Transient noise in LIGO E2 data

Julien Sylvestre
LIGO Project, Massachusetts Institute of Technology,
NW17-161, 175 Albany St., Cambridge, MA 02139, USA.
julien@ligo.mit.edu
(March 12, 2001)

INTRODUCTION

Results of an investigation of the transient noise in LIGO E2 data are presented here, together with the statistical methods employed. The focus of this report is on the analysis methods, rather than the results themselves, in an attempt to learn from preliminary data what the real challenges of dealing with the actual data will be.

I. METHODS

The analysis was divided into three distinct parts: (i) the detection of transients, (ii) the case-by-case detection of correlations between PEM and IFO channels, and (iii) the statistical detection of correlations for populations of transients. The following three subsections will describe the algorithms used in each part of the analysis.

A. Detection of transients

The detection, and to a certain extend the estimation of the parameters of transients in the data, is performed by a time-frequency cluster detector. For a given channel, the spectrogram is constructed in the usual manner by stacking power spectra computed over non-overlapping stretches of duration T. The spectra are simply approximated by the periodograms of the stretches, i.e. by the norm square of their discrete Fourier transforms. If the noise was Gaussian white, and no signal was present, the power at every pixel of the spectrogram would be independant of the power at every other pixel, and for every constant frequency row of the spectrogram the distribution of the power would be

$$p_{P(f)}(P) = \frac{e^{-P/P_0}}{P_0}, \tag{1}$$

where P_0 is a constant scale factor. If the noise is well-modelled by a colored Gaussian noise, then the independence conditions in time and frequency hold approximately if the auto-correlation time-scale of the noise is much smaller than T. In the present analysis, the noise was modelled by fitting a Rice distribution to every constant frequency row of the spectrogram:

$$p_{P(f)}(P) = \frac{1}{P_0(f)} \exp\left(-\frac{P+Q(f)}{P_0(f)}\right) I_0\left(\frac{2\sqrt{PQ(f)}}{P_0(f)}\right), \tag{2}$$

where I_0 is the modified Bessel function. This simply amounts to assuming independence in time and frequency, and to allow for the presence of a deterministic signal (i.e. constant power) at any frequency. The fit was done by maximizing a likelihood function numerically. Details are beyond the scope of this paper, but the shape of the log-likelihood function was simple enough to allow the use of a fast minimum-finding algorithm that was always converging to a global minimum.

Equation 2 was used to compute a frequency dependent threshold $P_*(f)$ on the power, using

$$\alpha = \int_0^{P_*(f)} p_{P(f)}(P) dP, \tag{3}$$

where α is the confidence level. A new representation of the data, the thresholded spectrogram (hereinafter TS), was formed by simply thresholding on the power; pixels with excess power were labelled 'black pixels', others were labelled

'white pixels'. Consequently, $1-\alpha$ is the black pixel probability, i.e., the probability for any pixel to be black when no signal is present.

The final step in the detection was the search for significant clusters in the TS. Loosely speaking, noise has a tendency to fill the TS uniformly with black pixels, leading to an almost exponential distribution of the number of clusters as a function of size. A threshold on the clusters size seems to be a good way to detect signals which are expected to form clusters in the time-frequency plane. This argument is made more precise by first defining clusters to consist of connected sets of nearest neighbors black pixels in the TS (pixels touching by a 'side'), and then by computing the expected number of clusters of size s per pixel, where s is the number of pixels contained in the cluster:

$$\langle n_s(p)\rangle = p^s D_s(p). \tag{4}$$

Here, p is the black pixel probability, and D_s is the perimeter polynomial:

$$D_s(p) = \sum_t g_{st}(1-p)^t, \tag{5}$$

where g_{st} is the number of shapes a cluster of size s and perimeter t can have. If only clusters of size $s \geq \sigma$ are considered significant, the false alarm rate is then simply given by

$$\mathrm{FAR} = B \sum_{s'=\sigma}^{\infty} \langle n_s\rangle, \tag{6}$$

where B is the bandwidth of the TS.

In general, it may be desirable to extend the definition of clusters to allow for more general configurations which could be expected for some types of signal. In particular, two clusters of size smaller than the threshold σ may be unlikely to occur in noise due to their proximity in the time-frequency plane; thresholding on the distance in addition to the size often leads to a better detection efficiency at fixed false alarm rate for signals that tend to form 'diagonal' clusters. The complete presentation is again beyond the scope of this paper, but it is possible to compute mainly by computer enumeration the function $\langle \nu_{s_1,s_2}(d)\rangle$, which is the average number per pixel of pairs of clusters of size s_1 and s_2, separated by a distance[1] d. A set $\{\delta_{s_1,s_2}\}$ of thresholds on the distance is defined, so that the FAR is

$$\mathrm{FAR} = B\left[\sum_{s'=\sigma}^{\infty}\langle n_s\rangle + \sum_{s_1,s_2}\sum_{\delta'=1}^{\delta_{s_1,s_2}}\langle \nu_{s_1,s_2}(d)\rangle\right] \tag{7}$$

when $s_1, s_2 < \sigma$.

As a final comment on this detector, it should be noted that it is not optimal in the usual Neyman-Pearson sense; the excess power statistics is an example of an optimal detector for signals of finite duration and bandwidth. However, the detector described above performs in general almost as well as the excess power detector; in the worse case scenarios, the minimum signal to noise ratio detectable with the same efficiency (i.e. probability of detection) at the same FAR is a factor of two less than that of the excess power detector. The main advantage of the TS detector is that it is the maximum likelihood estimator of the region of the time-frequency plane where there is non-zero signal power. This is especially important for characterization and coincidence studies.

B. Case-by-case coincidences

The study of case-by-case coincidences refers to the detection of coincident transients in any two channels using only information near the transients in question, by opposition to a detection of a significant correlation between two channels for a population of transients, which will be the subject of the next section. A number of methods are available to use in conjunction with the cluster detector of section I A, but only one of them will be described here.

Assuming that a cluster was detected in channel 1, and that it consists of the set Γ_1 of (t,f)-points, and that a similar set Γ_2 exists for channel 2, and assuming a linear finite transfer function, the two sets should have similar shapes, and the power at points in both sets should be linearly correlated for subsets of points of constant frequency

[1]The distance is defined here as $d = |f_1 - f_2| + |t_1 - t_2|$. Other metrics are obviously possible.

common to both Γ_1 and Γ_2. The correlation is detected with a non-parametric statistic, since the power is obviously not Gaussian; in the present analysis, the Kendall τ statistic was used.

The detection of the correlation is done in the usual way by thresholding τ with threshold τ_* such that the probability for τ to exceed τ_* when no correlation is present is small. No formal study of the efficiency of the present detector has been carried out, but it is not expected to be too spectacular; experience with data suggests that only signals with duration of at least $5T$ are detectable by this algorithm.

C. Population coincidences

The aim of this section is to describe a powerful algorithm that was used to detect abnormal correlations between channels during times where transients of a certain class were detected in one of the two channels. In the description of the analysis, it will also be necessary to touch on a pattern recognition algorithm that was used to identify members of transient classes.

The basic idea of the method is to compare the distribution $p_{X_{\rm on}}(X)$ of the weighted cross-correlation X during 'on' times, defined as times where a transient of class C is present in channel 1, to its distribution $p_{X_{\rm off}}$ during 'off' times, where the data in channel 2 are randomly shifted in time to produce a fake data set with no genuine correlations. The cross-correlation is defined by

$$X = \max_{\tau} | \int_0^T d\tau_1 d\tau_2 x_1(\tau_1 - \tau - t_1) Q_C(|\tau + \tau_1 - \tau_2|) x_2(\tau_2 - t_2)|, \tag{8}$$

where T is the transient duration, t_i is an arbitrary time shift for channel i, and Q_C is a weighting kernel specific to class C. Note that Q_C not only encodes knowledge about the signal shape, but also possibly knowledge about the transfer function between the two channels. The detection can be performed by comparing the mean of the distributions using a classical Student t test, or by just comparing the distributions themselves, using e.g. the Kolmogorov-Smirnov statistic.

A class C will be defined by a canonical set of points Γ_C in the TS, or more generally by a set of such sets, e.g. to allow for varying signal duration. Bayesian classification is performed by maximizing the likelihood ratio function over the classes for a certain transient Γ, using a prior distribution $p_{p_s}(p_s)$ for the probability to have a black pixel when signal is present. That is:

$$\hat{C} = \arg\max_C \prod_{\Gamma} \int_0^1 dp_s p_{p_s}(p_s) \left[p_s/p; \frac{1-p_s}{1-p} \right]_C^{\Gamma}, \tag{9}$$

where the special notation $[a;b]_C^{\Gamma}$ is used to mean "use a when $\Gamma_{ij} = \Gamma_{C,ij}$, use b otherwise". In practice, $p_{p_s}(p_s)$ was chosen to be $\delta(p_s - \rho_s)$ for ρ_s reflecting the approximate density of points during transients; obviously, this is somewhat of an oversimplification, but is a rather good approximation if the transients in the class all have very similar amplitudes. This classification scheme is optimal.

It should be noted that a case-by-case correlation detector can be constructed from a class which has been shown to have a significant effect; this detector consists of the class definition and of a threshold X_* on the cross-correlation. The probability of false alarm estimated from X_* and $p_{X_{\rm off}}$ will be the true probability of false alarm (up to error from the finite sample $X_{\rm off}$ which is irrelevant since the sample can always be enlarged) but the estimated probability of detection

$$P_D = \int_{X_*}^{\infty} p_{X_{\rm on}}(X) dX \tag{10}$$

will be an upper limit on the true probability of detection, with equality only when a perfect classification algorithm exists. This case-by-case correlation detector is very efficient, assuming that X_* was previously computed. It turns out that it is possible to compute X_* in an efficient manner, i.e. in at most the same number of operations required to compute X; the algorithm doing that will be presented elsewhere.

II. RESULTS

This section presents results from the analysis of E2 data, corresponding to the different methods of analysis of section I.

T	1 s	B	500 Hz
α	0.91036	σ	5
$\delta_{1,1}$	0	$\delta_{1,2}$	0
$\delta_{1,3}$	0	$\delta_{1,4}$	0
$\delta_{2,2}$	0	$\delta_{2,3}$	0
$\delta_{2,4}$	2	$\delta_{3,3}$	3
$\delta_{3,4}$	4	$\delta_{4,4}$	4

TABLE I. Parameters of detection algorithm. Refer to section I A for the definition of the variables.

A. Transients rate

Table I lists the parameters used when applying the detector of section I A to the data. The same parameters were used for all channels. In all cases, only frequencies between DC and the minimum of the Nyquist frequency and 500 Hz were considered, the principal reason for that being the non-stationarity on timescales of ∼15 minutes of the background noise above 500 Hz in control signals of the IFO.

Figures 1 and 2 show histograms of the number of clusters (or pair of clusters) of a certain size per second in the common and differential arm length control signals, respectively. These figures were generated from 38.7 and 38.9 hours of locked data, taken between GPS times 657890388 and 658070675. The blue crosses show the actual data, while the magenta circles show what was expected for a Gaussian noise, using parameters from table I.

B. Case-by-case coincidences

The correlation in power over the cluster shapes in two channels was measured using the Kendall τ statistics, as described in section I B. Calibrated data from the common and the differential arm length control signals (labelled L_+ and L_- respectively) were compared with a number of PEM channels, and all clusters with correlation significantly different than zero, at the 99% level, were removed from the L_+ and L_- sets of 'genuine' clusters. The PEM channels that were considered were: x and y axes of LVEA seismometer, microphones in PSL enclosure and HAM7 chamber, x axis of x-arm mid-station seismometer, y axis of y-arm mid-station seismometer, x and y axes of accelerometers on BSC7 and BSC8, x axis of accelerometer on BSC5, and y axis of accelerometer on BSC6. In general, LVEA seismometers and accelerometers were the best diagnostic channels, in terms of the number of rejected clusters.

The total number of coincidences found was small compared to the total excess over Gaussian noise. Figure 3 shows the total number of bursts in 3000 s of data (blue crosses), the expected Gaussian rate (magenta circles), and the number of bursts after the coincidences with the PEM as described above were taken (green crosses). This tells either that most of the transient noise was not correlated with the PEM channels examined, or that the algorithm used to spot coincidences was not efficient.

C. Population coincidences

In general, the control signals were too messy to allow a clean separation of different burst types, so the alternative approach of first selecting classes of bursts in the PEM channel, and then looking for cross-correlation with the IFO channels, was selected. Again, the best PEM channels for finding sources of noise were the LVEA 'mechanical' channels, e.g. LVEA seismometers and accelerometers. Table II shows a few classes that were identified that way. There was a number of classes that were apparently only visible in the PEM channels. For instance, the transient noise in accelerometers in the mid stations was of the same nature as the one in the LVEA accelerometers, but no genuine correlations were found between mid stations accelerometers bursts and IFO signals.

III. CONCLUSION

The time domain weighted cross-correlation is by far the most sensitive technique to detect coincident transients in two channels, and in particular it works well when only one of the two channels has a large transfer function to the

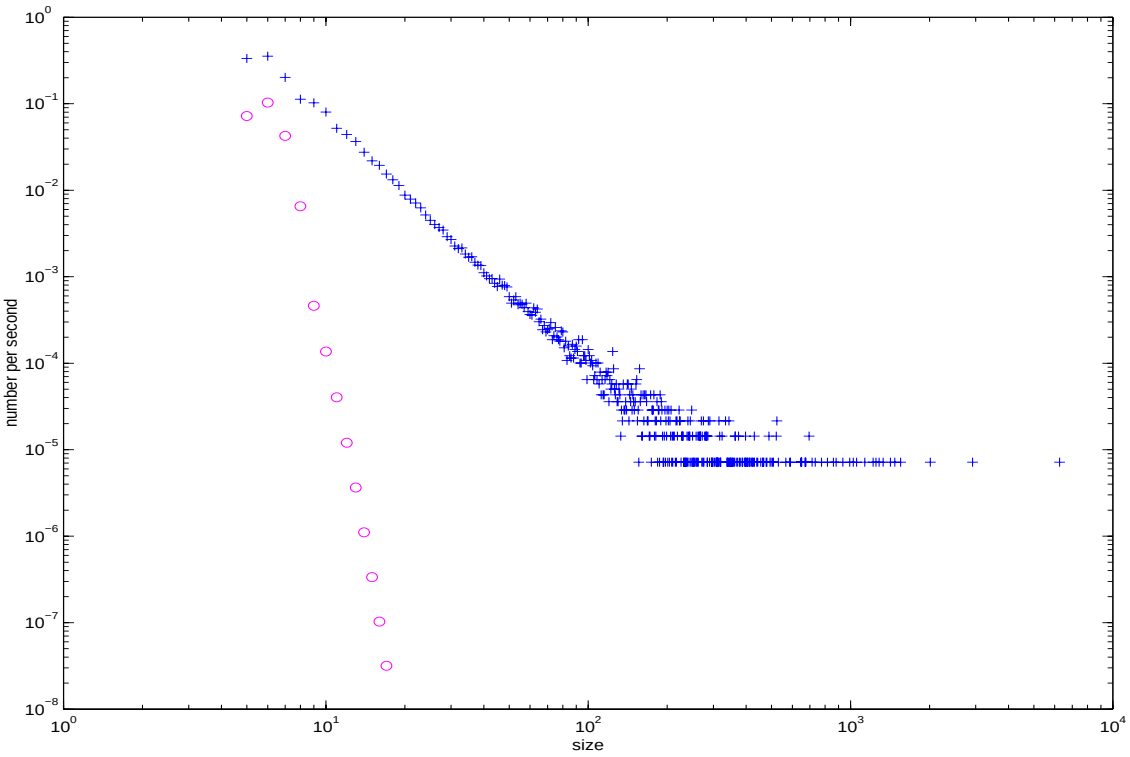

FIG. 1. Rate of bursts vs size in the common arm length control signal (H2:LSC-CARM_CTRL).

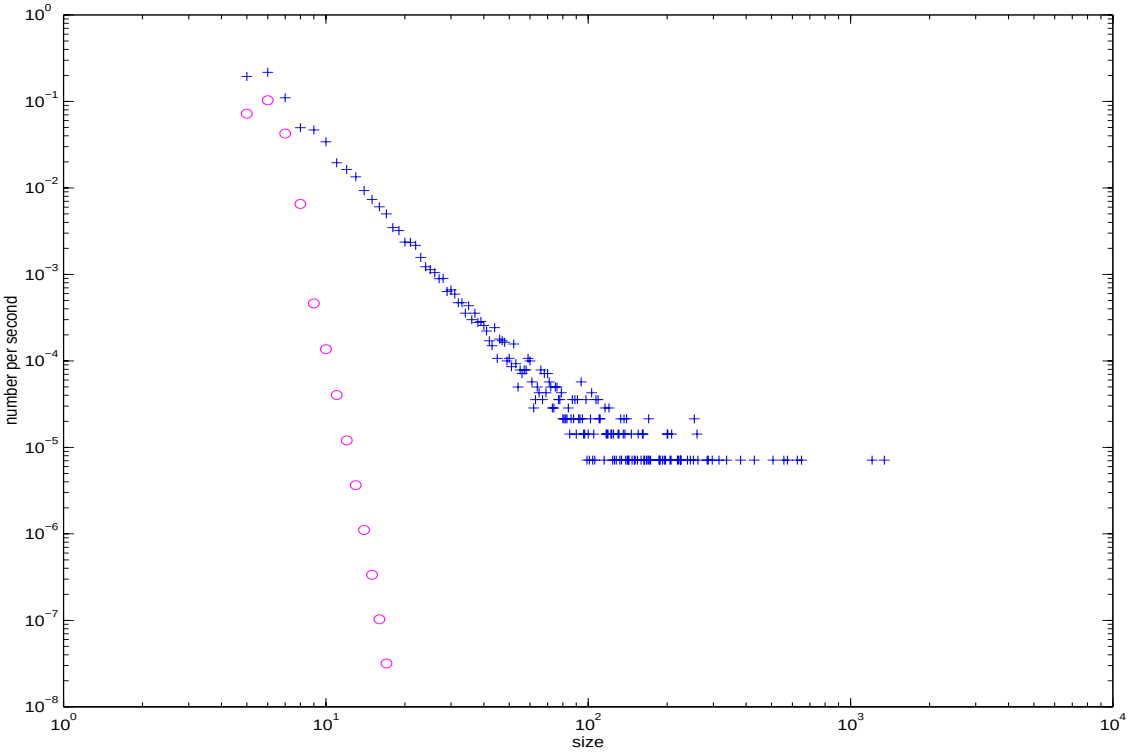

FIG. 2. Rate of bursts vs size in the differential arm length control signal (H2:LSC-DARM_CTRL).

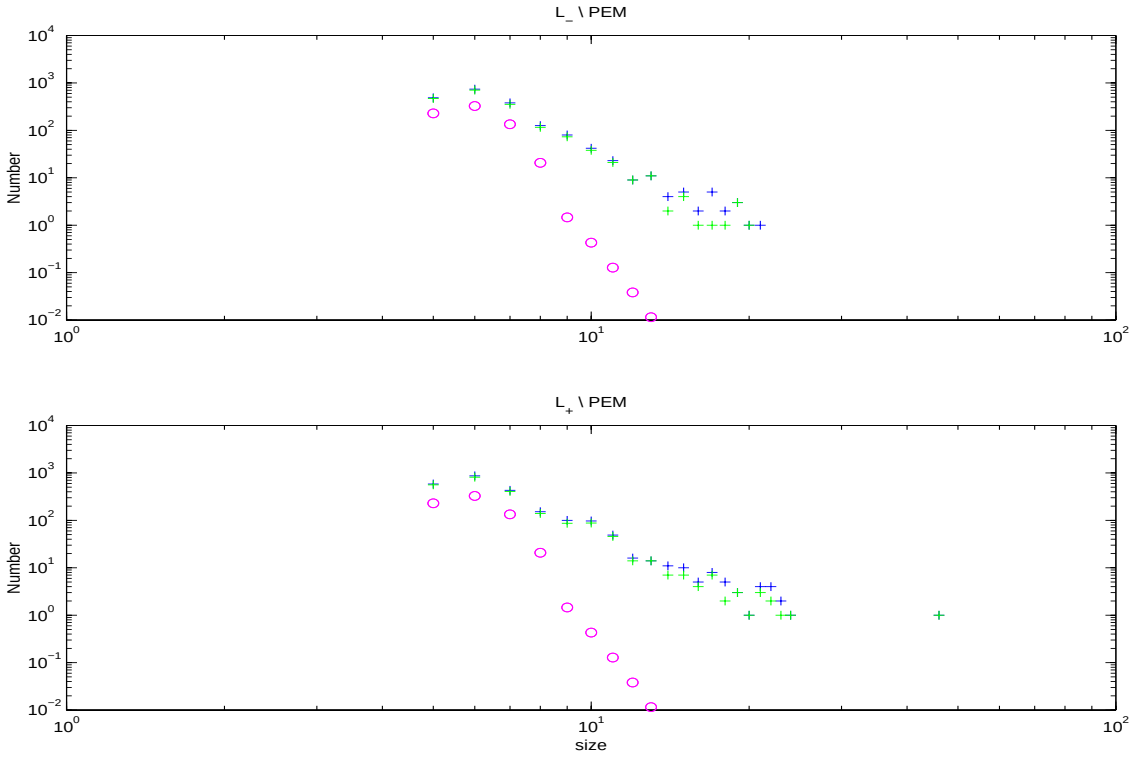

FIG. 3. Rate of bursts vs size in the differential and common arm length control signals (top and bottom, respectively).

class	visible in	frequency (Hz)	duration (s)	significance	rate (s^{-1})
narrowband I	LVEA seismometers LVEA accelerometers CARM_CTRL IOO-MC_F	55, 57, 71-72	100-300	$\ll 10^{-6}$	$6 \cdot 10^{-4}$
narrowband II	LVEA seismometers LVEA accelerometers CARM_CTRL IOO-MC_F	58-59	100-300	$\ll 10^{-6}$	10^{-4}
narrowband III	LVEA seismometers LVEA accelerometers CARM_CTRL IOO-MC_F	56-57	100-300	$\ll 10^{-6}$	$4 \cdot 10^{-4}$
broadband I	LVEA seismometers DARM_CTRL	15-35	1-30	$5 \cdot 10^{-3}$	$2 \cdot 10^{-4}$

TABLE II. Results of population (class) coincidences. Notes: (i) the "visible in" column is not necessarily exhaustive; it only lists channels that are known to be correlated. (ii) The frequency and duration columns refer to the range of parameters in the pattern matching algorithm, and may not reflect to exact population distribution. (iii) The significance is the probability that the "On" and "Off" time populations would be as different as observed (as quantified by the Kolmogorov-Smirnov statistics) due to random fluctuations only. (iv) The rate is measured over $1.2 \cdot 10^5$ s.

common source of noise. That channel is used to generate a best estimate of the burst waveform, and this estimate is used as a matched filter on the other channel. The methods of section I B are probably not useful for anything but the largest contributors to the transient noise with large transfer functions to both PEM and IFO channels; in that case, their principal advantage is their computational efficiency. The results reported in section II C correspond to cross-correlation analysis on classes of bursts, but work is in good progress to implement a cross-correlation analysis on a case-by-case basis. This is likely to lead to efficient algorithms for linear transfer functions, but much work needs to be done to deal with non-linear phenomena.