non-stationary and non-Gaussian detector noise.

Submitted:
Classification methods for noise transients in advanced gravitational-wave detectors II: performance tests on Advanced LIGO data. (by the authors)

What's next?

- Three different methods have been developed for the fast classification of noise transients.
- Transients are split in to types by waveform morphology first, and then can be split up in to further types by frequency and SNR.
- Results are similar for all methods.
- We plan to use Dictionary Based Algorithm.
- We plan to use Images Deep Learning Classification
- Next we plan on looking at how these codes perform when using data from multiple auxiliary channels.
- We are ready to apply WDF-ML to O1 data.
- We will apply the denoising and the machine learning procedures to the triggers produced by Omicron.
- We also want to test this tools for Virgo data.