

Using a cleaning technique for the search of continuous gravitational waves in LIGO data

Paola Leaci¹, Pia Astone², Maria Alessandra Papa^{1,3} and Sergio Frasca²

¹ Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, 14476 Golm, Germany

² University 'La Sapienza', 00185 Rome, Italy

³ University of Wisconsin-Milwaukee, WI 53201, Milwaukee

E-mail: paola.leaci@aei.mpg.de and maria.alessandra.papa@aei.mpg.de

Abstract. High frequency short events, due for instance to delta-like spurious disturbances, may affect the broad band noise level and thus produce a loss in the efficiency of detection of continuous gravitational waves. We identify such events, remove them from a set of LIGO fifth science run (S5) data and characterize the resulting sensitivity improvements. We use the same parameter values as used in a previous pilot study on Virgo data, but do not observe on LIGO S5 data the same sensitivity improvements.

PACS numbers: 04.80.Nn, 07.05.Kf, 95.55.Ym

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The detection of continuous gravitational wave (CW) signals is one of the challenges for ground based gravitational wave detectors [1, 2, 3, 4, 5]. The sensitivity of any CW search is determined by the level of noise and of spectral line contamination in the frequency band of interest and by the search technique. The noise level is entirely determined by the instrument and in fact the sensitivity of gravitational wave detectors is usually represented by graphing the power spectral density (PSD) as a function of frequency. Large short-lasting disturbances may affect the level of this noise floor. The sensitivity of CW searches may be improved by removing this type of disturbance before performing the search, irrespective of having identified their origin. Such a pre-processing procedure was first proposed in [6], which is the pilot study referred to in the abstract. In the present paper we characterize its performance on LIGO data from the 5th science run (S5). This run started on November 4 (14), 2005 at the LIGO Hanford (Livingston) Observatory and ended on October 1, 2007.

We stress that this paper reports on the results obtained using as input only science-mode data, i.e. data segments covering a time period for which the detector is optimized for data collection and ready to observe a possible astrophysical event.

2. Cleaning procedure

The cleaning procedure we describe here identifies disturbances in the time domain data and removes them from the data set. Such a data set is divided in 1800 s chunks and each of these is cleaned separately. After the cleaning step, each chunk is Fourier-transformed producing data in a format that is used as input to many CW searches. This Fourier-domain data is referred to as Short time-baseline Fourier Transforms, SFTs in short. The steps carried out to accomplish the cleaning on each chunk are:

- (i) A 10th-order *Butterworth filter* (at 38 Hz): it's a high-pass filter useful to minimize the contribution of the prominent low frequency component.
- (ii) *Tukey window*: the 0.1% of the data are effectively lost unless the data are interlaced.
- (iii) A 1st-order *high-pass filter* (at 100 Hz) and event identification: the events are identified on this data.
- (iv) *Time Domain cleaning procedure*: after finding events and registering some of their characteristics, their amplitude is subtracted from the amplitude of the original time series. Note that the high-passed time series, containing all the events, is not the same as the original time series because the high-passed time series does not have the low frequency component of the original data.

In order to identify the events, we set a threshold on the critical ratio (CR) of the high-passed data (produced in the third step above). The CR is defined as

$$\text{CR} = \left| \frac{y(i) - \mu_A(i)}{\sigma_A(i)} \right|, \quad (1)$$

where $y(i)$, $\mu_A(i)$ and $\sigma_A(i)$ are the data samples, the mean and the standard deviation, respectively. These last two quantities are estimated in an autoregressive way with an adaptive threshold, i.e. a threshold that changes with time as the sensitivity of the detector changes, as a consequence of the non-stationary noise. We chose a threshold value of 5. The value of the time constant for the autoregressive estimation was chosen to be 20 s, which is appropriate to track typical changes in the statistics of the data. An event starts when the amplitude of a sample goes above threshold ($\text{CR} > 5$). The last sample of an event is the last sample above threshold followed by a series of samples all lying below the threshold and covering a time longer than a given dead-time interval Θ . In other words, the event identification scheme takes as belonging to the same event all samples above threshold which are closer than Θ . Events begin and end with a sample over threshold. We have taken $\Theta = 0.1$ s as this is smaller than the average duration of the typical disturbances that we want to remove, which typically have a few threshold-crossings.

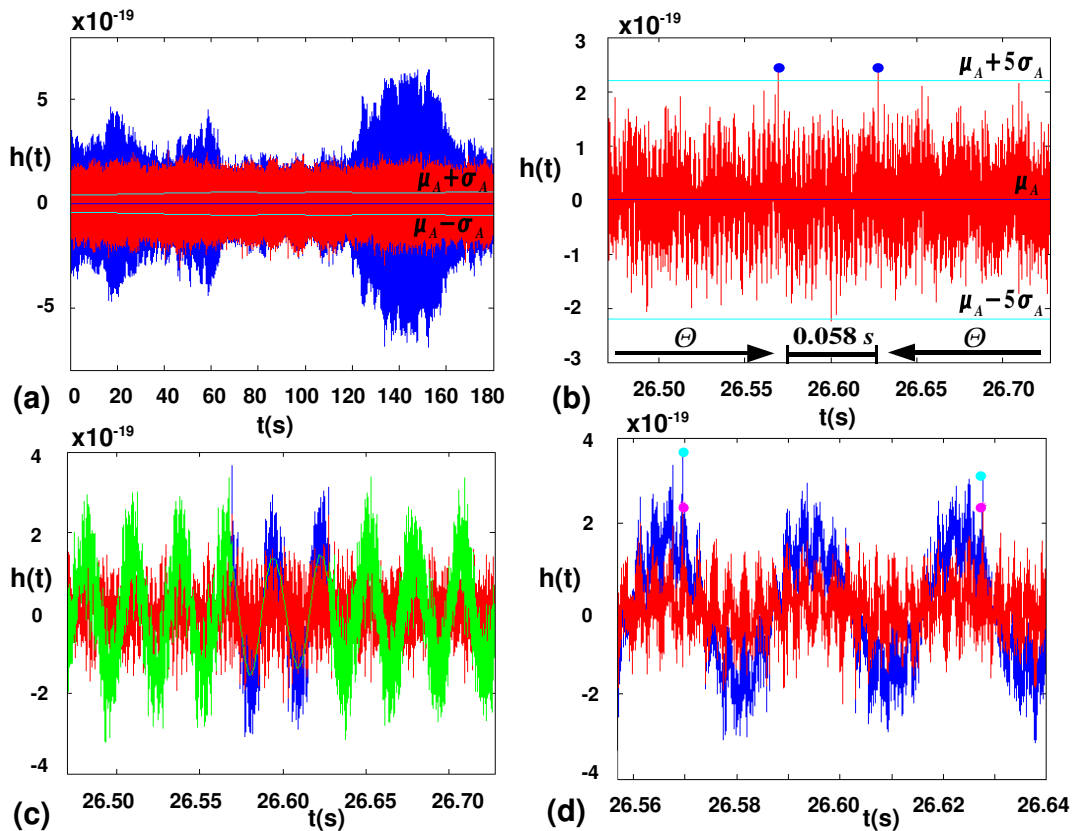


Figure 1. Strain amplitude $h(t)$ curves versus the time. The blue curve represents $h(t)$ after a Butterworth high-pass filter, with a knee frequency of 38 Hz. The red curve consists of the further application of a high-pass filter, with a cutoff frequency of 100 Hz to identify 'high frequency spikes'. The 3 horizontal lines in the panel (a) represent the autoregressive estimation, $\mu_A \pm \sigma_A$ (cyan lines) and μ_A (blue line). The green curve corresponds to the cleaned data set and all the circles in the subplots label the beginning and the end of the considered event, lasting for about 0.058 s.

After the events have been identified they are subtracted from the original series, as was explained above. To avoid discontinuities in the cleaned time series, a few samples before and after the event are used to smoothly connect the original and the cleaned data. We used ten samples, corresponding to 6×10^{-4} s, being the sampling time of $\sim 6 \times 10^{-5}$ s.

As an example, figure 1 (a) shows data from the 4 km Hanford (H1) detector taken during S5. The original time data series $h(t)$ is the blue curve and the time series on which the events are identified is the red curve. We further zoom in the region where an event has been identified. The event contains 954 samples, corresponding to roughly 0.058 s. As we can see in figure 1 (b), the data lies above the threshold ($CR > 5$) twice (blue circles). After the subtraction of the amplitudes of all the samples of the event, the cleaned time series is obtained and plotted in figure 1 (c) (green curve). A further zoom of the region where the subtraction is applied is shown in figure 1 (d). Note that the cleaned time series (green curve) is obtained as subtraction between the original

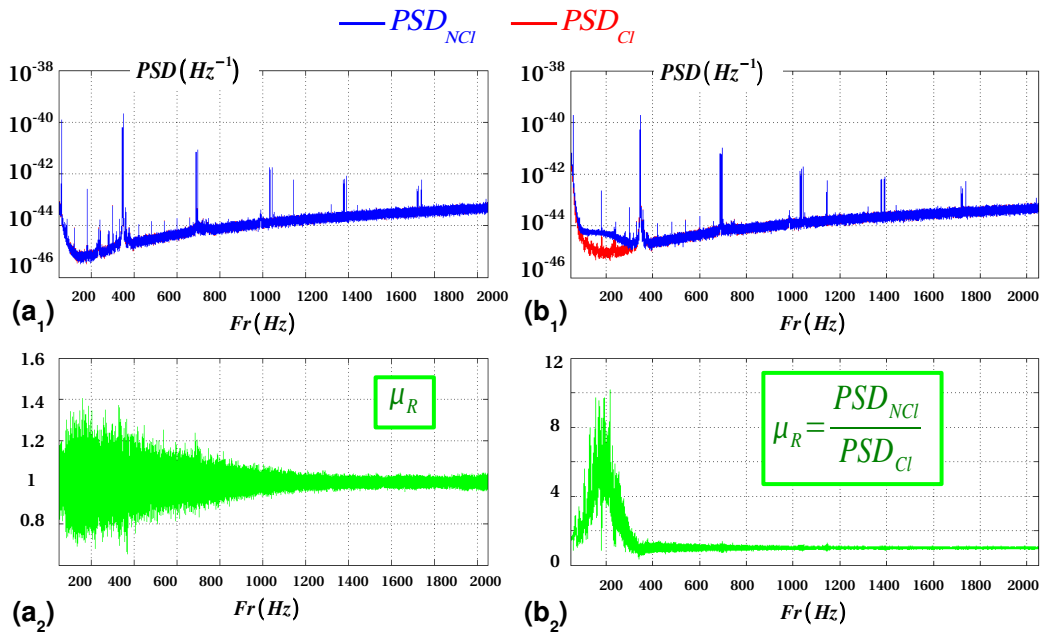


Figure 2. Two examples of noise floor estimation. Panels (a₁), (b₁): the PSD of uncleaned data (PSD_{NCI} , blue curve) is compared to the PSD of the corresponding cleaned data (PSD_{CI} , red curve). It is evident that the data analyzed in the panel (b₁) is more disturbed. Therefore, the cleaning procedure is more effective in this case. In the bottom panels (a₂), (b₂), the ratio μ_R between the PSDs is also shown versus the frequency.

time series (blue curve) and the high-passed series (red curve) only for the amplitudes of those samples labelled as part of an event. The green and blue curves coincide for the remaining samples that do not belong to any event.

The autoregressive mean μ_A (blue line) and standard deviation σ_A are also shown in figure 1 (a,b).

3. Results and conclusions

We have applied this cleaning procedure to data from the S5 run of the H1 detector taken between GPS time 846374761 (Nov 01 00:05:47 GMT 2006) and GPS time 875234357 (Mon Oct 01 00:39:03 GMT 2007). Contiguous segments of 1800 s science-mode data were selected and yielded 12590 SFTs of the data. We present here results only for data up to May 16 23:00 GMT 2007 (7007 SFTs).

The cleaning procedure, consisting of a set of C-code routines in the PSS package [7], has been integrated into some of the SFT-production software used by the LIGO Scientific Collaboration (LSC) [8].

We illustrate in figure 2 (a₁) and (b₁) the improvements that may be obtained when the data is particularly quiet and disturbed, respectively. We show the PSD of 1800 s of H1 data with and without cleaning. We can appreciate a decrease of the noise floor of a factor ~ 5 in the band around 200 Hz in the cleaned data with respect to the original

data. As shown in figure 3, the increase in the level of the noise floor is due to a single very loud disturbance (pointed by the arrow in the top panel of figure 3) with an energy that is about 30 times greater than the mean of the other events. The gain in sensitivity (of order 2) more than counter-balances the effective loss in observation time for having effectively decreased the observation time, ~ 1.1 s out of 1800 s.

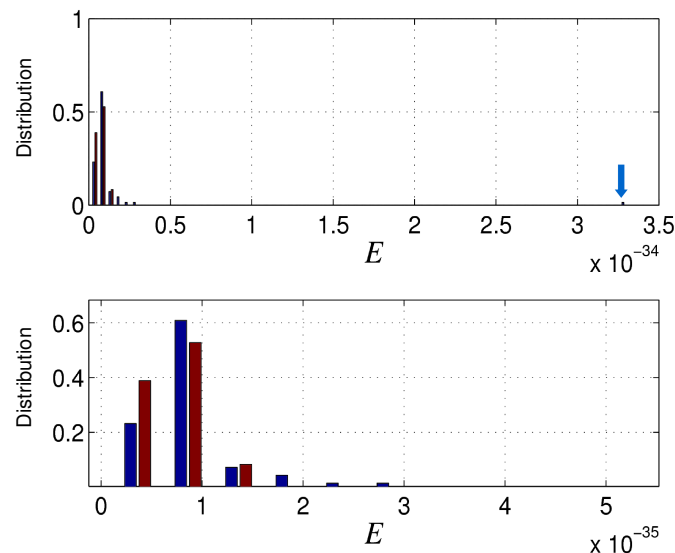


Figure 3. Distribution of the event energies $E = \sum_i |y(i)|^2$ of 36 (red bars) and 69 (blue bars) events found for the data whose PSDs are plotted in red in figure 2 (a₁) and (b₁), respectively. The bottom plot is a zoom of the top one. The two populations of disturbances have very similar energy distributions. What produces the increase in the noise floor of the PSD of figure 2 (b₁) is the very loud event at $E \sim 3.3 \times 10^{-34}$. We point out that the energy is preserved, i.e. the Parseval theorem is satisfied.

Figure 4 shows the cumulative results on 7007 SFTs. In particular, we pick four 1 Hz bands centered at 500, 857, 1370 and 1630 Hz and compute the mean value μ_R of the ratio between the PSD of 7007 uncleaned and cleaned data series. These bands do not contain spectral disturbances and have been chosen to be representative of the noise floor at nearby frequencies. The histograms of μ_R are shown in figure 4. The average value of μ_R is 1.0055, 1.0026, 1.0011 and 1.0009 for the different four bands, respectively.

The expected improvement in sensitivity for a CW search may be expressed by the lowest detectable strain amplitude h_0 . For a coherent search this is proportional to $\sqrt{PSD/T}$, where T is the coherent observation time. Hence we denote the gain in sensitivity from the cleaning procedure, G_{h_0} , with

$$G_{h_0} = \sqrt{\mu_R \frac{T - T_e}{T}}, \quad (2)$$

where T_e is the sum of the duration of all the events subtracted during the time T . Figure 5 shows the distribution of the duration of events of 100 SFTs uniformly selected from the original data set. The mean value of T_e is 0.55 s. If we use such a mean

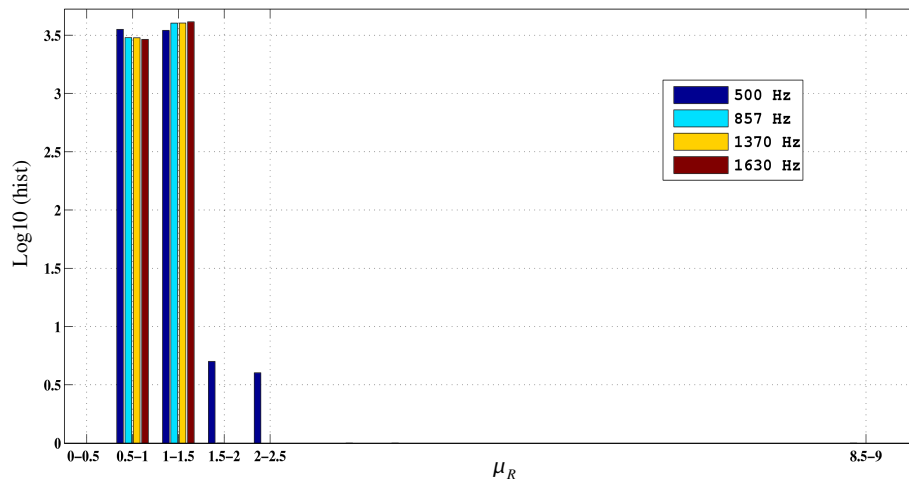


Figure 4. Log10 of the histogram of the mean value μ_R of the ratio between the PSD of the uncleaned and cleaned data series, computed in four 1 Hz frequency bands. For clarity the x-axis labels indicate the 0.5 wide bin intervals.

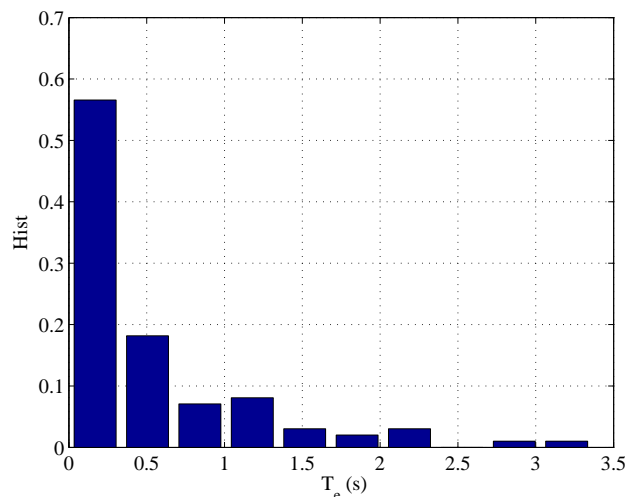


Figure 5. Histogram of the duration of events (T_e) identified by the procedure.

value in equation (2) and the 1-Hz-mean values of μ_R shown in figure 4, we obtain the distributions shown in figure 6. The effect of cleaning is to decrease the PSD but also, effectively, to decrease the observation time because also the signal is removed when one subtracts the disturbed events. Figure 6 shows that on typical LIGO science-quality S5 data the gain in sensitivity is marginal: it happens in a very small fraction of the total SFTs and not uniformly at all frequencies. The mean values of G_{h_0} for the bands of figure 6 are 1.0016, 1.0010, 1.0004 and 1.0003, respectively. On data from the first Virgo science run (VSR1) improvements of the order of 1.07/1.3 at high/low frequencies were observed analyzing roughly 10 days of data due to the fact that non science-mode were used, unlike what was done with the LIGO data. A further cleaning procedure is under

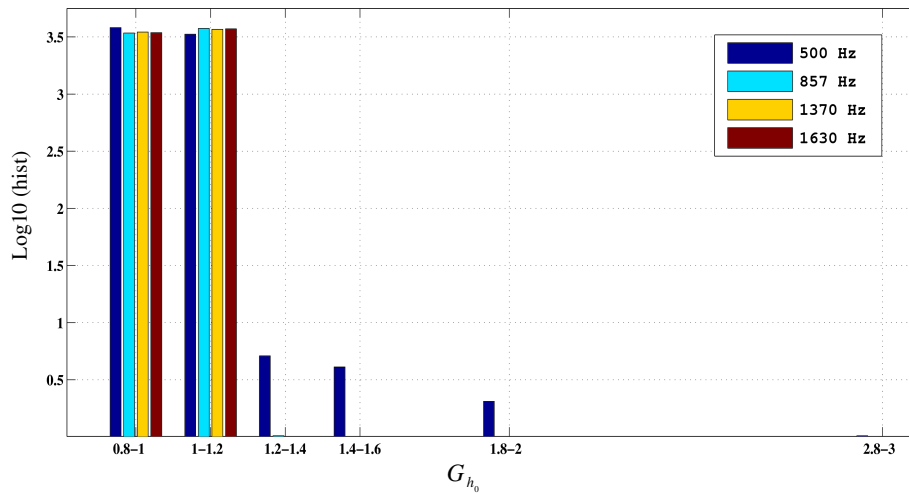


Figure 6. Log10 of the histogram of G_{h_0} computed assuming $T_e = 0.55$ s and using the average μ_R values on the four 1 Hz bands centered respectively at 500, 857, 1370 and 1630 Hz.

development [9] aiming at removing disturbances in small frequency sub-bands, a few Hz each.

Acknowledgments

The authors would like to thank B. Machenschalk for his ongoing work aimed at integrating the original PSS library cleaning routines into the LALAPPS SFT production routines. We also acknowledge the LSC-Virgo continuous waves search working group for useful discussions and input.

References

- [1] Abbott B P *et al* 2009 *Phys. Rev. D* **79** 022001
- [2] Abbott B P *et al* 2009 *Phys. Rev. D* **80** 042003
- [3] Abbott B P *et al* 2009 *Phys. Rev. Lett. D* **102** 111102
- [4] Owen B J 2009 *arXiv:0903.2603*
- [5] Owen B J 2009 *Class. Quantum Grav.* **26** 204014
- [6] Astone P *et al* 2005 *Class. Quantum Grav.* **22** S1197
- [7] <http://wwwcascina.virgo.infn.it/cgi-bin/cvsweb/cvsweb.cgi/>; <http://grwavsf.roma1.infn.it/PSS/>
- [8] <https://www.lsc-group.phys.uwm.edu/daswg/projects/lalapps.html>
- [9] Astone P *et al* (in preparation)