

## Chapter 7

# False Alarm Rate Detection Statistic

In this chapter, we describe the development of the false alarm rate (FAR) as the detection statistic for the Search for Low Mass CBCs in the First Year of LIGO’s Fifth Science Run (S5) Data, referred to as “this search.” This new detection statistic allows us to search over a large region of parameter space without being limited by a high background FAR from a smaller subregion. In reference [21], coincident triggers were ranked by combined effective SNR. However, for this search, we use a statistic derived from the background FAR, thus enabling a much improved efficiency for a given FAR.

The FAR is a statistic that can be defined for triggers when there is a measure of the background trigger rate. The time-shifted triggers provide an estimate of the FAR for each in-time coincident trigger. By counting the number of time-shifted triggers with a combined effective SNR greater than or equal to the in-time coincident triggers’ combined effective SNR, and dividing by the total amount of time we searched for time-shifted triggers, we calculate the FAR for each in-time coincident trigger. This procedure is done separately for different trigger categories, which have different background rates. The result is a varying mapping of combined effective SNR to FAR for the different trigger categories. The FARs from the different categories can then be recombined into a combined FAR (FAR<sub>c</sub>) for the final combined set of triggers.

Section 7.1 discusses how different trigger categories were chosen for this search, section 7.2 explains the FAR calculation, section 7.3 details how trigger categories are recombined, section

7.4 shows the comparison of different detection statistics, section 7.5 presents comparisons of in-time triggers to the many different background trials, and section 9.3.2 describes implications for measuring the background probability for the loudest triggers.

## 7.1 Definition of Trigger Categories

In this search, there are several aspects that influence the background rate. These include the type of coincidence a trigger was found in (triple coincident triggers versus double coincident triggers), the vetoes that were applied to the triggers, and the mass of the templates that correspond to the triggers. All of these aspects combine in order to define the trigger categories.

### 7.1.1 Coincidence Types

The first aspect we will focus on is the coincidence type. In this search we are searching for coincident triggers in time and two mass parameters between three GW detectors, the Hanford 4 km interferometer (H1), the Hanford 2 km interferometer (H2), and the Livingston 4 km interferometer (L1). A trigger is deemed a double coincidence if it passes the coincidence requirements between two detectors. A trigger is promoted to a triple coincidence if there are three double coincident triggers that are all found in coincidence with each other. Since triple coincident triggers have an extra coincidence requirement, there is a much lower background rate for triple coincident triggers than for double coincident triggers.

Say each detector has a probability of producing a background trigger for a particular mass template at a particular time of  $\epsilon_1$ . Adding the double coincident requirement then reduces that probability by an additional factor of  $\epsilon_1$ , so that  $\epsilon_2 = \epsilon_1 \epsilon_1 = \epsilon_1^2$ . Adding an additional coincident requirement added another factor of  $\epsilon_1$ , so the probability of getting a triple coincidence trigger at a particular time is then  $\epsilon_3 = \epsilon_1 \epsilon_2 = \epsilon_1^3$ .

This background rate reduction can be illustrated by looking at background H1H2L1 triggers and H1L1 triggers produced during triple coincident time (figure 7.1). For this search the different trigger types we decided to use were H1H2L1, H1L1, and H2L1 triggers. We exclude H1H2 triggers

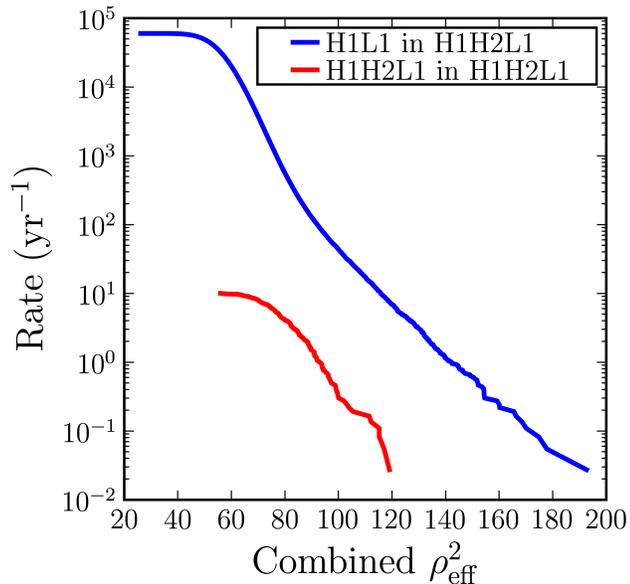


Figure 7.1: Effect of Coincidence on Background Trigger Rate

The cumulative rate versus combined effective SNR for background H1L1 double coincident triggers during triple coincident time (blue) and H1H2L1 triple coincident triggers during triple coincident time (red). The additional coincidence requirement for H1H2L1 triggers is seen to drastically reduce the background trigger rate.

from this analysis as their background cannot be effectively estimated (section 6.5).

### 7.1.2 Veto Differences

Another aspect that affects the background rate is the differences in vetoes that are applied to different triggers. One set of vetoes that are trigger dependent are the amplitude vetoes applied between H1 and H2 triggers. These vetoes compare the effective distances found between coincident H1H2 triggers in H1H2L1 coincidences or, for H1L1 or H2L1 triggers, determine whether the other Hanford detector should have also produced a trigger at that time. These vetoes can only be applied when both Hanford detectors were in operation resulting in background rate differences between H1L1 triggers from triple coincident time versus H1L1 triggers from double coincident time, and similarly for H2L1 triggers. This effect is shown for H2L1 trigger in figure 7.2, which clearly shows broader background distributions for H2L1 time where the veto is not applied even though there is

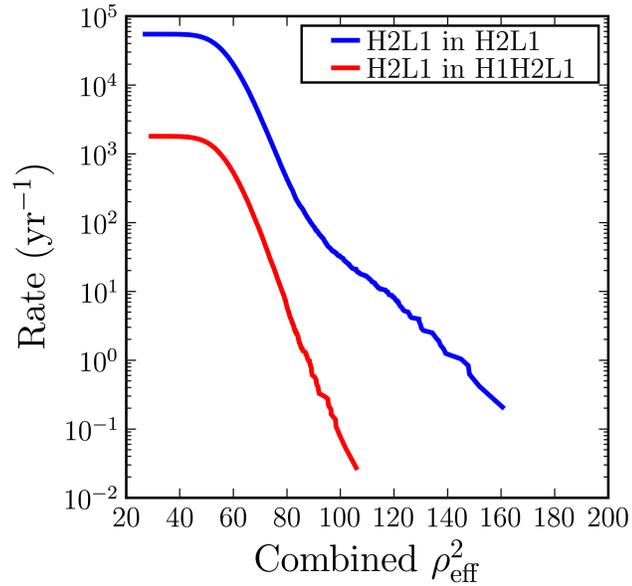


Figure 7.2: Effect of Amplitude Veto on Background Trigger Rate

The cumulative rate versus combined effective SNR for background H2L1 double coincident triggers during double coincident time (blue) and H2L1 double coincident triggers during triple coincident time (red). The Hanford amplitude veto drastically reduces the background rate during triple coincident times, even though there is an order of magnitude more triple coincident time than double coincident time.

an order of magnitude more triple coincident time. This effect is also stronger for H2L1 triggers in triple coincident time than for H1L1 triggers in triple coincident time (figure 7.3). This is due to the fact that the H2 detector is roughly half as sensitive as the H1 detector, which means that if there is a trigger in H2, most of the time H1 should also have produced a trigger at that time.

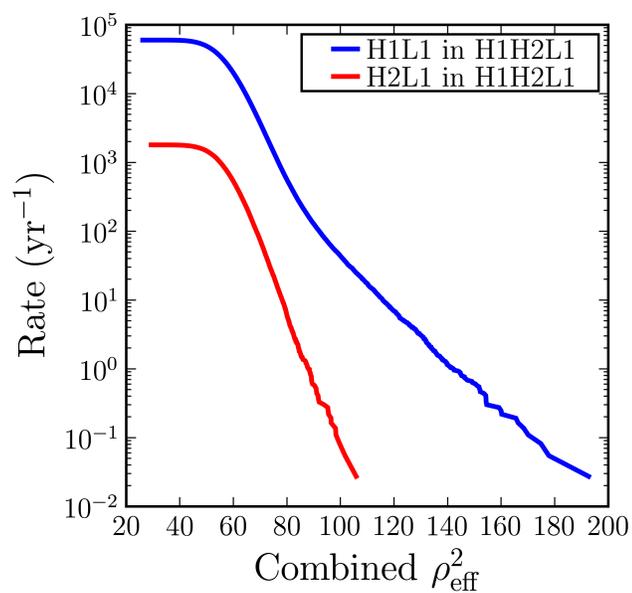


Figure 7.3: Effects of Different Amplitude Vetoes for H1 and H2

The cumulative rate versus combined effective SNR for background H1L1 double coincident triggers during triple coincident time (blue) and H2L1 double coincident triggers during triple coincident time (red). The Hanford amplitude veto is seen to be more stringent for H2L1 triggers than for H1L1 triggers due to the sensitivity asymmetry between H1 and H2.

### 7.1.3 Mass Differences

This search is over a much larger parameter space than a single search has covered before. Specifically, the mass space we cover with templates goes from a minimum total mass  $M = m_1 + m_2$  of  $2 M_\odot$  to a maximum total mass of  $35 M_\odot$  with a minimum component mass of  $1 M_\odot$ . In order to cover this large space, we use around 6000 templates whose time durations vary from over 44 seconds to under 0.34 seconds.

Due to the large variation in the length of the templates, we find that there is a larger variation in the templates' responses to noise glitches in the detectors where high mass templates tend to pick up the noise glitches as triggers with higher effective SNR. This can be seen in figure 7.4 where we plot the cumulative distributions of background H1L1 double coincident triggers during triple coincident time versus combined effective SNR separated by small bands in the mean chirp mass  $\mathcal{M} = M\eta^{3/5}$ , where  $\eta$  is the symmetric mass ratio  $\eta = m_1 m_2 / M^2$ . Three populations are found with different slopes for the distributions signifying different background rates, thus, we divide the mass space into three regions, for simplicity. The chirp mass divisions for these regions occur at  $\mathcal{M} = 3.48 M_\odot$  and  $\mathcal{M} = 7.40 M_\odot$ .

After dividing the triggers of a given type into these three different populations, we find that, indeed, the higher chirp mass templates have a broader combined effective SNR distribution for background triggers, signifying different background rates for the different chirp mass regions (figure 7.5).

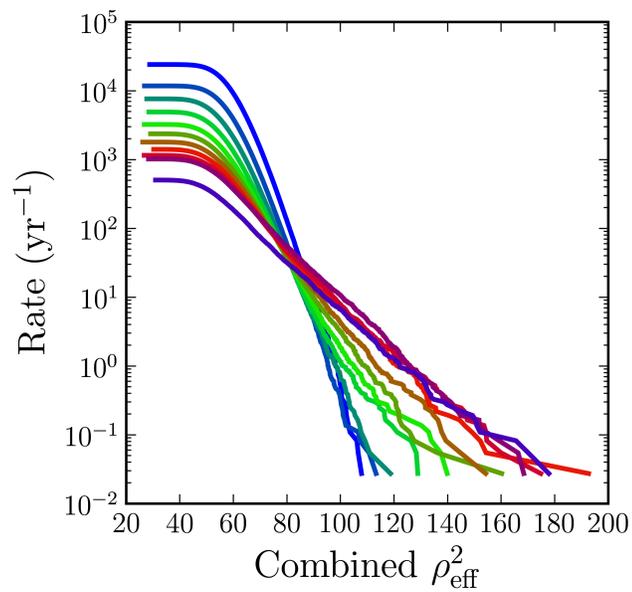


Figure 7.4: Combined  $\rho_{\text{eff}}^2$  Distributions for Background Triggers for Different Chirp Mass Bands

The cumulative rate versus combined effective SNR for background H1L1 double coincident triggers during triple coincident for eleven separate bands of “mean chirp mass” of equal widths. Lines colored toward the blue end of the spectrum are lower mass bands while line colored toward the purple end of the spectrum are high mass bands. Three populations are found with different slopes for the distributions signifying different background rates. Similar results are obtained by looking at plots for different coincident trigger types.

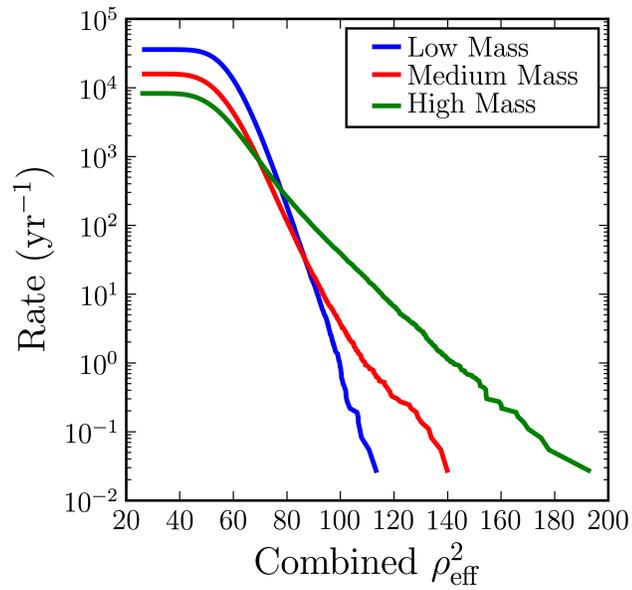


Figure 7.5: Combined  $\rho_{\text{eff}}^2$  Distributions for Background Triggers for Different Chirp Mass Bins  
 The cumulative rate versus combined effective SNR for background H1L1 double coincident triggers during triple coincident time for the low chirp mass bin (blue), the medium chirp mass bin (red), and the high chirp mass bin (green). Higher chirp mass templates have a shorter time duration, which leads to a larger response to noise glitches, and thus a broader combined effective SNR distribution for background triggers.

### 7.1.4 Trigger Categories

Due to all of the above effects, we end up dividing the triggers from this search into 15 different categories during our three different observation times. During triple coincident times we have 3 coincidence types subdivided into 3 chirp mass bins, resulting in 9 trigger categories. During each of the double coincident times we have 1 coincidence type subdivided into 3 chirp mass bins, resulting in 3 trigger categories for each double coincident time.

## 7.2 FAR: False Alarm Rate

The FAR can be calculated using any intermediate statistic ranking the “loudness” of triggers (such as the combined effective SNR for this search), where a higher intermediate statistic corresponds to a larger excursion for background noise behaviour. The FAR for any trigger is calculated using

$$\text{FAR} = \frac{N}{\sum_i T_i}, \quad (7.1)$$

where  $N$  is the total number of background triggers with intermediate statistic greater than or equal to the one in question, and  $T_i$  is the amount of analyzed time in background trial  $i$ .

For individual trigger categories, the maximum FAR is found by

$$\text{FAR}_{\max,j} = \frac{N_j}{\sum_i T_i}, \quad (7.2)$$

where  $N_j$  is the total number of background triggers in the  $j$ th trigger category. This can be converted into a minimum inverse false alarm rate (IFAR) trivially by taking the inverse.

The expected number of triggers below a particular FAR due to background is  $\text{FAR} \times T_0$  where  $T_0$  is the foreground time analyzed. Another way to say this is the expected number of triggers above a particular IFAR due to background is  $T_0/\text{IFAR}$  where  $T_0$  is the foreground time analyzed, as seen in figure 7.6.

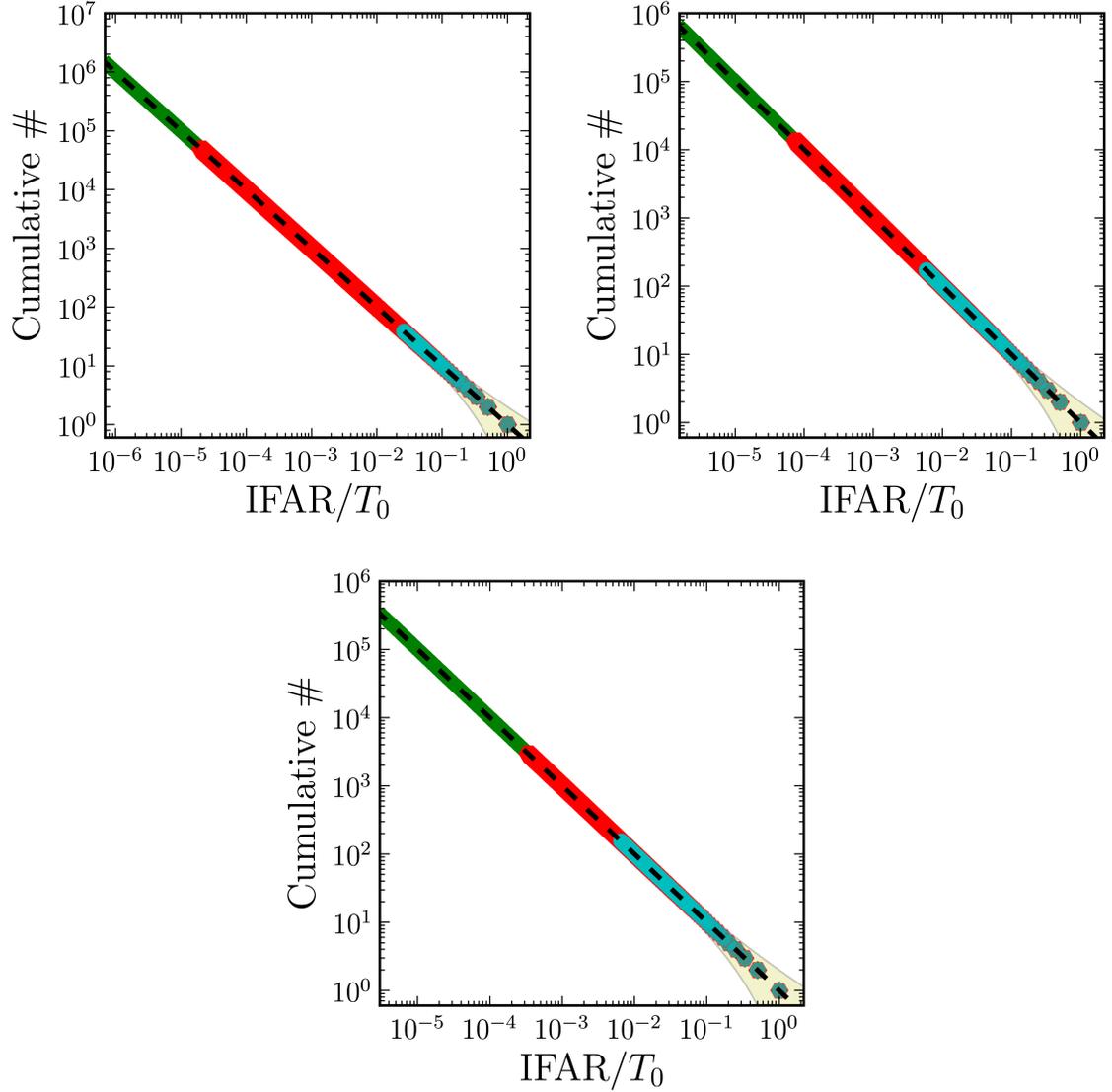


Figure 7.6: Background IFAR Plots

The cumulative number of triggers versus IFAR divided by the analysis time for the nine categories of background triggers in triple coincident time. The top left, top right, and bottom plots show the three different coincident trigger types for the low, medium, and high mass bins respectively. Because they are background triggers, they should, and do, fall exactly on the background line. The different category triggers in each plot all start at the lower right corner of the plot, but since they overlap, only the top color is shown.

### 7.3 FARc: Combined False Alarm Rate

When combining categories, we think of the different categories (coincidence type, trigger mass bin) as different trials. If there are  $m$  trials, then we expect that of those trials there will be  $m$  triggers with an IFAR  $\geq T_0$ . A simple way to normalize the FAR to bring the expected number of triggers with an IFAR of  $T_0$  back to one is

$$\text{FAR}' = m \times \text{FAR} . \quad (7.3)$$

This can be seen in figure 7.7.

The simple way of combining shows that, moving from right to the left, there is a kink in the background triggers whenever you reach a minimum IFAR for a trigger category. The reason these kinks occur is because once we reach a minimum IFAR for a trigger category, there is then one less “trial” for us to combine. A better thing to do at that point is to normalize the FAR by the remaining number of categories rather than the total. In analytic form, the FARc becomes

$$\text{FARc} = \left[ \sum_{j=1}^p \Theta(\text{FAR}_{\text{max},j} - \text{FAR}) \right] \times \text{FAR} + \sum_{j=1}^p [\Theta(\text{FAR} - \text{FAR}_{\text{max},j}) \text{FAR}_{\text{max},j}] , \quad (7.4)$$

where  $p$  is the number of categories of triggers and  $\Theta(x)$  is the Heaviside function. Calculating the FARc in this way can be seen in figure 7.8.

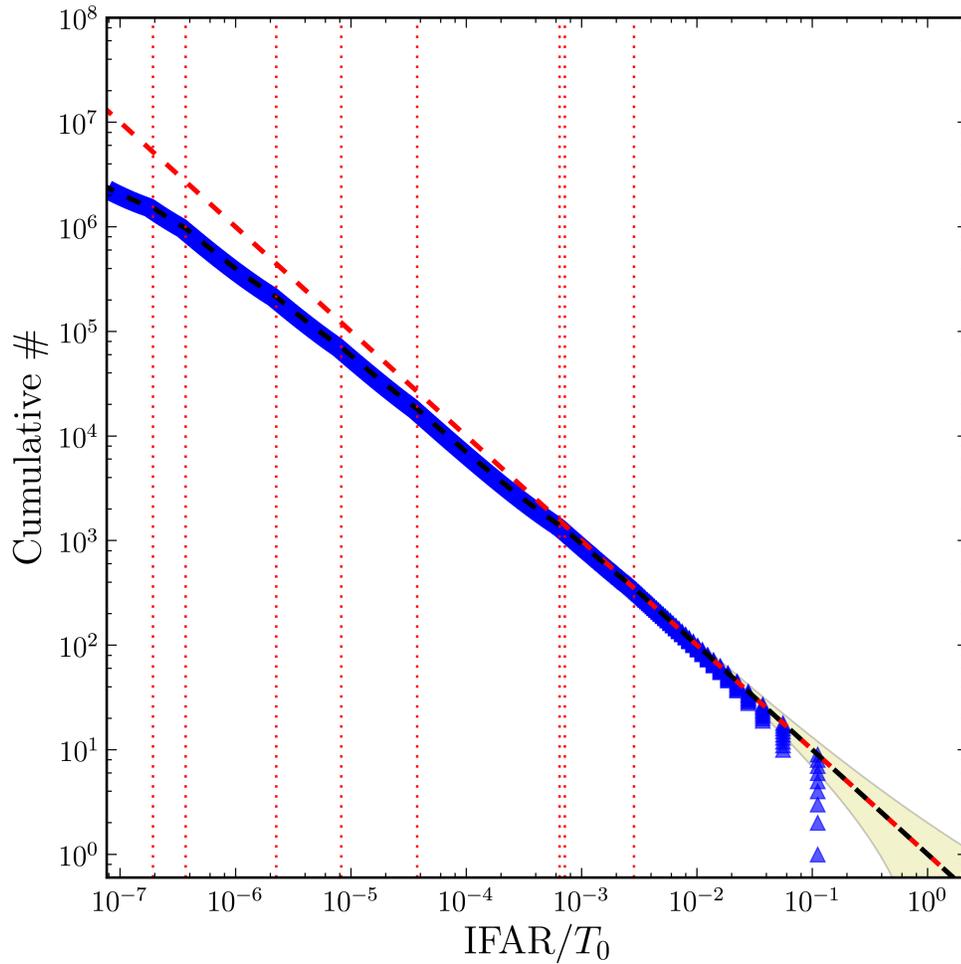


Figure 7.7: Combined Background IFAR Plot

The cumulative number of background triggers versus IFAR divided by the analysis time for nine categories of background triggers combined without correcting for the minimum IFAR of each category. As in the previous plot, because they are background triggers, they should, and do, fall exactly on the background line. The vertical, red, dotted lines denote the minimum IFAR for eight different trigger categories, which are the locations of the kinks in the background line. The ninth is not shown as that is where the background line terminates.

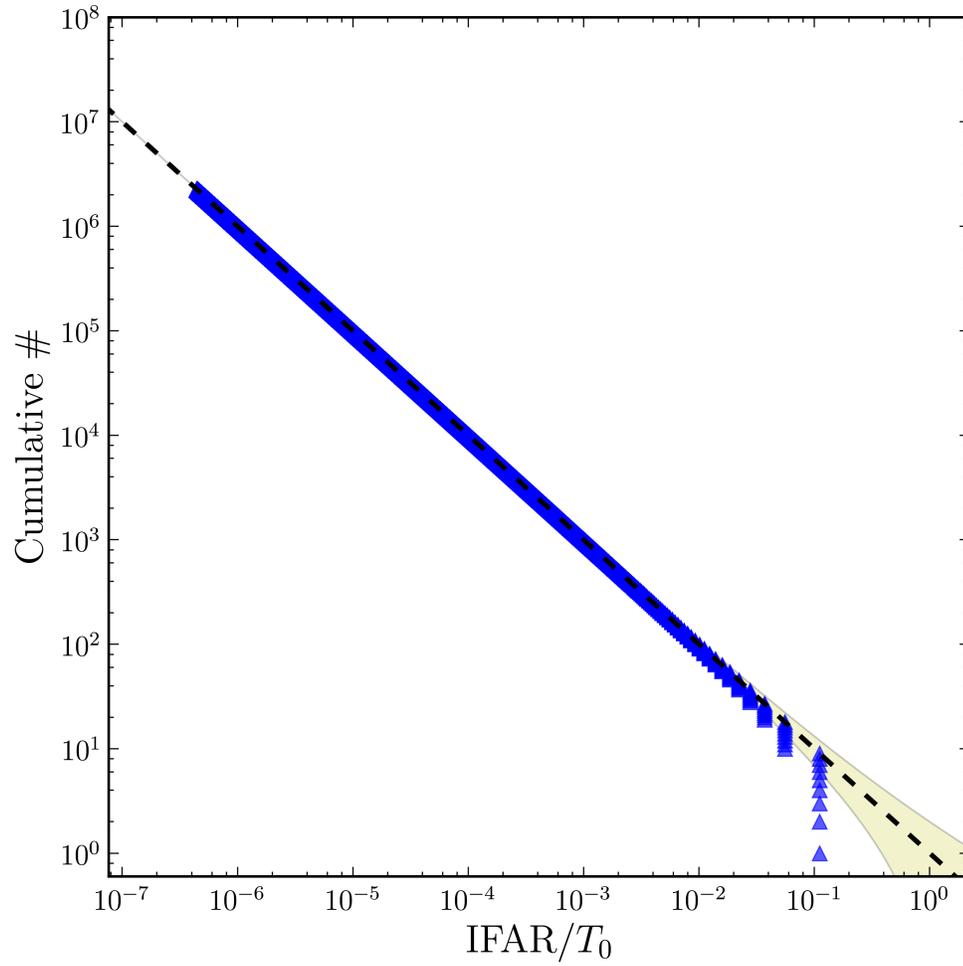


Figure 7.8: Background IFARc Plot

The cumulative number versus the combined IFAR (IFARc) divided by the analysis time for nine categories of background triggers. Again, because they are background triggers, they should, and do, fall exactly on the background line.

## 7.4 Comparison of Detection Statistics

Now that we have a method for calculating the FARc detection statistic, let us compare this statistic against the combined SNR and combined effective SNR in recovering software injections signals. This is done using a common comparison known as the Receiver Operating Characteristic (ROC) curve, shown in figure 7.9.

The FARc statistic out performs both the combined SNR and combined effective SNR for all mass regions. It is interesting to note that the largest improvement is seen for the low mass bin, and the least improvement in the high mass bin. This lends support to our assumption that the high mass templates have a higher false alarm rate.

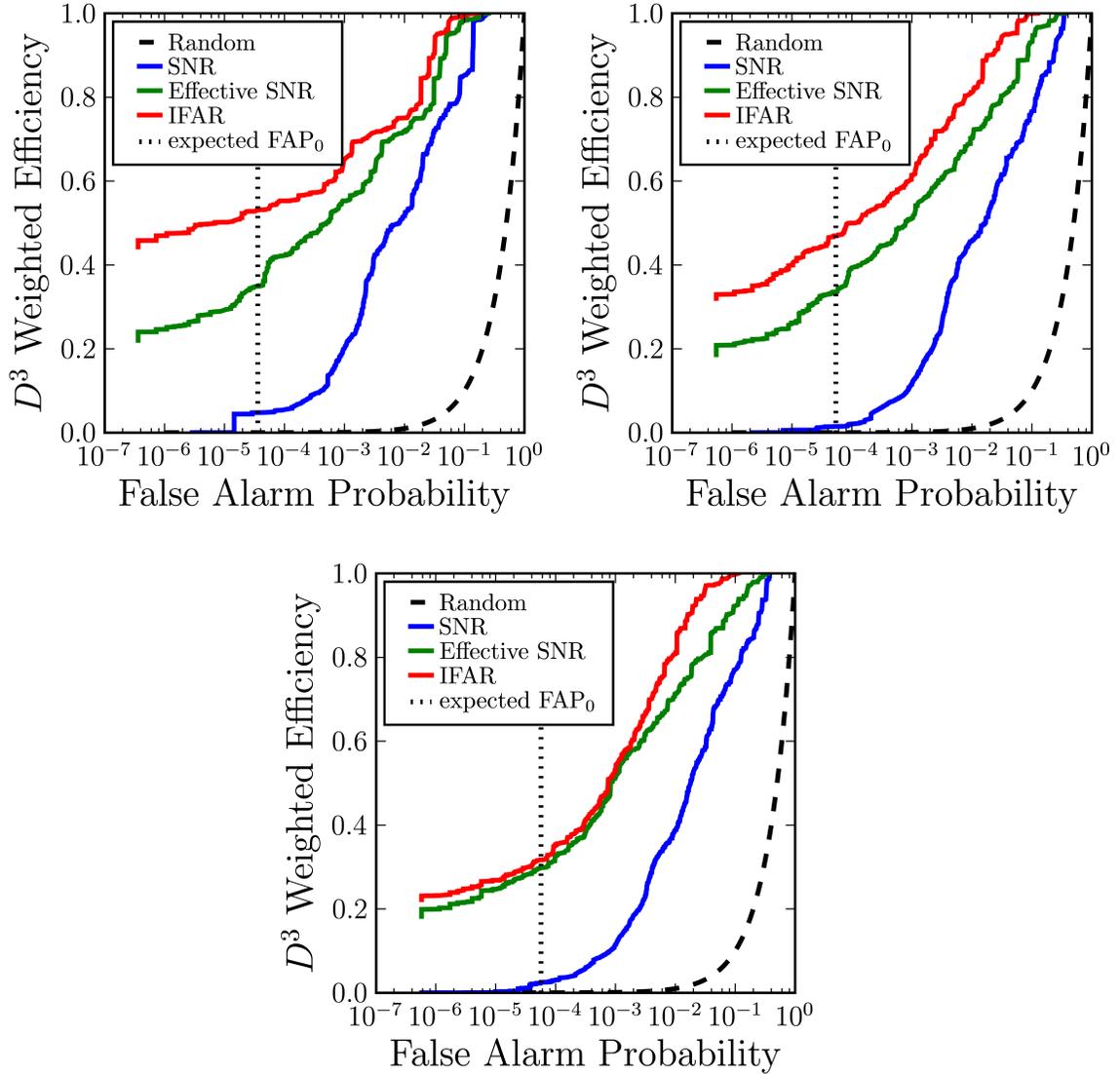


Figure 7.9: Comparison of Detection Statistics

We compare the efficiency of recovering injected signals versus false alarm probability (FAP) for three different detection statistics in the low (top left), medium (top right), and high (bottom) mass regions. The vertical line in each plot shows where the loudest in-time trigger is expected to occur when it is consistent with background. The injected signals were spaced uniformly in  $\log(D)$ . The plot has been made by reweighting each injection such that they represent a population uniform in the volume ( $D^3$ ).

## 7.5 Plotting Background Trials: Lightning Bolt Plots

In addition to plotting the IFARc of foreground triggers, we can also plot the IFARc of each of the individual background trials (i.e., time shifts). In this case, the IFARc has been normalized by

$$\text{IFARc}_i = \text{IFARc} \times \frac{T_0}{T_i}, \quad (7.5)$$

where  $i$  corresponds to the different background runs. An example plot of this is shown in figure 7.10.

The plots give 100 examples of the expected distribution of background-only triggers to better judge the significance (inconsistency with background) of the in-time coincident triggers.

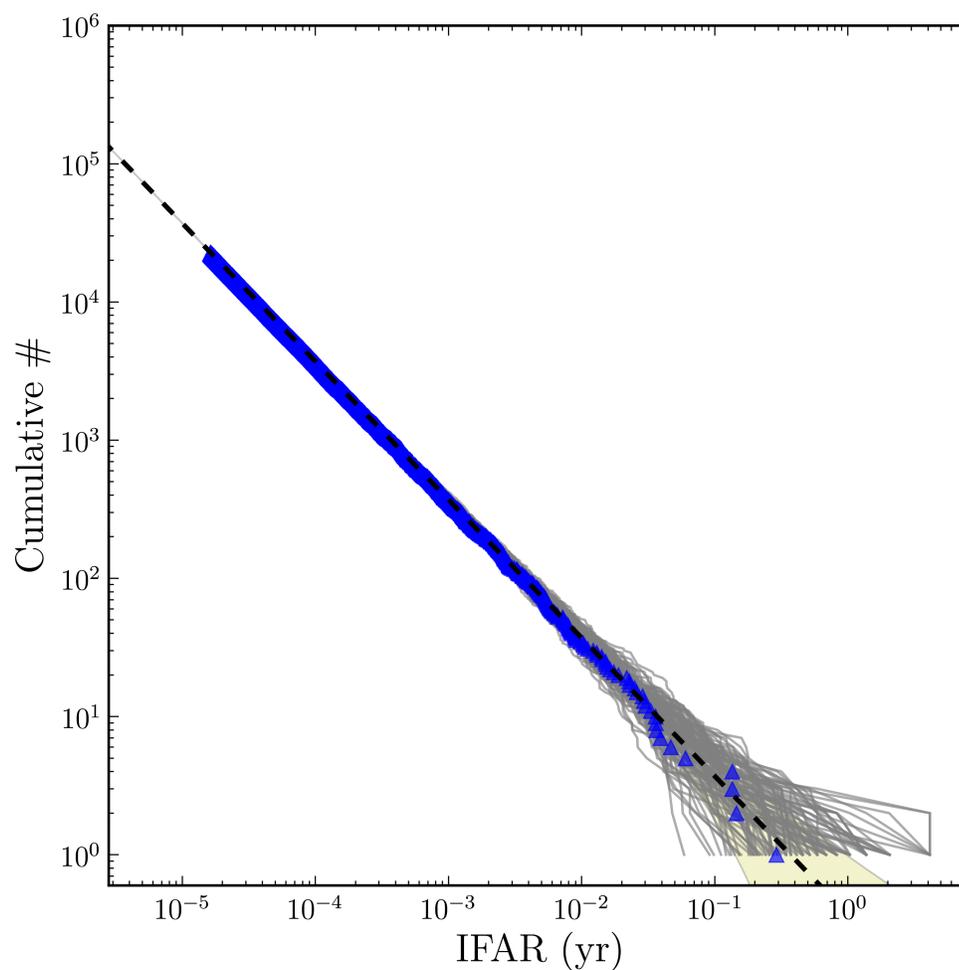


Figure 7.10: In-Time IFARc Plot

The cumulative of coincident triggers with IFARc greater than the value of IFARc on the  $x$ -axis for nine categories of foreground triggers from H1H2L1 observation time in this search. The dashed line shows the background distribution estimate averaged over all the time slides. The grey line give the distribution for each of the 100 background trial taken separately. The yellow band shows the  $N^{1/2}$  one-standard deviation errors on the average background estimate.

## 7.6 False Alarm Probability

Using the FARc as the detection statistic allows a very easy calculation of the false alarm probability (FAP) and the background probability  $P_0$  [130, 131]. Assuming a Poisson distribution, the FAP is the probability of getting any (zero) background triggers louder than the loudest zero-lag trigger (i.e., with a FARc lower than the FARc of the loudest zero-lag trigger since low FARcs are more significant) and the background probability is the probability of getting zero background triggers louder than the loudest zero-lag trigger, given by

$$\text{FAP} = 1 - e^{-x} \tag{7.6a}$$

$$P_0 = e^{-x} , \tag{7.6b}$$

where  $x = -\text{FAR} \times T$ , FAR is the FARc of the loudest zero-lag trigger, and  $T$  is the total analyzed time searching for zero-lag triggers. In-time coincident triggers with low FAP are GW detection candidates, and are subject to follow-up, as described in section 10.1.