Investigating data quality metrics for stochastic gravitational-wave detection

Caltech LIGO SURF 2021

Makenzi Fischbach
Wellesley College

Mentors: Derek Davis & Arianna Renzini
California Institute of Technology

July 2021

Investigating data quality metrics for stochastic gravitational-wave detection

Makenzi Fischbach[1]
Wellesley College[1]

Mentors: Derek Davis[2] and Arianna Renzini[2]
California Institute of Technology[2]
(Dated: July 30, 2021)

Abstract. The detection of gravitational waves has created the opportunity for many new discoveries. One such potential discovery is the stochastic gravitational wave background. In order to detect it, stochastic data must be properly monitored and analysed. Stochmon, a low latency stochastic data monitoring pipeline, works to monitor the quality of stochastic data. Stochmon has not been recently updated and is not well integrated with current gravitational wave data analysis tools. The goal of this project is to identify potential improvements to make to Stochmon's analysis functions, implement said changes, and integrate the system with existing analysis tools so that it can be used during the next observing run. A new feature of Stochmon, the stochastic detector sensitivity (SDS) has been implemented which calculates the energy density at which a detector can detect a stochastic signal.

I. INTRODUCTION

Ever since their initial detection in 2015, gravitational waves (GWs) have been at the forefront of scientific research. GWs are notably ripples caused by disruptions to the fabric of space-time typically traced back to high-energy events, such as binary black hole mergers, compact binary coalescence (CBC), and bursts. GWs have the potential to provide unprecedented insight into astrophysical phenomena and the primordial universe [1].

The Laser Interferometer Gravitational-Wave Observatory (LIGO) has the ability to directly detect the GWs permeating from high-energy events and has been doing so since the first successful GW detection on September 14th 2015 [2]. LIGO is a large interferometer consisting of two, four kilometer arms oriented in an L-shape. A laser beam is split using a beam splitter and the two resulting beams are sent down the arms of the detector. If the light beams go undisturbed by GWs, the light from both arms will arrive back at the detector at the same time and cancel each other out, resulting in no GW detection. If a GW is present, it will create a slight change in distance and the two beams will return to the detector at different times. In this instance, the two beams of light will have varying phases and will not cancel, providing evidence of the presence of a GW. There are LIGO detectors in Livingston, Louisiana and Hanford, Washington.

While the sources of GWs are isolated astrophysical events, currently detected GWs can be detected from the stochastic gravitational-wave background (SGWB) [3]. The SGWB is a stochastic signal composed of the weak GW signals from a large number of unidentified events [4]. For instance, the superposition of GW signals from a population of binary black holes would appear as a stochastic signal. The SGWB can also be credited to stochastic processes that occurred in the primordial stages of the universe. We expect a successful detection of the SGWB to occur in the near future.

Stochmon is a data-quality monitor which specializes in the analysis of LIGO and Virgo low-latency stochastic data [5]. The monitor has a variety of tools that provide us with useful data, such as estimates for the sensitivity at which stochastic data is being collected and analyzed as well as coherence estimates for the two LIGO locations and the noise stationarity of the detectors.

A. Problem & Objectives

The improvement of Stochmon will lead to a direct improvement in the analysis of stochastic data, the quality of data, and the overall ability of LIGO to detect the SGWB. With the current instrumentation, the detection of the SGWB, and GWs in general, is imperfect. This is especially evident in frequency bands where the data is corrupted by noise.

Improvement in stochastic data analysis could lead to a deeper understanding of the primordial universe and the stochastic events which may have occurred around the time of the Big Bang [6]. Additionally, stochastic data analysis can provide the ability to achieve a deeper understanding of what the universe is composed of and allow for a method of detection free of scientific models.

As of now, Stochmon exists and is operational. However, it has not been integrated into many of the analysis tools used by LIGO and is not actively being improved or monitored. We aim to improve the Stochmon system and its ability to investigate the performance of LIGO's detectors in detecting the SGWB. Improvement must be identified through an analysis of the current efficacy of the Stochmon system. Prior to beginning the project, the assumption is that all elements of Stochmon can be updated in some way to achieve a higher quality of stochastic data analysis. Another objective is to ensure that Stochmon is well integrated with other existing online data monitoring tools. These updates must then also be integrated so that they can be utilized during the next LIGO detection run.

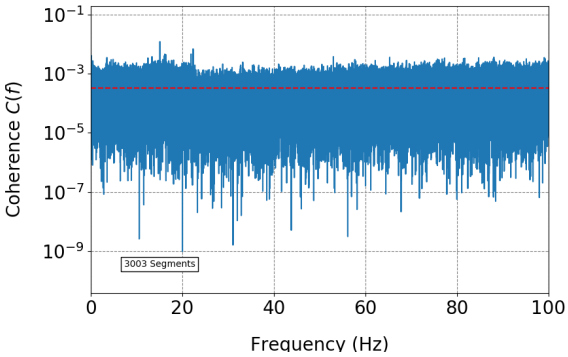

FIG. 1. The coherence between Livingston and Hanford with 1 mHz frequency resolution. The dashed red line signifies the expected level of coherence. This plot shows the coherence between the detectors is strongest from 0 Hz to about 22 Hz. Figure reproduced from the Stochmon summary page [7].

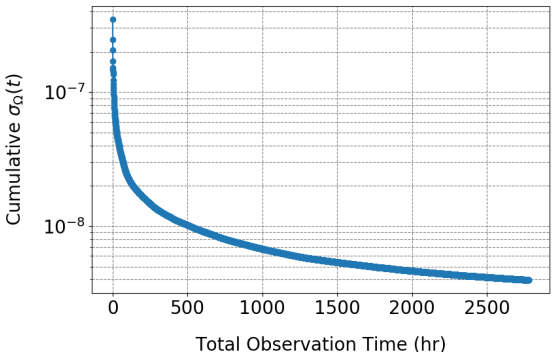

FIG. 2. Energy sensitivity vs. observation time. The search sensitivity decreases as observation time increases. The cumulative sensitivity is at its highest at the start of the observing period. The variance on Omega decreases as a function of time as $1/\sqrt{t}$, implying that the sensitivity to Omega increases through integrating over the whole observation time. Figure reproduced from the Stochmon summary page [7].

II. APPROACH

The approach for this project is entirely contingent on what elements of Stochmon ultimately need improving and how those updates can be accomplished. The first step is to identify the components of Stochmon which are the most beneficial and need the most revision. After the initial identification, the next step would be to identify the ways in which the component could be improved and how those improvements could be implemented. The length of each step in the process is dependent on what approaches are taken and how intensive those approaches may be. One area that has already been identified for improvement is the systematic integration of the output of Stochmon with the LIGO detector Summary Pages. While working on improvements, we will be working closely with the original developers and maintainers of Stochmon. They will provide us with guidance and support throughout the process.

Stochmon consists of many tools which aid in the stochastic data analysis. Stochmon provides a detailed analysis of cross-correlated data between the Hanford and Livingston LIGO locations. This H1-L1 coherence, shown in Figure 1, is determined by dividing the cross power of the two detectors by the product of the auto powers [5]:

$$coh(f) = \frac{|\overline{S_{12}(f)}|^2}{\overline{S_1(f)}\ \overline{S_2(f)}}. \tag{1}$$

Knowing the coherence aids in the cross-analysis of data and therefore in the process of separating the stochastic data from any disruptive external artifacts or noise from instrumentation.

Stochmon also provides an analysis of the cross amplitude density plots for both detectors [5]:

$$a(f) = |\widetilde{s}_I \widetilde{s}_I(f)|^{1/2}. \tag{2}$$

One of Stochmon's main features is the analysis of detector sensitivity to stochastic signals (Figure 2). The strain sensitivity (σ_h) is the sensitivity of what is measured with the detector. The energy sensitivity (σ_Ω), which is the cosmological quantity used in publications, is determined and is then compared to the aforementioned strain sensitivity [5]:

$$\sigma_\Omega(f) = \frac{10\pi^2}{3H_{100}^2}\frac{f^3}{\gamma(f)}\sigma_h(f)^2. \tag{3}$$

An analysis can be performed by taking a weighted average of both the sensitivity of time and frequency [5]:

$$\sigma = \sum_{t=1}^{n}\sum_{f=1}^{m}\left(\sigma(f,t)^{-2}\right)^{-1/2}. \tag{4}$$

The analysis of sensitivity provides a deeper understanding of the detectors' strengths and weaknesses, as well as how they can be improved.

III. STOCHASTIC DATA AND STOCHMON

An analysis of the detectors' sensitivities is one of Stochmon's central features which aids in the data analysis process. In order to best implement constructive change to the way in which the sensitivity is monitored, a deeper understanding of the sensitivity of the detectors as a whole had to be developed.

Polarization is the orientation in which a wave, such as a GW, oscillates. Each detector is most sensitive to different locations in the sky and has different polarization responses dependent on their location and orientation on

Earth. To best visualize these sensitivities, we can determine the detector polarization response functions of each detector in both the cross and plus polarization using built in Bilby functions and plot them using healpy (Figure 3) [8].

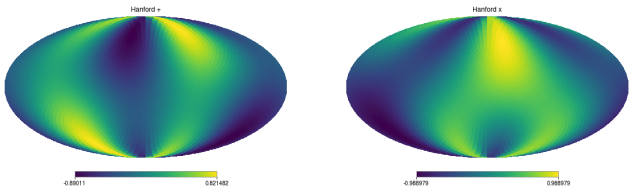

FIG. 3. The H1 cross (right) and plus (left) polarization response functions. The dark blue and yellow represent locations in the sky in which the detector is most sensitive. The cross and plus polarizations allow for a wider range of high sensitivity.

After finding the detectors' individual polarization response functions, an overlap function for the H1 and L1 detector pair as a function of time can be determined and plotted (Figure 4). When the overlap function is plotted as a function of time, the areas of high sensitivity appear to rotate around the map once per sidereal day.

Next, the initial overlap function is multiplied with the plane wave term, where Δt is the delay term between both detectors, to get a visualisation (Figure 5) for the full stochastic sky response dependent on frequency:

$$\gamma(f, n; t) = overlap(n; t) * e^{i2\pi f \Delta t}. \quad (5)$$

The overlap functions can also be visualized in three dimensions, where the radius of the plot correlates to the value of the overlap function at a given point in the sky.

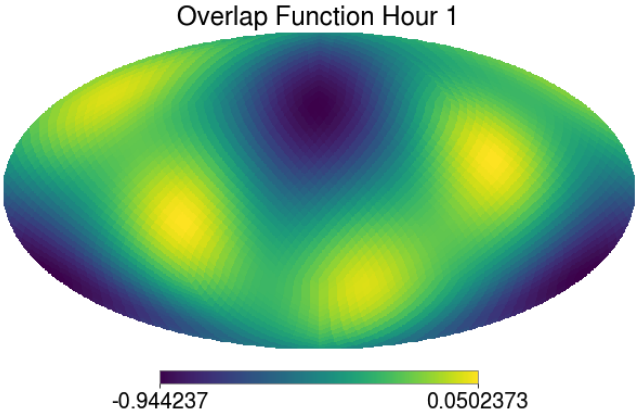

FIG. 4. The overlap function of H1 and L1 as a function of time at hour 1 of a sidereal day. The dark blue and yellow represent locations in the sky in which the detectors are most sensitive. As the sidereal day continues, the areas of high sensitivity rotate along with the detectors relative to a fixed point in the sky.

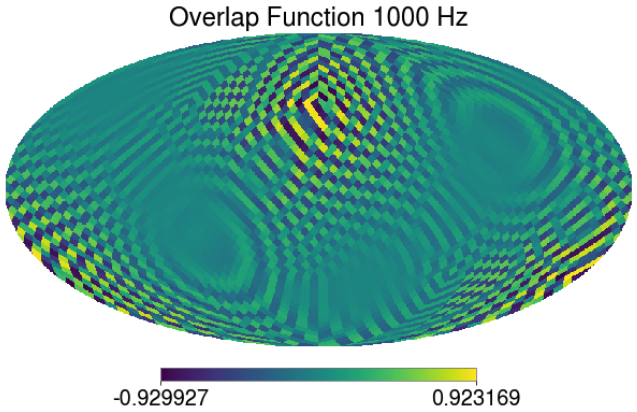

FIG. 5. The overlap function of H1 an L1 as a function of frequency at 1000 Hz. The dark blue and yellow represent locations in the sky in which the detectors are most sensitive. As the frequency increases, the number of waves that fit between the two detectors increases.

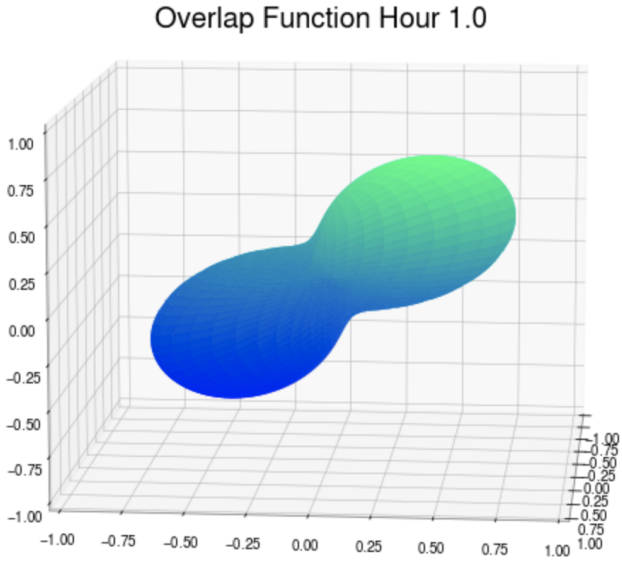

FIG. 6. 3D overlap function dependent on time at hour 1 of a sidereal day. In this plot, the coloring aids in the 3D visualization and has no further significance. The axes represent the value of the overlap function at a given location.

Figure 6 shows the 3D overlap function as a function of time and Figure 7 is the three-dimensional representation as a function of frequency.

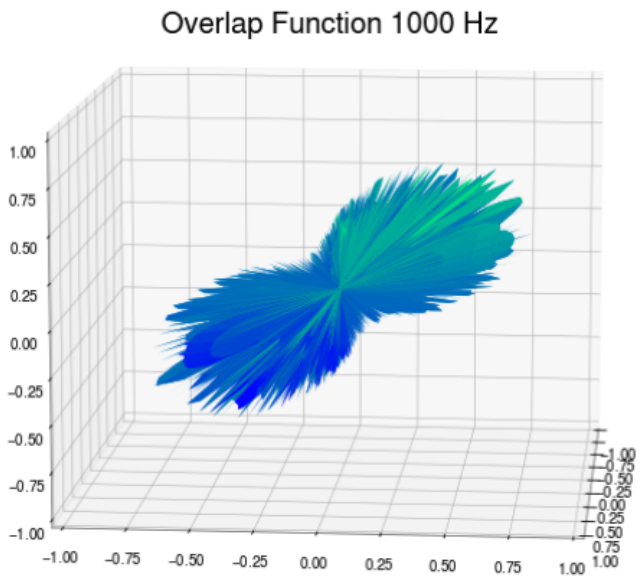

FIG. 7. 3D overlap function dependent on frequency at 1000 Hz. In this plot, the coloring aids in the 3D visualization and has no further significance. The axes represent the value of the overlap function at a given location.

Overlap functions are a helpful tool for both potential stochastic gravitational wave detection and stochastic data analysis since they allow us to identify the locations in which the detectors are most sensitive to detecting a stochastic signal.

IV. DETCHAR AND DATA QUALITY ANALYSIS

To implement stronger tools for stochastic data quality (DQ) analysis, it is beneficial to turn to general DQ analysis for guidance. At LIGO, one of the ways DQ is evaluated is through DQ shifts. DQ shifts are standard procedure where the designated 'shifters' for the week at both H1 and L1 are tasked with reviewing a week's worth of data and plots and writing a summary of any significant or notable changes. These can be either positive or negative changes.

While doing DQ shifts, shifters refer to the detector characterization (Detchar) summary pages, which show an overview of the data content for each day of an observing run [9]. On these pages, there are four main plots that shifters analyse. These plots are meant to detect and highlight any noise, glitches, or events in the data that may affect analysis done with said data.

The first plot is the spectrogram, which is a plot of time versus frequency for a day of observing. Glitches and noise in the data present themselves as red or blue lines. Red lines represent high noise relative to the median while blue lines represent a decrease from the standard noise. The binary neutron star (BNS) inspiral range plot shows the sensitivity of the detectors by providing

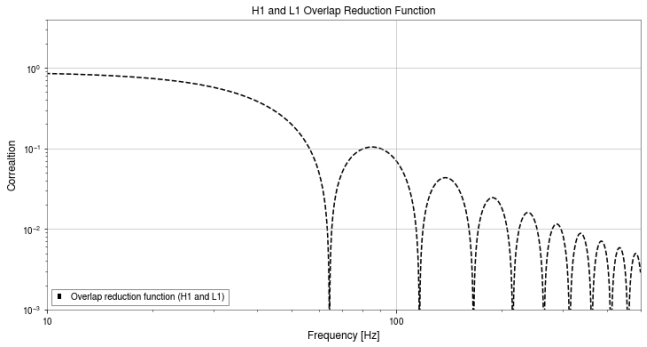

FIG. 8. Overlap reduction function for the H1 and L1 detector pair. The correlation drops to zero at several frequencies. Correlation steadily decreases as frequency increases.

the distance at which CBC events can be detected with two 1.4 solar mass objects at a signal to noise ratio (SNR) of 8. Shifters look for significant increases or decreases in the range which may be indicative of a change in the detector's functionality. 'Glitchgrams' and glitch rate plots both show potential glitches in the data. Shifters look for loud glitches that may be in clusters. These clusters of noise could have negative impacts on the data and impact its usability. The detchar summary pages also have Hveto and LASSO, tools that identify the potential channels which may be responsible for the noise and glitches in the data [10, 11].

V. DETECTOR CORRELATION

Measuring the correlation between a pair of detectors is necessary for a stochastic search. Stochastic data should be the same for each detector, up to the response of the individual detectors. Therefore, stochastic data is included in correlated data since all other noise and uncorrelated data is excluded.

Overlap reduction functions (ORFs), as shown in Figure 8, provide values for the frequency dependent correlation between data from a pair of detectors [4]. The correlation is dependent on the response functions of both detectors.

When calculating the correlation between a pair of detectors, the data being measured is divided into segments. The correlation of each individual segment is calculated and the segments are then averaged together. As the number of segments (N) increases, the measured correlation gets closer to the true correlation value (Figure 9). This is because as N increases, the correlated data remains constant while the uncorrelated data will decrease and approach zero at a rate proportional to $\frac{1}{\sqrt{N}}$. This provides us with the ability to detect very weak signals.

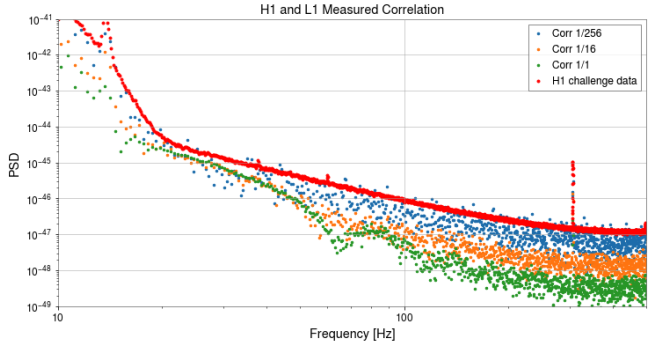

FIG. 9. As the number of segments increases, the measured correlation approaches the true correlation value.

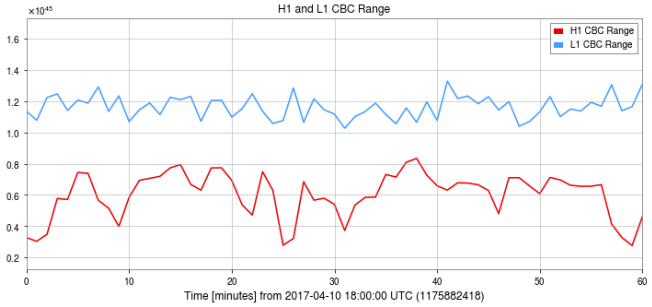

FIG. 10. CBC range for H1 and L1 using an hour of data. The plot shows the distance at which the detectors can detect a CBC event with two 1.4 solar mass objects at an SNR of 8. The y-axis is negligible due to the purposeful exclusion of the necessary constant. With the inclusion of the constant, the y-axis would represent distance in *Mpc*.

VI. STOCHASTIC DETECTOR SENSITIVITY

CBC ranges, such as the BNS inspiral range plot included on the Detchar summary pages, are useful in assessing the sensitivities of the detectors (Figure 10). Based on the CBC range, we implemented a new feature of Stochmon. This new feature is the stochastic detector sensitivity (SDS) (Figure 11). Similar to the CBC range, its primary function is to assess an individual detector's sensitivity for stochastic data. While the CBC range is conveyed as a function of distance (Mpc), the SDS is calculating the energy sensitivity at which stochastic data can be detected.

Due to stochastic data detection's heavy reliance on the use of correlated data, many stochastic analysis tools neglect the benefit of looking at the detectors separately. The SDS is beneficial since it allows for the assessment of the sensitivity of individual detectors as opposed to the sensitivity of an entire network and therefore provides the opportunity to run diagnostics on individual detectors. Since the ORF is the same for a pair of detectors, their individual PSDs operate as the differentiating factor between ranges.

The CBC range, as shown in Equation 6, is proportional to the integral of $f^{\alpha-3}$ over the PSD of a specific detector. The SDS is calculated in a similar fashion to the CBC range (Equation 7):

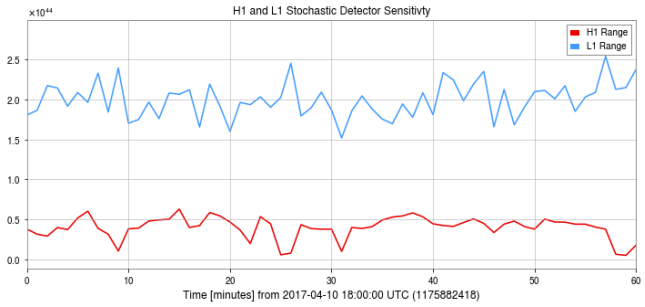

FIG. 11. stochastic detector sensitivity for H1 and L1 for an hour of data. The plot shows the energy sensitivity at which the detectors are most sensitive to stochastic data. The y-axis is negligible due to the the necessary, not yet calculated constant.

$$\propto \int \frac{(f^{\alpha-3})}{(PSD)}\, df. \tag{6}$$

For calculating the CBC range, α is equal to $\frac{2}{3}$, resulting in a numerator of $f^{-\frac{7}{3}}$. The constant typically included in the CBC range calculation has been purposefully excluded.

The main difference between calculating the CBC range and the SDS is the inclusion of the ORF for a pair of detectors in the numerator of the integral:

$$\propto \int \frac{(ORF)(f^{\alpha-3})}{(PSD)} df. \tag{7}$$

In Equation 7, the CBC range α value of $\frac{2}{3}$ is used for calculating the SDS. The value $\frac{2}{3}$ is attributed to the stochastic background's expected spectral shape given by the superposition of numerous inspiraling compact binaries. The constant needed for this calculation has been excluded as it has not yet been computed.

The correlation between the CBC range and SDS has been plotted and shows that there is a fairly strong correlation between the two values (Figure 12). This strong correlation is expected since both the CBC range and SDS are dependent on the sensitivity of the same detector. There are some unexpected outliers that should be further investigated.

VII. NEXT STEPS

One of the next steps for this project is to compute the missing conversion constant to re-normalize the fractional energy density for the SDS calculations. The constant can be found using an equation for energy density (Ω_0) derived using Equations 3 and 4:

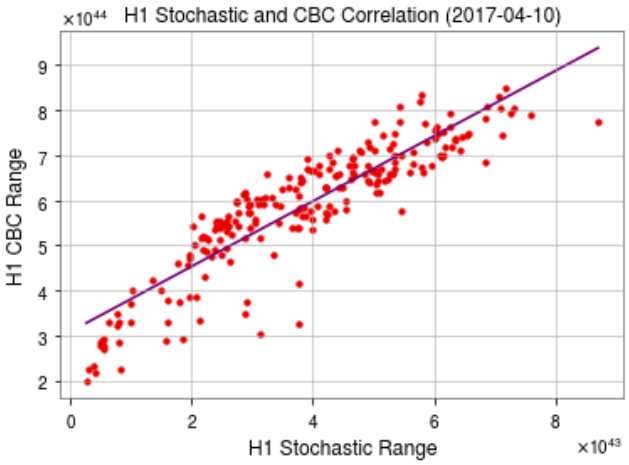

FIG. 12. The correlation between the CBC range and SDS for H1 for one hour of data. The correlation is fairly strong.

$$\Omega_0 = \frac{\rho}{T^{1/2}} f_0^2 \left(\frac{2\pi^2}{3H_0}\right)\left(\frac{1}{2\pi}\right)\left(\int \left(\frac{(ORF)(f^{\alpha-3})}{PSD}\right)^2 df\right)^{-1/2}. \quad (8)$$

Another next step is to turn the SDS Python script, which calculates and plots the SDS for a given section of data, into an executable so an entire year's worth of data can be assessed.

[1] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. Upper limits on the isotropic gravitational-wave background from advanced ligo's and advanced virgo's third observing run, 2021.

[2] J Aasi and et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, Mar 2015.

[3] Michele Maggiore. Gravitational wave experiments and early universe cosmology. *Physics Reports*, 331(6):283–367, Jul 2000.

[4] Joseph D. Romano and Neil. J. Cornish. Detection methods for stochastic gravitational-wave backgrounds: a unified treatment. *Living Reviews in Relativity*, 20(1), Apr 2017.

[5] G.Hernandez. Stochmon: A ligo data analysis tool, 2013.

[6] Tania Regimbau. The astrophysical gravitational wave stochastic background. *Research in Astronomy and Astrophysics*, 11(4):369–390, Mar 2011.

[7] Tom Callister et al. Stochmon summary page.

[8] Gregory Ashton and et al. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241(2):27, Apr 2019.

[9] D. Davis and et al. Ligo detector characterization in the second and third observing runs, 2021.

[10] Joshua R Smith and et al. A hierarchical method for vetoing noise transients in gravitational-wave detectors. *Classical and Quantum Gravity*, 28(23):235005, Nov 2011.

[11] Marissa Walker and et al. Identifying correlations between ligo's astronomical range and auxiliary sensors using lasso regression. *Classical and Quantum Gravity*, 35(22):225002, Oct 2018.