LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Technical Note **LIGO-T1900398-v1** 2019/07/01

Analisys of correlations between non-stationary noise and auxiliary channels in LIGO

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

[1]Universita' di Napoli Federico II, Complesso Universitario di Monte S.Angelo, I-80126 Napoli, Italy
[2]LIGO, California Institute of Technology, Pasadena, CA 91125, USA

Distribution of this document:

LIGO Scientific Collaboration

Draft

California Institute of Technology
LIGO Project, MS 18-34
Pasadena, CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project, Room NW17-161
Cambridge, MA 02139
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

LIGO Hanford Observatory
Route 10, Mile Marker 2
Richland, WA 99352
Phone (509) 372-8106
Fax (509) 372-8137
E-mail: info@ligo.caltech.edu

LIGO Livingston Observatory
19100 LIGO Lane
Livingston, LA 70754
Phone (225) 686-3100
Fax (225) 686-7189
E-mail: info@ligo.caltech.edu

`http://www.ligo.caltech.edu/`

1 Non-stationary noise in LIGO

Data from the LIGO detectors typically contain non-Gaussian and non-stationary noise which arise due to instrumental and environmental conditions. This non-stationary noise implies a time variations of the LIGO detectors sensitivity.
Operating the LIGO detectors at the sensitivity needed for gravitational-wave detection requires a large number of control loops and sensors to precisely control and monitor the instrumentation and environment. The channels that continually record information from those sensors allow to analyze the problematic variations in the sensitivity of the detector. However, due to the complexity of the system, the number of auxiliary sensor channels is around hundreds of thousands. The LIGO auxiliary channels include microphones, seismometers, magnetometers, photo-diodes, voltage monitors, etc.
Searching through such a large amount of data to determine which sensors are most highly correlated with variations in the considered frequency range is an important and difficult challenge. Characterizing these correlations from the data and improving the background of LIGOs searches is crucial for reducing non-stationarity in the detectors and increasing the statistical significance of gravitational-wave candidate events.

2 Computation of BLRMS and multi-dimensional linear regression

To perform quantitative analysis on the noise variations over time, it is useful to compute the band-limited RMS (BLRMS), that is the total power of the noise in a limited frequency band [1]. If this computation is performed periodically it is possible to track the time evolution of the noise.
In this way, considering n frequency bands and chosing a suitable time window t for the averaging of the power spectrum and a time T as the sampling rate to compute the BLRMS, the results are n new signals which give the noise power in each band as a function of time [2].
Once the n frequency bands have been selected, the time evolution of the BLRMS can be analyzed. In this time evolution, *glitches* could be observed. A glitch is a high-amplitude, short-duration excursion of unknown origin in the control loop used to sense the differential motion of the interferometer arms, which is calibrated to produce GW strain data [3]. To identify glitches, for each point the median value is computed and a dispersion (a sort of standard deviation of the data, but using median instead of mean) in a chosen time band. It is used the median value and not the mean value because the first is less dependent to the value of the glitch.
In this way, it is possible to put a threshold (median plus three times this "median-based-standard-deviation") on the BLRMS value to identify glitches and reject them using an interpolated value.
Using the same procedure, it is possible to identify *lines*, that are strong sinusoidal contribution at a given frequency in the GW power spectrum. The BLRMS of a band containing a line might be completely dominated by the power in the line itself [1]. Therefore every fluctuation of the noise floor is hidden by the line amplitude.
Since the aim is to analyze the non-stationary noise floor on long time scales (approximately days), glitches and lines are not relevant in this context. As in the same way of glitches, lines can be ex-

cluded and BLRMS can be computed over frequencies excluding the lines.
Obtaining the time evolution of the BLRMS, that is already computed [2], for a certain frequency band without glitches and lines is a first step toward finding correlation between the noise variation and auxiliary channels. The next step is the use of statistical method to verify these potential correlations, in particular applying a multi-dimensional linear regression.
The values at N different times y_i of the BLRMS in a given band can be viewed as the noisy version of a linear combination of the k auxiliary channels measured at the same times $x_{1\cdots k,1\cdots N}$ [4],

$$y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + \varepsilon_i\,, \tag{1}$$

where the unknown coefficients β have to be estimated and ε is a source of additional noise, assumed to be normally distributed and considered as coming from the measurement of the y or also from the auxiliary channels. Therefore the hypothesis is that the input x signals are deterministic and that the output is a random variable given by an unknown linear combination of the x plus some random noise. The best estimation of the coefficients is obtained by minimization of the sum of squared errors and if all the auxiliary signals are linearly independent, there is a unique solution. This technique is useful to understand which amount of the total noise non-stationarity can actually be predicted from a given set of auxiliary channels but it is easy to show that in some cases, standard linear regression can fail.
Combining some input signals, a standard linear regression can be performed to compare the estimated coefficients to the real ones. If there is some background noise, the regression coefficients are close to the real ones, but some other random coefficients also can get large. In addition, in the case of no-correlated input signals, a good recontruction can be obtained if the number of signals is large enough.
Because of the great number of auxiliary channels in the LIGO detector, this problem of ”overfitting” must be considered.

3 Goals for this project

To conclude, my work will analyze the noise non-stationarity on long time scales using the BLRMS and the correlations with the auxiliary channels of the detector via linear regression. Overfitting problems could occur due to the high number of auxiliary channels and solutions must be found.
Overfitting problems are also common in machine learning and probably it could be useful to consider some consolidated techniques in this field to figure out these problems, like regularization or dimensionality reduction.
Similar analysis can be found, for example, in [5] where it is described an algorithm that uses linear regression, namely LASSO (least absolute shrinkage and selection operator) regression, to analyze all of the auxiliary channels and identify a small subset of them that can be used to reconstruct variations in LIGO's astrophysical range.
In this context, my work can start analyzing a small subset of the auxiliary channels in order to implement a multidimensional linear regression. Then later, this subset can be enlarged to check the computational cost of the analisys.
Starting from BLRMS segments, the aims of the works are to find solutions to overfitting problems with a particular attention to the computational cost of the analisys in order to evaluate correlations between the time evolution of the noise and other detector channels.

Table 1: Planning

Weeks	Aims
week1	Study of the Python code to create the BLRMSs and of the time evolution in different band.
week2 - week3	Analysis of small subsets of the auxiliary channels in order to apply the linear regression method.
week4 - week7	Study of the correlated auxiliary channels in the chosen frequency bands and test of possible idea to reduce overfitting using simulated data.
week8 - week10	Application of the proposed solutions to real data.

4 Plan

In the Table 1, a general plan for the ten weeks work is described.

References

[1] G. Vajente, ”Analysis of sensitivity and noise soures for the Virgo gravitational wave interferometer”, Tesi di perfezionamento, Scuola Normale Superiore di Pisa

[2] G. Vajente, ”Band-limited RMS”, aLIGO LHO Logbook - posted 14:55, Monday 22 April 2019 (48668), https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=48668, https://git.ligo.org/gabriele-vajente/blrms

[3] Panchov et al. , ”Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817”, Phys. Rev. D 98(8), p.2 (2018)

[4] G.Vajente, ”Study of dark-fringe noises slow non-stationarities during VSR1”, Virgo Internal note, VIR-009A-08 (2008)

[5] M. Walker et al., ”Identifying correlations between LIGO’s astronomical range and auxiliary sensors using lasso regression”, Classical and Quantum Gravity 35(22),0264-9381 (2018)