Data quality vetoes and their effect on searches for gravitational waves from compact binary systems

Samantha Usman

August 12, 2014

Abstract

Data quality flags have been designed to prevent problems caused by noise transients in data coming from the LIGO and Virgo detectors. The vetoes, however, were created primarily to aid in the search for gravitational waves coming from burst sources, such as supernovae explosions. How, then, do these vetoes affect the search for inspiral signals? This paper looks to answer this question, then explores ways to improve both current data quality flags and the creation of new flags for coming science runs.

1 Introduction

Compact binary coalescences (called *CBC*) remain a likely source for gravitational waves. However, these gravitational waves are likely going to come from very distant binaries. By the time the waves reach Earth, the amplitude of the gravitational waves would be extremely small. For this reason, having extremely sensitive detectors and efficient analysis programs is of high importance. In an effort to improve the sensitivity of the LIGO and Virgo gravitational-wave detectors, *data quality vetoes* (*DQ flags*), or *vetoes*, have been created to remove data marred by background noise. The programs which generate these DQ flags compare excess power from auxillary channels to the main data channel. Repeated correlation between an auxillary channel and the main data channel could imply a repeated occurrence is causing the excess power to occur in both channels. Thus, the excess power in an auxillary channel can be used to remove gravitational wave candidates (called *triggers*) from the data. These correlations are then used to generate vetoes which remove triggers that occur at the time of this background noise. However, these vetoes were created to aid specifically to aid in the search for gravitational waves coming in bursts, like those from supernova explosions. This paper explores to what extent these DQ flags impact the CBC search pipeline, called *IHOPE*.

2 IHOPE Basics

In our research, we wanted to learn whether or not DQ flags were creating a significant effect on the background of the IHOPE pipeline. In order to understand these effects, we must first consider the basics of the IHOPE search pipeline. Below is a flowchart summarizing how the search pipeline functions.

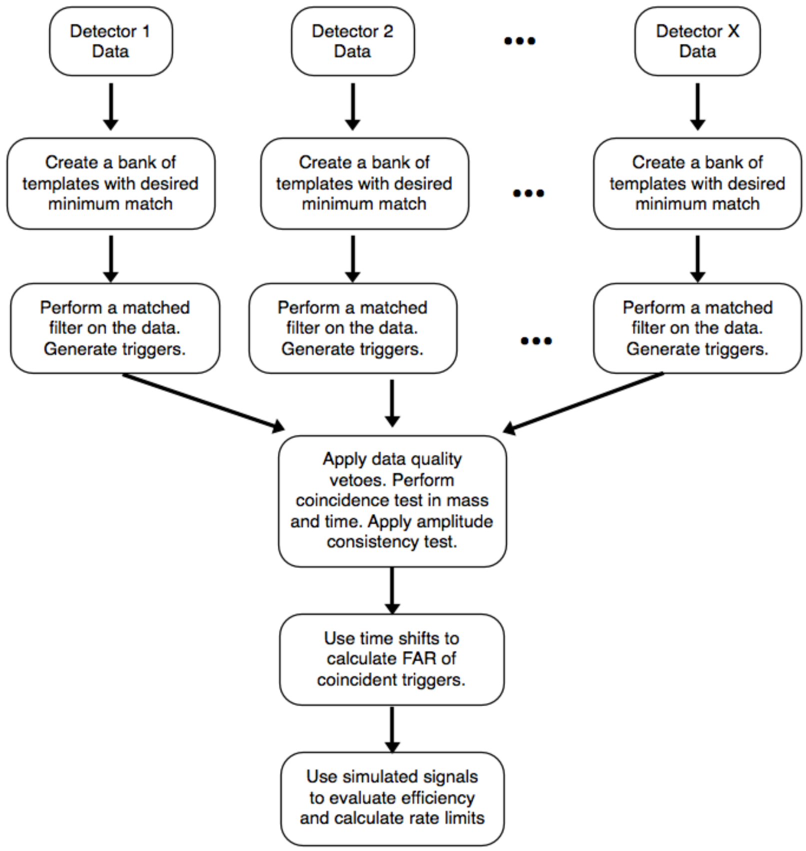


Figure 1: This figure is a flow chart describing the several stages that make up the IHOPE pipeline. Each step is described in more detail in this section.

The first part of the pipeline is the intake of data from the detectors running. In our analysis, we looked specifically at data taken during LIGO's sixth science run and Virgo's third science run, which occurred in autumn 2010.

Next, the pipeline compiles a set of theoretical gravitational waveforms. While waveforms are described by seventeen intrinsic parameters, it is reasonable to neglect orbital eccentricity and spin. Therefore, the current template bank used by IHOPE consists of waveforms whose phasing is determine almost exclusively by the mass of the two objects.

Following this, the waveforms are match-filtered against the data. Each waveform is tested for similarity against the data, comparing it to each second of data to identify areas which have a similar waveform. Each time a portion of data looks significantly similar to a waveform, it is marked as a gravitational wave candidate, called a *trigger*. The higher number the match-filter returns, the larger the trigger's signal-to-noise ratio (*SNR*). These triggers also undergo a χ^2 test. If the χ^2 test returns a high number, the trigger could have been caused by a noise transient. For this reason, we create a new parameter, aptly named new SNR, which in some sense divides the original SNR by the χ^2. This leaves triggers which were probably caused by glitches to have a very low new SNR while important candidates remain significant.

After collecting these triggers and their significances, we apply data quality vetoes. These vetoes are created in order to remove triggers which were most likely caused by background noise or simulations instead of a real gravitational wave. In the case of simulations, hardware injections are created by pushing mirrors in a way that mimics the changes made by gravitational waves. The times when these are performed are known and are used to test the performance of the detectors and the analysis programs, thus vetoes are used to remove these known injections from the analysis.

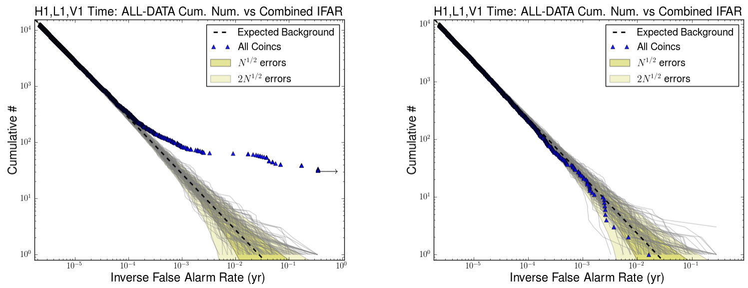
In the case of background noise, DQ flags are generated by comparing the main data channel to auxillary channels, which collect data from areas that would not be noticeably affected by gravitational waves. Seismometers, accelerometers and michrophones are all included in this set of auxillary channels. Algorithms compare these channels to the main data channel. If there is a correlation in excess power between these channels, there may be something occuring which causes noise in both the auxillary channel as well as the data channel. For example, a seismometer may pick up ground movement at the same time that the data channel believes there to be gravitational waves. We can use this correlation to identify the triggers as being generated by background noise instead of gravitational waves. These triggers are then removed, as they are no longer considered gravitational wave candidates.

After this step, we compare the remaining triggers between all the available detectors' data. A gravitational wave passing through the Earth should reach all of the triggers within a second, since the detectors are well within one lightsecond of each other. Thus, the detector looks for triggers whose templates share similar mass parameters and timing.

Lastly, since gravitational waves must occur practically simultaneously, we use time-shifting to estimate background noise in the detectors. One or more of the detectors' data streams are shifted with respect to the others. The triggers which survive these coincidence could not possibly occur physically, and are thus a good estimation of background noise. To study the effects of DQ flags, we looked largely at this noise. Assuming the vetoes are working effectively, the number of significant background triggers should decrease.

3 With and Without DQ Flags

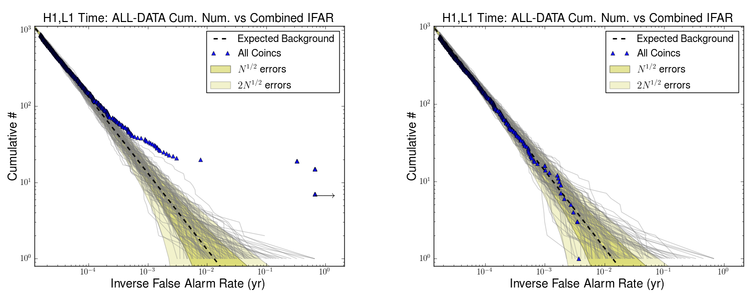
To begin our study, I first looked at the standard set of vetoes' effect on background noise. The first part of this consisted of visually comparing charts before and after DQ flags were applied.



(a) This plot contains no vetoes. Note that hardware injections are removed using vetoes, thus the significant triggers are all known injections.

(b) This plot contains the standard set of vetoes used in LIGO's sixth and Virgo's third science runs. Notice the hardware injections are now removed.

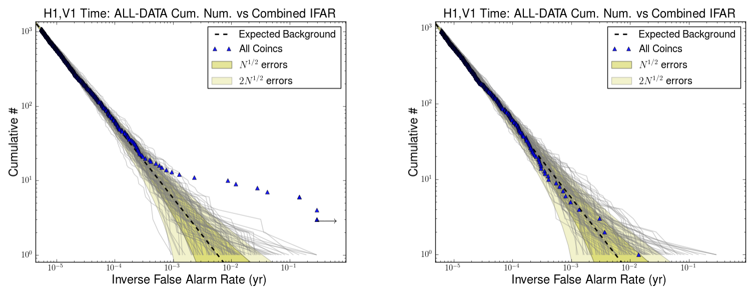
Figure 2: These plots describe detector data and the background noise contained therein. An in-depth description is included in this section. This pair is an analysis of data when all three detectors, Hanford, Livingston and Virgo, were on and collecting data.



(a) This plot contains no vetoes. injections.

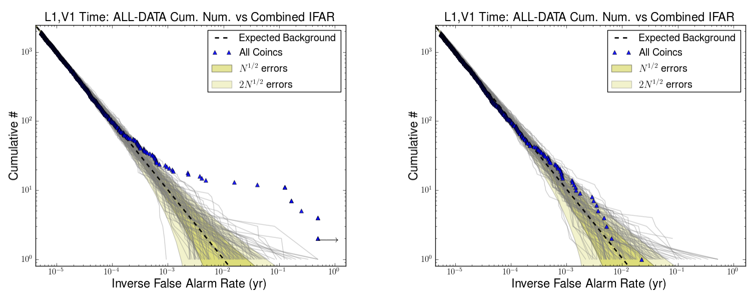
(b) This plot contains the standard vetoes.

Figure 3: This pair of plots is an analysis of data when the Hanford and Livingston detectors were on and taking data.



(a) This plot contains no vetoes. (b) This plot contains standard vetoes.

Figure 4: This pair of plots is an analysis of data when the Hanford and Virgo detectors were on and taking data.



(a) This plot contains no vetoes. (b) This plot contains the standard vetoes.

Figure 5: This pair of plots is an analysis of data when the Livingston and Virgo detectors were on and taking data.

On the x-axis, we have the *inverse false alarm rate* (*IFAR*). *False alarm rate*, also called *FAR*, estimates how often a trigger with a similar significance could occur given gravitational waves didn't exist and the detectors were simply measuring background noise. IFAR is often used in place of of FAR for ease of comprehension: larger numbers (in this case, IFAR) are assoiciated with more significant triggers. Thus this IFAR signifies how many years, approximately, one would have to wait to find a gravitational wave with the same significance from noise. The y-axis lists the total number of triggers that have the significance equal to or greater than the one listed along the x-axis. This is to say that when a coincidence test is performed, some triggers remain significant. These triggers would have a high IFAR and appear higher on the chart. It is to be expected that fewer triggers have high signficance, so the negative correlation is logical. Each gray line represents the triggers from a time-shifted coincidence test, while the black dotted line is their average. By moving the data sets with respect to each other in these time-shifts, we can only possibly measure background noise, since gravitational waves should occur almost simultaneously. The blue triangles represent triggers from the original coincidence without any time shifting. The triangles are thus generally indicative of possible gravitational waves. These figures, called *IFAR plots*, are often used to look at the estimation of background noise. For this reason, we look specifically at the gray lines, which represent triggers found from time-shifted coincidence tests.

As can be seen from the figures above, the DQ flags appear to have little to no effect on the background noise. The average inverse false alarm rate ($IFAR$) of the background noise for the times when Hanford, Livingston and Virgo were all on and running, for example, shifts from 0.03 to 0.025, a rather modest change. One explanation for this small change is the chi-squared test. This test effectively returns a large number when its corresponding trigger strongly resembles a delta-function glitch. When this test returns a number greater than the degrees of freedom (generally set to sixteen), new SNR is calculated, rendering these glitch-caused triggers relatively insignificant. However, when the standard set of vetoes had originally been chosen, the performance of the vetoes were based on the SNR, without taking the χ^2 into account. For this reason, these DQ flags may have performed well on the original triggers, but may not have performed when looking at them with respect to new SNR. Thus, we decided to test the performance of these vetoes specifically on their performance with respect to glitches with new SNR.

4 Re-testing DQ Flags

In order to examine the performance of the DQ flags, I wrote a program that calculates three particular properties of the vetoes, specifically their:

- *dead-time*, the percentage of time in which the triggers are removed out of the total amount of time analyzed (ideally this is as small as possible);
- *efficiency*, the percent of triggers removed out of the total number of triggers generated;
- *use-percentage*, the percent of veto segments which successfully remove at least one trigger; and
- *efficiency over dead-time*, the ratio of triggers being removed relative to the total amount of analyzable time removed (this would be approximately one if the veto is simply removing triggers at random).

Identifying these numbers allows us to judge how well the DQ flags are performing. I ran this program on the DQ flags with respect to both SNR and new SNR to see their performance before and after DQ flags were applied. It was easily shown that while most of the flags excelled at removing triggers with regular SNR, very few flags had a similar effect on triggers after the χ^2 test. For example, of Virgo's 77 category four flags, 36 were rated as "excellent" and 26 were rated as "good" before the application of the χ^2 test. After the test, none were.

Let's take a closer look at a specific example:
H1:DMT-SEVERE_LSC_OVERFLOW

SNR >	Triggers Removed/Total Triggers	Efficiency	Use-Percentage	Efficiency/Dead-Time
5	2495/134053	1.86%	97.5%	1.1
8	33/326	10.12%	5.0%	5.8
10	7/178	3.93%	0.4%	2.3
20	0/39	0.0%	0.0%	0.0
dead-time	1.735%			
DQperf	76.786%			

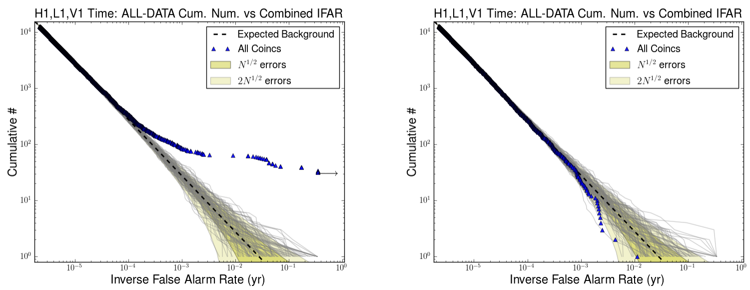
New SNR >	Triggers Removed/Total Triggers	Efficiency	Use-Percentage	Efficiency/Dead-Time
5	496/33259	1.49%	61.7%	0.9
8	0/25	0%	0%	0
dead-time	1.736%			
DQperf	21.429%			

Table 1: These tables look at the performance of the veto H1:DMT-SEVERE_LSC_OVERFLOW before and after the χ^2 test (top and bottom, respectively). DQperf is a parameter that estimates the veto's performance, with a high percentage corresponding to good performance.

As can be seen in the table above, this veto appears to perform excellently with respect to the standard SNR. However, once the χ^2 test is taken into account and the new SNR is calculated, the DQ flag no longer performs well. Since the χ^2 test reduces the significance of triggers, the total number of triggers above the given thresholds drop dramatically. A well-performing veto should, in theory, remove significant triggers that remain. A few of the vetoes did indeed continue to perform well after the χ^2 test. These remaining vetoes we used to create a new set high-performing vetoes to be used in an analysis.

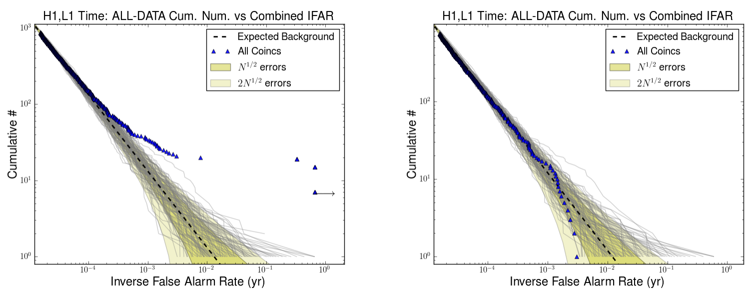
5 Analysis with High-Performing Vetoes

Using this performance-testing program, I was able to compile a list of well-performing vetoes. Then, I re-ran IHOPE with using this list of selected vetoes. We were hoping that these vetoes would remove the loudest vetoes while decreasing the amount of dead-time, thus removing the overall background noise. To analyze this, we return again to IFAR plots.



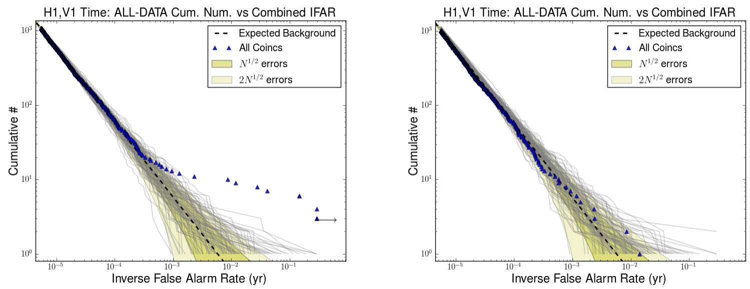
(a) Analysis without vetoes. (b) Analysis with high-performing vetoes.

Figure 6: For an explanation of the layout of these plots, see Figure 2.



(a) This plot contains no vetoes. (b) This plot contains selected vetoes.

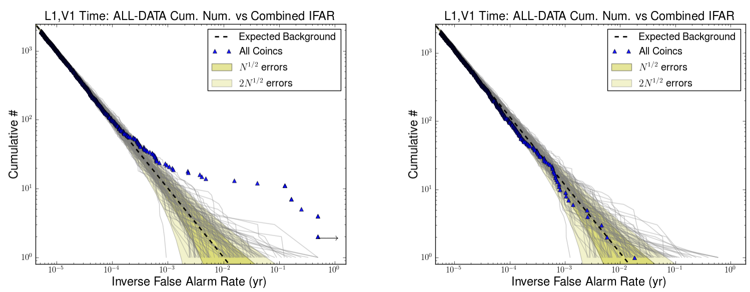
Figure 7: This pair of plots is an analysis of data when the Hanford and Livingston detectors were on and taking data.



(a) This plot contains no vetoes.

(b) This plot contains selected vetoes.

Figure 8: This pair of plots is an analysis of data when the Hanford and Virgo detectors were on and taking data.



(a) This plot contains no vetoes.

(b) This plot contains selected vetoes.

Figure 9: This pair of plots is an analysis of data when the Livingston and Virgo detectors were on and taking data.

As can be seen from the charts shown, even using the most well-performing of the vetoes have a negligible effect on the background noise of the CBC search. This lack of improvement highlights how effective the χ^2 test is in reducing the significance of triggers caused by glitches. Now is perhaps the time to wonder if there's a way to improve the implementation of the vetoes more efficiently. This is what we turned to next in our analysis.

6 New Implementation Techniques

Next, I looked at different ways to implement the vetoes. The time of a trigger is denoted by the time at which the frequency is at its peak, when the two bodies meet. Because of the matched-filtering process, it can be very easy to remove triggers caused by delta-function glitches, since the glitch will create the loudest trigger when the template aligns with the glitch at the end time. Other forms of glitches, however, may cause the glitches to occur at a slightly different time. For example, a sine-Gaussian glitch should create the loudest trigger when the template and the length of the sine-Gaussian glitch are aligned. Thus the trigger created will occur a bit later then the time of the glitch, whose time is measured as the center of the sine-Gaussian.

For this reason, I created a program to look at the efficiency and use-percentage of the DQ flags with different implementations. The standard use of the vetoes simply removes a trigger if its end time occurs in the middle of a veto segment. Other cases we looked at instead calculated what percent of template overlapped with the veto segment. My code specifically counted a trigger as removed in different cases, including whether there was any overlap, a 20%, 50%, 100% or any overlap, with or without the endpoint included. To get a rough estimate of how these vetoes performed compared to the veto acting at random, I also did a time-slide, shifting the veto set 30 seconds with respect to the data, then performing the same vetoes. In theory, the efficiency and use percentage should be high when the times aren't shifted, as the vetoes were designed to run at the same time as the data.

Thus, by looking at the difference of the normal and time-shifted analyses, we can see that the higher numbers indicate a better performance. Indeed, some vetoes performed better with these overlapping implementations instead of the standard implementation. Let's take a look at an example:

H1:DMT-OM1_OVERFLOW

Implmentation	Efficiency Diff.	Use-Percentage Diff.
Standard	0.09580%	9.85587%
Any Overlap, No Endpoint	0.11886%	8.96114%
20% Overlap, No Endpoint	0.13470%	13.9082%
50% Overlap, No Endpoint	0.09508%	7.12296%
20% Overlap, Endpoint Included	0.10697%	11.7915%
50% Overlap, Endpoint Included	0.10301%	9.07604%
100% Overlap	0.00000%	0.00000%

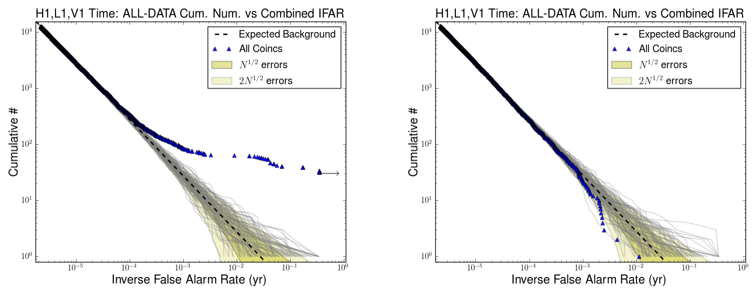
Table 2

As can be seen from this table, this particular veto has the highest efficiency and use-percentage compared to background noise when the trigger is removed if it has an overlap of 20% or greater, with no requirement on the endpoint. This veto, then, should be implemented in a different way than the current standard implementation.

Finally, for the end of my study, I looked into a new type of veto generator, called *Omicron*, and its performance with respect to the CBC search.

7 Omicron and the CBC Search

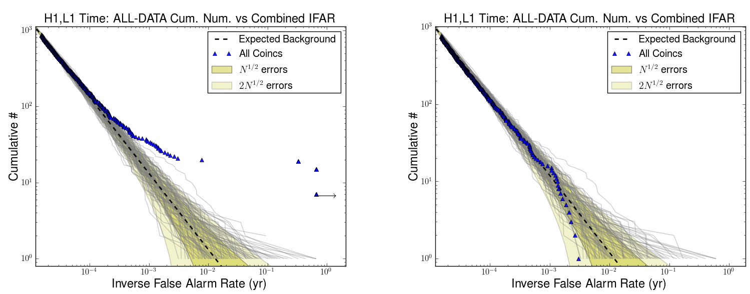
Omicron is a recently developed tool used to locate glitches in the data. This can be used to create new DQ flags to minimize background noise. Thus, the next step in our process was to test how well these new vetoes performed in the same analysis. To analyze this, we look again to the IFAR plots to compare their background noise.



(a) Analysis without vetoes.

(b) Analysis with omicron vetoes.

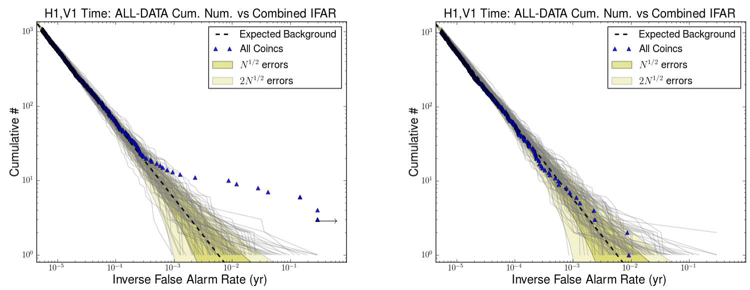
Figure 10: These are for times when the Hanford, Livingston and Virgo detectors were all on and collecting data at the same time.



(a) This plot contains no vetoes. injections.

(b) This plot contains omicron vetoes.

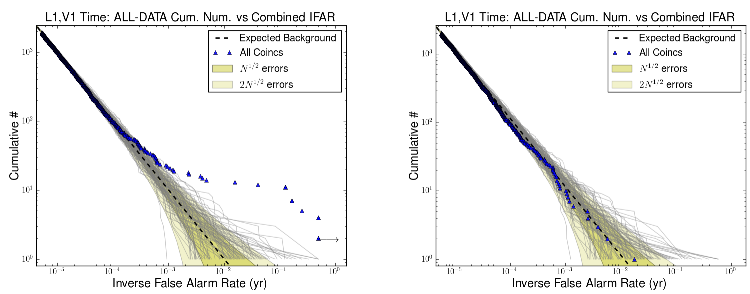
Figure 11: This pair of plots is an analysis of data when the Hanford and Livingston detectors were on and taking data.



(a) This plot contains no vetoes.

(b) This plot contains omicron vetoes.

Figure 12: This pair of plots is an analysis of data when the Hanford and Virgo detectors were on and taking data.



(a) This plot contains no vetoes.

(b) This plot contains omicron vetoes.

Figure 13: This pair of plots is an analysis of data when the Livingston and Virgo detectors were on and taking data.

In looking at these plots, we're able to see that even the analysis using omicron-generated vetoes saw negligible changes to the background slides. It appears that these vetoes, while seemingly helpful, actually have a minimal impact on the background noise in the CBC search.

8 Summary

8.1 Standard Vetoes

We can see that the standard set of vetoes used in LIGO's sixth science run and Virgo's third science run had little to no effect on the background noise. The vetoes were designed for the burst search, a search looking for gravitational waves coming from supernova explosions, to remove data believed to be caused by a local noise source. While the DQ flags successfully locate and remove triggers caused by these noise sources in the inspiral search, the majority of triggers cause by such noise sources are already rendered quite insignificant by the χ^2 test. Thus the vetoes end up removing insignificant triggers caused by noise and nearby gravitational-wave candidates. Therefore, the vetoes overall do not aid in the search.

8.2 High-Performing Vetoes

After testing the DQ flags, a short list of well-performing vetoes was compiled. However, when IHOPE was run using this new set of vetoes, the mitigation of background noise was again very minimal. This may be simply because the small list of vetoes was not enough to make a sizable dent in the background noise.

8.3 Omicron-Generated Vetoes

Lastly, we looked at the new generation of DQ flags, created using Omicron. While these were a hopeful proposal, upon further inspection, it was found that these, too, had barely any effect on the performance of the search. For this reason, we're inclined to believe these vetoes are not useful in the search for gravitational waves coming from compact binary coalescence.

9 Conclusion

As we have seen from this analysis, the standard set of vetoes currently used in the CBC search is not as useful as previously thought. Even when using the best-performing of the vetoes, the effect on the background noise remains negligible. Omicron-generated vetoes were a hopeful replacement, though upon further inspection, these were found to also do little in mitigating the effects of background noise. The χ^2 test is thus a powerful test which renders most DQ flags practically useless. While a few may be useful, these vetoes are on the whole useless if not detrimental to the search for inspiral signals.

10 Acknowledgements

I'd like to thank Bernard Whiting, Guido Mueller, Kristin Nichola, Andonis Mitidis and everyone else involved at University of Florida that made this REU possible. I'd also like to thank Duncan Brown and Peter Saulson for giving me the necessary last-minute letters of recommendation that allowed me to be accepted to the program. Lastly, I'd like to thank Florent Robinet, my boss for this past summer for whom it was a pleasure to work.