Event filtering prior to the core and angle fitters

Javier Bussóns Gordo

February 1, 2000

1 Introduction

This note discusses the need for a cleaning procedure to eliminate or reduce PMT hits due to electronic noise and overlapping triggers on an event-by-event basis. With such a procedure one can: i) improve the subsequent performance of the core and angle fitters and, therefore, the angular resolution, and ii) recover genuine events that would otherwise have not been fitted because of noise-induced image distortion.

2 Database and Analysis

I have chosen to analyse chunks of recent GRB*.dat files, that is, files which –unlike the Save*.dat files– contain all the triggers above $\sim$45 hits in the top layer (above $\sim$90 calibrated hits overall) and are not biased towards the better defined events. A very recent file (residing on /newman/gus/data/RAW), with raw event rate of 1159 Hz constitutes the dataset for the following analysis.

As Gaurang (pronounce Gau as in Gauss) has shown, by simply cutting on the recorded time of arrival one can filter out noisy hits due to either a smaller shower, an unassociated muon or any other creature accidentally overlapping with the main shower. My argument –nothing new under the sun– goes along the same lines and somehow follows the concept behind his famous speed-of-light-in-water t vs r diagrams:

Immediately after the CalibrateRaw routine and before any fit is done, I use the time stamps arrived at to work out how many PMTs are causally connected. I define the degree of association, DOA, of a particular hit PMT in a given event as the percentage of other hits it is associated to:

$$DOA = \frac{N_{ass}}{N-1} \times 100,$$

where N is the total number of calibrated hits in the event and N_{ass} (I couldn't resist the subscript) is the number of associated hits. The association criterion is based on causality: two PMTs are associated if they are separated in time no more than they are separated in distance (considering the speed of light in the medium as the maximum speed possible, of course). In other words, the maximum time

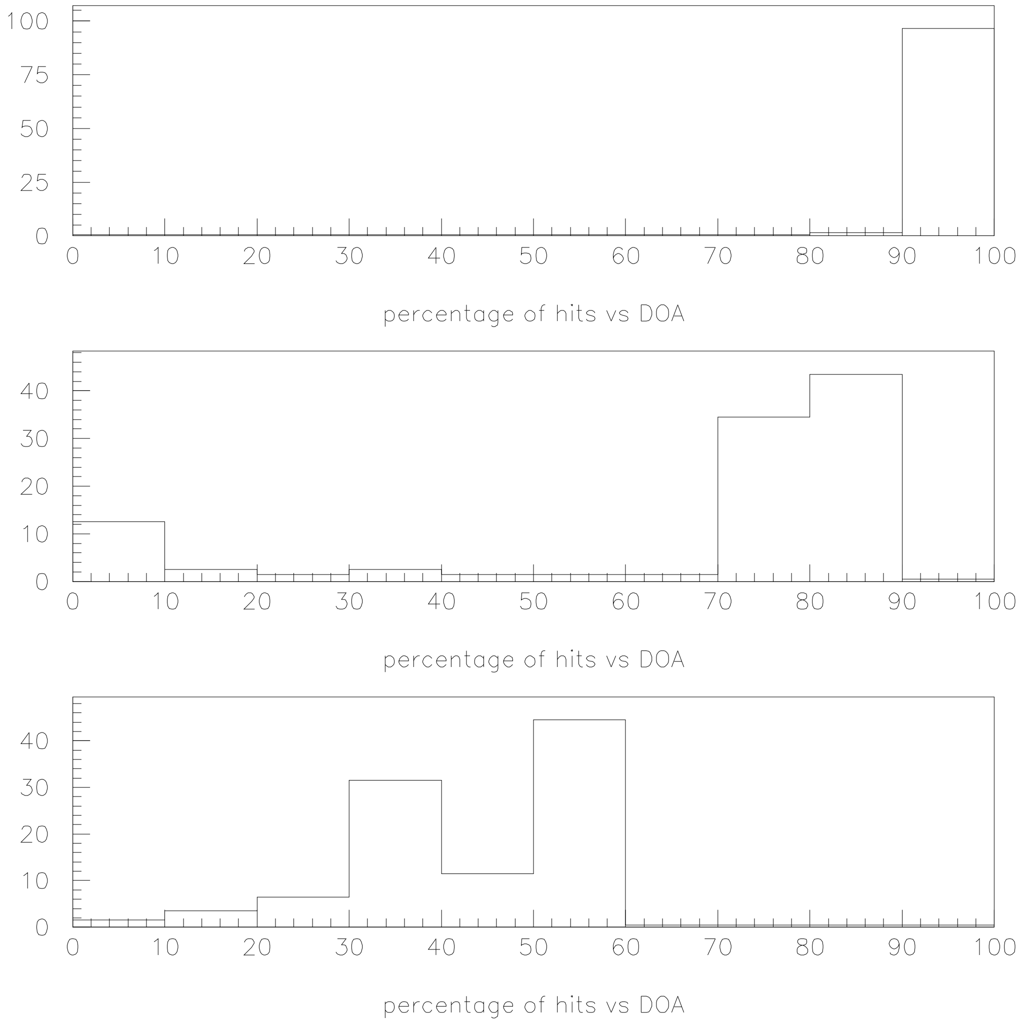

Figure 1: Event types (see text).

difference allowed between two tubes is the time it would take a flat light front to cover the intervening distance when its plane is perpendicular to the line linking the tubes. This is the slowest-case scenario since any other direction of travel for the front will result in a shorter time interval, including zero when the normal to the front is perpendicular to the line joining the tubes and they are, thus, hit simultaneously (note that any knowledge of the incoming direction of the shower may later be used to further constrain the association criterion).

Then, a cut is imposed so that those hits with a low degree of association (DOA, say, less than 50%) are thrown away. I have carefully studied how this degree of association is distributed among the hits of many individual events and have grown to picture the most typical specimens that make up the fauna of our data. Whatever the results, it is an interesting educative exercise (similar to looking at individual events in the Event Display).

3 Results

Lets look at a selected sample of event types: in Figure 1 I have plotted the percentage of PMTs as a function of the degree of association DOA. Each plot corresponds to a single event:

- Event 1 is a nice and well-behaved event, where most of the signal is in the 90–100% band (more than 90% of the hits are associated to 90–100% of the others).

- Event 2 is representative of a main shower overlapping with one or more small clumps: ~80% of the shower falls in the 70%<DOA<90% interval (the average DOA of this region, 80%, is consistent with the size of its population) whereas more than 10% of the signal falls in the DOA<10% interval.

- Event 3, however, seems to point towards two near-sized showers, with most of the signal in the middle region (degrees of association between 30% and 60%); this could be caused by, for example, two showers worth ~50% and ~40% of the image plus tiny noisy contributions.

Based on this concept and on the top distribution shown in Figure 2, I decided to apply a cut of DOA>50% to the data. The results are summarised in Tables 1 and 2 (results after a second-pass DOA>80% cut are included under DOA50+80):

Out of a total 10001 events, 676 were not fitted when no cuts were applied. The number of no-fits decreases to 148, 123 and 112 when the GBY, DOA50 and DOA50+80 cuts, respectively, are imposed. The top of Table 1 shows the results for all the events, whereas the middle and bottom portions show the results for the events not fitted initially but recovered by the relevant cut (RECOVERED) and for those events which were always fitted (ALWAYSFIT).

The fact that both the median Δ_{eo} and χ^2 improve with the cuts in the "ALL" population (ALL = RECOVERED $\bigcup$ ALWAYSFIT) but that they are not better for the subset of RECOVERED events implies that the angular resolution and the quality of the fit have indeed improved for those events that we used to fit in the standard procedure.

A much greater proportion of fitted events fail the Δ_{eo} half-fits in the standard procedure than is the case when the time cuts are applied and therefore the resulting Δ_{eo} values are not quite comparable. Table 2 makes a fairer comparison by using only the events which pass the Δ_{eo} half-fits in the standard procedure: the results get even better.

An indication of the effect of the cleaning procedure is given by the three distributions in Figure 2 where the mean degree of association grows from 84.1 to 94.6 (after the first DOA pass) to 96.8 (after the second pass).

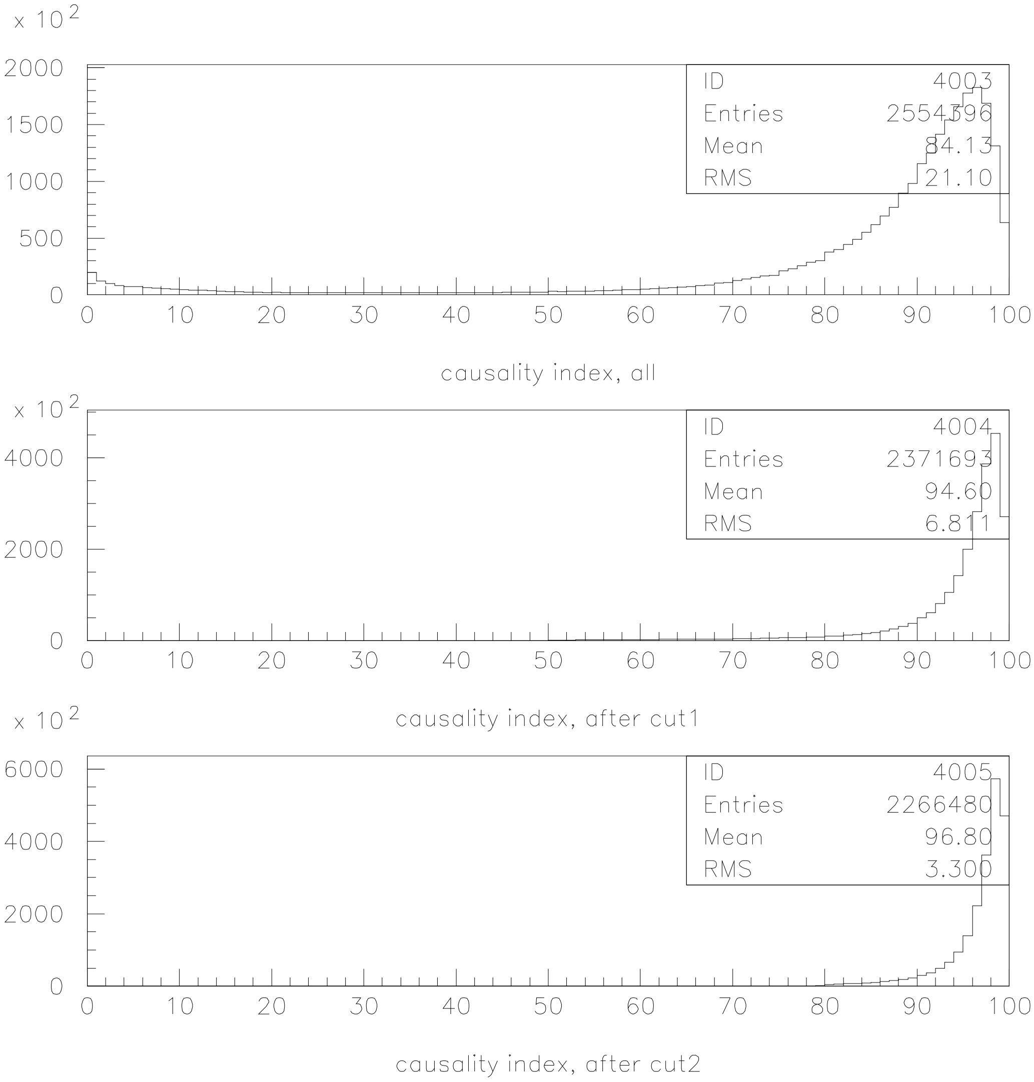

Figure 2: Distributions of the degree of association (DOA, called causality index in the plot) for 10000 events.

4 Conclusions

The results presented here call for some sort of cleaning method on an event-by-event basis. It can be argued which particular implementation or combination is better (the relative merits of two methods, GBY and DOA, have been exposed here) and what refinements must be added but there is no doubt about the necessity of such cleaning: we gain in angular resolution as a consequence of presenting the fitters with cleaner fronts. We do better on the events that we usually fit and, in addition, we recover lost events (some if not all of them may be useful).

	No cuts	DOA50 cut	GBY cut	DOA50+80
ALL				
events fitted	9325	9878	9853	9889
χ^2	3.35, 6.14	1.30, 1.63	1.50, 2.14	1.02, 1.04
Δ_{eo} failed	4153	578	663	472
$\Delta_{eo,0}$	3.54 [5848]	3.47 [9423]	3.49 [9338]	3.25 [9529]
$\Delta_{eo,40}$	2.37 [3224]	2.16 [4132]	2.18 [4131]	2.01 [4114]
$\Delta_{eo,50}$	2.09 [2549]	1.90 [3120]	1.92 [3127]	1.77 [3111]
RECOVERED		553	528	564
χ^2		1.68, 2.08	2.21, 3.60	1.16, 1.19
Δ_{eo} failed		90	91	88
$\Delta_{eo,0}$		5.16 [463]	5.33 [437]	4.53 [476]
$\Delta_{eo,40}$		2.67 [98]	2.80 [98]	4.51 [472]
$\Delta_{eo,50}$		2.38 [61]	2.47 [60]	3.32 [221]
ALWAYSFIT		9325	9325	9325
χ^2		1.28, 1.59	1.46, 1.99	1.01, 1.04

Table 1: Comparison of χ^2 (mean, rms) goodness of the fit as well as of the median Δ_{eo} (Nfit>0,40 and 50) for varios cuts. The comparison of Δ_{eo} is not totally fair because in each case a different number of events failed to produce a Δ_{eo} value (see Table 2 for a fairer comparison). Shown in brackets are the number of events remaining after the respective Nfit cut . Here, GBY is Gaurang's cut which rejects hits further than 150 ns from the median time of the event. RECOVERED refers to the subset of events not fitted in the standard procedure (no cuts) but fitted when the relevant cut was applied. ALWAYSFIT refers to the subset of events fitted both with and without cuts.

	No cuts	DOA50 cut	GBY cut	DOA50+80
χ^2	2.33, 4.12	1.24, 1.47	1.39, 1.89	0.97, 0.95
Δ_{eo} failed	0	24	38	37
$\Delta_{eo,0}$	3.54 [5848]	2.94 [5824]	2.97 [5810]	2.73 [5811]
$\Delta_{eo,40}$	2.37 [3224]	2.05 [3221]	2.08 [3220]	1.89 [3196]
$\Delta_{eo,50}$	2.09 [2549]	1.80 [2557]	1.82 [2558]	1.68 [2546]

Table 2: Comparison of χ^2 (mean, rms) and median Δ_{eo} (Nfit>0,40 and 50) for the 5848 events which do not fail the Δ_{eo} fits in the standard procedure. Shown in brackets are the number of events involved in each case.