Environmentally-induced nonstationarity in LIGO science run data

Robert Stone, Soma Mukherjee
The Center for Gravitational Wave Astronomy
The University of Texas at Brownsville and Texas Southmost College
80 Fort Brown
Brownsville, TX 78520, USA.
rstone@phys.utb.edu

PACS numbers: 95.55.Ym, 95.75.Wx, 95.75.Pq
LIGO-P0900054-v4

1 Abstract

NoiseFloorMon is a data monitoring tool (DMT) implemented at the LIGO sites to monitor instances of non-stationarity in the gravitational wave data that are correlated with physical environmental monitors. An analysis of the fifth science run is nearly complete, and test runs preceding the sixth science run have also been analyzed. These analyses have identified time intervals in the gravitational wave channel that indicate non-stationarity due to seismic activity, and these intervals are referred to as data quality flags. In the analyses conducted to date the majority of time segments identified as non-stationary were due to seismic activity at the corner station and the x-arm end station. We present the algorithm and its performance, and discuss the potential for an on-site pipeline that automatically generates data quality flags for use in future data runs.

2 Introduction

The fifth Science run (S5) took place at the Laser Interferometric Gravitational Wave Observatories (LIGO) between November, 2005 and September, 2007. This science run encompassed approximately two years of observation among the two detectors of the LIGO Hanford Observatory (LHO) located in Hanford, Washington, and the single detector of the LIGO Livingston Observatory (LLO) located near Livingston, Louisiana. The search for gravitational waves divides into four broad gravitational wave searches according to standard models of the astrophysical sources of the waves: a search for waves due to compact binary coalescence, a search for waves due to unmodeled signals which last from a few milliseconds to hundreds of milliseconds (bursts), a search for waves due to continuous signals, and a search for the stochastic background of gravitational waves.

The search for gravitational waves in the LIGO data requires knowledge of the state of the detector at any particular time. One aspect of detector characterization is the detector response to environmental conditions; another aspect is the detector response to its internal processes, e.g., electronic signals, thermal vibrations of detector components, radiation pressure of the laser on the mirrors, etc. A residual effect of the internal processes is the production of noise in one or more detector channels that couples to the signals in the gravitational wave detection channel. The combination of both environmental disturbances and the noise produced by the internal processes results in a time-dependent noise background that is classified as non-stationary. In contrast, a stationary signal is a manifestation of a stochastic process in which the statistical properties of a signal are invariant under a time translation. The joint probability density function $F(x_0, t_0; x_0, t_1; \ldots x_n, t_1)$ depends on the time intervals $t_n - t_0$ between measurements, not on the times $t_0, t_1, \ldots, t_n$ themselves [1]. Since the background noise in LIGO *does* depend on the measurement times, it is considered non-stationary.

One of the largest challenges in LIGO data analysis is to identify instances of non-stationarity present in the gravitational wave channel. Our work involves the identification of these instances in time as well as their correlation with channels designated as physical environmental monitors (PEM). The PEM data used in our analysis comes from seismometers located along the beam tubes at each observatory. The seismometers are located near the end-tube masses, near the mid-tube masses, and at the station where the beam tubes intersect.

Non-stationary noise can take the form of short-lived transients whose power exceeds the noise background(a "glitch"), or longer time periods when the statistical properties of the background data change substantially [2] [3]. The latter form, identified as slow non-stationarity to distinguish it from glitchy behavior, is particularly problematic. The identification of slow non-stationarity is an especially important consideration for certain triggered searches that involve the comparison of data around times of external astrophysical triggers (on-source data) with data from other time periods (off-source data) [5]. The search for gravitational waves correlated with gamma ray bursts is an example of such a triggered search [5] [4]. The non-stationarity of the background noise in a triggered search for an unmodeled source can produce the illusion of anomalous events in the data, necessitating the need to flag the non-stationary periods [2]. The goal of the work presented here is to provide an efficient means of flagging non-stationary data in the gravitational-wave channel.

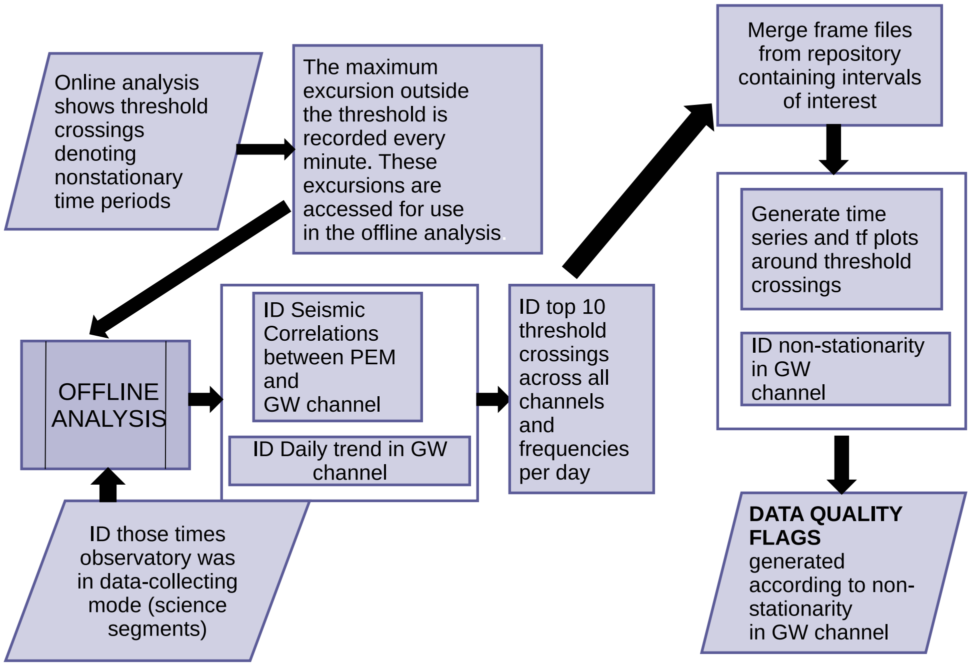

Figure 1: This figure describes the pipeline developed to flag non-stationarity. The steps are detailed in the Overview section.

3 Overview

The Noise Floor monitor (NoiseFloorMon) is a Data Monitoring Tool (DMT) [7] [8] implemented at the LIGO sites in Hanford, Washington and near Livingston, Louisiana to track non-stationarities during science runs. This monitor perform cross-correlations [6] between the PEM channels identified previously and the gravitational wave channels. The non-stationarities are immediately identified with this DMT and this data is transferred to the LIGO data repositories. The data is then accessed off-site for a subsequent off-line analysis. Figure 1 illustrates the steps in the entire analysis and the process is outlined here, with clarifications following.

1. The on-line monitor indicates instances of non-stationarity

   (a) The time series of the gravitational wave channel of each detector is first lowpassed at 4 kHz and resampled from 16 kHz to 2 kHz.
   (b) The data is whitened via a finite impulse response (FIR) filter [3].
   (c) Known instrumental and environmentally induced narrow-band resonances are notched out.
   (d) A threshold is generated on the conditioned data with a running median based on the algorithm Median-based Noise Floor Tracker (MNFT) as described in [3]. Those events that exceed the threshold are identified as non-stationary.
   (e) The most prominent event is recorded each minute, and the resulting data is archived at LIGO Data Analysis System sites [9]. Figure 2 shows an example of the output of the DMT as recorded at the observatories.
   (f) The data is stored at the LIGO data repositories.

2. The off-line analysis is used to generate data quality flags.
   (a) The on-line data is accessed from one of the data repositories.
   (b) The science mode segments are extracted for analysis from the data. These segments are generated when the detectors are optimized for data collection.
   (c) A threshold is generated on the data with a running median determined from every 20 sample points. This threshold is also based on the algorithm Median-based Noise Floor Tracker (MNFT) as described in [3]. Those events that exceed 2.5 times the threshold are identified as non-stationary.
   (d) Plots showing the cross-correlation between the PEM channels and the gravitational wave channel are produced.
   (e) The ten threshold crossings that have the greatest differences from the threshold across all PEM channels and frequency bands are identified. Due to the limitations of computational resources we examined the top ten threshold crossings, however we are evaluating whether this is a sufficient number to analyze.
   (f) Data files are generated that are centered around each threshold crossing.

    i. Two-minute segments of raw data are generated from the gravitational wave channel. The data is centered around the threshold crossing.
    ii. Two-minute segments of raw data are generated from the PEM channel that indicated the correlation. The data is centered around the threshold crossing.

(g) The gravitational wave data is downsampled and lowpassed or highpassed according to the frequency band indicating correlation, and whitened. Time series and time-frequency plots are produced.

(h) The PEM channel data is whitened and lowpassed or highpassed according to the frequency band indicating correlation, and whitened. Time series and time-frequency plots are produced.

(i) Points exceeding twice the standard deviation of the data in the modified gravitational wave data are identified as instances of non-stationarity requiring data quality flags.

A threshold is generated in both the DMT and in the off-line analysis using a running median based on the algorithm Median-based Noise Floor Tracker (MNFT) as described in [3]. The majority of the steps in the algorithm have been identified above: data is bandpassed and resampled, an FIR filter is applied to whiten the data, and known narrow-band resonances are removed. From [10], given a time series $\overline{V}$ the running median $\overline{\mathcal{V}}$ is defined as

$$\mathcal{V}(n) = \text{median}(\{V(r)\}), \text{ for } r = n - m_0, \ldots, n + m_0, \tag{1}$$

where m_0 is a fixed number such that $2m_0$ is defined as the blocksize of the median. In our analysis the median is then determined from a data segment that is approximately $\frac{1}{16}$ second long so that $2m_0 = 20$ (The sampling rate of the PEM channels is 256 Hz). The reason for using a running median is that it is less sensitive to the effects of transients in the data than a running mean; a running median is therefore a more valid benchmark to determine non-stationarity of the noise floor, i.e., the signal that remains after large transients and known resonances have been removed [10] [3]. Since our goal is to identify non-stationarities in the noise floor associated with seismic noise, we choose an estimator that will be least affected by transients that remain after the data conditioning.

The events that exceed the threshold are identified as non-stationary. The events that are the greatest distance from the threshold are considered the most prominent events. The most prominent event is recorded each minute, and the resulting data is archived at LIGO Data Analysis System sites [9]. Since threshold crossings are identified each minute the monitor readily measures short-term nonstationarity. Separate cross-correlations are performed between the time series of approximately 10 physical environment monitors (PEM) and the gravitational-wave channel AS_Q in four frequency bands for each detector: $0 - 16$ Hz, $16 - 32$ Hz, $32 - 64$ Hz, and $64 - 128$ Hz. These frequencies are analyzed because they are dominated by seismic noise.

4 Offline Analysis

The NoiseFloorMon *offline* analysis is also based on the algorithm Median-based Noise Floor Tracker as described in [3]. One day's allotment of the on-line DMT data is analyzed per PEM channel per frequency band, with the start of the day defined as 16:00 UTC. A running median is determined for each block size of 20 sample points, and this serves as the basis for a threshold of the NoiseFloorMon offline analysis. By comparing the NoiseFloorMon high-correlation events with other DMTs and offline analyses [11], we have gauged that a threshold of 2.5 times the running median offers sufficient sensitivity to anomalous events.

The offline analysis produces time series plots of the NoiseFloorMon DMT data per PEM channel per frequency band. The threshold is superimposed on these plots and points that cross the threshold are highlighted as can be seen in Figure 3. The output plots of the offline analysis show the cross-correlation values between the PEM channel labeled on the plot and the grav- itational wave channel. A threshold of two times the running median is indicated by the solid line, the circles above the line is a threshold of 2.5 times the running median (red in color versions). The points that exceed the threshold are topped with circles (green in color versions). The most prominent events are considered those points at the greatest differences above the threshold. Figure 4 shows the gravitational wave channel activity corresponding to the the threshold crossing indicated in Figure 3. An array is created per threshold crossing that contains the distance above threshold, the cross-correlation value assigned by the NoiseFloorMon DMT, the GPS time of the crossing (resolved to the nearest minute), the PEM chan- nel and the frequency band. The threshold crossings are sorted according to

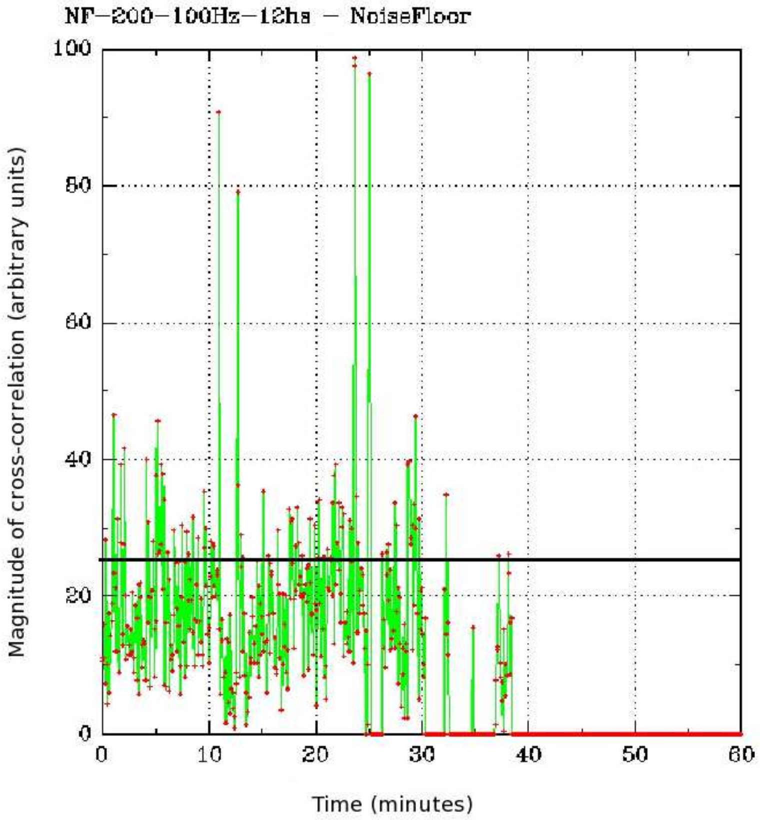

Figure 2: This figure shows a typical output of the NoiseFloorMon DMT that is run on-line at LHO and LLO. The cross-correlation value is plotted as a function of time in minutes. The line above 20 illustrates the threshold determined by the Median-based Noise Floor Tracker.

the greatest differences between the crossing and the threshold, and a list of ten most prominent threshold crossings is produced per day per observatory for S5. Figure 2 illustrates the connection between the on-line and off-line analyses.

The offline analysis was developed to highlight couplings between the gravitational wave and seismic activity measured by the PEM channels. Large-amplitude transient signals identified as glitches occurring simultaneously in both the gravitational wave channel and a PEM channel provide convincing evidence of coupling [11]. Since the NoiseFloorMon DMT records the loudest event each minute, and the gravitational wave channel sampling rate is 16 kHz, a systematic approach is taken to correlate a glitch seen in the gravitational wave channel with a threshold crossing in the offline analysis. Two-minute time series plots of the PEM channel and AS_Q are produced that are centered around the threshold crossing time. The AS_Q time series is then examined around a two-minute interval centered around the threshold crossing time, and the highest amplitude signal that is at least two sigma from the mean of the magnitude of the background is recorded.

In the majority of cases the non-stationarity identified in the NoiseFloorMon analysis corresponds to a glitch in the AS_Q channel at a point within the two-minute interval that is resolved to the nearest second. In addition to the two-minute time series plot of AS_Q that is generated, a whitened time series plot is also generated with previously identified transients (e.g., power line resonances) removed. Time-frequency plots of the PEM data and the AS_Q data are also generated.

5 Results

The overarching goal of this work is to identify correlations over years between seismic activity measured by the PEM channels and the gravitational wave channel. We anticipate that understanding the effect of seismic noise on the gravitational wave signal can lead to 1) possibly mitigating the effect through commissioning efforts, and 2) an explanation of the mechanism by which low-amplitude seismic disturbances are manifested as large glitches in the gravitational wave channel, a process known as up-conversion. To that end weekly and monthly trends are being cataloged to identify those PEM channels and frequency bands that register the greatest number of threshold crossings. For the months that have been analyzed approximately half of the threshold crossings were identified in the $0-16$ Hz band, while the other half were identified in the $32-64$ Hz band. Less than one percent of the crossings were found in the $64-128$ Hz band. In the analyses conducted to date the greatest number of threshold crossings have occurred due to seismic activity in the y-direction in the PEM channel seismometers identified as the LIGO Vacuum Equipment Area (LVEA or corner station) and the x-arm end station (EX) at both LHO and LLO. Lock loss (the state of the interferometer in which data acquisition is impossible due to excessive mirror displacements)or drops from science mode (the optimal state of the interferometer for data collection) were also correlated to those crossings that showed the greatest deviation from the threshold. [13].

Lock loss due to seismic activity is to some degree inevitable and there is an expected correlation between the seismic PEM channels and the gravitational wave channel for energetic events. We are especially interested in non-linear responses in the gravitational-wave channel to seismic activity. Nearly all of the threshold crossings that occurred in the $32-64$ Hz band were not associated with noise transients but rather with motion predominated by a particular frequency as seen in Figure 5. It is unclear where the motion originated. Threshold crossings due to noise transients that did not result in lock loss occurred primarily in the $0-16$ Hz band. Figure 6 shows an example of such a transient.

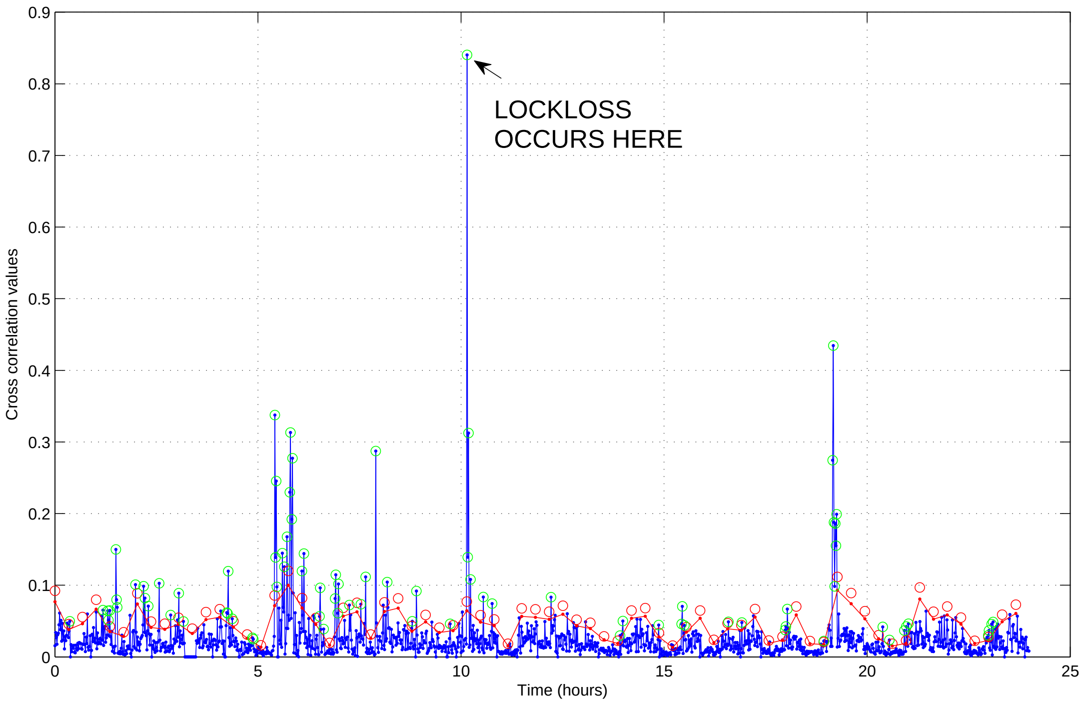

Figure 3: Sample output of the offline analysis indicating lock loss.

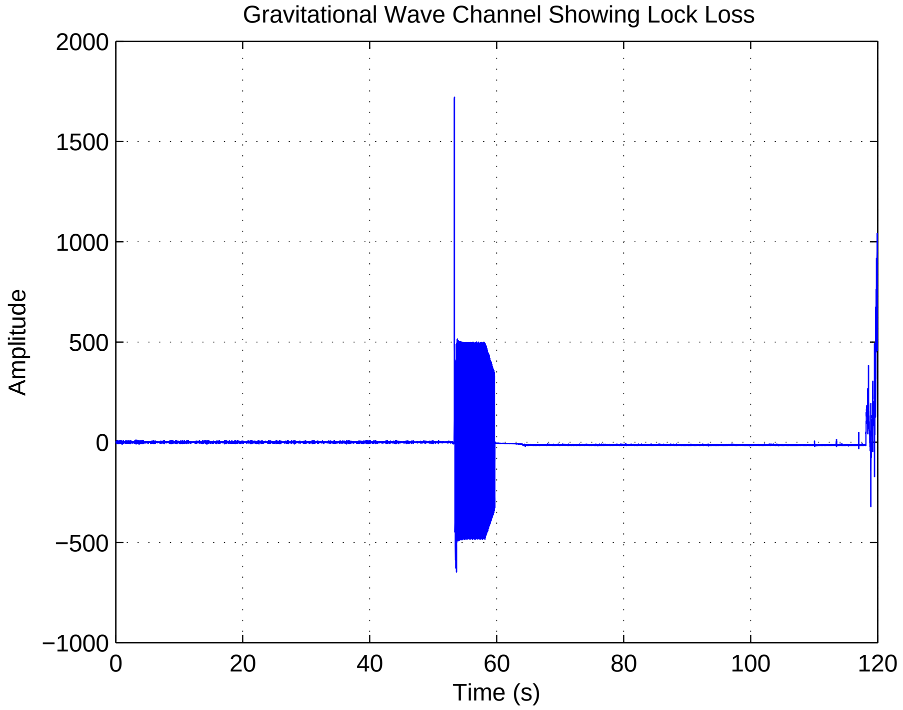

Figure 4: This plot shows the same lock loss as it appears in the gravitational-wave channel time series in a two-minute segment centered around the threshold crossing time. Subsequent investigation indicated that this lock loss event was caused by seismic activity.

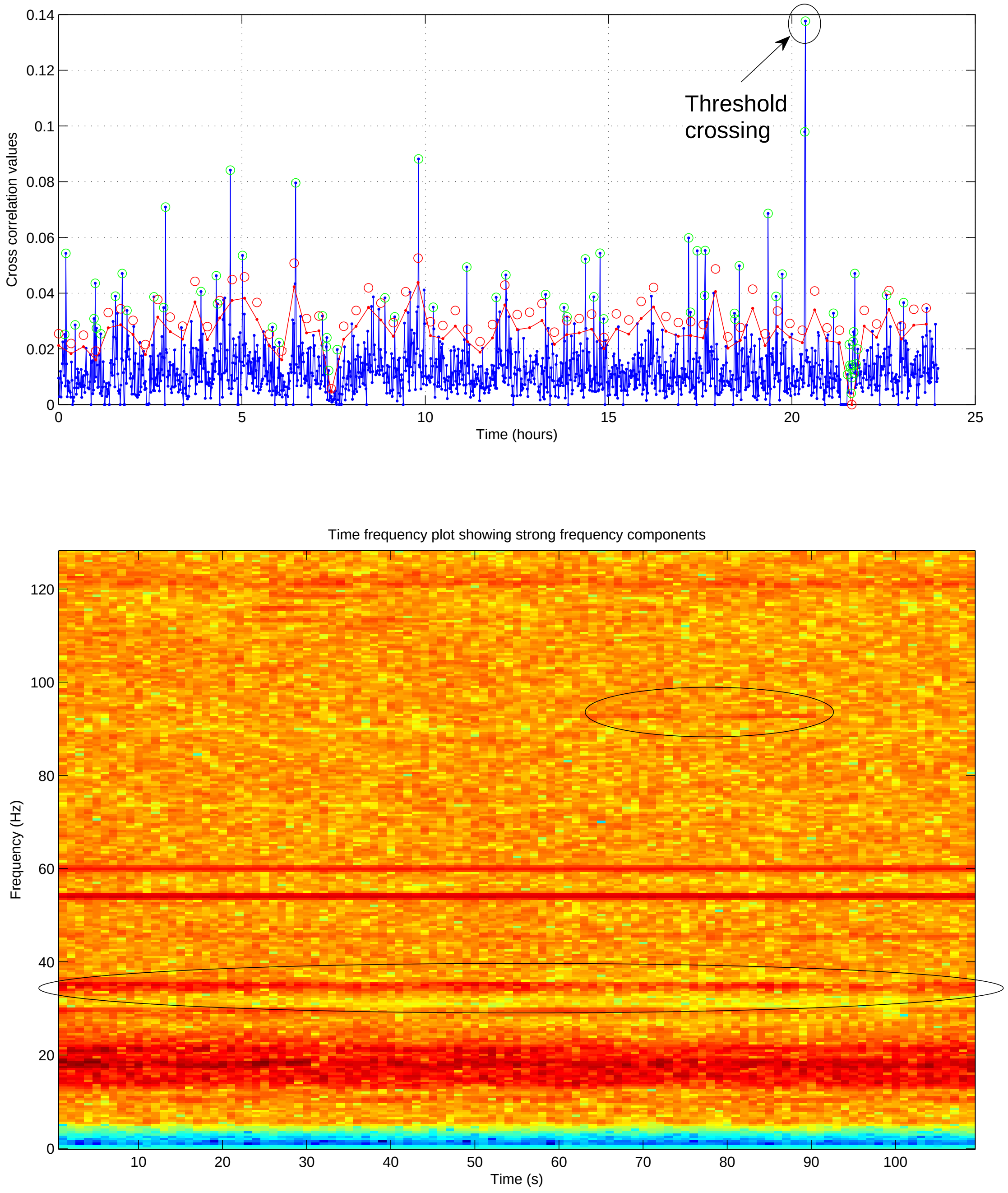

Figure 5: The bottom plots shows signals with strong components around 32 Hz centered around the threshold crossing.There is also a signal around 95 Hz. The top plot shows the crossing.

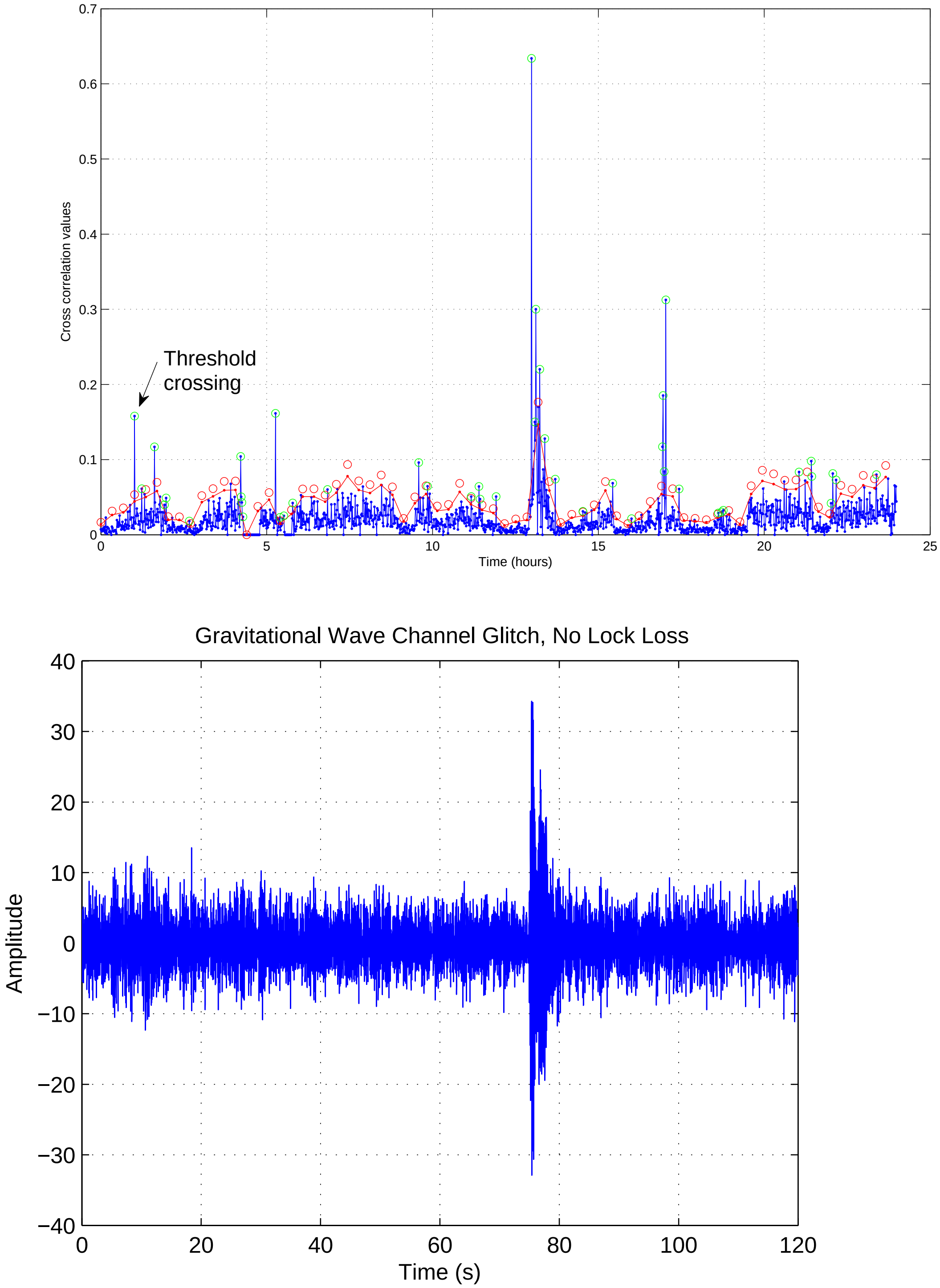

Figure 6: The bottom plot shows a glitch centered around the threshold crossing time. There was no lock loss in this case. The top plot shows the crossing.

6 Conclusions and future work

The NoiseFloorMon analysis identifies correlations between the seismic PEM channels and the gravitational-wave channel. The most prominent correlations occur in the $0 - 16$ Hz band and the $32 - 64$ Hz bands of the DMT. Many of the noise transients revealed by the analysis preceded lock loss. Many of the threshold crossings in the $32 - 64$ Hz band were the results of narrow-band resonances in the gravitational-wave channel. Although the analysis is on-going NoiseFloorMon appears to be a consistent indicator of non-stationarity in the gravitational-wave channel. Since the PEM channels constantly record data even if the interferometer drops from science mode our first analysis of S5 included non-science mode time segments. This has been corrected, and our immediate goal is to complete the full analysis for S5. Once all the threshold crossings have been cataloged we will work towards identifying those parameters that indicate non-stationarity in the data due to seismic activity and flag those events in the databases. After the entire S5 analysis is complete our goal is to contribute to a model of seismic up-conversion in the gravitational-wave channel. An algorithm is currently being tested that will generate the data quality flags with minimum lag times at each site. This algorithm should be implemented for use during S6 which will begin in mid-2009.

7 Acknowledgment

The author wishes to thank the LIGO laboratory for the use of data from the fifth science run. The LIGO technical document number assigned to this paper is LIGO-P0900054-v4. This work is supported by NSF grant PHY0555842 (2006).

The authors gratefully acknowledge the support of the United States National Science Foundation for the construction and operation of the LIGO Laboratory and the Science and Technology Facilities Council of the United Kingdom, the Max-Planck-Society, and the State of Niedersachsen/Germany for support of the construction and operation of the GEO600 detector. The authors also gratefully acknowledge the support of the research by these agencies and by the Australian Research Council, the Council of Scientific and Industrial Research of India, the Istituto Nazionale di Fisica Nucleare of Italy, the Spanish Ministerio de Educación y Ciencia, the Conselleria d'Economia, Hisenda i Innovació of the Govern de les Illes Balears, the Scottish Funding Council, the Scottish Universities Physics Alliance, The National Aeronautics and Space Administration, the Carnegie Trust, the Leverhulme Trust, the David and Lucile Packard Foundation, the Research Corporation, and the Alfred P. Sloan Foundation.

References

[1] Helstrom, C., *Statistical Theory of Signal Detection.* Oxford: Pergamon Press Ltd, 1968.

[2] Mukherjee, S; the LIGO Science collaboration, "Development of a DMT monitor for tracking slow nonstationarities present in LIGO science data.", *Journal of Physics: Conference Series*, 32 (1), (2006): 44-51.

[3] Mukherjee, S. "Median based noise floor tracker : robust estimation of noise floor drifts in interferometric data." *Classical and Quantum Gravity* 20 (2003): S925-S936.

[4] Abbott, B., et. al., "Implications for the Origin of GRB 070201 from LIGO Observations." *The Astrophysical Journal* 681 (2) (2008): 1419-1430

[5] Mohanty S.D., Marka S., Rahkola R., Mukherjee S., Leonor I., Frey R, Cannizzo J, Camp J "Gamma ray bursts and gravitational waves: triggered search strategy in the LIGO science runs." *Classical and Quantum Gravity* 21 (2004): S765-S774.

[6] Proakis, J.G. and Manolakis, D.G. 1996 Digital Signal Processing Third Ed. Prentice Hall, New Jersey.

[7] http://www.ligo.caltech.edu/∼jzweizig/DMT-Project.html

[8] Zweizig, J. "Data simulation for DMT." LIGO Technical document G030493-00-E, California Institute of Technology, Pasadena, 2003.

[9] http://www.ldas-sw.ligo.caltech.edu/cgi-bin/index.cgi

[10] Mohanty, S.D. "Median based line tracker (MBLT): model independent and transient preserving line removal from interferometric data." *Classical and Quantum Gravity* 19 (2002): 1513-1519.

[11] Blackburn, L., *et al.*, "The LSC Glitch Group : Monitoring Noise Transients during the fifth LIGO Science Run", *Classical and Quantum Gravity* 25 (2008): 184004.

[12] http://www.ligo.caltech.edu/∼shourov/q/qscan/

[13] Daw E J, Giaime J A, Lormand D, Lubinski M and Zweizig J. "Long-term study of the seismic environment at LIGO" *Classical and Quantum Gravity* 21 (2004): 2255-2273.