University of Massachusetts Dartmouth

Department of Mathematics

An Exploration of Statistical Methods in Determining the Gaussianity of LIGO Detector Data

A project in
Data Science
by
Ashish Thomas Mathew

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

May, 2025

Abstract

The Advanced Laser Interferometer Gravitational-wave Observatory (LIGO) is one of the most sophisticated instruments ever built, capable of measuring motion 10,000 times smaller than a proton, and was the first device to detect a Gravitational Wave (GW) event. Due to its high sensitivity, a common issue is its susceptibility to glitches: transient, non-Gaussian noise bursts occurring at a high enough rate to contaminate the time series data obtained from the detector. The current methods to detect and study glitches, though highly effective for analysing the strain data during the detection run, prove to be slow when working with a large throughput of detector data. Here we study a few possible statistical approaches to detecting these glitches in the time-amplitude domain. By sampling the strain data, conditioning it, and employing tests of normality, mainly the Shapiro-Wilks test, the Kolmogorov-Smirnov test, and the Anderson-Darling test, we study their effectiveness in determining the Gaussianity of a sample time-series strain data from the detector. The alternate hypothesis of these tests, i.e., the sample belonging to a non-Gaussian distribution, indicates the presence of a glitch or a signal. The results of this experiment explore in detail the viability of these tests as efficient alternatives to the current solutions used in detecting a glitch or astrophysical event in a timeseries signal.

1 Introduction

Nearly a century after Einstein's prediction of the existence of Gravitational Waves (GWs) in 1916, the first direct gravitational wave detection was achieved by the Advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) and Virgo collaboration during the binary black hole merger event GW150914 on September 14, 2015. The LIGO Livingston (L1) interferometer underwent several upgrades following this run - effectively a massive redesign - enhancing its sensitivity by 15% to 25% [1]. This improvement was evident during the O2 observing run during which L1, in conjunction with its Hanford (H1) counterpart and Virgo (V1), detected eleven new gravitational wave signals [2]. Subsequently, during the O3 run, all three detectors operated at their best possible sensitivity, leading to the first single-detector GW detection, GW190425, achieved by LIGO Livingston [3]. There have been over 90 GW events recorded at a high level of confidence since the LIGO-Virgo collaboration's inception to date.

A consequence of these detectors' sensitivities is their susceptibility to noise. During observation runs, various noise sources, such as seismic noise, suspension thermal noise, and sensing noise, affect the data collected by the interferometers. All these sources collectively produce time series signals which can be treated as stochastic processes with their corresponding joint probability distributions and statistical properties [4]. In the absence of astronomical signals or glitches, the probability distribution of the noise follows a normal distribution, also known as *Gaussian* noise. In the event of a Gravitational wave or any other signal, the noise exhibits a high signal-to-noise ratio (SNR), making it *non-Gaussian*.

Among these noise sources, the most problematic are **glitches**: transient events caused by non-astrophysical phenomena such as anthropogenic noise, weather conditions, or instrument malfunctions [4, 5]. Glitches manifest as localized bursts of excess power in interferometer time series data and often do not have well-defined sources. They can occur at energy levels and frequencies that overlap with GW signals, thereby mimicing them and increasing the number of false positive detections. During the first half of the third observing run (O3a) LIGO Hanford and Livingston recorded glitches rates of 0.29 min^{-1} and 1.10 min^{-1} respectively, which rose to 0.32 min^{-1} and

1.17 min^{-1} during the second half (O3b) [6]. Of the 90 candidate events detected in this run, 17 had a probability of astronomical origin (p_{astro}) below 0.5, suggesting significant ambiguity about their origin. A notable example of glitches posing such an issue was during the binary neutron star merger event GW170817 [7], during which instrumental noise transients were detected before the event's coalescence time, complicating its detection and subsequent analysis.

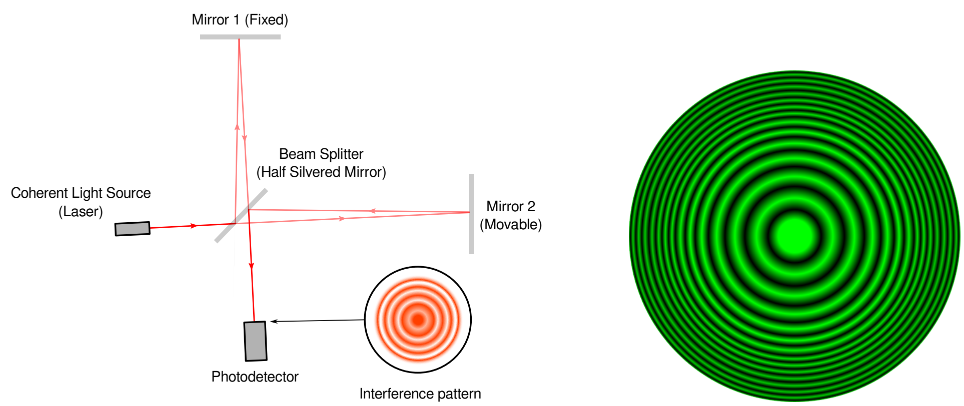
Detecting and mitigating the effects of glitches in interferometric strain data remains an active area of research within astrophysical data analysis, with several techniques proposed and implemented for the same [8–10]. However, many of these are computationally expensive and lack the efficiency required for near-real-time assessment of the signals for glitch activity. The more popular solutions incorporate the *q-transform* [11, 12], a modification of the standard Fourier transform where the analysis window scales inversely with frequency. This q-transform computes a p-value depicting the statistical significance of excess power in the data. Assuming the background data to be Gaussian, the excess power that does not follow a Gaussian distribution suggests the presence of a glitch. While this method is effective for post-detection run analysis, it is computationally intensive for real-time uses. Performing q-transforms on 10-second segments of interferometer data on a laptop requires between 20 to 45 seconds per segment. When scaled up to multiple samples, the computational time would increase significantly.

This project explores statistical hypothesis testing on time-amplitude domain data as a faster alternative to the current glitch detection methods working in the frequency domain. Building on the idea presented in [4], which treats detector noise as stochastic processes with joint probability distributions, it would be reasonable to assume that parametric and non-parametric statistical tests could be effectively applied to such data. This study focuses on applying (1) the Shapiro-Wilk test, (2) the Kolmogorov-Smirnov test, and (3) the Anderson-Darling test on preprocessed samples of clean and glitched data from the LIGO Livingston interferometer. The objective is to determine the normality of these samples and assess how well these tests differentiate between Gaussian (clean) and non-Gaussian (glitched) segments. This experiment also includes studying the distributions, waveforms, and frequency ranges, with particular emphasis on failure points for each statistical test across various glitch types. Furthermore, considering the frequency bands at which each glitch type occurs, this study also examines how each statistical test performs under band-pass filtering at various frequency ranges.

This report is organized as follows: Section 2 discusses how interferometer data is obtained and describes the properties of the time series data used in this study. It also outlines the process of acquiring clean and glitched samples and the preprocessing applied. Section 3 describes in detail the statistical tests of Gaussianity used to assess the sample data. The experimental procedures and results are presented in Sections 4 and 5. Finally, Section 6 summarizes the findings of our experiments, discussing the strengths and limitations of the methods, followed by the future scope in Section 7.

2 Data Acquisition and Conditioning

The Advanced LIGO and Virgo interferometers are large-scale, heavily modified versions of the Michelson interferometer (Figure 1a), originally invented by American physicist Albert A. Michelson in 1887 [6, 13]. LIGO operates in a vacuum, using a laser beam of light split into two orthogonal parts with the help of a half-silvered mirror mounted on horizontal seismometer suspensions. The orthogonally split beams are then sent through two arms of the interferometer, known as Fabry–Pérot cavities. Each of these arms is 4 kilometers long, consisting of mirrors at each end. The first of the mirrors is fixed in place, while the second one is movable with the help of precise micrometer drives. The laser beams, when reflected by the mirrors on either end of the arms, are combined again at the beam splitter and sent to photoelectric detectors. The aLIGO interferometer is designed such that if the beams travel equal distances, the recombination would lead to a destructive interference pattern, resulting in no light coming out of the instrument [14]. When a disturbance of either astrophysical or terrestrial origin is detected, an infinitesimally small change occurs in lengths of the detector arms, in the order of $10^{-8}m$ [15]. One of the arms is stretched, and the other arm is compressed in the perpendicular direction, altering the lengths of the reflected and combined beams, misaligning them and resulting in an interference pattern (Figure 1b). This interference pattern, coupled with results from several other sensors, provides detailed information on the *strain* of the GW or glitch event being observed.



(a) Construction of an Interferometer (b) Interference pattern

Figure 1: Illustrations of a basic Michelson interferometer [16, 17] and what an interference pattern would look like.

This project uses the strain data from LIGO Livingston (L1) during the first part of the third observing run (O3a). This data is in the time domain, as timestamps in the GPS time system at nanosecond precision, and records the amplitude of the noise event as a differential change in lengths of the interferometer arms. Figure 2 shows the steps taken to acquire signal data from the detector and condition it for statistical testing. We utilise two separate modules to load the glitched time series segments and the clean time series segments.

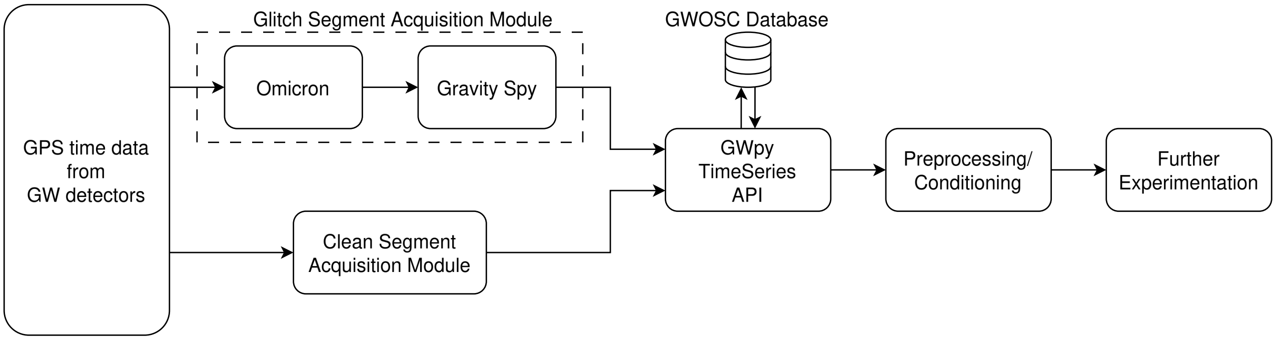


Figure 2: The Data Acquisition and Conditioning Pipeline

2.1 Glitch Samples

The GPS times of the glitch occurrences are obtained using *Gravity Spy* [18]. Gravity Spy is a large-scale citizen-science project combining astrophysics, machine learning, and human efforts to classify glitches in GW interferometer data. Based on the morphological characteristics, a total of 22 classes of glitches were identified at an SNR above 7.5 with the peak frequencies between 10Hz and 2048Hz.

The *Omicron* transient search algorithm was used under Gravity Spy's hood to generate q-transform spectrograms and calculate SNRs of the timeseries samples [8], and was crucial in determining the most useful samples for data classification and analysis.

The O3a dataset we are working with contains the following

Glitch Class	Count	Glitch Class	Count
Fast_Scattering	21749	Whistle	896
Tomte	18708	Low_Frequency_Lines	788
Blip_Low_Frequency	7549	Scratchy	207
Scattered_Light	5398	Repeating_Blips	164
No_Glitch	5358	Violin_Mode	164
Extremely_Loud	4319	Paired_Doves	155
Koi_Fish	4268	Light_Modulation	72
1400Ripples	2363	Helix	21
Blip	1947	Wandering_Line	20
Power_Line	1189	1080Lines	9
Low_Frequency_Burst	1187	Chirp	6

Table 1: The number of glitches

2.2 Clean Samples

This work uses the `GWpy` Python package to access the gravitational wave detector strain data. These strain readings can be queried as `TimeSeries` objects for further study [19].

TimeSeries objects are built upon `numpy` arrays, allowing most of the core numpy functions and utilities to be used with them. GWpy, additionally, provides functions for signal processing, tabular data filtering, and visualization, which this project uses to a great extent.

2.3 Data Conditioning

2.3.1 Whitening

2.3.2 Filtering

3 Methods

3.1 Motivation based on problem statement

3.2 Shapiro-Wilks Test

3.3 KS

3.4 AD

4 Experimentation with the complete frequency range

5 Experimentation with a band pass filter applied

6 Conclusions

7 Future Scope

8 Appendix

References

[1] Andrew Grant. "Advanced LIGO ramps up, with slight improvements". en. In: *Physics Today* 2016.11 (Nov. 2016), p. 12128. ISSN: 19450699. DOI: 10.1063/pt.5.9074. URL: https://pubs.aip.org/physicstoday/online/12128 (visited on 04/05/2025).

[2] The LIGO Scientific Collaboration, the Virgo Collaboration, et al. "GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs". In: *Physical Review X* 9.3 (Sept. 2019). arXiv:1811.12907 [astro-ph], p. 031040. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.9.031040. URL: http://arxiv.org/abs/1811.12907 (visited on 04/05/2025).

[3] R. Abbott et al. *GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run*. arXiv:2010.14527 [gr-qc]. Mar. 2021. DOI: 10.1103/PhysRevX.11.021053. URL: http://arxiv.org/abs/2010.14527.

[4] The LIGO Scientific Collaboration, the Virgo Collaboration, et al. "A guide to LIGO-Virgo detector noise and extraction of transient gravitational-wave signals". In: *Classical and Quantum Gravity* 37.5 (Mar. 2020). arXiv:1908.11170 [gr-qc], p. 055002. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/ab685e. URL: http://arxiv.org/abs/1908.11170 (visited on 04/12/2025).

[5] The LIGO Scientific Collaboration and the Virgo Collaboration. "Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914". In: *Classical and Quantum Gravity* 33.13 (July 2016). arXiv:1602.03844 [gr-qc], p. 134001. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/0264-9381/33/13/134001. URL: http://arxiv.org/abs/1602.03844 (visited on 04/11/2025).

[6] The LIGO Scientific Collaboration, the Virgo Collaboration, et al. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run*. arXiv:2111.03606 [gr-qc]. Oct. 2023. DOI: 10.1103/PhysRevX.13.041039. URL: http://arxiv.org/abs/2111.03606.

[7] The LIGO Scientific Collaboration and The Virgo Collaboration. "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral". In: *Physical Review Letters* 119.16 (Oct. 2017). arXiv:1710.05832 [gr-qc], p. 161101. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.119.161101. URL: http://arxiv.org/abs/1710.05832 (visited on 04/17/2025).

[8] Florent Robinet et al. "Omicron: A tool to characterize transient noise in gravitational-wave detectors". In: *SoftwareX* 12 (2020), p. 100620. ISSN: 2352-7110. DOI: https://doi.org/10.1016/j.softx.2020.100620. URL: https://www.sciencedirect.com/science/article/pii/S2352711020303332.

[9] Duncan M. Macleod et al. "GWpy: A Python package for gravitational-wave astrophysics". In: *SoftwareX* 13 (2021), p. 100657. ISSN: 2352-7110. DOI: https://doi.org/10.1016/j.softx.2021.100657. URL: https://www.sciencedirect.com/science/article/pii/S2352711021000029.

[10] D. Davis et al. "Subtracting glitches from gravitational-wave detector data during the third observing run". In: *Classical and Quantum Gravity* 39.24 (Dec. 2022). arXiv:2207.03429 [astro-ph], p. 245013. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/aca238. URL: http://arxiv.org/abs/2207.03429.

[11] S. Chatterji et al. "Multiresolution techniques for the detection of gravitational-wave bursts". In: *Classical and Quantum Gravity* 21.20 (Oct. 2004). arXiv:gr-qc/0412119, S1809–S1818. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/0264-9381/21/20/024. URL: http://arxiv.org/abs/gr-qc/0412119 (visited on 04/15/2025).

[12] Leah Vazsonyi and Derek Davis. "Identifying glitches near gravitational-wave signals from compact binary coalescences using the Q-transform". In: *Classical and Quantum Gravity* 40.3 (Feb. 2023). arXiv:2208.12338 [astro-ph], p. 035008. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/acafd2. URL: http://arxiv.org/abs/2208.12338 (visited on 04/14/2025).

[13] R. Weiss et al. *Quarterly Progress Report No. 105*. Apr. 1972. URL: https://dspace.mit.edu/handle/1721.1/56271.

[14] The LIGO Scientific Collaboration. *What is an Interferometer?* URL: https://www.ligo.caltech.edu/page/what-is-interferometer (visited on 04/07/2025).

[15] Sudarshan Ghonge et al. *Assessing and Mitigating the Impact of Glitches on Gravitational-Wave Parameter Estimation: a Model Agnostic Approach*. arXiv:2311.09159 [gr-qc]. Oct. 2024. DOI: 10.48550/arXiv.2311.09159. URL: http://arxiv.org/abs/2311.09159.

[16] Stannered. *Simple Michelson interferometer diagram*. 2007. URL: https://commons.wikimedia.org/wiki/File:Interferometer.svg.

[17] WiredSense. *The Michelson Interferometer - A Laser Lab Alignment Guide*. WiredSense Tutorials. URL: https://www.wiredsense.com/tutorial/the-michelson-interferometer-a-laser-lab-alignment-guide.

[18] M Zevin et al. "Gravity Spy: integrating advanced LIGO detector characterization, machine learning, and citizen science". In: *Classical and Quantum Gravity* 34.6 (Feb. 2017), p. 064003. ISSN: 1361-6382. DOI: 10.1088/1361-6382/aa5cea. URL: http://dx.doi.org/10.1088/1361-6382/aa5cea.

[19] D. M. Macleod et al. "GWpy: A Python package for gravitational-wave astrophysics". In: *SoftwareX* 13 (2021), p. 100657. ISSN: 2352-7110. DOI: 10.1016/j.softx.2021.100657. URL: https://www.sciencedirect.com/science/article/pii/S2352711021000029.