BACKGROUND ESTIMATION IN A GRAVITATIONAL WAVE EXPERIMENT

PIA ASTONE

INFN, sezione "La Sapienza", Physics Department, P. Aldo Moro, 2
Rome, I-00185,Italy

SERGIO FRASCA

INFN, sezione "La Sapienza" and Physics Department, P. Aldo Moro, 2
Rome, I-00185,Italy

GUIDO PIZZELLA

INFN, Laboratori Nazionali di Frascati and University of Rome "Tor Vergata" Physics Department, Via della Ricerca Scientifica
Rome, I-00133,Italy

The problem to estimate the background due to accidental coincidences in the search for coincidences in gravitational wave experiments is discussed. The use of delayed coincidences obtained by orderly shifting the event times of one of the two detectors is shown to be the most correct

Keywords: PACS number 04.80.Nn

1. Introduction

When searching for coincidences due to short bursts of gravitational radiation (GW) we are faced with the problem that the coincidences found at zero delay could be casual. In order to measure the background due to the accidental coincidences, the most common procedure adopted since the beginning of the gravitational wave experiments[1] consists in shifting the time of occurrence of the events of one of the two detectors a certain number of times. The distribution of the delayed coincidences gives the statistical properties of the background and allows to estimate, in the case of a coincidence excess, the probability that the excess was accidental.

In this contribution we shall try to catch the problems which may arise in this procedure and suggest how to cope with them.

2. Time delay histogram

For the sake of simplicity we consider here only coincidences between pairs of detectors, but all the considerations apply to the general case on N detectors. The outputs of the background estimation procedure, obtained by off-timing techniques, are the "time delay histograms".

There are several types of delay histograms. Two typical cases are shown in Fig.1. The upper part of the figure shows the delay histogram obtained with the 1998 IGEC data of the detectors NAUTILUS[2] and EXPLORER[3]. The lower part of the figure shows the time delay histogram obtained with 100 days of data recorded by EXPLORER and NIOBE in 1995[4].
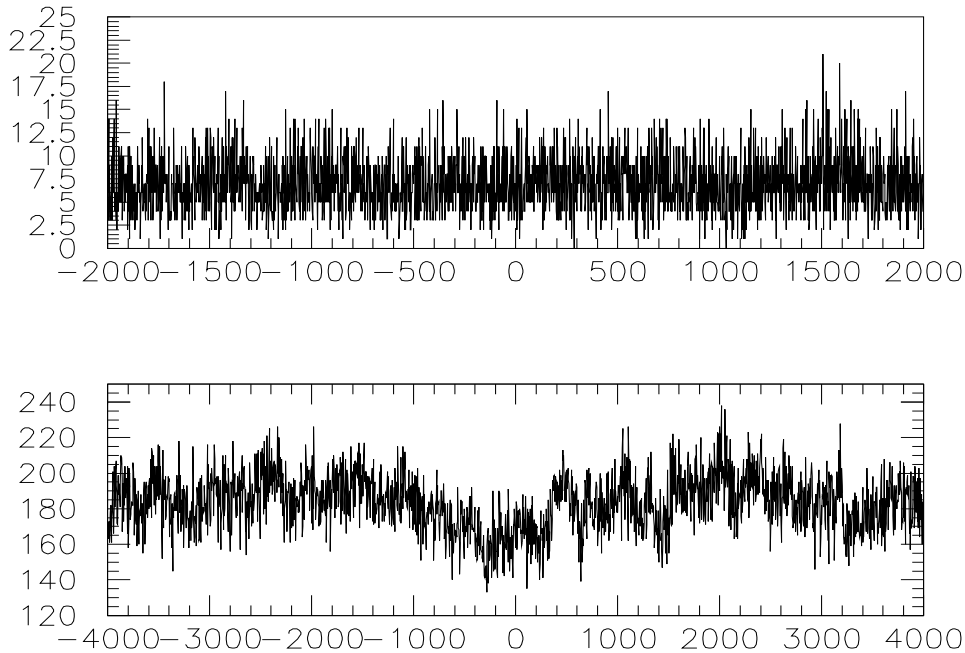

Figure 1: In the upper figure we show the number of coincidences versus the time delay in seconds (delay histogram), for EXPLORER/NAUTILUS 1998. In the lower figure we show the delay histogram for EXPLORER/NIOBE 1995.

From the figure we note that the upper histogram can be considered "good", as no particular structures in the data appear. This lead to the prediction that the distribution of the accidental coincidences is well described by a Poisson distribution, as verified in many cases. This reflects the fact that -in this particular case- over the whole observation time T_{obs} the noise process is stationary or -more specifically- the event occurrence times fulfill the conditions which define a Poisson process.

The lower figure shows a particular structure, that reflects same non stationary noise in one or both the detectors. In particular, around the zero delay the number of off-timing coincidences is clearly systematically lower compared to the behaviour at ± 4000 s. In this case, the standard procedure of comparing the n_c coincidences at zero delay with the average number $\bar{n}$ of shifted coincidences will lead to underestimate -if any- a physical effect. On the contrary, suppose, even if here it is not

the case, n_c=210 events. Given the local background (the background estimated in an interval ±500 s, for example) the suspicion may arise that same physical effect has been observed. The use of the more robust but here meaningless estimation over ±4000 s would mask the effect. If forced, by evident non stationary noise, to use "local shifts" the final estimation will clearly result to be less accurate.

One could expect that the non-stationary noise would give a non-poissonian distribution of the delayed coincidences. Instead we find, for the data of the lower part of Fig.1, the distribution shown in Fig. 2, well fitted with a Gaussian curve.
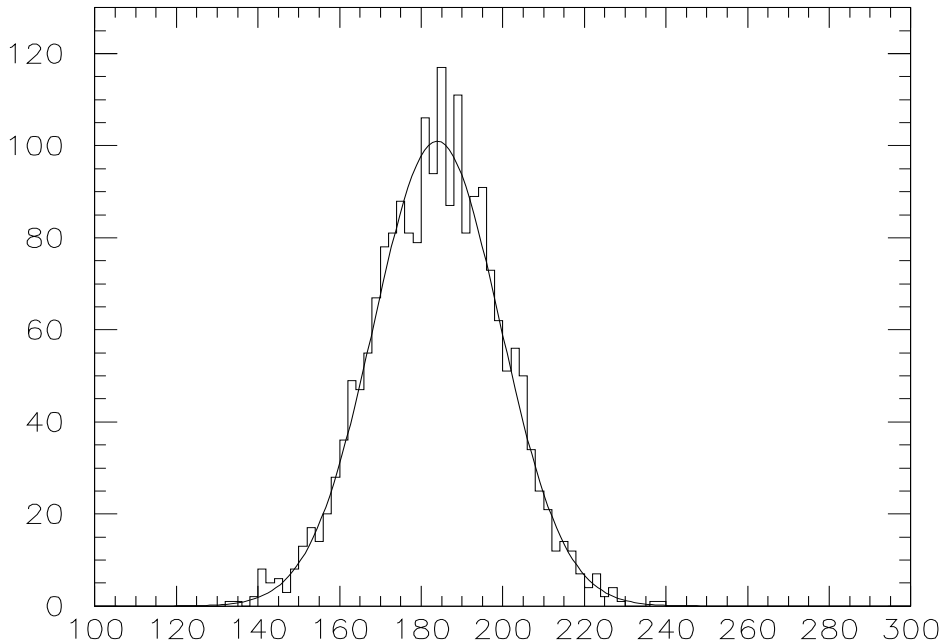

Figure 2: Distribution of the EXPLORER/NIOBE delayed coincidences. The line is a gaussian fit.

This is because the number of the background coincidences at the different delays here is so high that (due to the central limit theorem) the observed distribution is a Gaussian one.

Therefore we cannot use, in general, the distribution of the delayed coincidences to validate a statistical result. The natural suggestion is that a coincidence result should always be presented as a time delay histogram plus the number n_c of coincidences and the average $\bar{n}$ of the delayed coincidences.

3. Remarks on the previous examples and the use of the moving threshold to select the events

Going into a more detailed study of the data we noticed that the effect in Fig.

1 is not due to "stop runs" in the data (see the discussion on this point in section 5). It is due to the non stationary noise, which produces in a detector a highly varying number of event per hour. In the upper part of Fig.1 for both detectors, EXPLORER and NAUTILUS, we had made use of a moving energy threshold adapted to the noise that keeps nearly constant the event rate. In the lower part of Fig.1 we had made use of NIOBE events obtained with a fixed threshold, that produced an event rate from a few events up to two hundred events per hour. Thus the use of a moving threshold adapted to the noise is recommended.

The use of a moving threshold reduces the effect of the non stationary noise, but we still have a problem when the detectors have very different sensitivities. This has to be considered, if possible, when comparing the two events lists. We use here the terms "events" to indicate the quantities measured by the detectors and "signals" to indicate the physical quantities we aim to infer (see ref.[5]). The Explorer and Nautilus detectors during 1998 had very different noise. Indicating the noise with the effective temperature T_{eff},* only a very small number of hours of Explorer have a sensitivity better than 15 mK, which is the worst Nautilus sensitivity. This means that the sensitivity of the global analysis is set by Explorer.

In general, we do not make assumptions on the signals amplitudes and the standard analysis is done using all the data. Our remark is intended to note that it is worth, in addition to the standard analysis, to do separate analyses, considering the different detectors sensitivities and the possible signals amplitudes. If the signal amplitudes are expected to be so large, compared to the noise of the worst detector, that its detection efficiency (the fraction of events detected at a given level) be $\epsilon \simeq 1$, then the standard analysis can be applied without particular care.

But, as it usually seems the case, when the signal amplitudes are expected to be "small", then proper additional analyses must be done using only the data corresponding to similar sensitivities. Clearly this will reduce the observation time T_{obs}, but the combination of data which are measuring different physical effects may produce an artificially spoiled result[6].

4. Random coincidences to estimate the background

The procedure so far described uses "shifts" to estimate the background. One might envisage alternative procedures, such as a random reshuffling of the times of one of the two sequences. Then the events will be distributed in a random way over the entire T_{obs}.

The same arguments we used in the previous section and, in particular, the lesson we learn with Fig.1, show eloquently what happens in such a case. In fact, while with the shift procedure we maintain the data structure and so we can do considerations and derive conclusions from it, when we have randomly reshuffled the times important information contained in them is lost forever. We can loose a

---

*T_{eff} is a parameter that is related to the event amplitude h by a simple equation (i.e. 10 mK means $h = 8 \cdot 10^{-19}$).

genuine effect or we can claim for a possible -false!- discovery.

It is easy to convince ourselves that the result of a random reshuffling may overestimate or underestimate the true background, depending on the relative positions of holes in the two data streams. This is easily illustrated in Fig.3

$$n_c = 0$$
$$\overline{n} = 0$$
$$n_r > 0$$

$$\frac{n_c}{\overline{n}} \quad 0 \atop > n_r$$

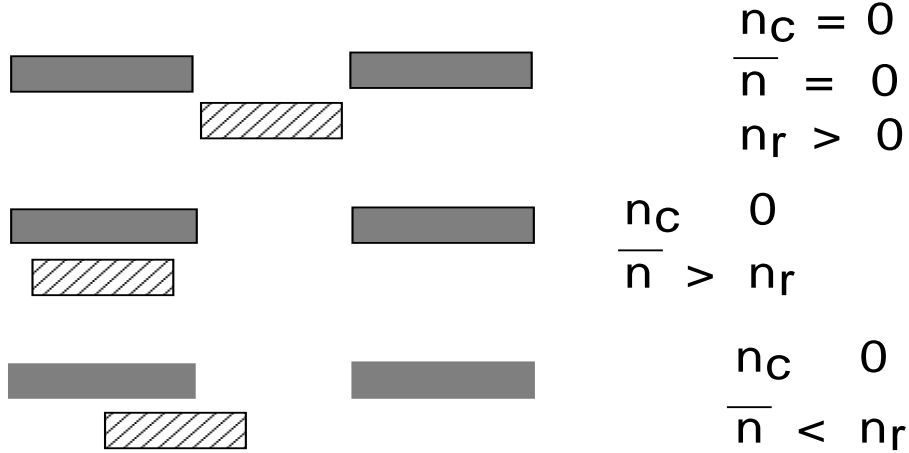$$\frac{n_c}{\overline{n}} \quad 0 \atop < n_r$$

Figure 3: The time runs horizontally. The black regions indicate the event coverage of the first detector, the diagonal marked regions indicate the event coverage of the second detector. The upper case occurs when the events of the second detector fall in a hole of the first detector. In this case we have $n_c = 0$, $\bar{n} = 0$ for the average estimated with the shifts and $n_r > 0$ for the average estimated with randomly changing the event times. The other two cases clearly follow.

5. The presence of "stop runs" in the data

Usually the data of both the sequences will contain "holes", that is missing data due to "stop runs". In total the events of the two detectors will cover a common period of time T_{obs}. When we apply the time shift procedure for the determination of the accidental coincidences the observation time will be different for each shift, since the events of the detector with the shifted times might overlap with a hole in the event list of the other detector (or viceversa).

If the holes for each detector are randomly distributed, this change in the observation period turns out to be negligible, since the decrease of it due to a hole will be compensated by the increase due to another hole. In any case it is always possible

to estimate this change and normalize the numbers of accidental coincidences to the real observation time T_{obs}. This normalization is important if the holes are not randomly distributed, but always occur for the two detectors at the same times.

6. Conclusion

To summarize, the final recommendation is:

- to use the "shifts", to maintain the information on the noise structure

- not to use "random" data reshuffling to estimate the background

- always give as final result the time delay histogram, plus n_c and $\bar{n}$

- use with care detectors with different sensitivities.

- start/stop times are necessary to take into account different periods of overlapping when shifting

- use the adapted threshold to select the events, to make the events occurrence more stationary

1. J. Weber, Phys. Rev. Lett. 22, 1320 (1969)
2. P. Astone et al, Astroparticle Physics, 7 (1997) 231-243
3. P. Astone et al., Phys. Rev. D. 47, 362 (1993).
4. P.Astone et al, Astroparticle Physics 10 (1999)83-92
5. P.Astone, S.D'Antonio and G.Pizzella, This Proceedings (1999)
6. P.Astone and G.D'Agostini, CERN-EP/99-126 and `hep-ex/9909047`