Classification methods for noise transients in advanced gravitational-wave detectors II: performance tests on Advanced LIGO data

Classification methods for noise transients in advanced gravitational-wave detectors II: performance tests on Advanced LIGO data

Jade Powell[1], Alejandro Torres-Forné[2], Ryan Lynch[3], Daniele Trifirò[4,5], Elena Cuoco[6,7], Marco Cavaglià[5], Ik Siong Heng[1] and José A Font[2,8]

[1] SUPA, Institute for Gravitational Research, School of Physics and Astronomy, University of Glasgow, Glasgow, G12 8QQ, UK
[2] Departamento de Astronomía y Astrofísica, Universitat de València, Dr. Moliner 50, 46100, Burjassot (València), Spain
[3] Massachusetts Institute of Technology, 185 Albany St, 02139 CA, USA
[4] Dipartimento di Fisica E. Fermi, Università di Pisa, Pisa 56127, Italy
[5] Department of Physics and Astronomy, The University of Mississippi, University, MS 38677, USA
[6] European Gravitational Observatory (EGO), Via E Amaldi, I-56021 Cascina, Italy
[7] Istituto Nazionale di Fisica Nucleare (INFN) Sez. Pisa Edificio C—Largo B. Pontecorvo 3, 56127 Pisa, Italy
[8] Observatori Astronòmic, Universitat de València, C/ Catedrático José Beltrán 2, 46980, Paterna (València), Spain

E-mail: j.powell.2@research.gla.ac.uk

Abstract

The data taken by the advanced LIGO and Virgo gravitational-wave detectors contains short duration noise transients that limit the significance of astrophysical detections and reduce the duty cycle of the instruments. As the advanced detectors are reaching sensitivity levels that allow for multiple detections of astrophysical gravitational-wave sources it is crucial to achieve a fast and accurate characterization of non-astrophysical transient noise shortly after it occurs in the detectors. Previously we presented three methods for the classification of transient noise sources. They are Principal Component Analysis for Transients (PCAT), Principal Component LALInference Burst (PC-LIB) and Wavelet Detection Filter with Machine Learning (WDF-ML). In this study we carry out the first performance tests of these algorithms on

gravitational-wave data from the Advanced LIGO detectors. We use the data taken between the 3rd of June 2015 and the 14th of June 2015 during the 7th engineering run (ER7), and outline the improvements made to increase the performance and lower the latency of the algorithms on real data. This work provides an important test for understanding the performance of these methods on real, non stationary data in preparation for the second advanced gravitational-wave detector observation run, planned for later this year. We show that all methods can classify transients in non stationary data with a high level of accuracy and show the benefits of using multiple classifiers.

Keywords: gravitational waves, LIGO, machine learning

(Some figures may appear in colour only in the online journal)

1. Introduction

The advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) detectors are two 4 km interferometers at Hanford, Washington (H1) and Livingston, Louisiana (L1) [1, 2]. The Italian 3 km interferometer Virgo is expected to join the advanced detector network early in 2017 [3]. The detector duty cycle and sensitivity to astrophysical signals will be determined by noise sources created by the instruments and the environment. In particular, as the detector noise is non-Gaussian, short-duration transients will limit the sensitivity of searches for transient astrophysical sources such as compact binary coalescences [4].

The first aLIGO observation run (O1) began autumn 2015. On the 14th September 2015 the aLIGO and Virgo teams detected gravitational waves from the binary black hole system GW150914 [5]. A second binary black hole detection was made on the 26th of December [6]. An extensive study of the noise transients, which occurred in the data containing the detections, was carried out for the validation of the signals [7]. As the advanced detector network approaches its design sensitivity, the number of detections is expected to increase. Adding more detectors to the network increases the number of possible noise sources and the time it will take to identify their origin. Transients which occur in any one detector will limit the joint analysis time for the network. Understanding the sources of noise transients in the detectors will become increasingly more important with a latency of a few hours.

The detectors contain many environmental and instrumental sensors, which produce auxiliary channels of data that can be used to monitor the detector behaviour and track the causes of short-duration noise artifacts. Auxiliary channels that are not sensitive to gravitational waves can be used to identify noise transients, also known as 'glitches', in the detector output and veto those events [8–10]. Classification and categorization of transients using individual channels of data may provide valuable clues for the identification of their sources, which can aid in efforts to eliminate them [11, 12]. So far classification has mainly been achieved by visual inspection of spectrograms of the transients, but automatic classification is essential for future detections of astrophysical gravitational-wave signals.

Three methods for fast classification of transients have been developed for the analysis of aLIGO and Virgo data. They are Principal Component Analysis for Transients (PCAT), Principal Component LALInference Burst (PC-LIB) and Wavelet Detection Filter with Machine Learning (WDF-ML). Previous work has shown that these methods can classify artificial data sets with an efficiency up to 95% [11]. In this paper we evaluate the performance of these algorithms using glitches in real data from aLIGO. In section 2 we provide details of

the detector data. In section 3 we give a brief overview of the three different algorithms and details of any improvement they underwent since the previous study. In section 4 we present the results for the three algorithms on glitches from aLIGO L1 and H1 detector data. This is followed by a discussion in section 5 of our plans for future improvements and classification during the second aLIGO run (O2) and first Virgo observation run.

2. The data

In this study we use 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 H1 and L1 detectors in data analysis mode during ER7 was 50–60 Mpc [13]. The mean range for both detectors is shown in figure 1.

2.1. Livingston

In the period analyzed, data from L1 consists of 48 segments where the interferometer was locked and in data analysis ready mode. These data segments vary in length from 1 second to ∼7 h. We discard any segments of data that are less than a minute in duration as a longer segment of data is required to measure the power spectral density (PSD). The total discarded amount was 49 s of data. The total length of L1 data analysed is ∼87 h.

Glitches of different types are often recognised by their shape in a spectrogram such as those shown in figure 2. A description of the most common glitch types, which have occurred in aLIGO data, are described in [15]. Figure 2(a) shows glitches characterized by a tear drop shape. Figure 2(b) shows longer duration transients known as 'whistles', which are caused by radio frequency beats [15]. Only a small number of whistles (∼11) were found in the frequency and SNR range used in this study. Some other glitches in the data that are not shown in figure 2 include those below 10 Hz and scattered light. Glitches span the entire frequency range considered in this study. Some transients may have occurred due to the increased microseism created by tropical storm 'Bill' in the Gulf of Mexico [13].

A number of hardware injections were also made during ER7. An example is shown in figure 2(c). Hardware injections are artificial signals simulated by inducing a motion of the optics that can be used to test which auxiliary channels are sensitive to gravitational waves [8, 9].

2.2. Hanford

In the period analyzed, data from the H1 detector consists of 50 segments where the interferometer was locked and in data analysis ready mode. The data segments vary in length from 1 s to almost 14 h. As with L1 we discard any segments of data that are less than a minute in duration, which was a total of 116 s of data. The total length of Hanford data analysed is ∼141 h.

The H1 data is highly non-stationary and contains many more transients than the aLIGO L1 data. In particular, the H1 data contains many high SNR transients that caused a significant drop in the binary neutron star inspiral range. An example is shown in figure 3(b). It was suspected that these large transients were caused by cleaning of the beam tube [13]. A few other examples of common transients found are shown in the other spectrograms displayed in figure 3. As with the L1 data, H1 data also contains a number of hardware injections.
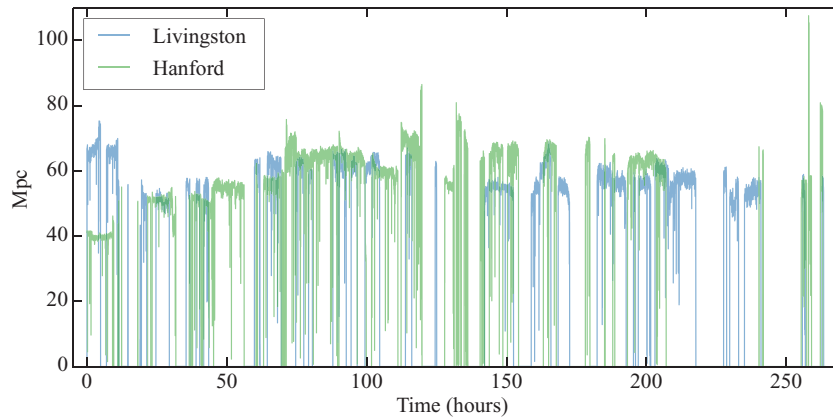
Figure 1. The mean binary neutron star inspiral range for the two aLIGO detectors during ER7. The Hanford detector had a higher range but also a higher glitch rate. The average range was 50–60 Mpc.

3. Transient classifying algorithms

Three different classifying algorithms were developed for the fast classification of noise transients in the detectors. Most of the technical details have been described in [11]. Here we give a brief outline of the three methods and describe any changes that have been made to improve their performance and latency. Figure 4 outlines the classification procedures for all three methods. More details are given in the following sections.

To find transients in the data we use event trigger generators (ETGs). ETGs typically search for excess power in individual interferometers and output the time, SNR, frequency, duration and other parameters of transients found in the data. PC-LIB uses Omicron, the main ETG used by the LIGO Scientific Collaboration's (LSC) detector characterization group [16, 17]. WDF-ML and PCAT have their own internal ETGs.

3.1. PCAT

PCAT uses a technique called Principal Component Analysis (PCA) that allows for dimensional reduction of large data sets [11, 18]. In the first stage of the PCAT analysis, the data are downsampled to 8192 Hz, whitened and high-pass filtered at 10 Hz. Then PCA is applied to all of the noise transients found by the ETG in all the analyzed segments of data. PCAT uses a 0.125 s window around each GPS time as glitches are typically of ms duration. This can lead to a loss of sensitivity to longer duration glitches. However, this effect can be safely neglected as longer duration glitches do not occur very often during observing runs, when the data is generally more stable than the ER data.

A projection of the original waveforms on to the principal components (PCs) allows for the calculation of scale factors for each PC called PC coefficients. Noise transients of different types are separated in the PC coefficient parameter space. This allows PCAT to classify the transients by applying a gaussian mixture model (GMM) machine learning classifier to the PC coefficients [19].
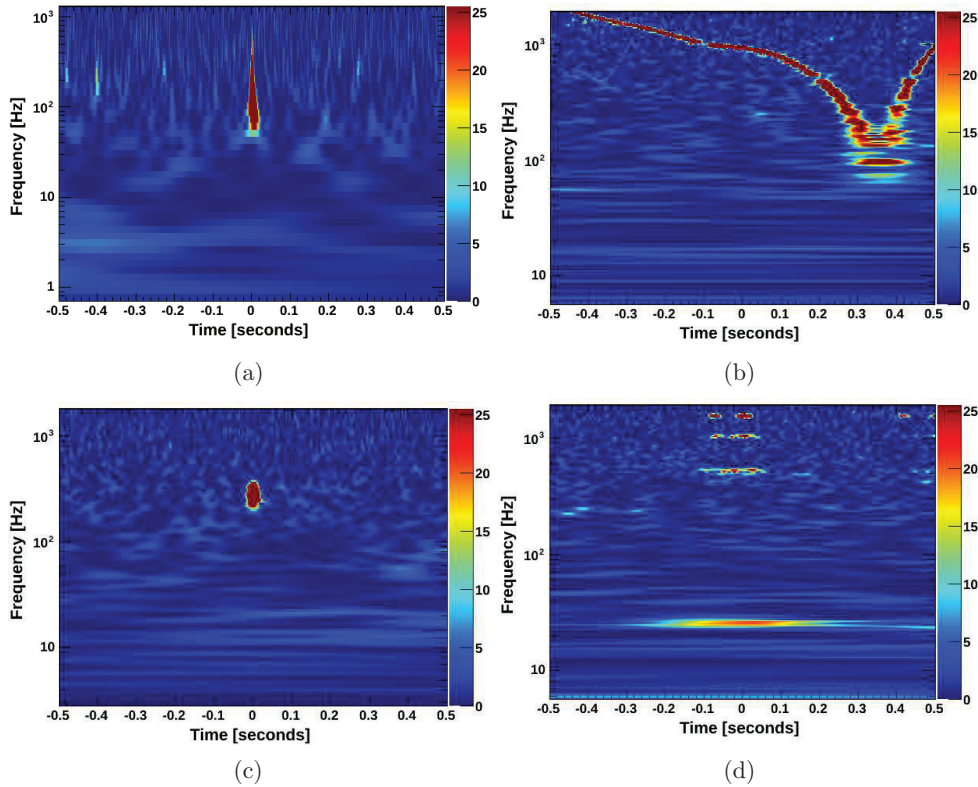
Figure 2. Spectrograms of typical transient types found in the aLIGO Livingston ER7 data. They are generated using the Omega scan tool in LigoDV-Web [14], which matches the data to sine Gaussians. (a) A transient characterized by a tear drop shape in the spectrogram. (b) A 'whistle' glitch that often has a long duration and occurs at high frequencies. (c) A hardware injection. (d) A transient type characterized by high frequency lines and lower frequency features.

3.2. PC-LIB

LALInference Burst (LIB) is a Bayesian parameter estimation and model selection tool, which uses a sine-Gaussian signal model to estimate parameters of gravitational-wave bursts [20]. It can also be combined with Omicron to be run as a search [21]. PC-LIB adapts LIB for the classification of transients by replacing the LIB sine-Gaussian signal model with a new signal model created from a linear combination of PCs calculated from the waveforms of known transient types [22, 23]. These known transients may have been previously classified by examining spectrograms of the transients or by one of the other methods. Thus PC-LIB can only classify transients that have occurred in the data many times before. When transients of a new type start to appear in the data new signal models must be created.

In our previous study we created signal models using fifty transient waveforms. In this study we only use ten waveforms. This change will allow us to start classifying new transient types more quickly as they start to appear in O2 data, without any loss in sensitivity, as glitch
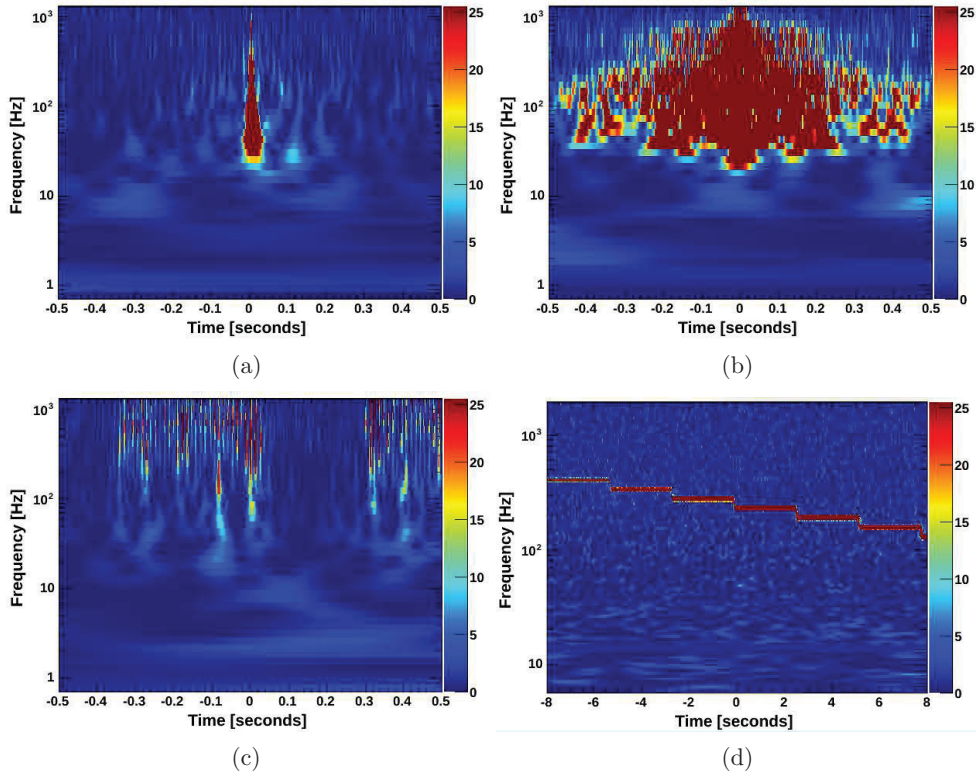
Figure 3. Examples of some of the most common transient types found in the Hanford ER7 data (a) A tear drop glitch. (b) Transients of this type have a large SNR and duration. They created significant drops in the detectors range. (c) A high frequency transient type. (d) A longer duration line occurring at the beginning of a number of data segments.

waveforms for specific types do not have much variance in shape. Bayesian model selection can then be used to determine what population of noise transient each new glitch belongs to [21, 24, 25]. First, one second of data around the trigger time is downsampled to 8192 Hz and a 10 Hz high pass filter is applied. Nested sampling is then used to calculate Bayes factors to determine the correct transient type [24].

3.3. WDF-ML

Wavelet detection filter (WDF) is an ETG that is part of the noise analysis package (NAP), developed by the Virgo collaboration [26, 27]. It is combined with a machine learning classifier for transient classification (WDF-ML).

In order to reduce the number of wavelet coefficients produced by WDF-ML, the data are downsampled before any data conditioning in the time domain to prevent border effects introduced by the fast fourier transform (FFT). The downsampling is a new feature of WDF-ML that was not implemented in the version of the algorithm used in our previous study. The data are then whitened using parameters estimated at the beginning of each locked segment. After whitening, a wavelet-transform is applied, using a bank of wavelets, as described in [11].
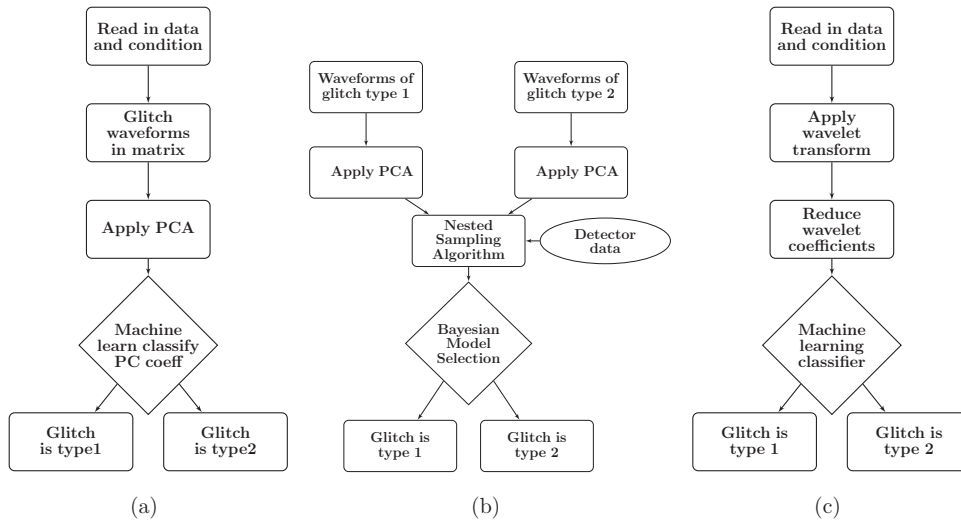
Figure 4. Classification procedures for the three different methods used in this study. (a) PCAT applies principal component analysis to all transients detected in a stretch of data and then applies a machine learning classifier to the principal component coefficients. (b) PC-LIB uses a combination of principal component analysis and Bayesian model selection to determine the glitch type. (c) WDF-ML applies a machine learning classifier to wavelet coefficients obtained by applying a wavelet transform to the transients in the data.

We use a window of 2048 points, with an overlap of 1968 points, which corresponds to a duration of 0.25 s, as transients are typically of a short (ms) duration.

The wavelet coefficients identified by the WDF-ML ETG are further cleaned using a wavelet de-noising procedure where only wavelet coefficients above the noise level are retained [11]. WDF-ML produces a list of wavelet coefficients, frequency, duration and SNR for each transient. The dimensions of the wavelet coefficients are then reduced by applying PCA and spectral embedding [28, 29]. The transient classification is then performed by applying a machine learning GMM classifier to the reduced wavelet coefficients [19].

4. Classification

In the following sections we show the classification results obtained by PCAT, PC-LIB and WDF-ML on aLIGO H1 and L1 data. All algorithms are run with the same configurations that we expect to use during O2 to better understand our performance during the future observing runs. To determine if the glitches are classified correctly spectograms of all glitches are made and visually inspected.

4.1. Livingston

To find the transients in the L1 data we look for triggers that are coincident within half a second in the outputs of all ETGs. The WDF-ML ETG was run with an SNR threshold of 10 at a sampling rate of 8192 Hz. Omicron was run with a lower SNR threshold of 5. We then look for transients that are coincident between both WDF-ML and Omicron, above SNR 20, and find a total of 426 coincident transients. As the PCAT ETG cannot find the lower frequency
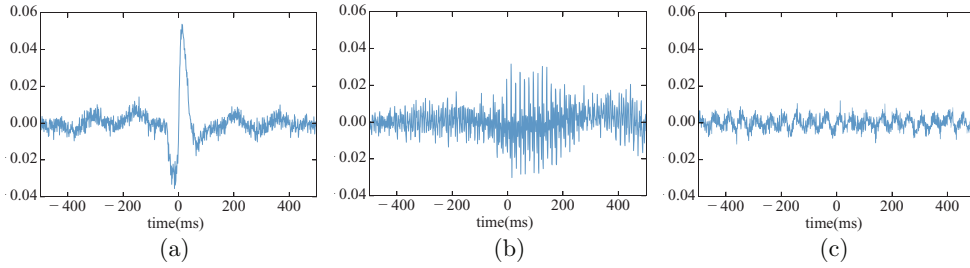
Figure 5. The typical high pass filtered and whitened time series waveforms for three of the most common transient types found in the Livingston detector. (a) A spike which appears as a tear drop in a spectrogram. (b) The time series waveform of the glitches shown in figure 2(d). (c) The time series of a whistle glitch.

(below 10 Hz) triggers and some longer duration triggers we still classify transients that are coincident between Omicron and WDF, but missed by PCAT, as those triggers would still be classified when running in low latency.

4.1.1. PCAT. Applied a threshold on the SNR of the transients of 4.5 and the maximum possible number of transient types was set to 10. The ideal number of PCs can be estimated by finding the knee of the data set variance curve. This gives a total number of 20 PCs.

PCAT classifies all the transients into 10 different classes. 90 triggers that were coincident between the Omicron and WDF-ML ETGs were missed by the PCAT ETG. Included in these missed triggers are all of the whistles, as their duration is longer than the PCAT analysis window, and 17 transients that are not visible in a spectrogram. 20 of the lower SNR hardware injections are also missed.

The data contains three main types of transients with examples shown in figures 2(a)–(c). As PCAT does not detect any of the whistles shown in figure 2(b) the remaining glitches are classified into two main types, further split into sub-types.

PCAT classes 1, 4 and 10 contain the transients which appear as a spike in the time series, as shown in the figure 5(a) and in the spectrogram in figure 2(a). Class 4 contains only 2 transients, class 1 contains 123 transients and class 10 contains 100 transients. Class 1 and 10 contain 11 and 20 hardware injections respectively. The three sub classes are characterized by different duration of the transients. Triggers in class 1 have the lowest ($\leqslant 0.005$ s) duration, class 10 have a larger ($\leqslant 0.01$ s) duration, and class 4 contains two longer ($\geqslant 0.01$ s) duration spikes. Two of the transients in class 10 were incorrectly classified.

Classes 3, 5, 6, 7 and 8 contain the transients with a time series waveform shown in figure 5(b) and a spectrogram shown in figure 2(d). Triggers in classes 5, 7 and 8 all have SNR values between 20 and 25 and durations of ∼0.01 s. Class 3 contains triggers of the same transient type but with larger durations ($\leqslant 0.02$ s) and SNR values up to 50. Class 6 contains only one transient, also of the same type, but with an SNR value of 57 and a duration value of 0.005 s.

PCAT classes 2 and 9 contain 11 and 7 glitches respectively. As these transients are not visible in a spectrogram it is not possible to determine what their type is and if they are classified correctly. Overall 95% of the transients are correctly classified by PCAT.

4.1.2. PC-LIB. Classifies all transients into four different types. To create the signal models the first 5 PCs for each glitch type are used as determined by the knee of the explained variance curve. Class 0 contains 33 transients that are not detected by PC-LIB and are thus

classified into a noise class. Most of the noise class transients occur at frequencies lower than the 10 Hz cutoff used by PC-LIB.

Class 1 contains 249 transients that correspond to PCAT class 4 and 5 and appear as a spike in the time series. Two of the transients in this type are mis-classified. All of the hardware injections in the data are classified in this type. Class 2 contains 131 transients which correspond to PCAT sub-classes 3, 5, 6 and 7 and have a time series waveform shown in figure 5(b). There are no incorrectly classified transients in this class.

Finally, class 3 contains 13 transients. Most of the transients in this class are the whistle transients shown in figure 2(b) and in figure 5(c). Three of the transients in this class are mis-classified and should be in class 2. Overall PC-LIB classifies 98% of the detected transients correctly.

4.1.3. WDF-ML. Classifies all transients into five different classes. The 5 classes consist of two different types of transients as WDF-ML cannot accurately classify the longer duration whistles due to the short analysis window. Sub-classes are determined by the wavelet family of the transients rather than split by duration or SNR as for PCAT.

WDF-ML class 0 contains 195 transients and class 3 contains 86 transients that appear as a spike in the time series. These correspond to PCAT types 1, 4 and 10 and PC-LIB's class 1. The two sub-classes contain 29 hardware injections. They also contain 8 of the whistle glitches, as WDF-ML cannot accurately classify longer duration transients. Four of the class 0 and one of the class 3 transients are incorrectly classified.

The second main transient type found by WDF-ML corresponds to PC-LIB's class 2 and contains transients characterized by the typical spectrogram shown in figure 2(d). The transients were split into three sub-classes, namely class 1 that contains 46 transients, class 2 that contains 70 transients and class 4 that contains 29 transients. Class 1 contains three incorrectly classified transients and class 3 contains two of the whistles glitches. Class 4 contains 4 hardware injections that are mis-classified. Overall WDF-ML classifies 95% of the L1 transients correctly.

4.1.4. Comparison. Figure 6 shows a comparison of the classifications made by all three methods. All methods are able to classify transients with a high level of accuracy in real non-stationary data. WDF-ML performs better at classifying very low frequency transients as it does not need to use a lower frequency cutoff. The Omicron SNR, duration and frequency of all 426 transients are shown in figure 6(d). The constant lines are due to the Omicron's method for measuring frequency [16].

Only PC-LIB is able to separate the whistle transients into a separate class due to the longer 1 s time window used by this method. The efficiency in classifying these transients for the other algorithms could be improved by using a longer time window. However, this could lead to multiple shorter duration transients occurring in the same time window. As PC-LIB looks for specific known transient types it could be used to add labels to the classifications of the other methods so that it will make it easier to find out which class corresponds to which transient type, defined in [15], and which classes are new types that have not occurred previously. As WDF-ML and PCAT can classify new transient types as soon as they appear in the data they can be used to provide waveforms for PC-LIB to use to create new signal models.

4.2. Hanford

As for the L1 data transients coincident within 0.5 s between all ETGs are classified. A higher SNR threshold of 30 is used for H1 as the data contains many more transients than the L1 data and is more non-stationary. A total of 1865 coincident transients are classified in H1.
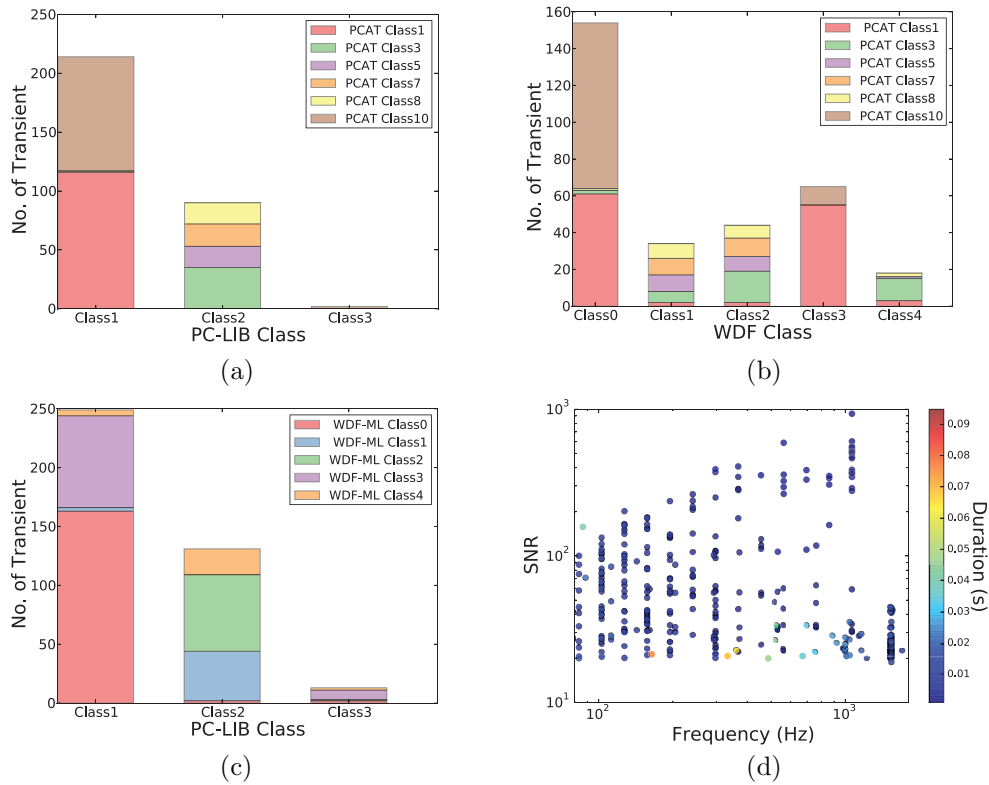
Figure 6. Classification comparisons for the three different methods on the data from LIGO Livingston. (a) Compares the classification results of PCAT and PC-LIB. PCAT class 2, 4, 6 and 9 are not shown as they contain less than 15 transients. (b) Compares the classification results of PCAT and WDF-ML. (c) Compares the classification results of PC-LIB and WDF-ML. (d) The SNR and frequency of all the transients classified in the data.

4.2.1. PCAT. Uses 20 PCs to classify the transients into 7 different types. 120 of the transients coincident between WDF-ML and Omicron ETGs are not detected by the PCAT ETG. They are transients below 10 Hz or triggers from the long duration lines, shown in figure 3(d), which are not really glitches. The detected transients are split into 7 different classes.

The data contains two main types of transients. The first type is characterised by a typical spectrogram shown in figure 3(a) and a time series waveform shown in figure 7(a). PCAT splits this type into 6 different sub-classes with 267, 603, 648 44, 1, and 64 transients respectively. Class 1 has 9 mis-classified transients. Classes 2, 3 and 5 all have one mis-classified transient. Class 2, 3 and 6 contain lower duration ($\sim$0.005 s), but with different Q and frequency ranges, where Q is defined as $Q = \text{duration} \times 2\pi \times \text{frequency}$. Class 1, 5 and 16 contain relatively longer duration waveforms ($\sim$0.01 s) which also have different Q and frequency ranges.

The second type of transient has a typical time-frequency morphology shown in figure 3(c) and time series waveforms shown in figure 7(b). This type is found in PCAT class 4 that contains 117 transients that are all classified correctly. Overall PCAT classifies 99% of the detected H1 transients correctly.
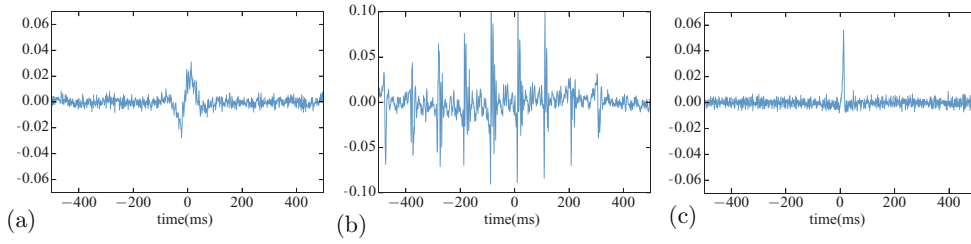
Figure 7. The typical whitened and high pass filtered time series waveforms for three of the most common transient types found in the Hanford detector. (a) A spike in the time series that appears as a tear drop in a spectrogram. (b) The time series of second most common glitch found in the data, as shown in figure 3(c). (c) The time series waveform of a hardware injection.

4.2.2. PC-LIB. As with the L1 data we use 5 PCs to create signal models for the H1 transients. PC-LIB splits the transients into two different classes. A noise class contains the 6 transients shown in figure 3(d) as they cannot be detected.

Class 1 contains 1651 transients that correspond to a spike in the time series as in PCAT sub-classes 1, 2, 3, 5, 6 and 7. This class also contains 13 hardware injections. 23 transients are mis-classified and should be in class 2.

Class 2 contains 207 transients, which have a typical spectrogram shown in figure 3(c), and correspond to PCAT class 4. This class also includes 4 hardware injections that are more similar to a sine-Gaussian in shape than those classified into class 1. This class includes 61 transients that are mis-classified and should be in class 1. Overall PC-LIB classifies 95% of the detected H1 transients correctly.

4.2.3. WDF-ML. Splits the H1 data into three different classes. Class 1 contains 1358 transients, which appear as a spike in the time series, and correspond to PC-LIB class 1 and the 6 PCAT sub-classes. This class contains all hardware injections and all very low frequency transients that can not be detected by PCAT and PC-LIB. 10 of the transients in this class are mis-classified. WDF-ML class 2 contains 145 transients that are characterized by spikes in the time series, but have longer durations and lower SNR values than the transients in WDF-ML class 1.

WDF-ML class 0 contains 326 transients corresponding to PCAT class 4 and PC-LIB class 2. This class also contains 122 mis-classified transients. As before, this is because all of the transients in the class have a duration ($\sim$1 s) which is much longer than the time window used in the WDF-ML analysis. Overall WDF-ML classifies $\sim$92% of the H1 transients correctly.

4.2.4. Comparison. The results obtained by all three methods for the H1 data are compared in figure 8. The Omicron SNR, duration and frequency of the transients is shown in figure 8(d). As WDF-ML uses a small time window of 0.25 s the efficiency of the classification is reduced when the data are highly non-stationary and contain many long ($\sim$1 s) duration transients. Even with 137 mis-classified transients the overall accuracy of the WDF-ML H1 results is $\sim$92%. WDF-ML estimates the PSD at the beginning of each locked segment. This may introduce errors towards the end of the segment if the data is highly non-stationary. Machine learning methods perform better when the data set analysed is large. Therefore, the larger number of glitches in H1 may have improved the classification efficiency.
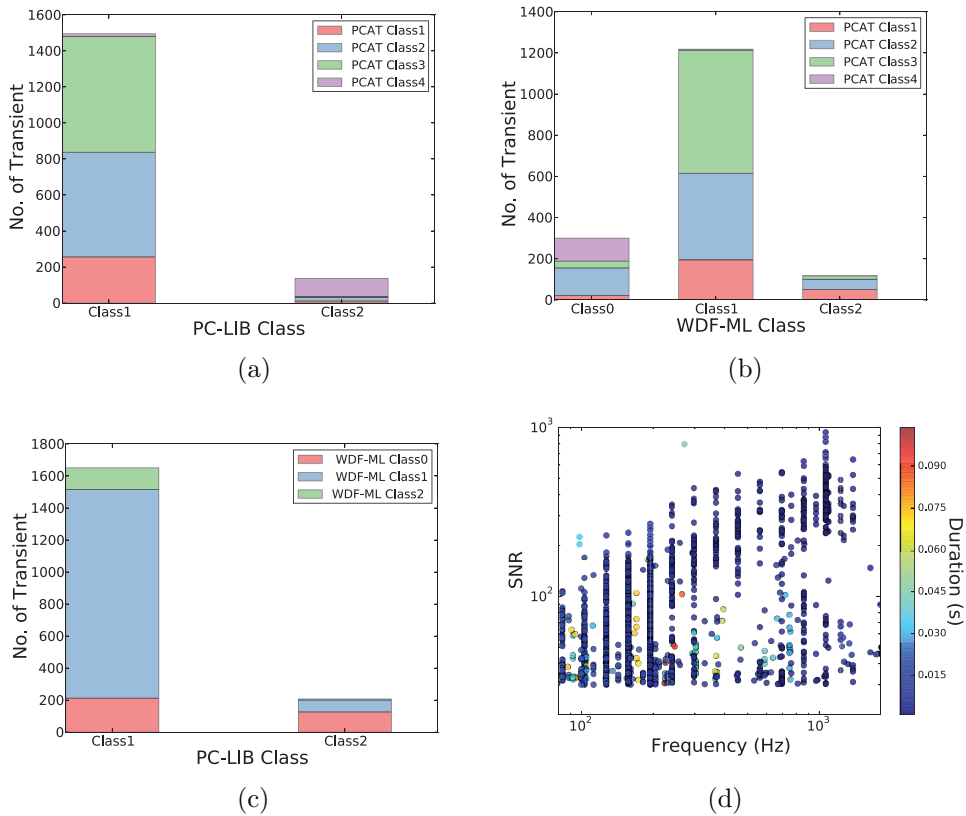
Figure 8. Classification comparisons for the three different methods for aLIGO Hanford data. (a) PC-LIB splits the transients into two classes. PCAT can split different types into sub-classes. (b) PCAT and WDF-ML comparison. WDF-ML has difficulty with transients which have a larger duration than their analysis window. (c) Comparison of PC-LIB and WDF-ML classifications. (d) The Omicron SNR, duration and frequency of all the transients classified in the data. The discreteness in frequency is due to Omicron.

5. Discussion

Non-Gaussian noise in the aLIGO and Virgo detectors can potentially mimic a gravitational-wave signal, reduce the duty cycle of the instruments and decrease the sensitivity of the detectors. Classification of different noise transient signals may help identify their origins and lead to a reduction in their number. We have developed three methods for noise classification and have previously demonstrated their performance on simulated transients in simulated Gaussian aLIGO noise [11]. However, as real noise from the advanced detectors is non-stationary and non-Gaussian, a better understanding of how our methods will perform during the upcoming observation runs of the advanced detectors is required. Although the detectors are typically more stable during observing runs than during ER7, we expect the types of glitches investigated in this work to be representative of the glitch classes in the upcoming observing runs.

In the ER7 data from the L1 detector 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 95% of them were correct. In the H1 data 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.

The efficiency of the WDF-ML algorithm is reduced for the Hanford glitches because the duration of the transients becomes much larger than the analysis window, which reduces the efficiency of the overall classification. This could be prevented by applying a high duration cutoff to the transients found by the ETG before classification. Most high duration and SNR transients are removed by data quality vetoes. Conversely, short duration transients will be more important as they have a higher impact on the gravitational-wave search backgrounds. Since they are rarely removed by vetoes their accurate classification is crucial to improve gravitational-wave searches as an accurate categorization will allow us to search for couplings within the detector [7, 30]. In the future the whitening performed by WDF-ML will be improved by using a technique known as adaptive whitening [31].

Because of the different strengths and weaknesses of the different methods having multiple classifiers is a winning strategy. WDF-ML can classify lower frequency transients than the other two methods. PC-LIB is better able to classify longer duration transients due to its longer analysis window. PCAT can classify new types of transients as soon as they appear in the data and thus provide transient waveforms for PC-LIB's signal models.

Further improvements could also be made by using a training set of pre-classified waveforms or exploring the use of dictionary learning algorithms for glitch classification [32]. The aLIGO gravity spy project aims to build these data sets through a citizen science program [33, 34].

Acknowledgments

References

[1] Harry G M and LIGO Scientific Collaboration 2010 *Class. Quantum Grav.* **27** 084006
[2] Aasi J *et al* 2015 *Class. Quantum Grav.* **32** 074001
[3] Acernese F *et al* 2008 *Class. Quantum Grav.* **25** 114045
[4] Allen B, Anderson W G, Brady P R, Brown D A and Creighton J D E 2014 *Phys. Rev.* D **85** 122006
[5] Abbott B P *et al* 2016 *Phys. Rev. Lett.* **116** 061102
[6] Abbott B P *et al*, LIGO Scientific Collaboration and Virgo Collaboration 2016 *Phys. Rev. Lett.* **116** 241103
[7] The LIGO Scientific Collaboration and the Virgo Collaboration 2016 *Class. Quantum Grav.* **33** 134001
[8] Smith J R, Abbott T, Hirose E, Leroy N, MacLeod D, McIver J, Saulson P and Shawhan P 2011 *Class. Quantum Grav.* **28** 235005

[9] Ajith P, Isogai T, Christensen N, Adhikari R X, Pearlman A B, Wein A, Weinstein A J and Yuan B 2014 *Phys. Rev.* D **89** 122001

[10] Aasi J *et al*, The LIGO Scientific Collaboration and The Virgo Collaboration 2015 *Class. Quantum Grav.* **32** 115012

[11] Powell J, Trifirò D, Cuoco E, Heng I S and Cavaglià M 2015 *Class. Quantum Grav.* **32** 215012

[12] Mukherjee S, Obaid R and Matkarimov B 2010 *J. Phys.: Conf. Ser.* **243** 012006

[13] McIver J *et al* 2015 https://wiki.ligo.org/DetChar/DataQuality/PreO1DQShifts#ER7

[14] Areeda J and Smith J 2016

[15] Smith J https://dcc.ligo.org/LIGO-G1500642/public

[16] Robinet F 2015 https://tds.ego-gw.it/ql/?c=10651

[17] Chatterji S, Blackburn L, Martin G and Katsavounidis E 2004 *Class. Quantum Grav.* **21** S1809–18

[18] Hastie T, Tibshirani R and Friedman J 2001 The Elements of Statistical Learning (Berlin: Springer)

[19] Pedregosa F *et al* 2011 *J. Mach. Learn. Res.* **12** 2825–30

[20] Essick R, Vitale S, Katsavounidis E, Vedovato G and Klimenko S 2015 *Astrophys. J.* **800** 81

[21] Lynch R, Vitale S, Essick R, Katsavounidis E and Robinet F 2015 arXiv:1511.05955

[22] Heng I S 2009 *Class. Quantum Grav.* **26** 105005

[23] Logue J, Ott C D, Heng I S, Kalmus P and Scargill J H C 2012 *Phys. Rev.* D **86** 044023

[24] Sivia D S 1996 *Data Analysis: a Bayesian Tutorial* (Oxford: Oxford University Press)

[25] Veitch J *et al* 2015 *Phys. Rev.* D **91** 042003

[26] Acernese F *et al* 2005 *Class. Quantum Grav.* **22** S1041

[27] Acernese F *et al* 2007 *Class. Quantum Grav.* **24** S671

[28] Ng A Y, Jordan M I and Weiss Y 2001 On spectral clustering: analysis and an algorithm *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press) pp 849–56

[29] Belkin M and Niyogi P 2003 *Neural Comput.* **15** 1373–96

[30] The LIGO Scientific Collaboration and The Virgo Collaboration 2016 *Phys. Rev.* D **93** 122004

[31] Cuoco E, Cella G and Guidi G M 2004 *Class. Quantum Grav.* **21** S801

[32] Torres-Forné A, Marquina A, Font J and Ibáñez J 2016 *Phys. Rev.* D at press (arXiv:1612.01305)

[33] Zevin M *et al* 2016 Gravity Spy: Integrating advanced LIGO detector characterization, machine learning and citizen science *American Astronomical Society Meeting Abstracts* (*American Astronomical Society Meeting Abstracts* vol 228) (arXiv:1611.04596)

[34] Simpson R, Page K R and De Roure D 2014 Zooniverse: Observing the world's largest citizen science platform *Proc. of the 23rd Int. Conf. on World Wide Web* (New York: ACM) pp 1049–54