Classical and Quantum Gravity

Noise in the LIGO livingston gravitational wave observatory due to trains

To cite this article: J Glanzer *et al* 2023 *Class. Quantum Grav.* **40** 195015

OPEN ACCESS
IOP Publishing

Classical and Quantum Gravity

Class. Quantum Grav. **40** (2023) 195015 (17pp)

https://doi.org/10.1088/1361-6382/acf01f

Noise in the LIGO livingston gravitational wave observatory due to trains

J Glanzer[1,*], **S Soni**[2], **J Spoon**[3], **A Effler**[4]
and G González[1]

[1] Department of Physics, Louisiana State University, 202 Nicholson Hall, Baton Rouge, LA 70803, United States of America
[2] LIGO Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America
[3] Department of Physics, University of Omaha Nebraska, 6001 Dodge Street Durham Science Center, Omaha, NE 68182, United States of America
[4] LIGO Livingston Observatory, Livingston, LA 70754, United States of America

E-mail: jglanz2@lsu.edu

Received 21 April 2023; revised 21 July 2023
Accepted for publication 14 August 2023
Published 1 September 2023

Abstract

Environmental seismic disturbances limit the sensitivity of LIGO gravitational wave detectors. Trains near the LIGO Livingston detector produce low frequency (0.5–10 Hz) ground noise that couples into the gravitational wave sensitive frequency band (10–100 Hz) through light reflected in mirrors and other surfaces. We investigate the effect of trains during the Advanced LIGO third observing run, and propose a method to search for narrow band seismic frequencies responsible for contributing to increases in scattered light. Through the use of the linear regression tool Lasso (least absolute shrinkage and selection operator) and glitch correlations, we identify the most common seismic frequencies that correlate with increases in detector noise as 0.6–0.8 Hz, 1.7–1.9 Hz, 1.8–2.0 Hz, and 2.3–2.5 Hz in the LIGO Livingston corner station.

Keywords: gravitational waves, detector characterization, scattering

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

---

[*] Author to whom any correspondence should be addressed.

1. Introduction

The detection of gravitational waves from extreme astrophysical events has opened the door to exciting discoveries. The Advanced Laser Interferometer Gravitational-wave Observatory (aLIGO) and advanced VIRGO (AdV) detectors have detected many gravitational waves from coalescing binaries of black holes and neutron stars [1–4]. Since the first detection in 2015, there has been an influx of more events detected. By the end of the first observing run (O1), both the LIGO Scientific Collaboration and the Virgo collaboration had reported three binary black hole gravitational wave events [5]. The second observing run (O2) not only detected seven binary black hole mergers, but the first merger of two neutron stars [6, 7]. In the third observing run (O3), a total of 74 events were admitted into the GWTC-3 catalog [2, 3]. New optics, increased laser input power, and the introduction of squeezed light were among the many changes responsible for improved detector sensitivity in the third observing run [8].

Short duration (0.1–3 s) non-astrophysical transients known as 'glitches' adversely impact the detector's data quality and complicate the process of identification of gravitational waves in the data [9–11]. Environmental and instrumental noise are common sources of these brief transients. Ground motion due to earthquakes, ocean waves, trains, and human activity can couple to the detector components. For aLIGO, quadruple suspension systems and passive/active seismic isolation are used to dampen the effects. The in-vacuum optical tables incorporate an active vibration isolation system providing attenuation of environmental seismic noise below 1 Hz. Above 1 Hz, the quadruple pendulum optic suspensions provide passive isolation in the horizontal and vertical degrees of freedom [12]. Although the ground motion is largest below 10 Hz, it produces transients in gravitational wave strain data in the 10–150 Hz range due to non-linear coupling. At the LIGO Livingston Observatory (LLO), 50% of the noise transients in the LLO detector in O3 were due to light reflected in mirrors and other surfaces ('scattered' or 'stray' light) [13].

In this paper, we investigate scattering noise caused by trains passing near LLO. Noise from trains can couple to the interferometer by causing displacements of particular scattering surfaces in the detector. In section 2, we introduce how scattered light produces noise in the detector. In section 3, we describe the methods used to investigate the effect of trains in producing transient noise in the detector. In section 4, we present the results of our analysis showing the advantages of the methods used, and discuss possible surfaces producing the scattering glitches.

2. Seismic activity and scattered light

The motion of the ground where a gravitational wave detector is located can lead to bad data quality. Earthquakes and wind shake the instrument in 0.03–0.1 Hz band, ocean currents in the Gulf of Mexico are the main source of increased microseismic motion in 0.1–0.3 Hz band and human activities such as logging, construction work and vehicles can cause increased ground motion in 1–6 Hz anthropogenic band. Although all of these frequency bands are not in the sensitive detector gravitational wave band (10–5000 Hz), ground motion can couple non-linearly into that band [8, 14]. For example, during O3 transients due to scattered light were the most frequent source of transient noise at both LLO and LIGO Hanford Observatory (LHO), and were a result of increased ground motion in one or more of these seismic bands [13].

Depending on the surface of the mirror, a small fraction of incident light can scatter off from the surface. This stray light may hit another moving surface in the vicinity and a part of it can rejoin the main beam. This leads to introduction of time dependent phase modulation to the main laser beam. The additional phase shows up as noise h_{ph} in the detector data, shown in equation (1).

$$h_{\mathrm{ph}}(f) = \frac{K}{2}\frac{\lambda}{4\pi L}\mathcal{F}[\sin\delta\phi] \tag{1}$$

where

$$\phi(t) = \phi_0 + \delta\phi_{\mathrm{sc}}(t) = \frac{4\pi}{\lambda}[x_0 + \delta x_{\mathrm{sc}}(t)], \tag{2}$$

here, K is the ratio of stray light amplitude to the amplitude of light in the main beam (usually unknown but very small), λ is laser wavelength (1064 nm) and L is the length of interferometer arms (4 km).

$\mathcal{F}$ indicates a Fourier transform, x_0 is the static path which corresponds to static phase ϕ_0, and δx_{sc} is the time-dependent displacement of the scattering surface which gives rise to the additional phase $\delta\phi_{\mathrm{sc}}$. If the phase modulation $\delta\phi_{\mathrm{sc}}$ is small, $\sin\phi \approx \phi$ and the noise couples linearly; if δx_{sc} is a large fraction of λ, the noise couples non-linearly. In this latter case, as a result of fringe wrapping, the phase noise $h_{\mathrm{ph}}(f)$ associated with scattered light can show up as arches in $h(t)$ spectrogram [15, 16].

If we differentiate the phase $\phi = 2\pi ft$ in equation (2) with respect to time, we obtain the frequency of the noise:

$$f(t) = |\frac{2v_{\mathrm{sc}}(t)}{\lambda}| \tag{3}$$

where v_{sc} is the velocity of the scatterer and f is the peak frequency of the transients.

If the surface excited by ground motion is in approximately periodic motion, the noise will appear as arches in a time-frequency spectrogram shown in the left plot of figure 1. From equation (3) we can see that a scatterer moving with higher velocity will lead to transients at a higher frequency. A scatterer receiving or reflecting too little light or moving with small velocity amplitude will cause transients below the gravitational wave band [14].

Transient scattered noise can be classified into two main categories, depending on the frequency of the ground motion producing it: ‘slow scattering’ and ‘fast scattering’, shown in Figure 1. Slow scattering is usually a result of increased ground motion in earthquake and microseismic band, whereas fast scattering transients are usually a result of high ground motion in the anthropogenic band. In this paper, we focus on fast scattering transients caused by increased anthropogenic motion due to trains [17]. For details on slow scattering and its reduction during O3, we refer to [14].

3. Methods

3.1. Detector seismic couplings

Physical environment monitoring (PEM) is a system of sensors monitoring and recording aspects of the physical environment surrounding the interferometer, allowing studies of the impacts of environmental noise on the detector [18, 19]. One methodology used to find the potential location and intensity of noise coupling in the detector is to inject a known form of disturbance into the detector and study the differential arm length or h(t) response. These

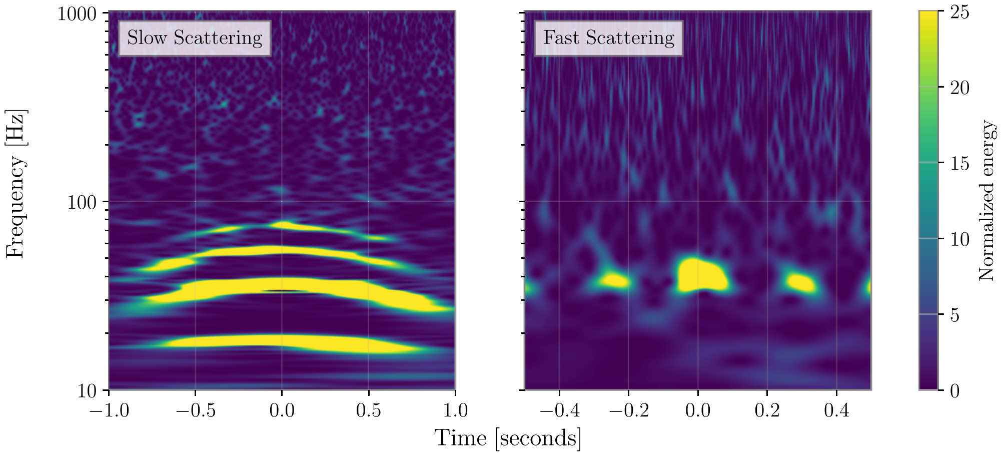


Figure 1. Time-frequency spectrograms of the different types of scatter in the main gravitational wave data channel. Slow scatter has multiple arches, indicating several light reflects. Higher frequency harmonics are not present for fast scatter.

Table 1. Resonant frequencies of various cryobaffles and arm cavity baffles.

Detector component	Resonance frequencies
ITMY Cryobaffle	3.82 Hz, 4.19 Hz [23]
End Y Cryobaffle	3.49 Hz, 4.62 Hz [24, 25]
End X Cryobaffle	4.10 Hz [26]
End Y Arm Cavity Baffle	1.6 Hz [21]
ITMY Arm Cavity Baffle	1.5 Hz [27]

injections, known as PEM injections include magnetic, seismic and acoustic injections. The site of injection, its amplitude, the frequency band, etc are varied to study the impact on h(t) of these changing parameters. Vibrational seismic injections have been regularly performed at LLO and LHO to characterize the coupling between ground motion in anthropogenic band and detector hardware [18, 19]. Vibrational shakers are used to perform these seismic injections and the response is monitored using accelerometers. If a detector component in the vicinity of the injection location has stray light incident on it, and has a resonance that falls within the frequency band of the injection, then its excitation would make noise in h(t).

Baffles are an example of detector hardware designed for the purpose of absorbing and/or redirecting any incident stray light on them. These components are installed at multiple locations in the detector. Depending on the surface, location and geometry of the baffle, they can backscatter stray light towards the test mass mirror and into the main beam [20]. The injections carried out since the end of O3 have revealed the presence of resonances in multiple baffles across the detector. At these resonant frequencies, the baffles can inject scattered light noise back into the gravitational wave readout up to as high as 100 Hz [21]. Table 1 lists the dominant cryobaffle and arm cavity baffle resonances found at input test mass Y arm (ITMY), X and Y end test mass (ETM) at LLO during shaker injection tests in 2020 and 2022. These resonances have been mechanically damped [22].

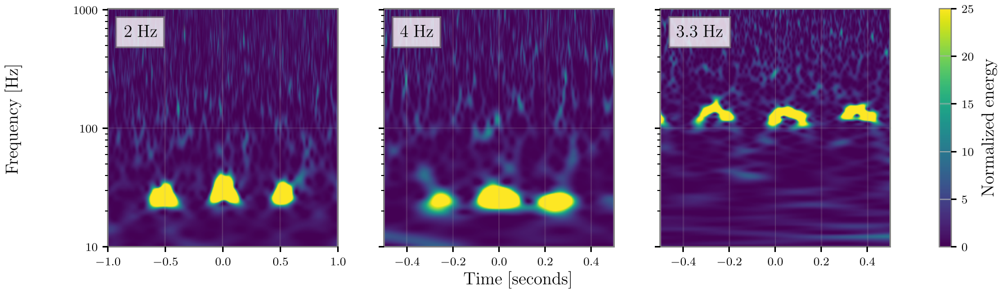


Figure 2. Time-frequency spectrograms of the most common types of fast scatter observed during and post O3. The different types are characterized by the spacing of arches in time [28].

3.2. Identifying fast scattering glitches

During O3, Fast Scattering glitches were the most common glitch type at LLO. Fast scatter shows up as short duration arches in the time-frequency spectrograms of h(t).

About 27% of all glitches classified by the image classifier Gravity Spy [29] with a confidence of 90% were Fast Scattering at LLO. Fast Scattering glitches have been observed to occur when there is an increase in ground motion, specifically, when there is an increase in the microseismic frequency range (0.1–0.3 Hz) and the anthropogenic frequency range (1–6 Hz). Ocean waves and currents in the Gulf of Mexico increase microseismic activity, whereas construction work and logging, thunderstorms, wind, and trains near the detector site leads to an increase in anthropogenic seismic activity at LLO [14, 30].

Several different types of Fast Scattering have been observed, characterized by the frequency of their repeating arches. The most common types seen during O3 are referred to as 4 Hz followed by 2 Hz fast scatter. In the data taken between O3 and O4, 3.3 Hz fast scatter was observed, with a higher peak frequency than in O3. Figure 2 shows the different types observed during O3.

3.3. Identifying trains and characterizing the ground motion they produce

There are several software tools used to detect, characterize and classify noise transients. These tools include but are not limited to Omicron, GWpy Omega Scans, and Gravity Spy [29, 31–36].

Omicron is used to search for excess power in the gravitational wave data. These transients detected by omicron are colloquially known as Omicron triggers or just triggers. Each of these triggers is assigned parameters such as event time, frequency, duration, signal to noise ratio (SNR) etc. Omicron triggers are further used downstream by a number of other detector characterization tools including Gravity Spy.

Gravity Spy is an image classifier based on convolutional neural networks, used to classify transient noise into different categories or classes depending on the time-frequency morphology of the transients in the data; see figure 3 for examples of some common glitch Gravity Spy glitch classifications [29, 37].

One of the major sources of increased noise in the anthropogenic frequency range is trains passing by the LLO Y end station, see figure 4. We use data from three seismometers (Guralp® CMG-3 T [18]) at LLO, one located at each end station and one in the corner station. Each seismometer measures ground motion in the *X*, *Y*, and *Z* directions.

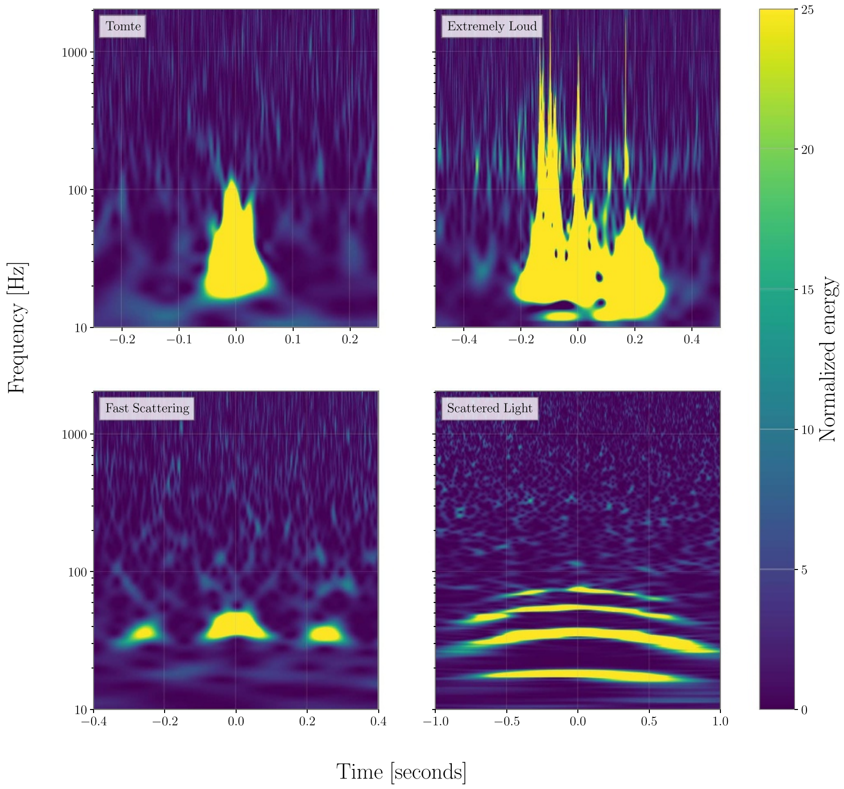


Figure 3. Examples of different glitch classifications from Gravity Spy. Different glitches are visualized in time-frequency space.

We developed a PYTHON tool in an effort to define, systematically, a cut on the root mean square ground motion above which we will consider a train to be affecting the interferometer. The tool finds the peaks and widths of trains based on a specified ground motion threshold. For our analysis, ground motion larger than a 600 $\mathrm{nm\,s^{-1}}$ threshold in ETMY_Y 1–3 Hz for more than 2.5 min were considered trains; see figure 5. This is typically caused by trains, but if there is any other source producing large motion in the same band, it will have a similar effect to noise produced by trains. This identified a total of 791 trains in O3.

Earthquakes and other seismic activity may occur during the trains. In order to investigate the effect of solely due to seismic noise due to trains, we also set cuts on earthquake and microseismic activity. For earthquakes, the threshold was ground motion greater than 100 $\mathrm{nm\,s^{-1}}$ in ITMY_X, and for high microseismic activity the threshold was ground motion greater than 1500 $\mathrm{nm\,s^{-1}}$ in ITMY_X.

In O3, there were 199 trains that did not occur during either an earthquake, high microseismic activity, or already high anthropogenic noise. During the day time hours near LLO, the anthropogenic noise sometimes increases due to activities such as logging and construction work. The cut for what was considered already high anthropogenic motion was bigger than 200 $\mathrm{nm\,s^{-1}}$ in ETMY_Y. We applied this limit approximately 30 min before each train to catch if there was a rise in the motion levels.

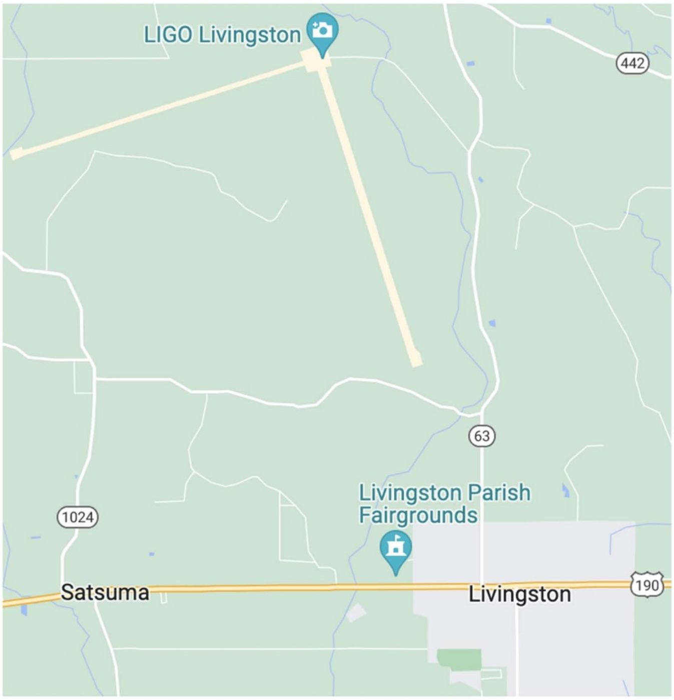


Figure 4. Location of LLO relative to the train track in Livingston, Louisiana. The track runs parallel to highway 190, which passes through Livingston. The end of the Y arm is approximately two miles from the track. *Image: Google Maps*.

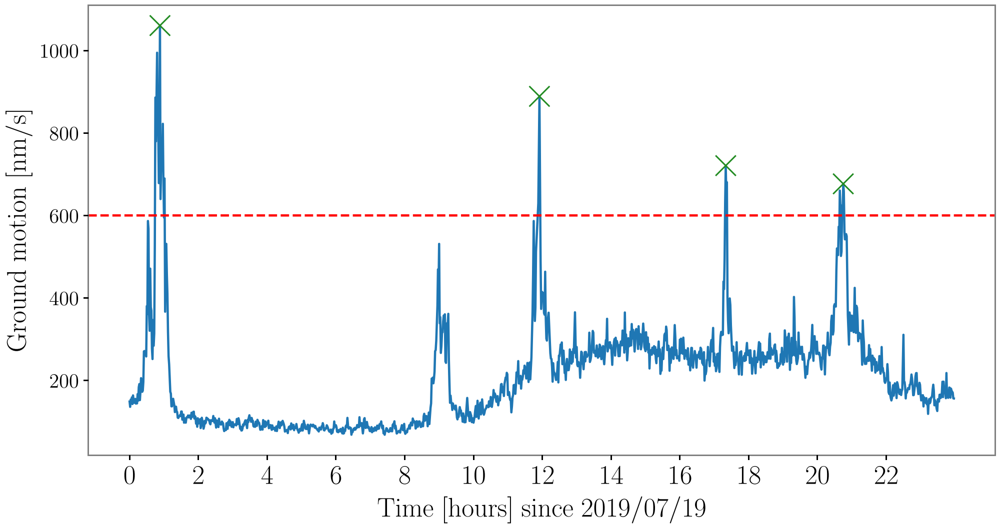
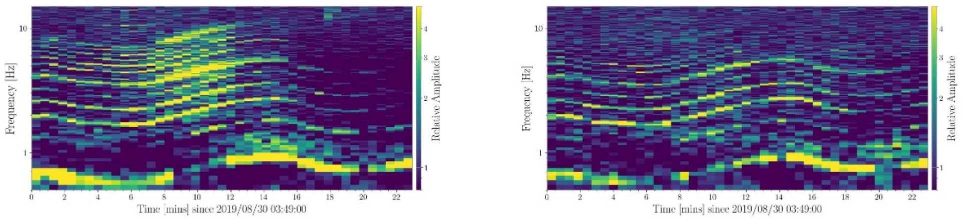


Figure 5. Example output of the python script used to select trains via seismometer data from the Y channel of ETMY in the 1–3 Hz frequency range. The horizontal bar is the cut above which data is associated with a train.



(a) Spectrogram of data from Y channel of the seismometer at ETMY

(b) Spectrogram of data from Y channel of the seismometer at ITMX

Figure 6. Visible harmonics during a train in O3. The fundamental frequency can be seen towards the lower end, ranging from 0.5–0.7 Hz. While there are more visible lines at higher frequencies near the Y end station as that is closer to the train track, the frequencies are identical in all stations. (a) Spectrogram of data from *Y* channel of the seismometer at ETMY. (b) Spectrogram of data from *Y* channel of the seismometer at ITMX.

To characterize the ground motion produced by trains, we produced spectrograms of data acquired from the seismometers; see figure 6. There are harmonic lines with changing frequency in the spectrograms. The harmonics can be explained by the physical structure of train wagons [38]. Train wagons distribute their weight along four axles, connected in pairs of two to the bogies on each wagon end. The load of each train axle results in a periodic force on the ground. The frequency of this periodic source depends on the geometric makeup of the axles and the train speed. The repeated force of the axles on the ground is the cause for the spectral line spacing in the spectrograms. The train accelerating/decelerating produces smoothly changing frequencies (the Doppler effect is negligible). For trains, the fundamental harmonic typically lies around 0.5–1.0 Hz. If we take the typical cargo train wagon to be approximately 18 m, this suggests a train speed of 9–18 ms^{-1} per second. In figures 6(a) and (b), the fundamental harmonics are around 0.7 Hz.

Although the amplitude and frequency of the ground motion changes smoothly during the time of a train, which is visible in the seismometers, the noise in the gravitational wave readout shows several different short ‘bursts’ of increased amplitude as shown in figure 7. We suspect each burst is produced by the seismic noise exciting mechanical resonances in different scattering surfaces.

3.3.1. Lasso correlation analysis. Overfitting is one of the primary concerns in model selection. Complex models run the risk of following the noise instead of the signal and thus overfitting. Multiple regularization techniques exist to help with reducing overfitting and increasing the interpretability of the models. The regularization works by adding a term to the cost function that depends on the weights associated with the model features. Large values of these weights correspond to increasing complexity and overfitting. By adding a term that is some function of these weights, regularization techniques aim to penalize the cost function for large weight values. Two common regularization techniques used in data analysis are Ridge [39] and Lasso (least absolute shrinkage and selection operator) [40]. Ridge regression uses L2 normalization, adding the term $\alpha\sum_{j=1}^{m}\beta_j^2$ to the cost function, whereas Lasso regression uses L1 normalization and adds the term $\alpha\sum_{j=1}^{m}\beta_j$ to the cost function. Here, α is a parameter that can be tuned to change the amount of penalty imposed on the cost function and β_j refers to the weight associated with the jth variable. Lasso regression can be used for feature reduction

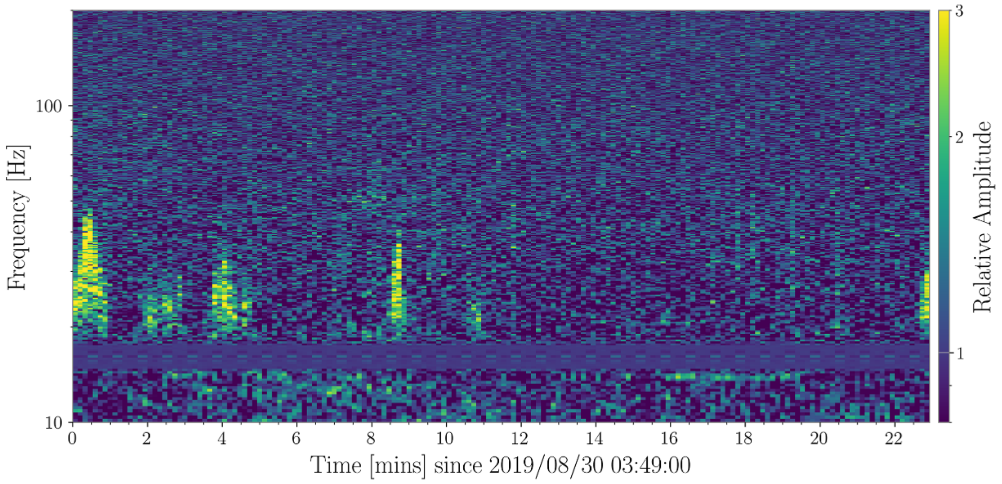


Figure 7. Spectrogram of increased detector noise during an O3a train on 30 August 2019. Bursts of increased noise are observed in the h(t) data.

and selection, as it can drive the coefficients to zero [41]. Ridge regression can only shrink the coefficients to zero. We selected Lasso due to its feature selection capability in an effort to find physical meaning of correlations between increased seismic and strain noise. We selected an alpha value of 0.003, and restricted the coefficients to be positive. The positive restriction helps prevent over-fitting of the data. To determine an appropriate alpha value, we initially started this analysis by using a specific version of Lasso called Lasso CV from the PYTHON package scikit-learn [42]. This version estimates the optimal alpha parameter by using a K-fold cross-validation. This method divides the data into K subsets, iteratively trains the model on K-1 folds, and evaluates different alpha values to determine the best average performance. The default is a five-fold cross-validation.

For our analysis, we use data from the seismometers to create a model of the strain noise during the time of trains. We bandpassed the seismic data from 0.3–10 Hz in steps of 0.2 Hz, and then create a time series with the rms of each band with a time step of 5 s. This gives us approximately 900 narrow seismic bands to correlate with h(t) for each the three seismometers in each direction. From this, we are able to look at specific band-passed frequencies in a particular station as well as direction where we can further analyze how they fit with the calibrated strain data. With these parameters, we have approximately 30–40 narrow seismic bands that correlate with the strain noise per train. An example output of this correlation is shown in figure 8.

To confirm which bands were the most frequently correlated with the gravitational wave channel, we identified a control group of times in which there was no noise in the detector for comparison. We selected 59 quiet times near trains (within ±1 h), with durations of 15 min and ran Lasso on them. During these quiet times, we would not expect any bursts of power (due to the trains) and therefore we can use this to check the validity of Lasso correlations during trains.

3.3.2. Spearman correlation analysis. To study the relationship between train induced anthropogenic ground motion and transient noise in the gravitational wave readout, we developed an algorithm that calculates the Spearman correlation between the two variables

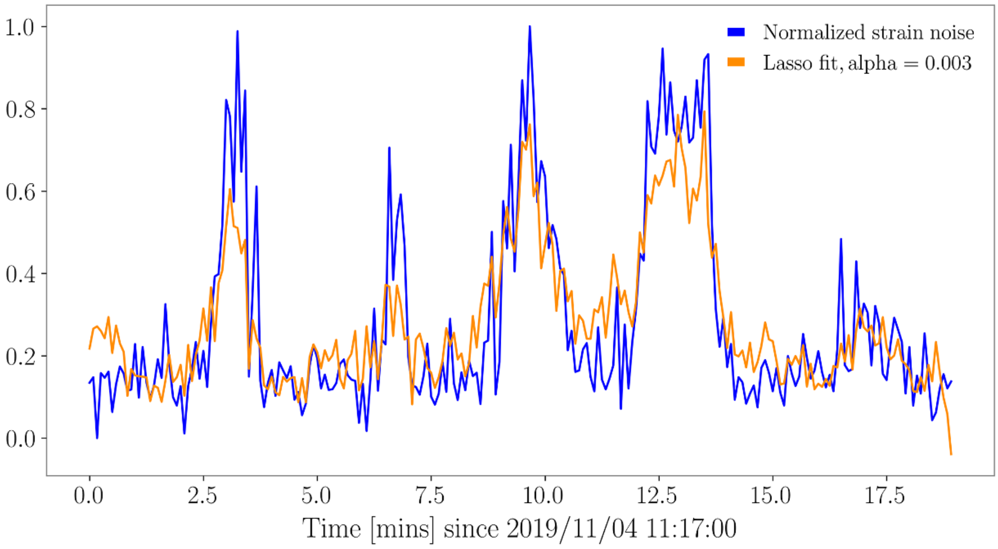


Figure 8. An example Lasso fit, with an r-square value of 0.72 for train 04 November 2019 11:17:00 UTC. In blue, we have the normalized strain noise which serves as our primary model. The orange represents a linear combination of data from all three seismometers.

of interest [43]. These variables are ground motion velocity measured in $\mathrm{nm\,s^{-1}}$ and the transient noise rate. We use Spearman correlation because we assume the relationship between how much the ground is moving and the rate of transients is monotonic, i.e. if one increases so does the other and vice versa. Days with higher ground motion for example, are associated with increased rate of transients [14]. The exact relationship between the ground motion and transient noise rate is complicated and depends on multiple factors such as how much of the ground motion translates to scatterer motion, the amount of scattered light hitting the scatterer etc.

For each train, we start with raw ground motion velocity data sampled at 512 Hz in the X, Y and Z directions recorded by the 3 seismometers located at End X, End Y and corner station. We then band pass this timeseries data in frequency bands from 0.3 Hz to 4.8 Hz in steps of 0.3 Hz and calculate the root-mean-square of 30 s of data. Next, we calculate the rate of transient noise for each of these trains, by calculating the total number of Omicron triggers between the start and end time of the train, and dividing that by the duration of the train. For these triggers, we apply the SNR threshold ($5 < \rho < 50$) and lower and upper frequency cutoffs at 10 Hz and 200 Hz respectively. For any given train, we have a total of 135 time-series streams, since the data from 3 seismometer locations along three axes is bandpassed in 15 frequency bands. Next, we calculate the Spearman correlation coefficient between the median ground motion value of these 135 time-series and the transient noise rate.

From the collection of 791 trains we randomly selected 330 trains and ran the algorithm to calculate the correlation coefficients defined above. This gives us, for each train, 135 coefficients, each one showing the strength of monotonic relationship between ground motion in a

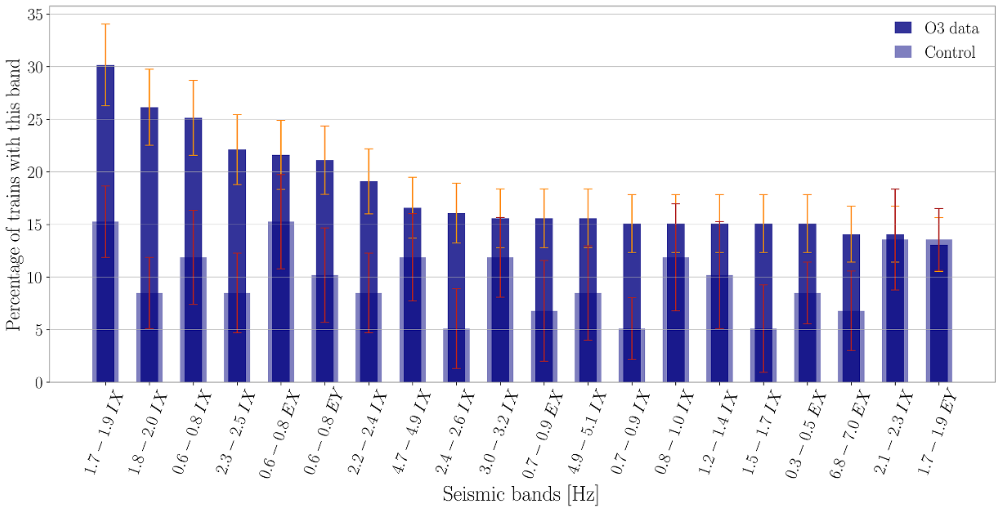


Figure 9. Lasso results displaying the top 20 most common seismic frequencies that have correlations with increased strain noise during trains in O3. The corner station (ITMX) is the most common detector location.

particular band and the location and axis of that ground motion in the detector. We then calculate the median of these coefficients, across all the trains analyzed. The results are shown in table 2.

4. Results

4.1. O3 Lasso results

With Lasso, we can see which narrow seismic frequencies are the most correlated with increased detector noise during the time of trains. In particular, we want to determine which frequencies and detector locations are the most common in an effort to locate potential scattering surfaces. As described in section 3, 199 trains during O3 were analyzed with Lasso.

The most common frequency ranges that correlate with increases in detector noise are 0.6–0.8 Hz, 1.7–1.9 Hz, 1.8–2.0 Hz, and 2.3–2.5 Hz as shown in figure 9. Frequently seen in the spectrograms of these trains is different narrow seismic frequencies correlating with the various bursts of increased power. This can be seen in figure 12, and suggests that there may be multiple scattering surfaces. Figure 10 shows the significance between the Lasso data and the control group for the top twenty most common seismic frequencies, whereas figure 11 shows the significance for just ITMX. More examples similar to figure 12 can be found in the appendix in figures 13–16.

4.2. O3 results from correlation study

Table 2 shows the results for Spearman correlation analysis between ground motion and rate of transients in the strain channel for trains in O3. The ground motion at each location and

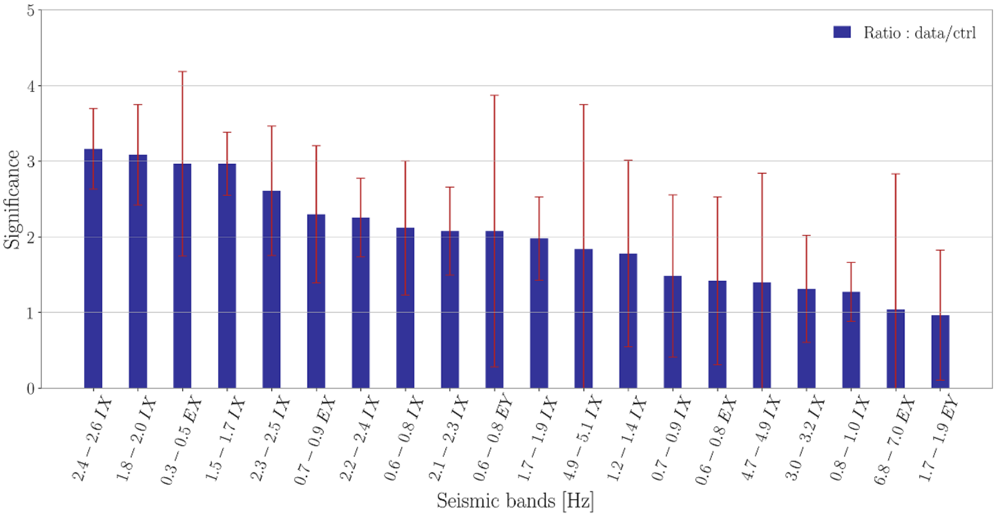


Figure 10. Significance of each seismic band as compared to the control group.

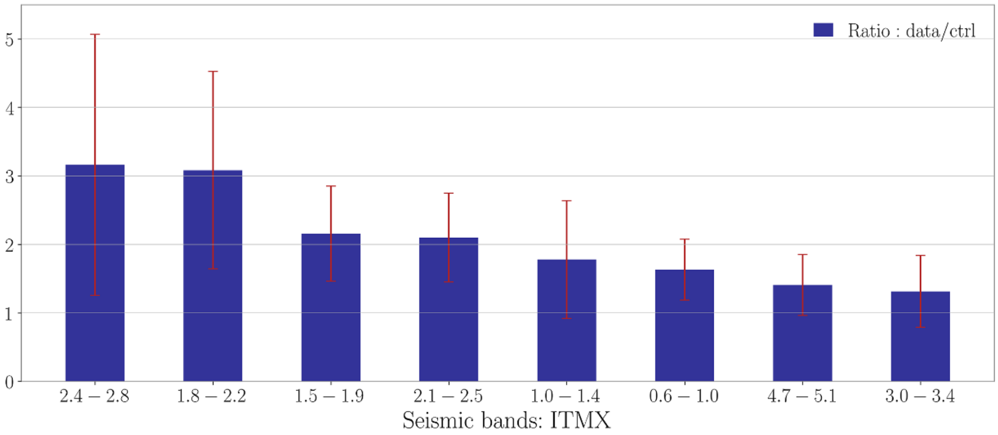


Figure 11. Significance of wider frequency grouping of only the ITMX Lasso data as seen in figure 10.

along each axis, is divided into several different frequency bands. In this analysis, we find the fast scatter rate correlates best with corner station in 1.8–2.0 Hz. Remarkably, as can be observed from this table, the corner station has the highest correlation with the transient rate in all the bands above 1.5 Hz. Its also worth stressing that the corner station couples only slightly better than other locations. However, the ground motion frequency bands excited by the trains are very similar at all locations and so even small differences in coupling correlations can be meaningful.

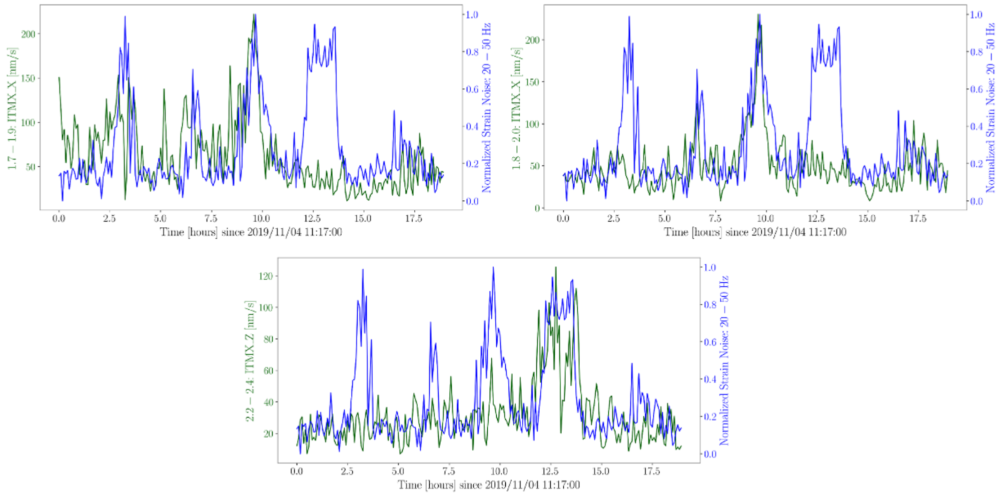


Figure 12. 1.7–1.9 Hz, 1.8–2.0 Hz, and 2.2–2.4 Hz seismic band passes in ITMX_X and ITMX_Z compared to the 20–50 Hz strain. Different frequencies correspond to the different bursts in power, suggesting multiple scattering surfaces.

Table 2. Spearman correlation coefficients between band-passed ground motion data from corner, X and Y end stations seisomemeter and transient noise rate in the strain data channel during trains. The highest coefficient within each band is shown in bold. The highest correlation co-efficient across all the bands, shown in magenta, is for corner station motion in the 1.8–2.1 Hz band.

Frequency band	IY_X	IY_Y	IY_Z	EX_X	EX_Y	EX_Z	EY_X	EY_Y	EY_Z
0.3 – 0.6	0.16	**0.18**	0.12	0.16	0.17	0.12	0.15	**0.18**	0.11
0.6 – 0.9	0.36	0.42	0.18	**0.49**	0.47	0.22	0.36	0.28	0.15
0.9 – 1.2	0.28	0.39	0.31	0.39	0.37	**0.41**	0.18	0.12	0.27
1.2 – 1.5	0.46	0.53	0.44	0.51	0.51	**0.56**	0.26	0.18	0.30
1.5 – 1.8	**0.66**	0.64	0.60	**0.66**	0.65	**0.66**	0.45	0.37	0.48
1.8 – 2.1	**0.69**	**0.69**	0.64	0.66	0.63	0.65	0.54	0.56	0.56
2.1 – 2.4	0.65	**0.66**	0.59	0.63	0.60	0.58	0.53	0.53	0.52
2.4 – 2.7	0.65	**0.67**	0.61	0.62	0.61	0.58	0.54	0.58	0.51
2.7 – 3.0	0.60	**0.61**	**0.61**	**0.61**	0.59	0.56	0.49	0.52	0.47
3.0 – 3.3	0.61	**0.62**	0.61	0.61	0.58	0.57	0.54	0.55	0.52
3.3 – 3.6	0.59	**0.61**	0.57	0.60	0.57	0.56	0.49	0.51	0.47
3.6 – 3.9	0.60	**0.61**	0.45	0.58	0.57	0.56	0.46	0.51	0.47
3.9 – 4.2	0.56	**0.62**	0.54	0.58	0.58	0.56	0.45	0.47	0.49
4.2 – 4.5	0.62	**0.68**	0.62	0.59	0.56	0.52	0.41	0.43	0.50
4.5 – 4.8	0.58	**0.63**	0.37	0.60	0.50	0.51	0.39	0.39	0.46

5. Discussion

A few conditions impact the rate of transients due to ground motion in the vicinity of the instrument. This involves the intensity of ground motion, resonant frequencies of the detector components and scattered light amplitude. As mentioned earlier, the amount of ground motion is captured by the seismometers at End X, End Y, and the corner station, and it shows a clear correlation with scattered light transients in h(t). As expected, higher ground motion leads to more noise for a given band of importance. The second key factor is the resonant vibrating motion of different detector components from which light can get scattered. Furthermore, these resonances are the reason why we have fast scatter at specific frequencies such as 2 Hz, 4 Hz, 3.3 Hz. The spacing between the fast scattering arches helps us narrow down the list of possible suspects. So similar degree of ground motion in a band that contains vibration resonances will create more noise in h(t) than in a band that does not. Finally, the third factor is how much light is scattered by the moving detector component, which depends on how much light is incident on it and what fraction of it is reflected towards the mirrors. This is often difficult to measure, and commissioners use different techniques, including taking photographs and videos of detector hardware to assess the amount of light on them.

Since the coupling between ground motion and strain noise depends on all these aspects, it can get very difficult to find the exact source of noise. A component with large motion amplitude may not have enough light amplitude, another component receiving sufficient light may be damped properly and thus would require too much motion to create noise above the strain sensitivity.

Trains near LIGO Livingston routinely create strain noise, mostly in 10–200 Hz frequency band. The ground motion amplitude due to these trains is highest at End Y but *a priori* it cannot be assumed to be location of noise coupling, as per the above discussion. In this paper, we answer two main questions with regards to noise due to trains. What ground motion frequency band correlates the most with strain noise? Which location (out of End Y, End X and Corner) has highest ground motion coupling with strain noise? To answer these questions we use two methods, the Lasso analysis and the Spearman correlation analysis. Using both these approaches, we find a high degree of correlation between ground motion in 1.8–2.2 Hz at the Corner station and strain noise due to trains.

Identification of frequency bands that correlate well with strain noise is an important step as it allows commissioners to narrow down the list of potential suspects with resonances within the band. These resonances may then be damped leading to reduction in the noise. We suspect the motion of ARM Cavity Baffle in the corner station could be responsible for increased noise in $h(t)$ during trains.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation and operates under Cooperative Agreement PHY-1764464. Advanced LIGO was built under Grant No. PHY-0823459. The authors acknowledge support from NSF PHY-2110509. S S acknowledges support from the United States National Science Foundation (NSF) under Award PHY-1764464. In this paper, we used the LIGO computing clusters and data from the Advanced LIGO detectors. This document has been assigned LIGO-P2300032.

Appendix

Here we provide more examples of seismic band pass plots of various trains.

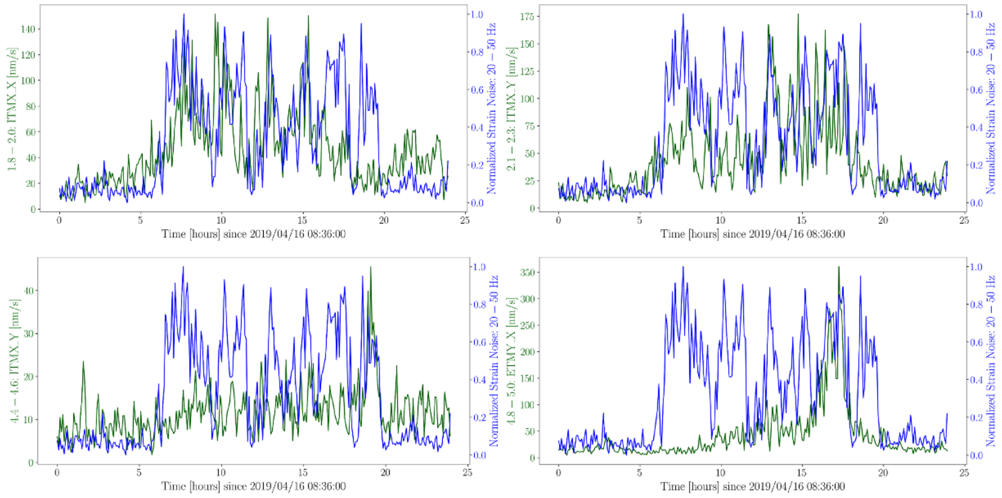


Figure 13. Train 16 April 2019 at 08:36:00.

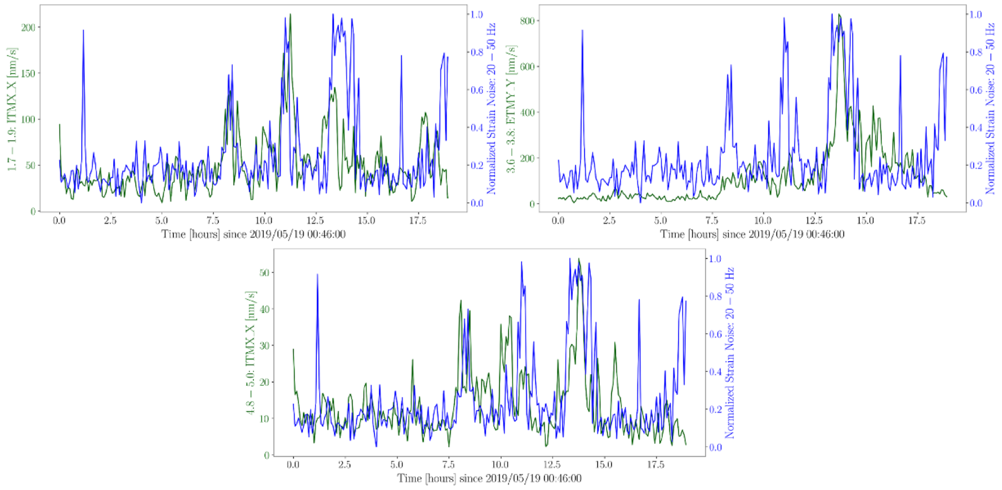


Figure 14. Train 19 May 2019 at 00:46:00.

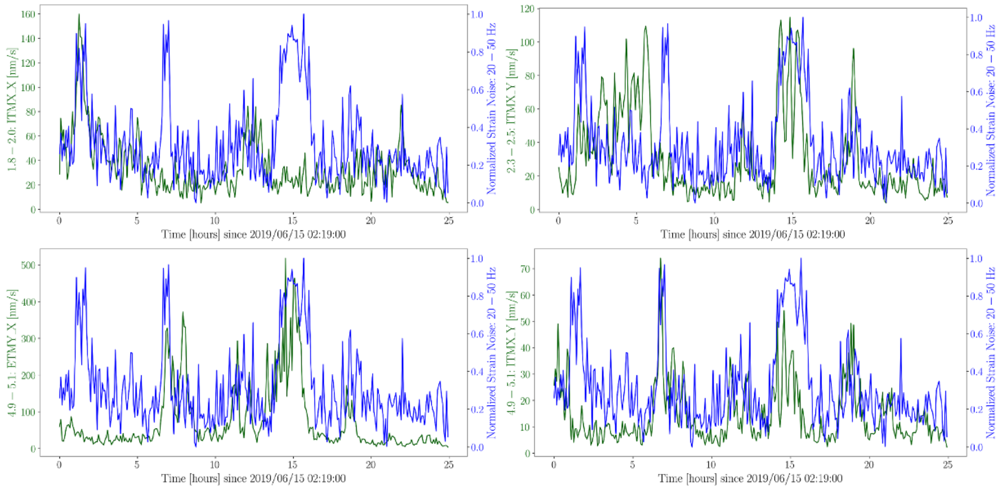


Figure 15. Train 15 June 2019 at 02:19:00.

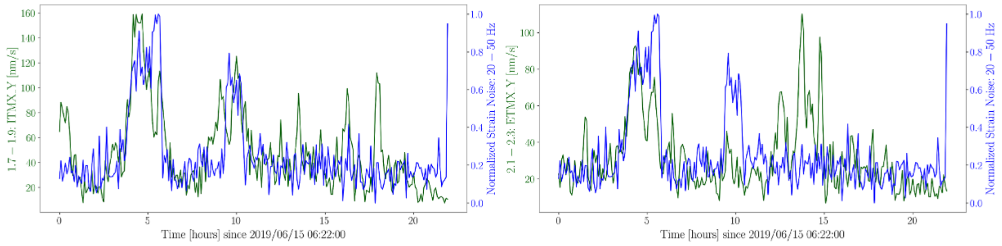


Figure 16. Train 15 June 2019 at 06:22:00.

ORCID iDs

J Glanzer https://orcid.org/0009-0000-0808-0795
S Soni https://orcid.org/0000-0003-3856-8534
G González https://orcid.org/0000-0003-0199-3158

References

[1] Aasi J *et al* 2015 *Class. Quantum Grav.* **32** 074001
[2] Abbott R *et al* (LIGO Scientific, Virgo and KAGRA Collaboration) 2021 GWTC-3: compact binary coalescences observed by LIGO and Virgo During the second part of the third observing run (arXiv:2111.03606)
[3] Abbott R *et al* (LIGO Scientific and Virgo Collaboration) 2021 GWTC-2.1: deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run (arXiv:2108.01045)
[4] Acernese F *et al* 2014 *Class. Quantum Grav.* **32** 024001
[5] Abbott B P *et al* 2016 *Phys. Rev.* X **6** 041015
[6] Abbott B P *et al* 2017 *Phys. Rev. Lett.* **119** 161101
[7] Abbott B *et al* 2019 *Phys. Rev.* X **9** 031040
[8] Buikema A *et al* 2020 *Phys. Rev.* D **102** 062003
[9] Abbott B P *et al* 2018 *Class. Quantum Grav.* **35** 065010

[10] Abbott B P *et al* 2020 *Class. Quantum Grav.* **37** 055002
[11] Soni S and Massinger T 2020 Worst noise offenders in O3a/O3b *Technical Report* G2001161 LSC (available at: https://dcc.ligo.org/LIGO-G2001161/public)
[12] Matichard F *et al* 2015 *Class. Quantum Grav.* **32** 185003
[13] Glanzer J *et al* 2023 *Class. Quantum Grav.* **40** 065004
[14] Soni S *et al* 2020 *Class. Quantum Grav.* **38** 025016
[15] Accadia T *et al* 2010 *Class. Quantum Grav.* **27** 194011
[16] Ottaway D J, Fritschel P and Waldman S J 2012 *Opt. Express* **20** 8329–36
[17] Soni S *et al* 2021 *Class. Quantum Grav.* **38** 195016
[18] Nguyen P *et al* 2021 *Class. Quantum Grav.* **38** 145001
[19] Effler A, Schofield R M S, Frolov V V, González G, Kawabe K, Smith J R, Birch J and McCarthy R 2015 *Class. Quantum Grav.* **32** 035017
[20] Flanagan E and Thorne K 1994 Noise due to backscatter off baffles, the nearby wall, and objects at the fare end of the beam tube; and recommended actions *Technical Report* T940063 LSC (available at: https://dcc.ligo.org/T940063/public)
[21] Effler A 2022 aLIGO LLO logbook 60927 (available at: https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep = 60927)
[22] Soni S 2023 aLIGO LLO logbook 63569 (available at: https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep = 63569)
[23] Effler A 2022 aLIGO LLO logbook 53364 (available at: https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep = 53364)
[24] Effler A 2022 aLIGO LLO logbook 53025 (available at: https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep = 53025)
[25] Effler A 2022 aLIGO LLO logbook 53057 (available at: https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep = 53057)
[26] Effler A 2022 aLIGO LLO logbook 53185 (available at: https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep = 53185)
[27] Effler A 2022 aLIGO LLO logbook 60943 (available at: https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep = 60943)
[28] Soni S 2021 Identification and reduction of scattered light noise in LIGO (available at: https://digitalcommons.lsu.edu/gradschool_dissertations/5498/)
[29] Zevin M *et al* 2017 *Class. Quantum Grav.* **34** 064003
[30] Daw E, Giaime J, Lormand D, Lubinski M and Zweizig J 2004 *Class. Quantum Grav.* **21** 2255
[31] Davis D *et al* 2021 *Class. Quantum Grav.* **38** 135014
[32] Macleod D, Urban A L, Smith J and Massinger T 2019 gwdetchar/hveto: 1.0.1 (available at: https://doi.org/10.5281/zenodo.3532131)
[33] Urban A L 2019 gwdetchar/gwdetchar: 1.0.2 (available at: https://doi.org/10.5281/zenodo.3592169)
[34] Macleod D, Urban A L, Coughlin S, Massinger T, Pitkin M, rngeorge P, Areeda J, Singer L, Quintero E and Leinweber K 2020 gwpy/gwpy: 2.0.1 (available at: https://doi.org/10.5281/zenodo.3973364)
[35] Robinet F, Arnaud N, Leroy N, Lundgren A, Macleod D and McIver J 2020 *SoftwareX* **12** 100620
[36] Essick R, Godwin P, Hanna C, Blackburn L and Katsavounidis E 2020 *Mach. Learn.: Sci. Technol.* **2** 015004
[37] Coughlin S *et al* 2021 Gravity spy machine learning classifications of LIGO glitches from observing runs O1, O2, O3a, and O3b (available at: https://doi.org/10.5281/zenodo.5649212)
[38] Fuchs F and Bokelmann G (the AlpArray Working Group) 2018 *Seismol. Res. Lett.* **89** 56–66
[39] Hoerl A E and Kennard R W 1970 *Technometrics* **12** 55–67
[40] Tibshirani R 1996 *J. R. Stat. Soc. Ser.* B **58** 267–88
[41] Walker M, Agnew A F, Bidler J, Lundgren A, Macedo A, Macleod D, Massinger T J, Patane O and Smith J R 2018 *Class. Quantum Grav.* **35** 225002
[42] Pedregosa F *et al* 2011 *J. Mach. Learn. Res.* **12** 2825–30
[43] Boslaugh S and Watters P 2009 *Statistics in a Nutshell: A Desktop Quick Reference* (O'Reilly Media, Inc.)