GEO S1 Burst Search group report

GEO600 Burst search group :

Babak,S., Balasubramanian,R., Cadonati, L., Churches,D., Colacino,C.N., Cutler,C., Dupuis,R., Heng,I.S., Hewitson,M., Hough,J., Koetter,K., Krishnan,B., Malec,M., Mohanty,S.D. ,Mukherjee,S., Papa,M.A., Pitkin,M., Robertson,D., Sathyaprakash,B., Schutz,B.F., Sintes-Olives,A.M., Strain,K., Ward,H., Weiland,U., Wilke,B., Zhu,R.

AEI Golm and Hannover, Germany;
Cardiff University and Glasgow University, UK;
M.I.T., Cambridge, USA;
Universitat de les Illes Balears, Spain.

LIGO Tech Doc. No. T030180-00-0
August 22, 2003

1 Introduction

The GEO detector conducted its first science data run (S1) in coincidence with the three LIGO detectors. The overall duty cycle was 98.52 %. The longest locked stretch was approximately 121 hours. 44 channels were recorded. A section of data between GPS 715608013 and 715618813 second was chosen to be the 'playground' data for initial analysis. There is one out of lock period in this data stream , that between GPS 715611796-715611965 second. A G1-H1-H2 triple coincidence was recorded for 2116 seconds, of which 1800 seconds would be effectively used, given the present 360 sec quantization in LDAS jobs: 715609677-715611793 s. L1 however was never in science mode during this time.

2 Search pipeline

2.1 Excess Power

2.1.1 Description

This pipeline [1] implements the LAL Excess Power DSO in a stand-alone mode. The search pipeline consists of code to (1) read data from frames, (2) fill the appropriate structures specified in the LAL wrapper, (3) call the four DSO functions (InitSearch, ConditionData, ApplySearch, FinalizeSearch) and (4) extract the triggers from the LAL wrapper output structure. At present only the Excess Power DSO has been included in this pipeline. Of the four DSO functions, ApplySearch was modified to remove all MPI related code. No LAL functions were touched in the modifications. The output trigger exists as a set of ascii files and has a form identical to the snglBurst table of LIGO.

2.1.2 Validation

An initial validation of the pipeline is done by comparing the output triggers with a spectrogram by eye. An example of such a spectrogram on which the triggers from the search pipeline are superimposed can be seen in figures 1. The search pipeline output coincides with the triggers seen by eye in the spectrogram. This has

Parameter	Entry
numPoints	65538
numSegments	8
ovrlap	32769
overlapFactor	3
minFreqBins	4
minTimeBins	4
flow	1000.0
deltaF	2.0
length	1024
numSigmaMin	3.0
alphaDefault	0.5
segDutyCycle	1
alphaThreshold	1.0e-10
events2Master	5
channelName	G1:DER_H_HP-EP
simType	0
specType	useMedian
winType	2

Table 1: List of Excess Power input filter parameters.

been tested over long stretches of data and have been found to be consistent. *Due to technical reasons, this algorithm was not run on the whole of GEO S1 and thus, events generated from this pipeline has not been used for further coincidence analysis with LIGO events.*

2.1.3 Input parameters

The input filter parameters used in the pipeline that processed the GEO S1 playground data are given in table 1.

2.2 GEO's implementation of the TFCluster : HACR (Hierarchical Algorithm for Curves and Ridges)

2.2.1 Description

The basic idea of the GEO++ [2] TFCluster monitor is to construct a spectrogram and identify regions in the spectrogram where an excess power is observed. The algorithm should be suitable for detection of short bursts typically a few milliseconds upto about a few tenths of a second and with a frequency resolution of about a few 10s of Hz. The algorithm is implemented as a monitor in the GEO++ environment (and recently renamed "HACR") under which GEO online detector characterization tasks will be carried out. This is a non-LAL implementation.

The algorithm to identify the regions or "clusters" is mostly similar to the one employed by Sylvestre in his tfcluster idea. The main difference is that in the implementation of HACR [3] pixels diagonally adjacent are also used to connect pixels into clusters. There is also no attempt as in Sylvestre's tfcluster algorithm to identify "generalized clusters". We use dual thresholds to determine clusters. The higher threshold is used to determine cluster starting points and the lower threshold is used to determine adjacent excess power pixels. This makes the algorithm more robust. The identification of black pixels in Sylvestre's algorithm is computed using the distribution function of the pixel amplitudes. We use a somewhat simpler strategy. We compute the variance of the pixel amplitudes after throwing away outliers. We compute the ratio of the deviation of the pixel value from the mean to the variance and compare this with a user specifiable threshold. The clusters so obtained are then classified as curves, glitches, lines, or simply blobs.

2.2.2 Validation

The thresholds for HACR were set using synthetic Gaussian and stationary noise. The thresholds were tuned to obtain a false alarm rate of about 2 events in every three hours. The monitor was then run on GEO S1 data, GEO injection period data (where Inspiral signals were injected using hardware) and a stretch of 3.5 days of S1 data. The algorithm was found sensitive enough to capture many of the injected inspiral events and could also classify these events as curves in the time frequency plane. The event rate observed for GEO S1 data is about 2 per second where the events are generated between a 100 and 3000Hz.

The events are available as a set of ascii files.

3 Data exchange with LIGO

In a data exchange exercise with LIGO, GEO playground data (calibrated h(t) channel) was also successfully passed through LDAS DSOs viz. Slope and TFClusters at MIT.

Two sets of TFCluster triggers were produced, corresponding to the following: G1tfcluster11: p=0.008 ; average 1Hz rate and G1tfcluster13: p=0.16 ; average 3Hz rate, p being the black pixel probability (probability that a pixel in the time-frequency plane was accidentally above threshold). For each event trigger, the GPS start time, the duration (seconds), the central frequency and the bandwidth (Hz), the number of black pixels in the cluster (labeled "amplitude") and the average power per pixel in the cluster (labeled "SNR") were listed. The power of an event was defined to be the product of "amplitude" and "SNR".

The Slope parameters were the same used in LIGO, chosen to have optimal sensitivity in the 400-500 Hz range. Threshold was tuned to obtain a manageable rate (chosen by hand). This parameter needs further tuning, either via simulations or by looking at the amplitude histogram.

The results were imported as xml files. An xml to ascii parser has also been made available by LIGO.

The results from the Slope and TFCluster are summarized at [4].

4 Veto

The veto scheme uses data from the channels G1:LSC_MID_EP-P_HP and G1:LSC_MID_EP-Q_HP (high power detector error point in phase and quadrature respectively). The ratio of the power of a signal in P (P_{cal}) to one in Q (Q_{cal}) is determined continuously from the calibration peaks i.e. $R_{cal} = P_{cal} / Q_{cal}$. If a glitch is detected in P (P_g), the ratio (R_g) of the power in P and Q (Q_g) is detected at this time. The duration of the glitch can be vetoed if R_{cal} differs significantly from R_g. Due to measurement errors associated with calibration peaks and glitches, a limit is set on R_g. The veto pipeline runs within the GEO detector characterization pipeline (GEO++) as a monitor (PQMon) [6].

By design, the PQMon can only directly veto events produced by a specific glitch finding algorithm that is coded *inside* itself. Hence, a modified scheme of vetoes is proposed by Heng [5] that can veto events from the output of the search pipelines.

In this scheme, the number of coincidences between vetoed PQMon glitch times for a glitch window of 10 samples and search pipeline (e.g. Excess Power or modified TFCluster) event times is calculated. Also, the PQMon vetoed glitch times are time shifted and the number of coincidences counted at these times shifts to determine the level of the "accidental" coincidence background. Two triggers with times t1 and t2 are considered to be coincident if

$$|t1 - t2| < dt \qquad (1)$$

where dt is the temporal range over which one searches for coincidences. dt is referred to as a "coincidence window".

For the analysis of GEO S1 data, a coincidence window of 0.2 second was chosen. The number of vetoed TFCluster events was determined by subtracting the mean number of accidental coincidences from the peak at 0 time shift.

The results described above are summarized at [5].

With the present S1 data from the GEO detector, a list of useful veto channels could not be identified. Thus, GEO could not use the knowledge from the auxiliary chanels to veto events from the h(t) channel. This was however a very educative lesson in the sense that GEO could identify the need for higher sampling rates and much larger signal levels for many of the channels that could serve as reliable veto channels.

5 Vetoed list of events

The EP search pipeline has run on six days GEO S1 data and has been vetoed using the above veto scheme. The list of vetoed events from the GEO burst search pipeline has been made available to LIGO. Figure 2 shows the distribution of the duration, bandwidth, central frequency and snr for a section of vetoed event list.

In addition, the same veto scheme has also been applied to the output of HACR events and the vetoed list of events from this has also been made available to LIGO. Figure 3 shows the distribution of the duration, bandwidth, central frequency and snr for one section of vetoed event list.

Note that in making the SNR histograms, some of the outliers has been rejected.

In figure 4, the histogram of time differences between consecutive HACR events is plotted. The graph is interesting since it shows that there are periodicities in the occurrence of events. The fundamental periodicity seems to be about .11 seconds or 110 milliseconds. There is of course a huge peak at 0 since there can be many events which are at the same time but different frequency bands. There are also many more peaks at multiples of fundamental 110 milliseconds beyond the range of the figure. This could be attributed to presence of fast non-stationarities throughout the GEO S1 data. In an example with data from G1:LSC_MID_EP-P_HP after bandpass filtering in the band 1600 to 2400 Hz, one sees the "periodic" glitches with period 0.1 sec [5]. In general, in most of the GEO S1 data, this fast non-stationarity is present in the 0 to 1500 Hz band. However there appear to be some "genuine" transients amidst the non-stationarity. Note the 0.1 sec periodicity bumps are also present in the specgram. It remains to be resolved whether the "genuine" glitches are offshoots of the bump glitches or are they from a separate class.

6 Injection and Simulation

6.1 Introduction

To distinguish between genuine burst gravitational wave events and spurious environmental excitations, a search for coincident events is performed on candidate event lists from two or more widely-spaced gravitational wave detectors. Since different detectors have different responses to incoming burst gravitational waves, signals must be injected (via both hardware and software) into the data and passed through the analysis pipeline to determine not only the detection efficiency of the single detector but also that of the network of detectors. Examining the arrival times of the injected signals that have been passed through the analysis pipeline allows one to determine the uncertainty in the estimate of the signal arrival time. This quantity is then used to determine the width of the time window within which events are considered to be coincident.

6.2 Injections into GEO S1 data

Sine gaussians with 4 central frequencies were injected post-acquisition via software into data from the GEO S1 run and passed through the HACR pipeline [7]. A plot showing the locations of the central frequencies of the sine gaussians is shown in figure 6. One should note that during the S1 run, the GEO600 noise spectrum was highly non-stationary below 1.5kHz. We expect this non-stationarity to increase the uncertainty in estimates of the amplitude and time of arrival of the observed signal. These signals, lasting about 100ms, were injected every 18 seconds into the data. The corresponding HACR SNR for each series of injected sine gaussians fluctuates due to interactions between the injected signal and the detector noise. The injected signal responses were identified in the HACR output by choosing the strongest event within a ± 0.1 second window about the signal injection time. This step was necessary due to the presence of background events identified by HACR close to the injection time.

The HACR SNR of events identified injected signals are plotted in figure 7. The error bars correspond to 1 standard deviation in the distribution of the observed SNR. For sine gaussians with central frequencies 2419Hz and 3927Hz, the standard deviation is about 5% of the mean SNR. The standard deviation is proportionally greater for sine gaussians with central frequencies of 529Hz ($\sim$ 55%) and 1423Hz ($\sim$ 15%). This is most likely due to the presence of non-stationarities in the noise at those frequencies. Most of the injected signals had a very high HACR SNR and were easily identified. However, for injected sine gaussians with central frequencies of 2419Hz, the observed HACR SNR is less than or equal to that of background events in a ± 0.1 second window for signal amplitudes smaller than 5×10^{-17}. The presence of these background events introduced an uncertainty in the HACR SNR estimation of the injected signal. The same uncertainty is observed for 529Hz sine gaussians with amplitudes less than 10^{-16}. The region of uncertainty is represented by the rectangles in figure 7.

The uncertainty in the arrival time estimate can also be determined through signal injections. The standard deviation of the difference between the arrival time as estimated by HACR and the time when the sine gaussian was actually injected is plotted in figure 8. The overall uncertainty increased as the central frequency of the injected sine gaussian decreased. As with the observed HACR SNR, the presence of background events introduced an uncertainty in the arrival time estimation. Taking the upper value of the for the 2419Hz injected sine gaussians, we can limit the minimum size of the coincidence window for amplitudes of 2×10^{-17} or greater to 0.05 seconds. This same minimum bound on the coincidence window should be applied to 529Hz sine gaussians with amplitudes greater than or equal to 10^{-16}

6.3 Summary and discussion

The analysis of sine gaussian injections into GEO600 S1 run data confirms the presence of non-stationarity in frequencies below 1.5kHz. There were greater fluctuations in the observed HACR SNR for the 529Hz and 1429Hz sine gaussians than that for the 2419Hz and 3927Hz sine gaussians. A minimum bound for the width of the coincidence window has been set at 0.05 seconds.

Acknowledgement : The GEO Burst group members thank the LIGO BUL members for the cooperation extended at all levels, and for help in passing GEO playground data through LDAS.

References

[1] http://www.aei.mpg.de/$\sim$ mohanty/EPBURSTS/epbursts.html

[2] http://www.astro.cf.ac.uk/pub/R.Balasubramanian/geo++/index.htm

[3] www.astro.cf.ac.uk/pub/R.Balasubramanian/geo++/godcsResults/newtfcluster/tfcluster.html.

[4] http://emvogil-3.mit.edu/ cadonati/S1/GEO/GEO-MIT.html

[5] http://weber.astro.cf.ac.uk/cgi-bin/burstnote.pl pp 4

[6] K. Koetter, I.S. Heng, M.Hewitson, K.A. Strain, G. Woan and H. Ward, Classical and Quantum Gravity,**20**, S895 2003.

[7] Heng, I.S., Amaldi Conference, Pisa, 2003 and Classical and Quantum Gravity, Submitted, 2003

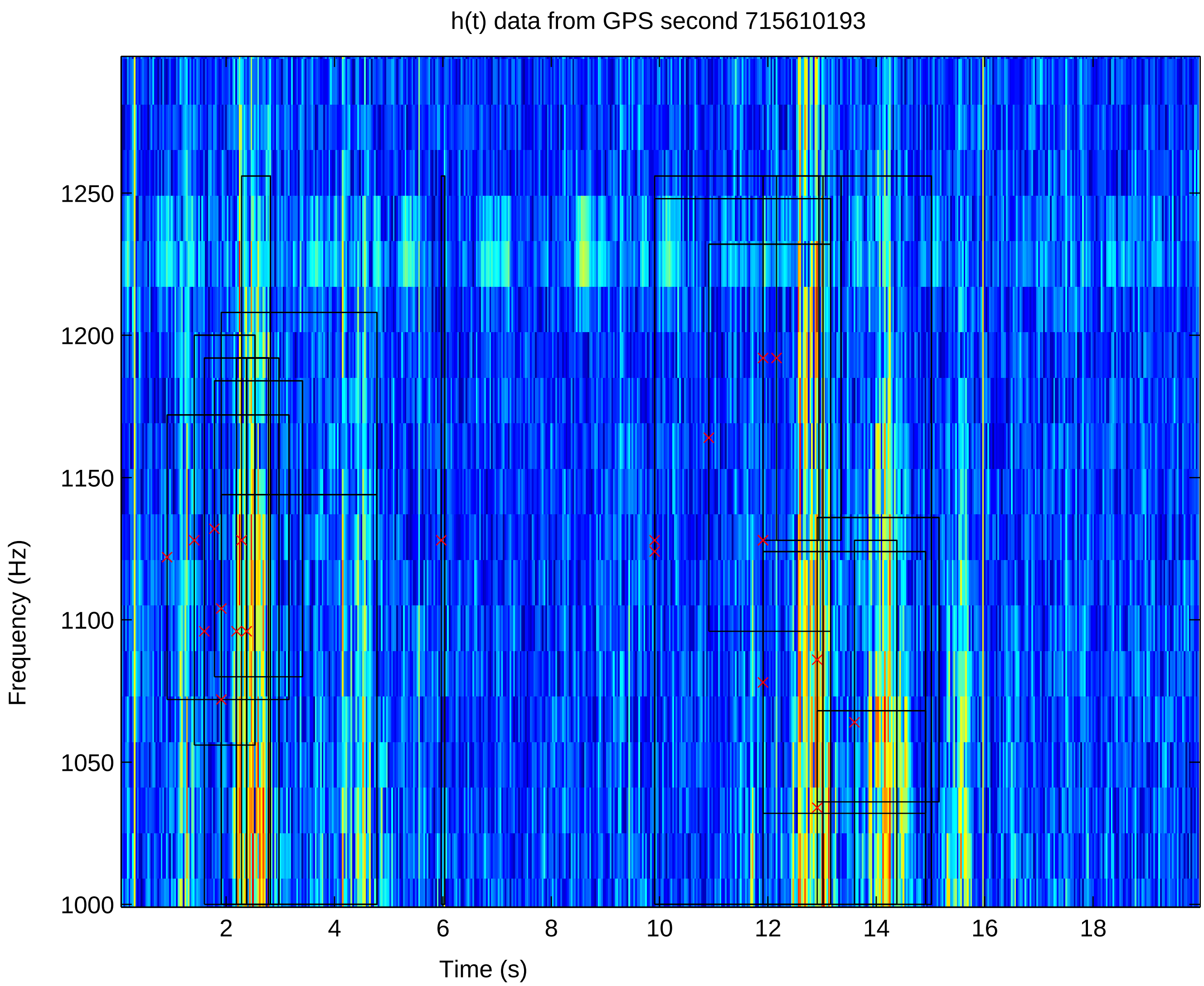

Figure 1: This figure shows the triggers from GEO's search pipeline that ran on the playground data set superimposed on a spectrogram. The search pipeline output coincides with the triggers seen by eye in the spectrogram. The red crosses correspond to the start time and central frequency of each trigger.

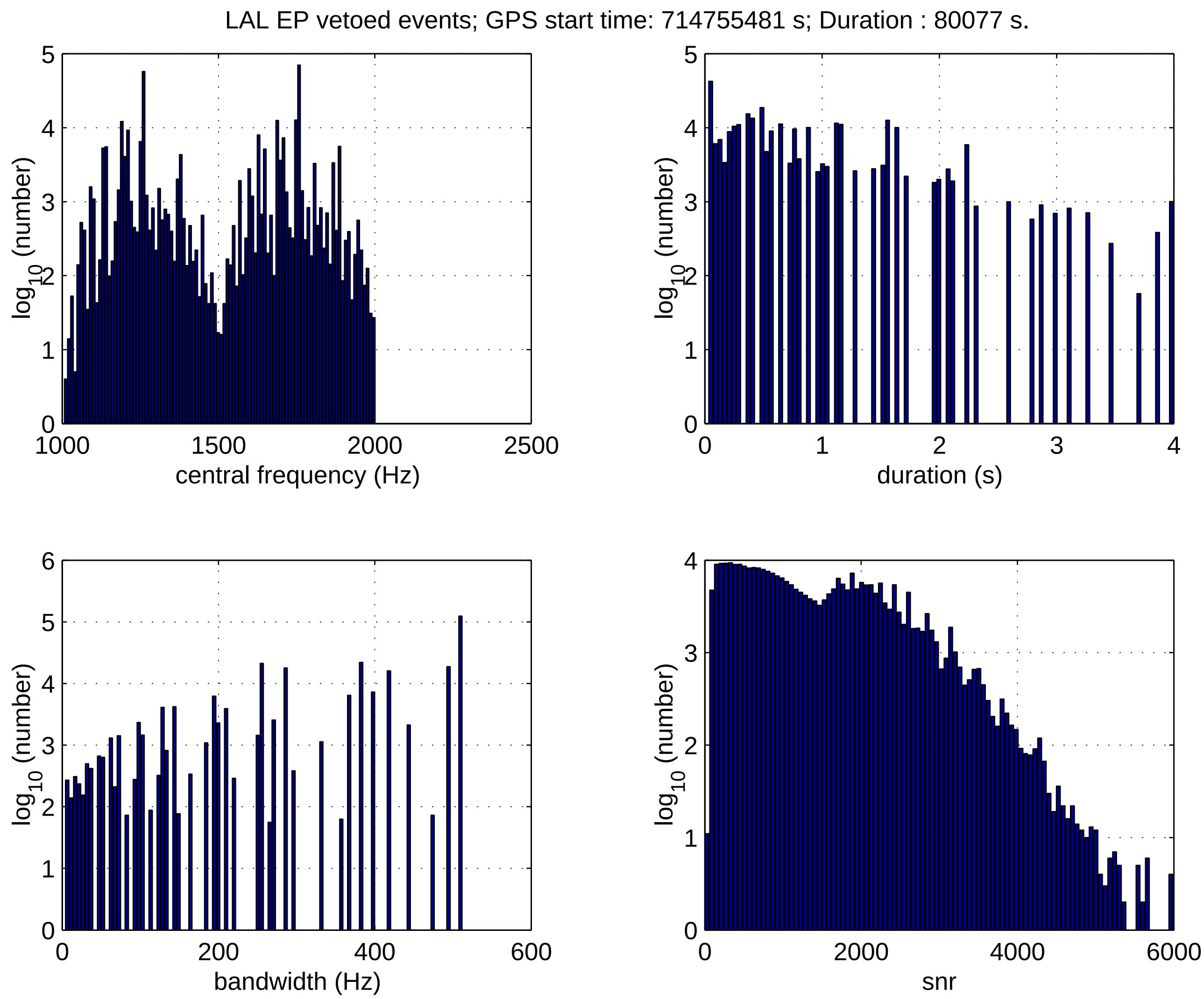

Figure 2: This figure shows the histogram of the trigger duration,bandwidth, central frequency and SNR from the vetoed list of events that has been generated by the LAL Excess Power pipeline on 6 days of GEO S1 data from GPS 714755481 s for a duration of 80077 s.

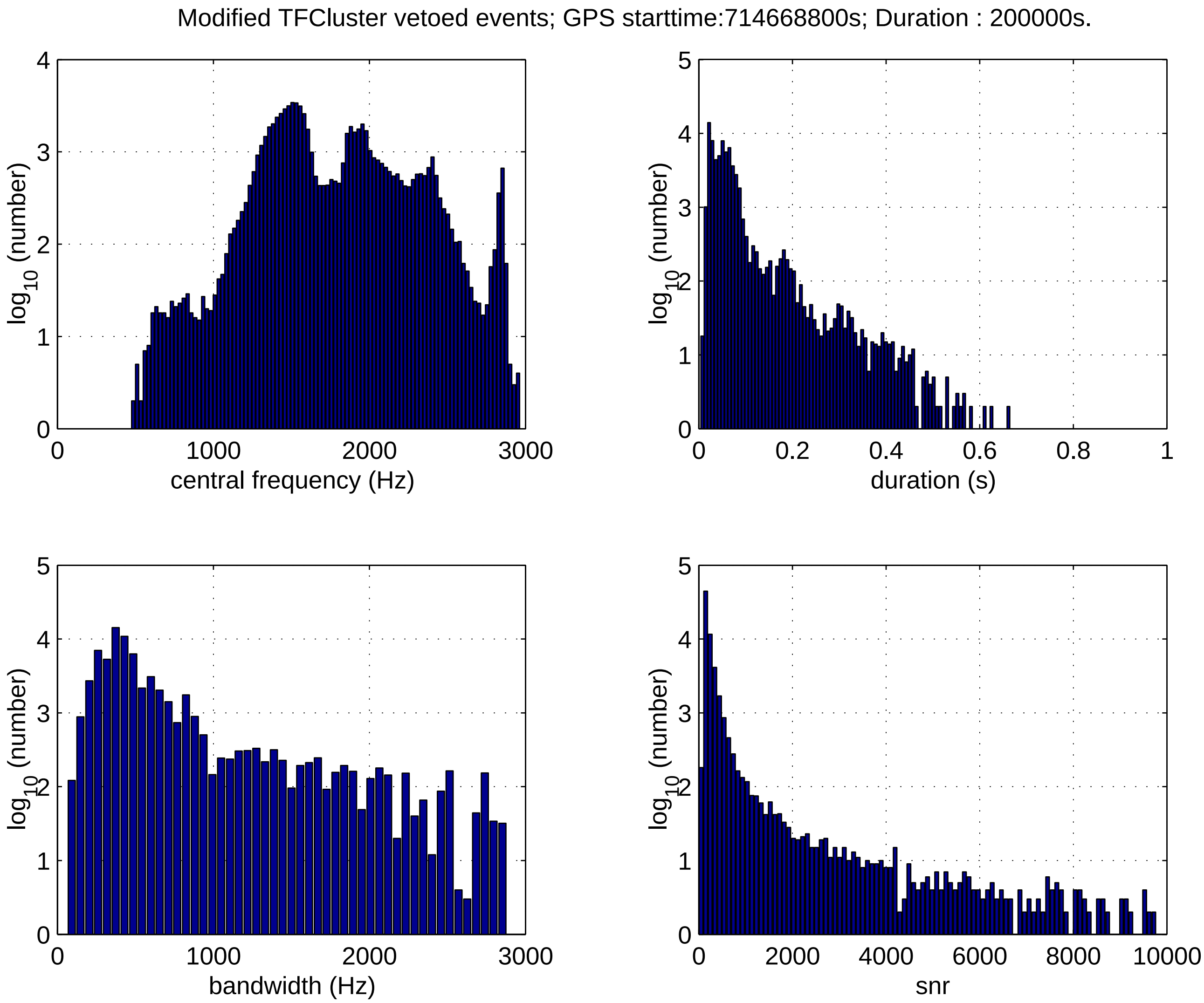

Figure 3: This figure shows the histogram of the trigger duration, bandwidth,central frequency and SNR from GPS 714668800 s for a duration of 200000 s from the vetoed list of events that has been generated by HACR pipeline on entire GEO S1 data.

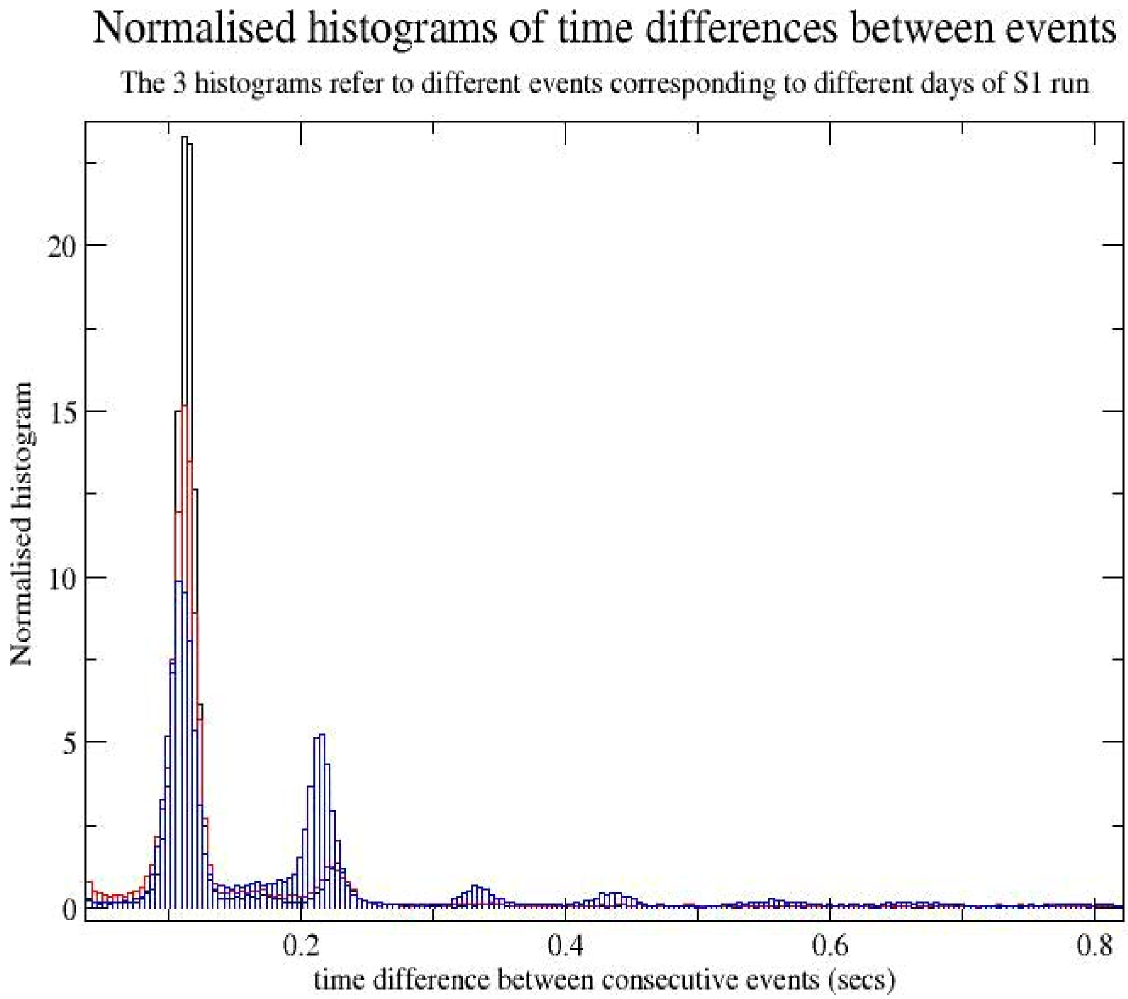

Figure 4: In this figure the histogram of time differences between consecutive events is plotted. The graph is interesting since it shows that there are periodicities in the occurrence of events. The fundamental periodicity seems to be about .11 seconds or 110 milliseconds. There is of course a huge peak at 0 since there can be many events which are at the same time but different frequency bands. Also there are many more peaks at multiples of fundamental 110 milliseconds beyond the range of the diagram.

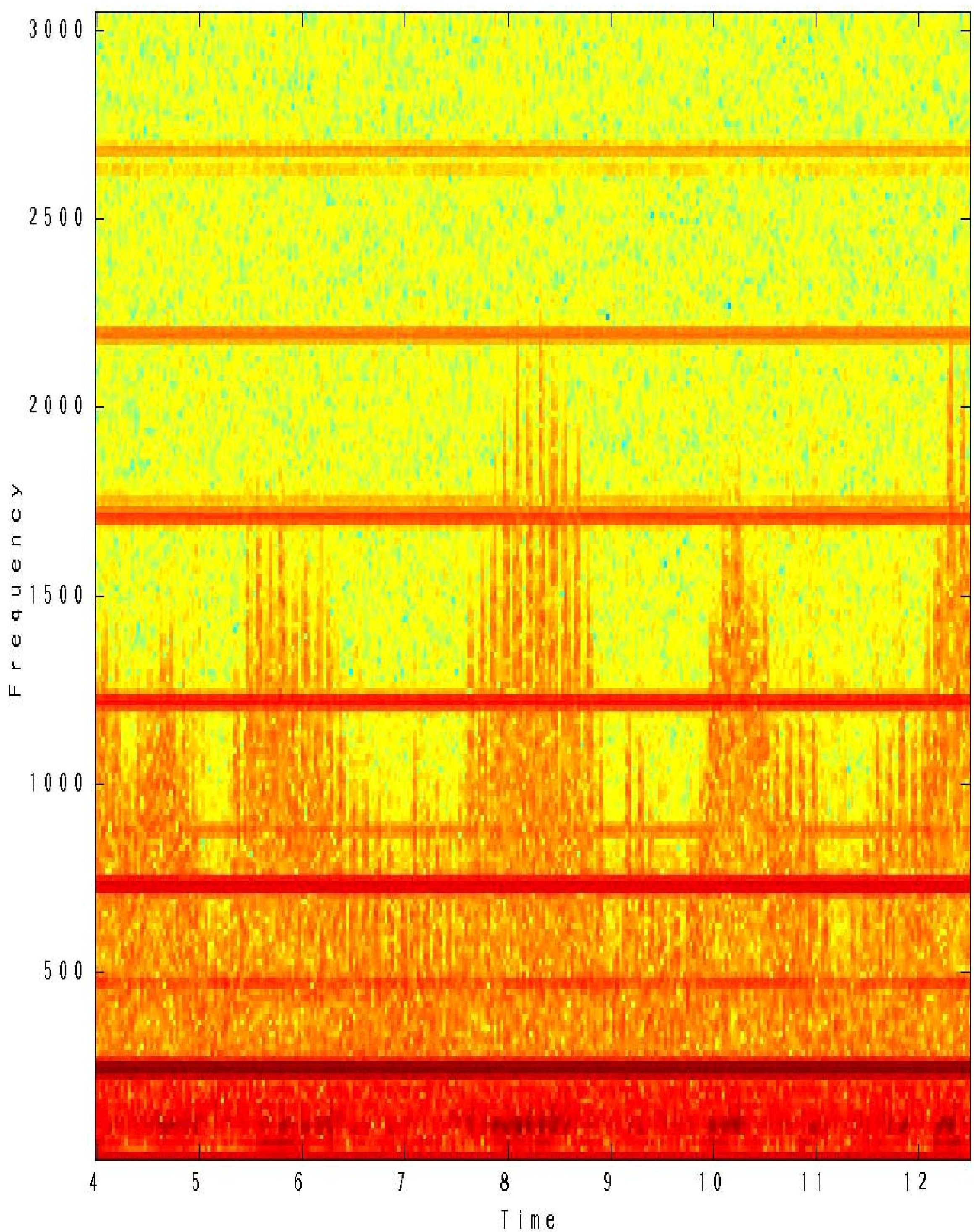

Figure 5: In general, in most of the GEO S1 data, this fast non-stationarity is present in the 0 to 1500 Hz band. Note the 0.1 sec periodicity bumps are also seen in the spectrogram.

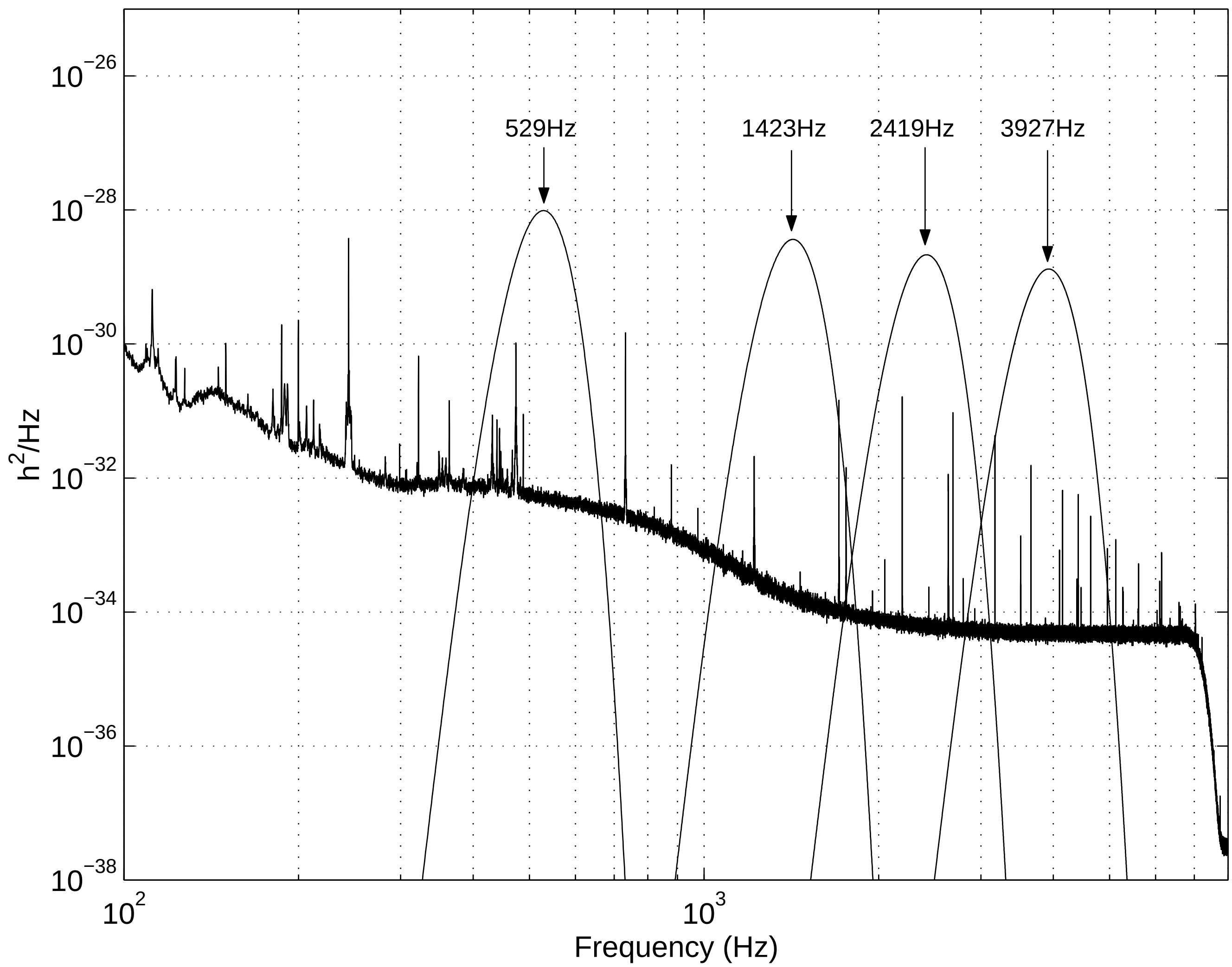

Figure 6: The power spectral densities of GEO600 S1 data from GPS second 715608013 to 715609013. The square of the fourier transforms of the sine gaussian signals are also plotted on the same graph.

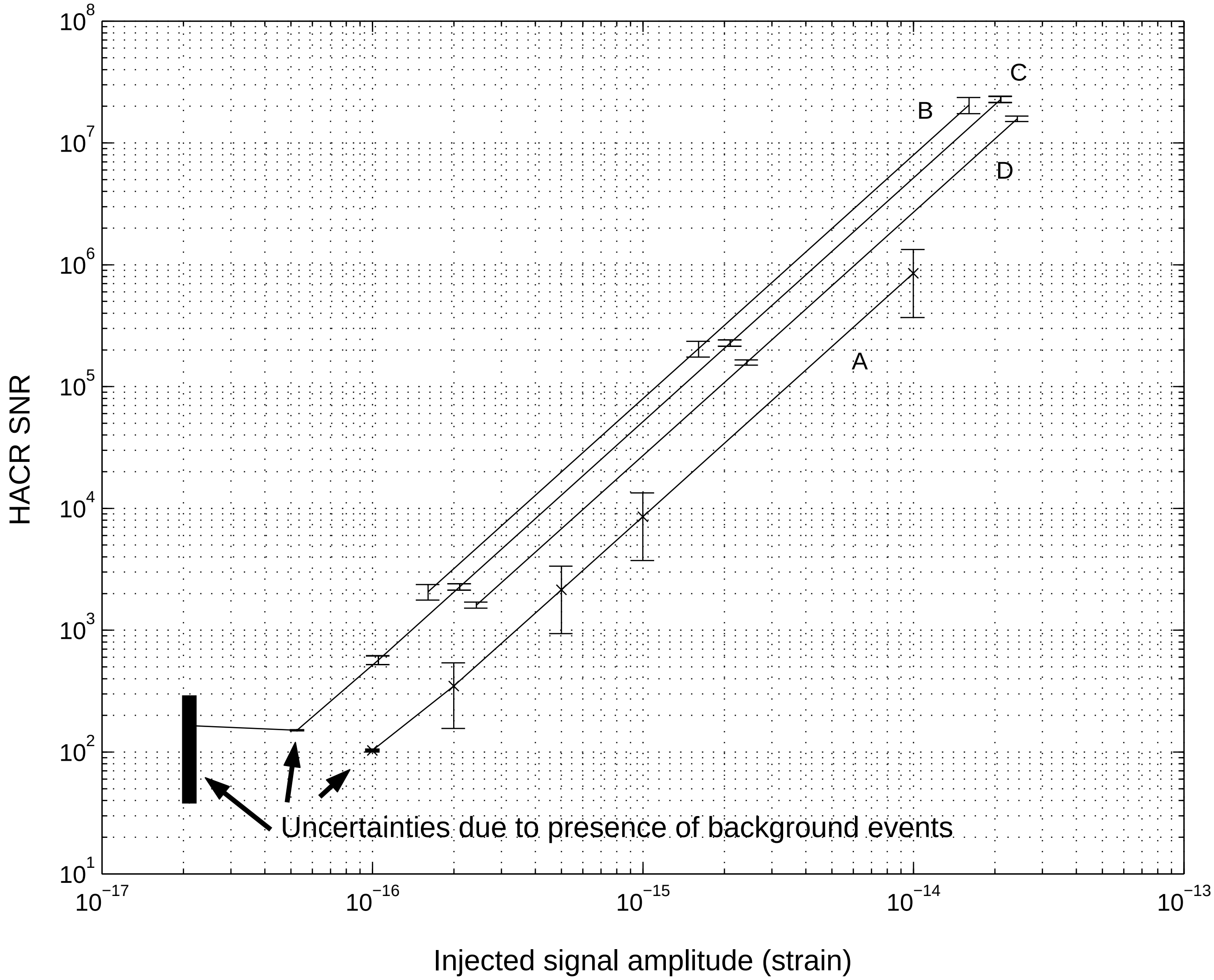

Figure 7: The observed HACR SNR against the amplitude of the injected sine gaussians with 4 different central frequencies. At lower SNRs, the presence of background events within the 0.1 sec window creates an added uncertainty in the estimation of arrival times as shown by the rectangles.

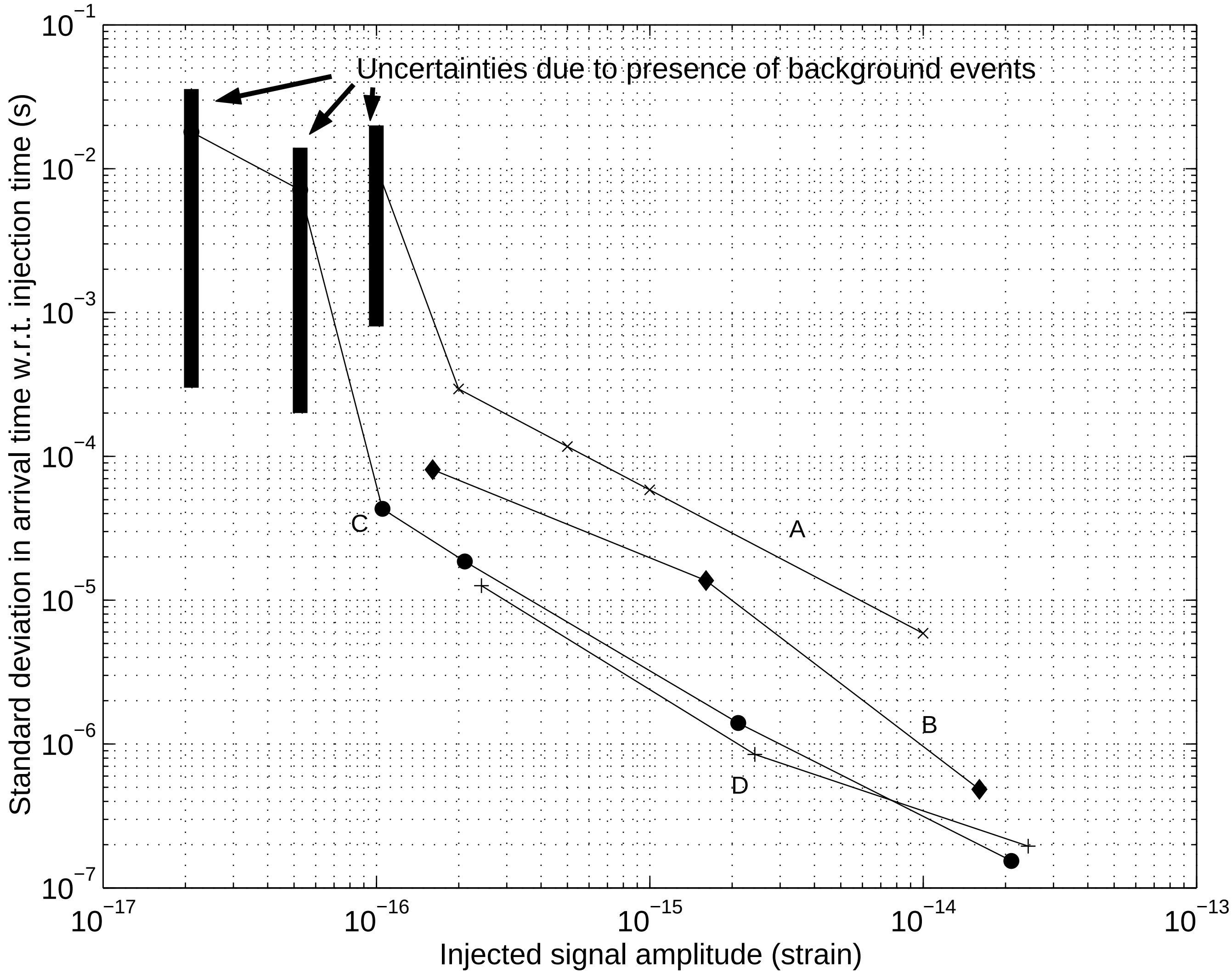

Figure 8: The uncertainty in the estimate of the arrival time with respect to the signal injection time for sine gaussians with central frequencies (A) 529Hz, (B) 1423Hz, (C) 2419Hz and (D) 3927Hz.