IOP PUBLISHING CLASSICAL AND QUANTUM GRAVITY

Class. Quantum Grav. **24** (2007) S415–S422 doi:10.1088/0264-9381/24/19/S05

Data quality studies for burst analysis of Virgo data acquired during Weekly Science Runs

F Acernese[1], P Amico[2], M Alshourbagy[3], F Antonucci[4], S Aoudia[5], P Astone[4], S Avino[1], D Babusci[6], G Ballardin[7], F Barone[1], L Barsotti[3], M Barsuglia[8], Th S Bauer[9], F Beauville[10], S Bigotta[3], S Birindelli[3], M A Bizouard[8], C Boccara[11], F Bondu[5], L Bosi[2], C Bradaschia[3], S Braccini[3], F J van den Brand[9], A Brillet[5], V Brisson[8], D Buskulic[10], E Calloni[1], E Campagna[12], F Carbognani[7], F Cavalier[8], R Cavalieri[7], G Cella[3], E Cesarini[12], E Chassande-Mottin[5], N Christensen[7], C Corda[3], A Corsi[4], F Cottone[2], A-C Clapson[8], F Cleva[5], J-P Coulon[5], E Cuoco[7], A Dari[2], V Dattilo[7], M Davier[8], M del Prete[3], R De Rosa[1], L Di Fiore[1], A Di Virgilio[3], B Dujardin[5], A Eleuteri[1], M Evans[7], I Ferrante[3], F Fidecaro[3], I Fiori[7], R Flaminio[7,10], J-D Fournier[5], S Frasca[4], F Frasconi[3], L Gammaitoni[2], F Garufi[1], E Genin[7], A Gennai[3], A Giazotto[3], G Giordano[6], L Giordano[1], R Gouaty[10], D Grosjean[10], G Guidi[12], S Hamdani[7], S Hebri[7], H Heitmann[5], P Hello[8], D Huet[7], S Karkar[10], S Kreckelbergh[8], P La Penna[7], M Laval[5], N Leroy[8], N Letendre[10], B Lopez[7], Lorenzini[12], V Loriette[11], G Losurdo[12], J-M Mackowski[13], E Majorana[4], C N Man[5], M Mantovani[3], F Marchesoni[2], F Marion[10], J Marque[7], F Martelli[12], A Masserot[10], M Mazzoni[12], L Milano[1], F Menzinger[7], C Moins[7], J Moreau[11], N Morgado[13], B Mours[10], F Nocera[7], C Palomba[4], F Paoletti[3,7], S Pardi[1], A Pasqualetti[7], R Passaquieti[3], D Passuello[3], F Piergiovanni[12], L Pinard[13], R Poggiani[3], M Punturo[2], P Puppo[4], S van der Putten[9], K Qipiani[1], P Rapagnani[4], V Reita[11], A Remillieux[13], F Ricci[4], I Ricciardi[1], P Ruggi[7], G Russo[1], S Solimeno[1], A Spallicci[5], M Tarallo[3], M Tonelli[3], A Toncelli[3], E Tournefier[10], F Travasso[2], C Tremola[3], G Vajente[3], D Verkindt[10], F Vetrano[12], A Viceré[12], J-Y Vinet[5], H Vocca[2] and M Yvert[10]

[1] INFN, sezione di Napoli and/or Università di Napoli 'Federico II' Complesso Universitario di Monte S. Angelo, and/or Università di Salerno, Fisciano (Sa), Italy
[2] INFN, Sezione di Perugia and/or Università di Perugia, Perugia, Italy
[3] INFN, Sezione di Pisa and/or Università di Pisa, Pisa, Italy
[4] INFN, Sezione di Roma and/or Università 'La Sapienza', Roma, Italy
[5] Departement Artemis—Observatoire de la Côte d'Azur, BP 42209 06304 Nice, Cedex 4, France
[6] INFN, Laboratori Nazionali di Frascati, Frascati (Rm), Italy
[7] European Gravitational Observatory (EGO), Cascina (Pi), Italy
[8] LAL, Univ Paris-Sud, IN2P3/CNRS, Orsay, France
[9] NIKHEF, NL-1009 DB Amsterdam and/or Vrije Universiteit, NL-1081 HV Amsterdam, The Netherlands
[10] Laboratoire d'Annecy-le-Vieux de Physique des Particules (LAPP), IN2P3/CNRS, Université de Savoie, Annecy-le-Vieux, France
[11] ESPCI, Paris, France

[12] INFN, Sezione di Firenze/Urbino, Sesto Fiorentino, and/or Università di Firenze, and/or Università di Urbino, Italy
[13] LMA, Villeurbanne, Lyon, France

E-mail: elena.cuoco@ego-gw.it

Received 13 April 2007, in final form 27 July 2007
Published 19 September 2007
Online at stacks.iop.org/CQG/24/S415

Abstract
Virgo started collecting science data during weekends in order to not interfere with commissioning activities. The goal of Weekly Science Runs is to ease the transition between commissioning periods and data taking periods, in addition to providing data sets exploiting the stationary behavior of the detector. The detection of gravitational wave (GW) bursts emitted by core collapse of supernovae is one of the most difficult tasks for the GW community due to the fact that there are uncertainties in the exact shape of the waveforms, as we do not have complete models. A major task for this kind of detection effort is the *cleaning* of the event triggers found by the detection pipelines, namely the removal of accidental transient signals due to noise source events. In order to clean our data from false GW events, we need to define a strategy for data quality cut and veto of auxiliary and environmental monitoring channels. In this paper we report on the analysis we performed on data acquired during Weekly Science Runs to explore and define the data quality cut and veto studies for burst analysis.

PACS numbers: 07.05.kf, 04.80.Nn, 95.55.Ym

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The interferometric gravitational wave (GW) detector Virgo [1] is completing the preliminary phase of commissioning before starting collection of scientific data. After the two long commissioning runs performed in 2005, the C6 run from July 29th to August 12th and the C7 run from September 14th to 19th, we started regular collection of science data during weekends, with minimal impact to commissioning activities. The idea is to collect data to test detection pipelines and check the evolution of our strain-equivalent noise sensitivity. Moreover, we can set up strategies to remove periods of bad data quality and analyze channels which acquire signals from control loops or which monitor the environmental noises to be used for vetoes. In 2006 we had six Weekend Science Runs (WSR), 2.5 days long each:

- WSR1 duty cycle 87.7%
- WSR2 duty cycle 71.2%
- WSR3 failed
- WSR4 failed
- WSR5 duty cycle 64.2%
- WSR6 duty cycle 80.5%.

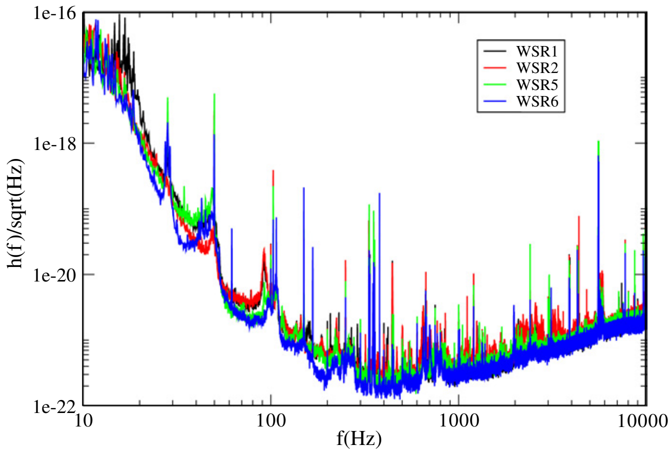

Figure 1. The strain-equivalent noise sensitivity during the WSR runs

This is a monthly program. In figure 1 we plot the sensitivity curves obtained during the last successful WSRs.

The burst group is devoted to the search of events which we generally call transient-like signals, which could be generated by core collapse or supernovae. The theoretical models for such a phenomenon leave big uncertainties on the produced GW signals; so, in principle, our pipelines must be able to detect all the signals which have a release of energy in short time durations. Many noise events could mimic a generic transient signal, so we have to be capable of distinguishing a non-real GW event from noise events. We analyze the events seen in the channels which monitor probable noise sources and check the coincidences with events seen in the RF-demodulated channel, called the dark fringe (DF), which should contain the GW signal. We can clean our burst trigger list by using a flag given by a data quality cut or veto strategy.

The data quality cut identifies the list of periods for which the interferometer (ITF) is malfunctioning, due to, for example, photodiode saturations, problems with calibration and problems in the data acquisition (DAQ) chain (see section 3). In the category of vetoes we include those events which could be analyzed in a statistical way (see section 4). To identify a channel which could be used as a veto for the events trigger list, we also analyze the triggers in the noise monitoring channel and check how many of these events happen in coincidence with the events in the DF, trying to statistically characterize the properties of these noise events. In the following section we report on the strategy we use to clean our data of these fake events.

2. Burst analysis

The Virgo burst search group uses a number of different techniques [2, 3] in order to identify transient-like events. In the following sections we report on the analysis executed on the WSR data using the mean filter and wavelet detection filter. These filters, used on the analysis of WSR5 data, are described as follows:

- *Mean filter.* The mean filter (MF) computes the value of the mean in a given time window of whitened data and looks for an excess of this quantity. Ten different analysis windows, with duration varying from 0.5 ms up to 10 ms, have been used [2].
- *Wavelet detection filter.* The wavelet detection filter (WDF) searches for excess power in a wavelet map obtained with a bank of discrete wavelets applied to whitened data.

A thresholding operation is applied on the wavelet coefficients with respect to the value of the RMS of the noise, and only those with the largest values are retained. These large coefficients are supposed to be linked to the transient signals that exceed the noise background [3].

Both filters have been applied either in the channel containing the gravitational wave signal or the channels which monitor the noises.

3. Data quality

We report below a list with the description of common data quality flags we use [6]:

- To reconstruct the strain signal the effects of the mirrors controls are removed by subtracting to the dark fringe the contribution coming from the mirrors motion due to the injected control signals on the reference mass coils drivers. To do this subtraction, the correction signals are filtered by the mechanical response of a pendulum plus the cavity filtering. Then this control free dark fringe signal is corrected for the optical transfer function (the cavity response) and finally converted to meter and strain ($h(t)$).
  When the reconstruction [7] of the calibrated gravitational channel (h) fails because the calibration lines are too weak, then the h signal, used by the detection algorithm, is missing; so we have holes in the data to be analyzed.
- During white noise injections (for calibration purposes) h is not reconstructed. So there is no calibrated gravitational signal channel to be analyzed. These periods are flagged in order not to run on them.
- The DF signal is the sum of two photodiode outputs. It happens that one of these photodiode channels saturates at ± 10 V. Then the periods for which the maximal absolute value is above 9.9 V are considered bad periods. The starting time and the end time are rounded to the smallest and biggest integer values, time being expressed in seconds.
- The photodiode shutter has been opened by mistake during some segments of WSR2. At the opening and closing of the shutter, some (acoustic) noise is induced in the DF signal. In addition to that, when the shutters are open some diffused light is suspected to occur. From the analyzed data we choose to flag the opening and closing periods ± 5 s with respect to the opening/closing time. This time window has been found necessary to protect the burst pipeline against the huge transient induced by the shutter opening and closure.
- The laser frequency noise is reduced thanks to an analog control loop. The loop sometimes has some malfunctions which can last up to few seconds. These periods were flagged.
- Forces on North End, West End and Beam Splitter mirrors are applied acting on coil drivers [8], which move the mirrors acting on magnets mounted on them. The coil currents can saturate at ± 10 V. A threshold at 9.9 V on the value of the correction signals is applied. This is necessary because close to the saturation coil drivers response is no more linear. It has been observed that it is reasonable to suppress 1 s before and 1 s after the saturation instant in the data to be analyzed.
- During WSR5, few problems in the Virgo timing system occurred. These periods were clearly identified in the data acquisition signals, because of the induced losses of data.

In figure 2 we plot the signal-to-noise ratio (SNR) values of triggers with respect to the time of the events, detected by the mean filter (MF) pipeline during WSR1 and WSR5. In the same plot we put in evidence the events which have been cleaned applying the data quality flags listed above.

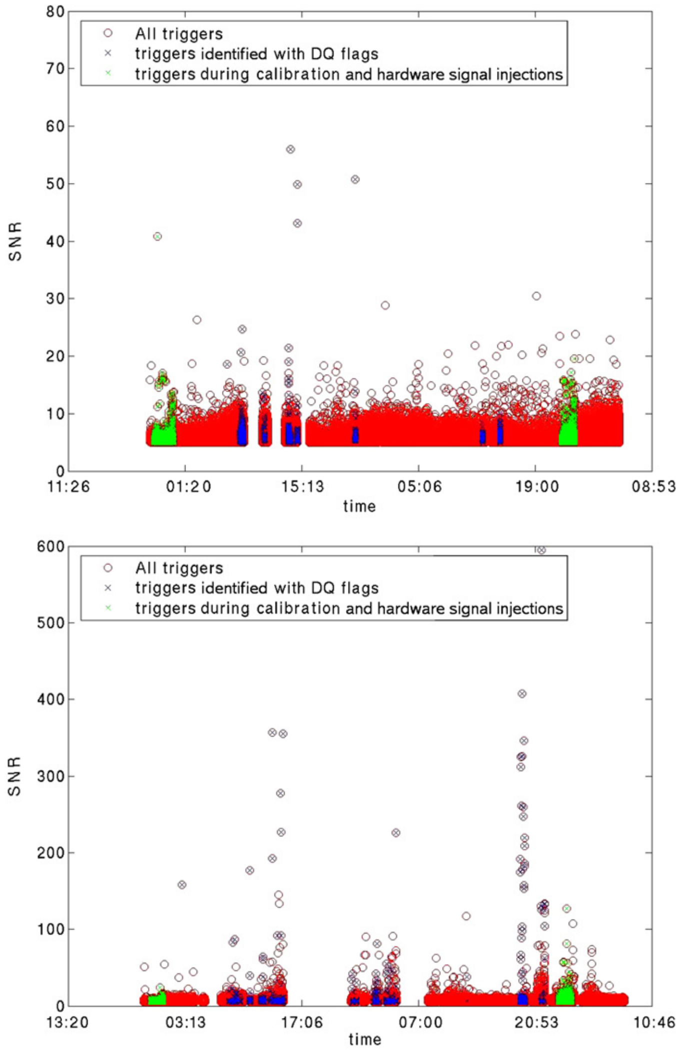

Figure 2. Mean filter triggers during WSR1 and WSR5. In the plot we put in evidence how by applying the DQ flags we can remove some high signal-to-noise ratio (SNR) events.

4. Veto study

We run our event trigger generator (ETG), that is the mean filter or WDF, on the DF channel to produce a list of triggers due to transient-like events, that is, events which release energy in a short time window. To clean our trigger list from fake events, we run the ETG on a set of auxiliary signals that we acquire to monitor the environmental noises and on channels used in the detector control loops, producing lists of events due to noises. Firstly we build the event distribution for the noise triggers, to identify the lower cut for the SNR value to use as the starting threshold value for our veto analysis.

To characterize a channel as a good veto for the fake events in the DF, we define some parameters which give indications on the quality of the veto procedure. The important

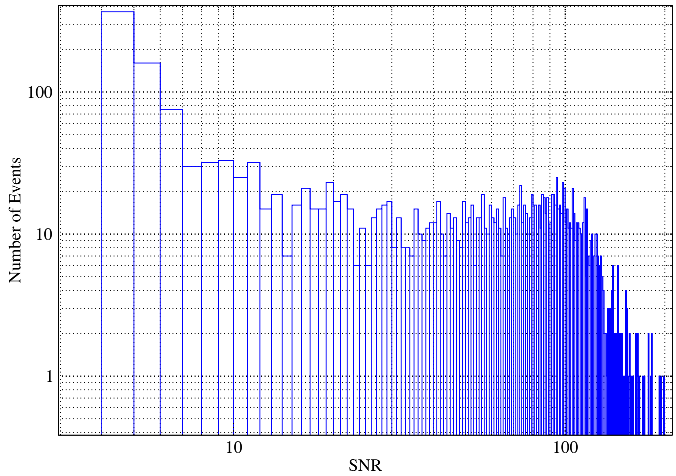

Figure 3. Event distribution versus SNR for the signal monitored by a piezoelectric accelerometer.

quantities to be studied before accepting a noise channel as useful for veto are listed below [4, 5].

(i) Use percentage: it is the percentage of auxiliary channel triggers, above the selected threshold, that vetoes at least one event found in the DF channel. Two events are defined as coincident if they happen inside a chosen window Δt.
(ii) Veto efficiency: it gives the percentage of glitches in the DF channel that are vetoed in the analysis.
(iii) Dead time: the time-window of data, in science mode, lost due to the veto procedure.

We study the dependence of the veto parameters (use percentage, veto efficiency and dead time) on the value of the veto window (Δt) for the coincidence and on the threshold for the SNR of the noise events. We choose the smallest value of Δt for which we can have reasonable values for the parameters defined above (see the following sections).

To establish the veto safety we check that the veto analysis does not remove hardware injected signals.

To understand the percentage of accidental vetoing of the DF signal, we time-shift the DF and analyze what happens in the use percentage versus Δt. The value of use percentage computed in this way is due to the casual coincidences and we want it is as small as possible.

4.1. Characterization of the veto channel

We report on the procedures followed to identify an auxiliary channel as a good veto to clean the DF trigger list of events. In particular, we analyzed the signal acquired by a vertical accelerometer located on the optical bench, in the detection laboratory.

The same ETG generator we applied on the DF was also applied on the auxiliary channel signal. In this way, we obtain a list of triggers for events seen in the noise monitoring channel.

The first step is to create the distribution of the events with respect to the SNR of the events themselves, in such a way as to guess which SNR threshold to select for the analysis. This is identified as the lowest SNR above which the events fall in the tail of the distribution.

In figure 3 we plot the events distribution for the signal acquired by the accelerometer described above. It seems that a threshold of SNR = 10 is a good starting value for the analysis.

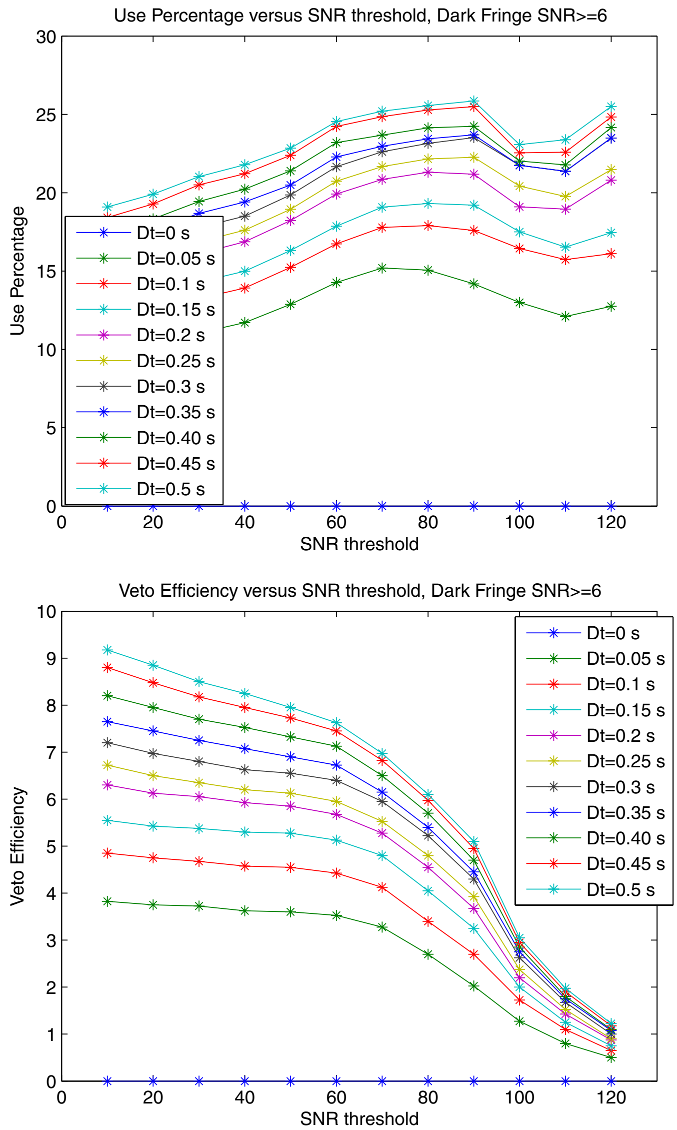

Figure 4. The use percentage and veto efficiency versus the SNR threshold. These plots are used to determine the parameters to be used for the veto procedure. We are looking either for the best value of the SNR threshold for the events in the auxiliary channels or for the best value for coincidence time window Δt with DF events.

The SNR threshold is then varied from SNR = 10 to SNR = 120 in steps of 10, and we estimate the value of the use percentage, veto efficiency and dead time with respect of the SNR threshold. In figure 4, the first row plots the use percentage, the second row plots the veto efficiency. Also the behavior with respect to the veto window width Δt for the coincidence is analyzed, by varying it from 0.05 s to 0.5 s in steps of 0.05 s.

The process of determining the values for the parameters associated with a channel to be used as a veto is in some way arbitrary. We try to find the optimal values for the number of DF events that are vetoed by the veto channel. The basic veto definitions and characteristics are defined in [5]. A good veto would have a large use percentage (percentage of veto triggers that veto at least one DF event), a large veto efficiency (percentage of DF triggers eliminated) and a small dead time (percentage of science-data time when veto is on). The veto channel may need to be appropriately filtered, and an event size (for example, the SNR from a burst search pipeline applied to the environmental or auxiliary channel) threshold set. The time window about a veto trigger is another parameter in the veto study. The filter frequency band, the event size threshold and the length of time about the veto trigger will all affect the use percentage, veto efficiency and the dead time.

For this channel the values of the use percentage are quite large for $\Delta t \geqslant 0.1$ s and reaches a plateau for larger Δt. We thus choose $\Delta t = 0.1$ s for the veto coincidence window.

We then selected a value of SNR threshold about 70 where the plateau for the veto parameters seems to also be reached. In this condition the use percentage is about 18%, the veto efficiency is 4.5–3.5% and the dead time is 0.5%.

In order to quantify the effects of the accidental coincidences in our study, we shifted the time of the DF triggers by 2.0 s, performing the veto analysis with the same parameters of SNR and veto windows.

The values obtained for the veto parameters we selected are: use percentage = 3.8%, veto efficiency = 0.9% and dead time = 0.5%. This shows that, given the parameters we chose, the events we flagged are real events in coincidence with that channel and not accidental ones.

5. Conclusion

In 2006 Virgo started the Weekly Science Runs program to acquire data in science mode during weekends, without interfering with the commissioning activity. The objectives were to check the evolution of the sensitivity of the interferometer and the data quality. In this paper we report on the work we performed to set up data quality flags and veto procedure for the burst detection analysis [6]. We identify some categories of data quality flags due to the malfunctioning of the detector or data acquisition system and to control loops. Moreover, we began to set up a strategy to clean the burst triggers list of fake events due to noise, using a statistical approach. These studies are fundamental, considering the recent agreement for data exchange with another project for the detection of the GW signal, where information on data quality also has to be exchanged.

References

[1] Acernese F *et al* 2007 Status of Virgo detector *Class. Quantum Grav.* **24** S381

[2] Arnoud N *et al* 2003 *Phys. Rev.* D **67** 062004

[3] Cuoco E 2005 *VIR-NOT-EGO-1390-308* (internal note)

[4] Di Credico A (for the LIGO Scientific Collaboration) 2005 *Class. Quantum Grav.* **22** S1051

[5] Christensen N (for the LIGO Scientific Collaboration) 2005 *Class. Quantum Grav.* **22** S1059

[6] Bizouard M A *et al* 2005 *VIR-NOT-LAL-1390-337* (internal note)

[7] Veziant O *et al* (the Virgo collaboration) 2003 *Class. Quantum Grav.* **20** S711

[8] Gennai A 2004 *VIR-SPE-PIS-4900-121* (internal note)