Sampling and curvature corrections from Milagro data itself

Javier Bussóns Gordo

December 21, 1999

1 Introduction

I originally set out to simply calculate a new set of curvature and sampling corrections from the data itself. These would substitute those estimated in the past for Milagrito and would become the defaults for Milagro in its stable configuration. However, the scope of the work has broadened a tad bit and the following list of topics is discussed in this memo:

- Estimate of individual time pedestals from the data by centering the medians of individual PMT time residual (T_χ) distributions.
- Calculation of a convergent set of PE-level-dependent weights to be applied in the fit to the shower front (Angle Fit). This is done by studying the rms spread of the T_χ distributions as a function of light level.
- Selection of events with good core location for the reconstruction of the shower front curvature.
- Sampling corrections.
- Effects of playing with the time cut (SIGcut) imposed on events arriving late with respect to the front fit as well as with the light level (PEcut) required for a PMT to be considered part of the cleaned event.
- Angular resolution achievable with optimum parameters (Δ_{eo} vs Nfit curves, etc.).

2 Database and Analysis

Although a number of engineering-mode runs were used during the early stages of the analysis, the results presented here accrue from the following four "Save" files taken after the 1999 summer repairs: 1667/0037, 1668/0007, 1692/0001 and 1697/0207. All in all, they comprise 304,594 events.

A brief recapitulation of the standard Milagro reconstruction code here will make the explanation of my analysis clearer:

After the dead PMTs are searched for (GetBadPMTs), a series of defaults are set (InitControl), such as core and angle fitters, sampling corrections, TOT-to-PE conversion, etc., before the fiducial PMT coordinates are retrieved (GetCoords) and the data read loop (MilagroEvent) is started. Each event in the loop is then calibrated (Calibrate) via edge-finding (GetEdge) and time-correcting (ElecSlew) routines. Here, by the way, is were the time pedestals are read in and used. Finally, each event goes through a reconstructon suite (Recon) which includes the core fitter, the sampling and curvature corrections to the arrival times, the angle fitter, the combination of the top and bottom angle fits, muon search, gamma-hadron separation, ...

For the purpose of this memo, the following are the relevant settings, both as I found them (standard code) and as I chose to use them:

* Electronic Slewing: the default corrections (the slewing curve of a typical PMT, which was assumed to hold true identically for all PMTs) are untouched. They must be changed to newer values as soon as these are available.

* Time pedestals: the offsets of the individual PMT "clocks" are estimated from the data. Values from laser calibration should replace these.

* Weights: they are basically designed to give brighter PMT pulses more weight in the Angle Fit than that given to dimmer signals. The rms spread in the arrival times (with respect to the fitted value or, for that matter, to the time of the hottest tube) is broader for dimmer than for brighter signals (top right panel in Figure 4) so the quantity:

$$w = \frac{1}{(RMS)^2}$$

can be used as the weight. The default was:

$$RMS = A + B\ R^{\alpha} + C\ (PE)^{\beta} + D\ (PE)^{\gamma}\ R^{\delta},$$

with A = 0.90404, C = 4.4354, β = -0.64704 and the rest are zero and where R is distance to the core. In this memo, the following functional form is used instead:

$$RMS = a + b\ \log(PE) + c\ \log^2(PE) + d\ \log^3(PE).$$

An iterative process starting with a = 1, b = c = d = 0 is found to converge to a particular 3rd-order polynomial.

*SIGcut: Another use of the RMS time spread is the following: For late hits (again with respect to the fitted value), that is for $T_\chi < 0$, a PMT signal will be kept if:

$$|T_\chi| \leq RMS\ SIGcut.$$

Here, in each of the –at most five– passes, the SIGcut represents the number of RMS's in time that a signal is allowed to deviate from the fitted event time to be kept as part of the event. A larger SIGcut allows more late light in. The defaults were: 4.0, 3.5, 3.0, 3.0, 2.5 (for the five passes). These values were quite conservative due to the lack of calibrations. In this memo, various sets of values are tried and their implications studied.

*PEcut: in each of the five passes, this is the minimum number of photoelectrons a pulse must contain for it not to be rejected. The defaults were: 1.75, 1.50, 1.25, 1.00, 0.75. In this work, other combinations are tried out.

*Curvature correction: the default, as per MonteCarlo simulations, was a correction –which is linear with distance to the core– of 4ns/100m. Here, both linear and second-order polynomials are considered.

*Sampling correction: the default, calculated using the peaks of the T_χ distributions as a function of PE, was:

$$\Delta T = 7.02\ \log(PE) - 41.8\ (PE)^{0.097} + 35.5.$$

In this memo, the medians rather than the peaks are used, the reason being that the former do not depend on the function proposed for the fit. Only histograms with more than 1000 entries are considered for good measure. The functional form chosen in this memo is, again, a polynomial in log(PE).

*Other changes:

- TOT-to-PE conversion: for tubes missing in the ADC_TOTPE look-up table, fake values (those of the next tube in the table) are made up so that tubes fixed in the summer'99 repairs can be utilised.

- Core fit: the standard fitter is used but, since the information from the bottom layer is never used, the bottom pass is removed.

- Angle fit: the standard fitter is used in the top layer (AngleFit_FlatTop) but a fit to the bottom layer (AngleFit_ChiSq) is added. Subsequently, the top and bottom fits are weighted by the relative number of tubes fit in each layer (Nfit_T, Nfit_B).

- Four alternatives to the traditional Δ_{eo} analysis are added for a better evaluation of the angular resolution: Δ_{vr} (alternating vertical rows), Δ_{hr} (alternating horizontal rows), Δ_{X1}, Δ_{X2} (rectangular detector is divided by each one of its two diagonals).

- A "clump" cut was devised in order to search for events with core on the pond. The series (n,R1,R2) corresponds to asking that the "n" hottest tubes are all contained in the circle of radius "R1" meters with centre in the pond centre and that the farthest all "n" tubes can be from one another is "R2" meters: in other words, I am looking for clumps of the hottest tubes.

- In my analysis, a preselection of events near the zenith and with a high-enough number of PMTs in both the top and the bottom angle fit is carried out in order to minimise the bias introduced by the lack of proper corrections (which are the unknowns) in the first iterations of the process (we start with no corrections at all, remember). See detailed cuts in the Results section.

Surviving reconstructed events are then used to produce two types of histograms:

- T_χ distributions for each PMT (only hits with $PE > 20$ are used here), from which we determine the time pedestals and correct them (shift the peaks –or in my case, the medians– so that all of them match up at T_χ=0).

- T_χ distributions for each [log(PE),R] bin, where R is the distance of a hit PMT to the fitted core: the PE/R plane is quantised as a grid with 20 R bins (covering the range 0 to 100m in steps of 5m) and 40 log(PE) bins (from -1 to 3 in steps of 0.1) and each of the resulting 800 T_χ histograms is searched for its mean, rms, peak and median, among other variables. Histograms with less than 1000 entries are obviated and for the survivors I take the standard deviation (rms) from the mean [peak, median] as a measure of the error in the mean [peak, median]. Remember that the standard deviation from the value q is:

$$\sigma_q = \sqrt{\frac{\sum_{j=1}^{N}(q_j - q)^2}{N-1}}.$$

Since the RMS itself is one of the variables we are after, we must place an error estimate on it as well. This is:

$$\sigma_{RMS} = \sqrt{\frac{\sigma^2}{N-1}},$$

where σ^2 is the variance.

Now we are ready to average the interesting variables (means, rms, peaks and medians) over the full log(PE) range for each R interval and thus obtain each variable as a function of R (and conversely to obtain them as a function of log(PE)). These averages are weighted by the relevant errors:

$$X_{mean} = \frac{\sum_{i=1}^{bins} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{bins} \frac{1}{\sigma_i^2}},$$

and carry an uncertainty:

$$\sigma_w = \sqrt{\frac{1}{\sum \frac{1}{\sigma_i^2}}},$$

that I will take into account in the fits.

The point here is that we can think of the deviations from a planar shower front (which is what we try to get in our Angle Fits) as due to two separable effects: something which is only R-dependent, namely the curvature and something which only depends on PE-level (I will call it sampling, not really knowing what other effects may be masked in it).

At this point one can use:

- the "RMS vs log(PE)" curve to determine the weights,
- the "medians (or means or peaks) vs R" curve to study the curvature of the shower front, and
- the "medians (or means or peaks) vs log(PE)" plots to correct for sampling effects.

In the fairly non-gaussian regime we find ourselves in, I prefer the much more stable medians to fit-dependent peaks and to unpredictable means.

3 Results

*Preselection Cuts:

$$\Theta_T < 25, \;\; \Theta_B < 25, \;\; |\Theta_T - \Theta_B| < 3,$$

$\mathrm{Nfit}_T > 50$, $\mathrm{Nfit}_B > 50$.

*Time Pedestals: in the very first iteration, all time corrections (offsets, curvature, sampling) are set to zero and all the weights are assumed equal (RMS = 1, independent of PE level). The medians of the T_χ distributions are computed for each PMT and taken as the time pedestals to be used thereon. In Figure 1 I have plotted the distribution of time pedestals (tpeds) before any correction is applied, where:

$$\sum_{i=1}^{653} |tped_i| = 1960ns, \ \sum_{i=1}^{653} tped_i = -1857ns.$$

These values, which measure the magnitude and direction of the offsets, improve to (630ns, -164ns) and (582ns, -1ns) in subsequent iterations (Figures 2 and 3 respectively).

*Weights: with time offset corrections in place, the resulting "RMS vs log(PE)" curve is best fitted to a 2nd-order polynomial which serves as expression for the RMS (new weights) in the following iteration. See Figure 5 for convergence of weight calculation (3rd-order polynomials are used after the second iteration). Different sets of weights have been seen to dramatically change, for instance, the Nfit distributions (particularly in the bottom layer), hence the importance of arriving at a robust, unbiased set. This convergence has been observed in every dataset used since the inception of this analysis. The final expression arrived at is:

$$RMS = 11.53 - 4.98\ \log(PE) - 1.30\ \log^2(PE) + 0.84\ \log^3(PE),$$

which is used hereon.

Just by using optimum pedestals and weights, a greater number of events pass the selection and the average χ^2 of the Angle Fit drops from a horrendous 7.8 (rms=12.1) to a much better 1.6 (rms=3.3), with the corresponding improvement in angular resolution. Figures 6 and 7 show the total T_χ distributions before and after the pedestal corrections and optimum weights are implemented.

*Curvature: as shown at the November 1999 meeting in New Hampshire, no curvature is appreciated with the current cuts and therefore I devised a cut (clump cut) that, no matter how stringent, would find me nice events with a well-located core in the pond. It is surprising to see how few events pass a decent such cut: only 1,748 out of 304,594 pass clump (9, 25, 17) and 3,806 pass clump (5, 30, 12). Note: refer back to "Other changes" in the previous section for the interpretation of this cut. This seems to indicate that we normally have a very small sample of well-located cores in our pond. Different clump cuts have been tried ([9,30,18],[7,30,18],[5,30,18],[9,25,18],[5,30,15]) with the results given in Table 1).

The bottom panels of Figure 8 show the curvature of the shower front as extracted using the clump (9,25,18) cut. A linear fit yields:

$$(\Delta T)_{curv} = 0.650 - 0.063\ R.$$

Caveat: statistics are poor in some regions of the "median vs R" curve and I need to study a much bigger sample (problems with kahuna and disk space don't let me

Clump:	9,25,18	9,30,18	7,30,15	7,30,18	5,30,15	5,30,18
Attrition:	1364	2174	2462	4039	6029	8715
Θ:	649	1038	1130	1834	2595	3719
Nfit_T:	599	959	1016	1640	2248	3194
Nfit_B:	598	958	1016	1639	2244	3186
D (m):	7.1	7.5	8.5	9.7	11.4	12.5
χ^2:	1.47±3.17	1.35±3.17	1.30±2.92	1.27±3.06	1.43±3.28	1.43±3.17
χ^2_B:	1.02±1.46	1.02±1.50	1.04±1.46	1.04±1.53	1.02±1.38	1.04±1.49
Δ_{eo}:	1.55±2.51	1.39±2.28	1.43±2.26	1.37±2.19	1.50±2.34	1.51±2.36

Table 1: Results of sevaral clump cuts: events surviving the cut (attrition), the zenith angle cut, the Nfit cuts, distance (D) from reconstructed core to hottest tube, χ^2 for the combined Angle Fit and for the bottom layer fit and mean Δ_{eo}.

do it before the Christmas break). However, application of this correction yields the planar front seen in the bottom panels of Figure 9.

*Sampling: here we don't need to be as severe as we were when searching for the curvature. In fact, the "median vs log(PE)" curve is very insensitive to the clump cut used. Therefore I decided to use the least severe of the tried cuts ([5,30,18], which can be relaxed much more, I believe). This way, a 2nd-order polynomial fit in the range $0. < \log(PE) < 2.5$ yields the following sampling correction:

$$(\Delta T)_{median} = -6.84 + 6.72\ \log(PE) + 1.62\ \log^2(PE).$$

Comparison of respective top panels of Figures 8 and 9 shows the effect of the sampling correction.

*SIGcuts and PEcuts: Finally, Tables 2 and 3 list the results of the optimisation of these parameters. A set of SIGcuts(1.5,1.25,1,1,1) are taken as optimum and used in the PEcut optimisation. Here, as expected, the angular resolution improves as the cut becomes more demanding. Intuitively, it is not surprising to see values of Δ_{X1}, Δ_{X2} consistently much larger than those of Δ_{eo}, Δ_{hr}, Δ_{vr}, given the greater connectivity of the former sets of sub-detectors (the latter are much more interleaved).

4 Conclusions

Optimum values of time pedestals, angle-fit weights, curvature and sampling arrival-time corrections, SIGcuts and PEcuts have been found for the current Milagro detector data. With them, median values of Δ_{eo} of the order of 0.7^o are achieved (0.69 for a typical PEcut[4,3,2,1,1,]; see Figures 10 and 11).

From here, we should plug the parameters arising from laser calibrations as soon as possible, re-do this analysis on an appropriately large and typical dataset and implement the new values. A joint strategy that takes care of all different aspects of the detector calibration (laser measurements, MC simulations and measurements from the data itself) must be agreed upon so routine calibration checks are performed with the desired frequency.

SIGcut	Δeo	Δvr	Δhr	ΔX1	ΔX2
1, 0.75, 0.5, 0.25, 0.25	0.75	0.74	0.77	1.56	1.72
1.5, 1.25, 1, 1, 1	0.72	0.70	0.73	1.48	1.63
2, 1.75, 1.5, 1.25, 1	0.73	0.74	0.76	1.51	1.61
3, 2, 1, 1, 1	0.72	0.71	0.74	1.49	1.65
4, 3, 2, 1, 1	0.74	0.74	0.76	1.53	1.64
5, 4, 3, 2, 1	0.78	0.82	0.82	1.60	1.72
6, 5, 4, 3, 2	0.86	0.89	0.88	1.70	1.79
6, 6, 6, 6, 6	0.98	1.02	0.98	2.02	1.95

Table 2: SIGcut optimisation (Δ values in degrees).

PEcut	Δeo	Δvr	Δhr	ΔX1	ΔX2
1, 1, 1, 1, 1	0.74	0.76	0.76	1.52	1.61
2, 1, 1, 1, 1	0.73	0.72	0.72	1.48	1.61
2, 2, 2, 2, 2	0.73	0.72	0.72	1.48	1.61
3, 2, 1, 1, 1	0.72	0.70	0.73	1.48	1.63
3, 3, 3, 3, 3	0.71	0.70	0.73	1.47	1.64
4, 3, 2, 1, 1	0.69	0.70	0.72	1.49	1.63
4, 4, 4, 4, 4	0.69	0.70	0.72	1.50	1.63
5, 4, 3, 2, 1	0.67	0.68	0.69	1.47	1.58
5, 5, 5, 5, 5	0.67	0.68	0.68	1.47	1.58
6, 5, 4, 3, 2	0.66	0.66	0.67	1.46	1.54
6, 6, 6, 6, 6	0.66	0.66	0.67	1.47	1.55
9, 9, 9, 9, 9	0.62	0.64	0.64	1.50	1.48

Table 3: PEcut optimisation (Δ values in degrees).

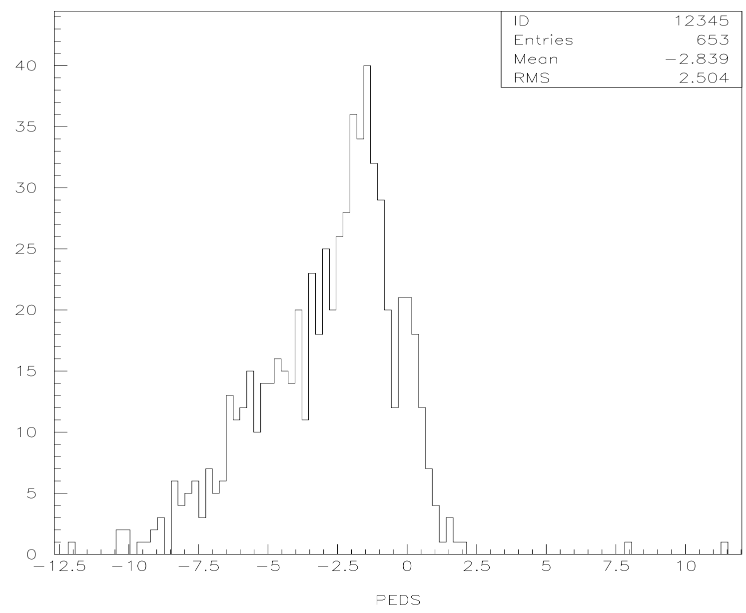

Figure 1: PMT Time Offsets.

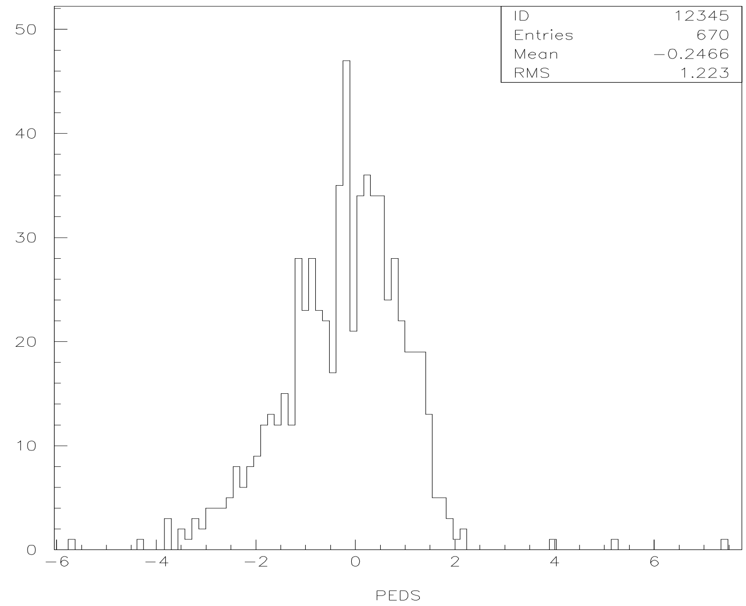

Figure 2: First correction of time pedestals.

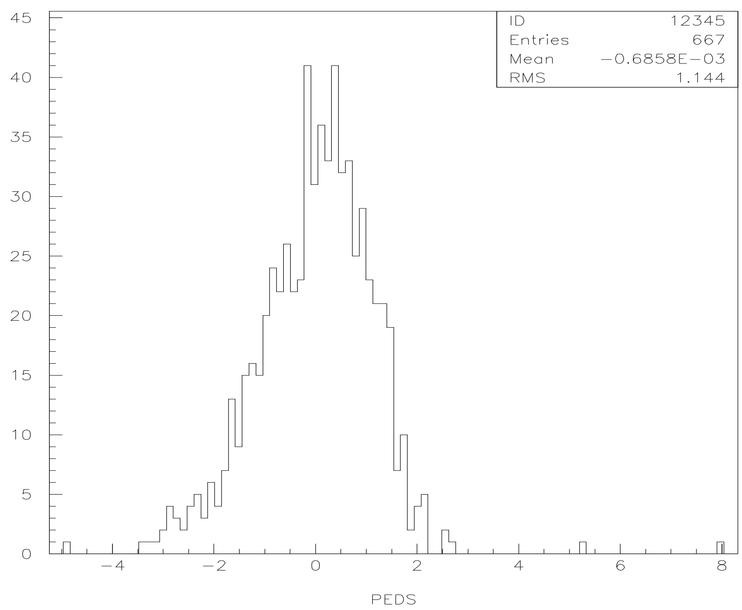

Figure 3: Second correction of time pedestals.

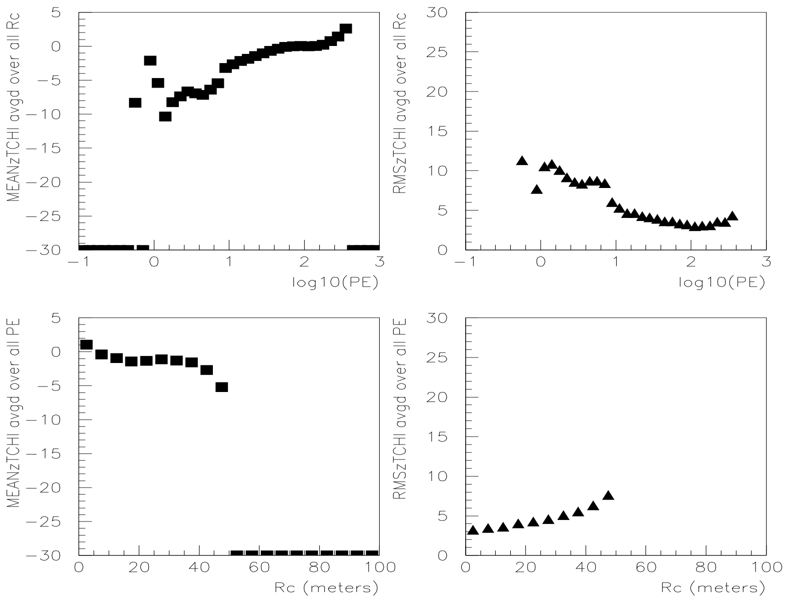

Figure 4: Typical mean and RMS curves. For convenience, points where the histograms are not sufficiently populated are artificially set to -30 or skipped altogether.

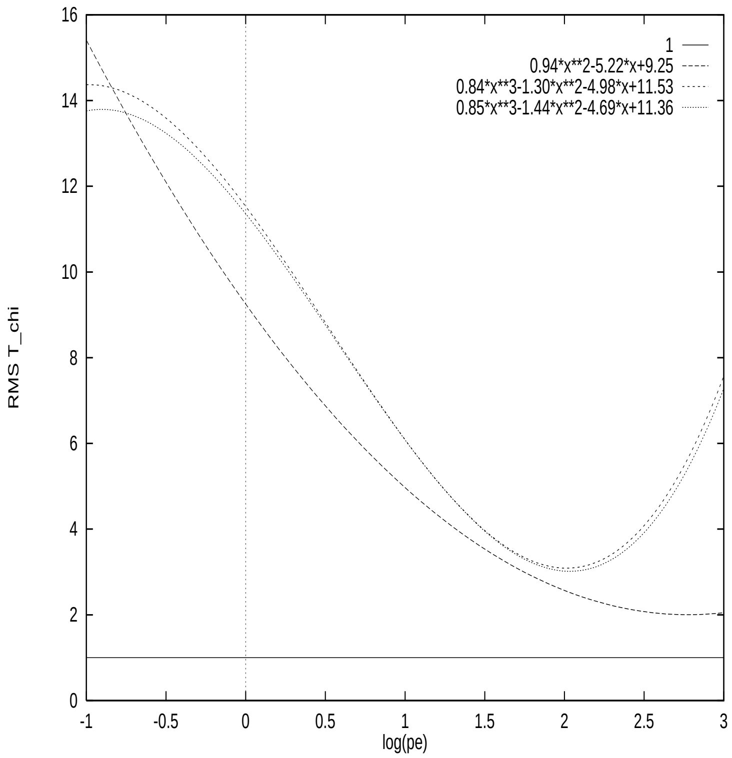

Figure 5: Fits to RMS curves. Starting with a flat linear dependence (RMS = 1), subsequent iterations rapidly converge to the 3rd-order polynomial shown here.

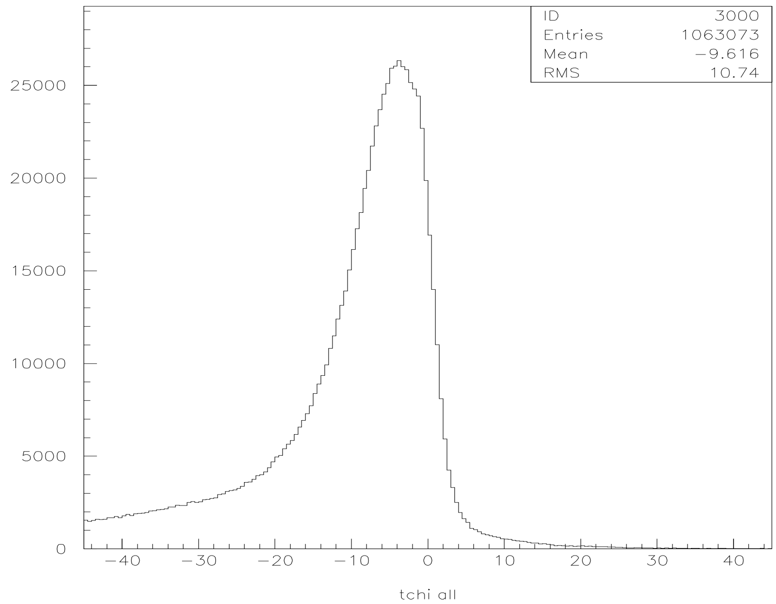

Figure 6: Initial distribution of time residuals.

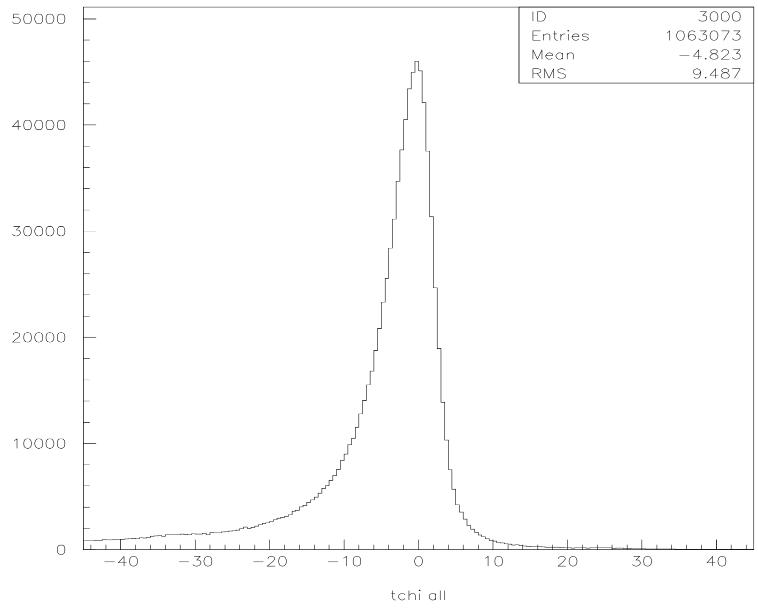

Figure 7: Distribution of time residuals corrected for time pedestals and with optimum weights.

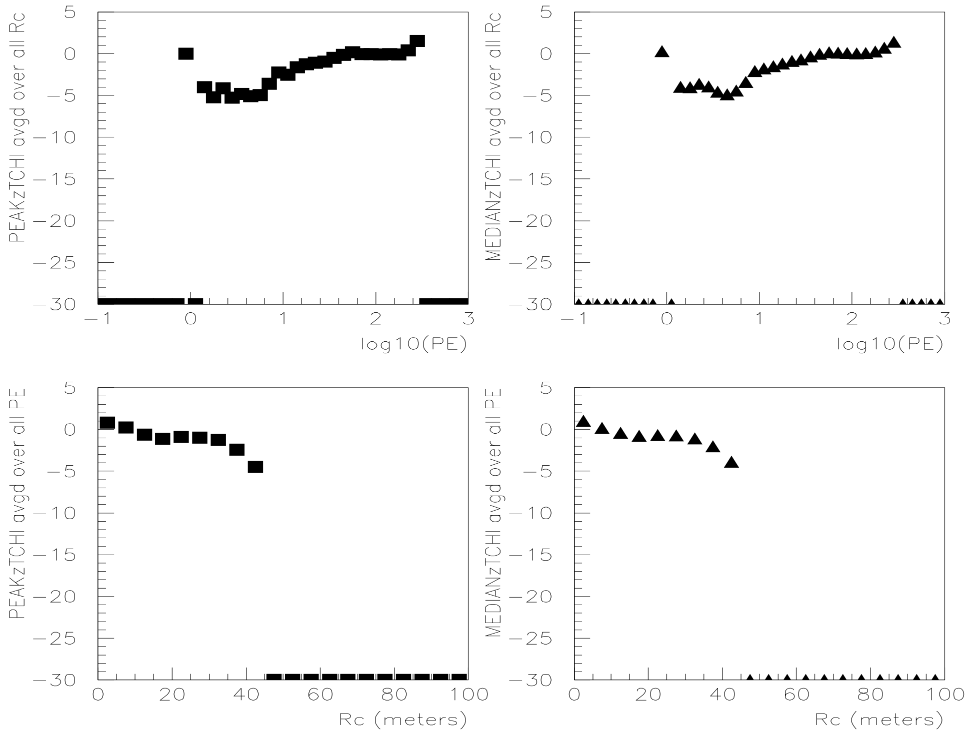

Figure 8: Distributions without curvature and sampling corrections: appreciable shower-front curvature is seen in the bottom panels by virtue of the clump cut. Like before, non-usable points are artificially set to -30 for convenience.

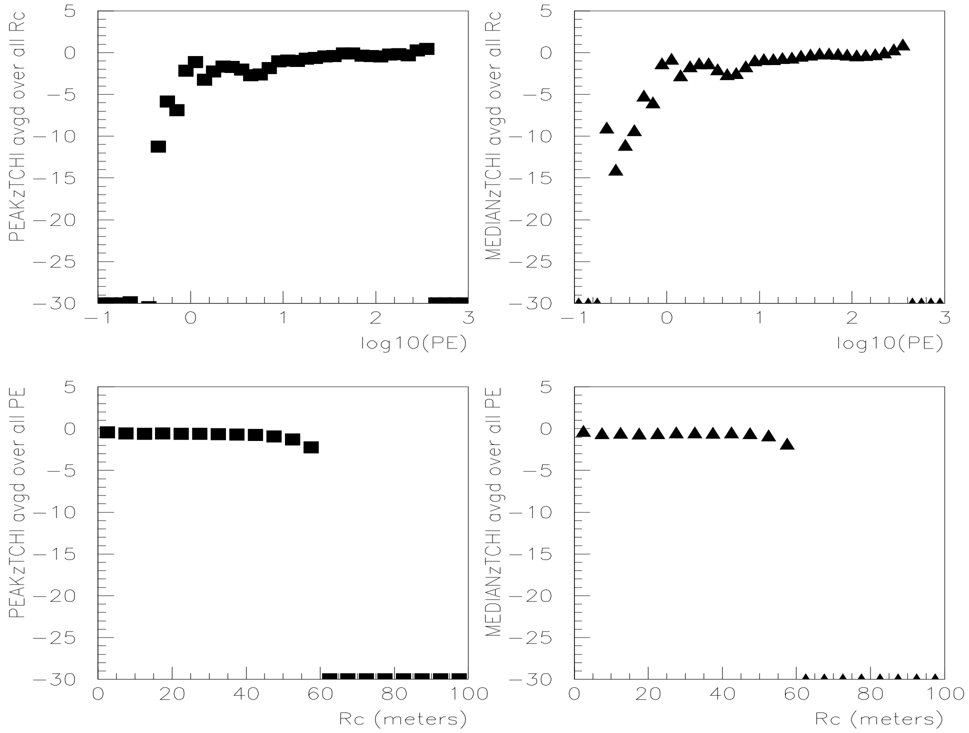

Figure 9: Distributions corrected for curvature and sampling effects.

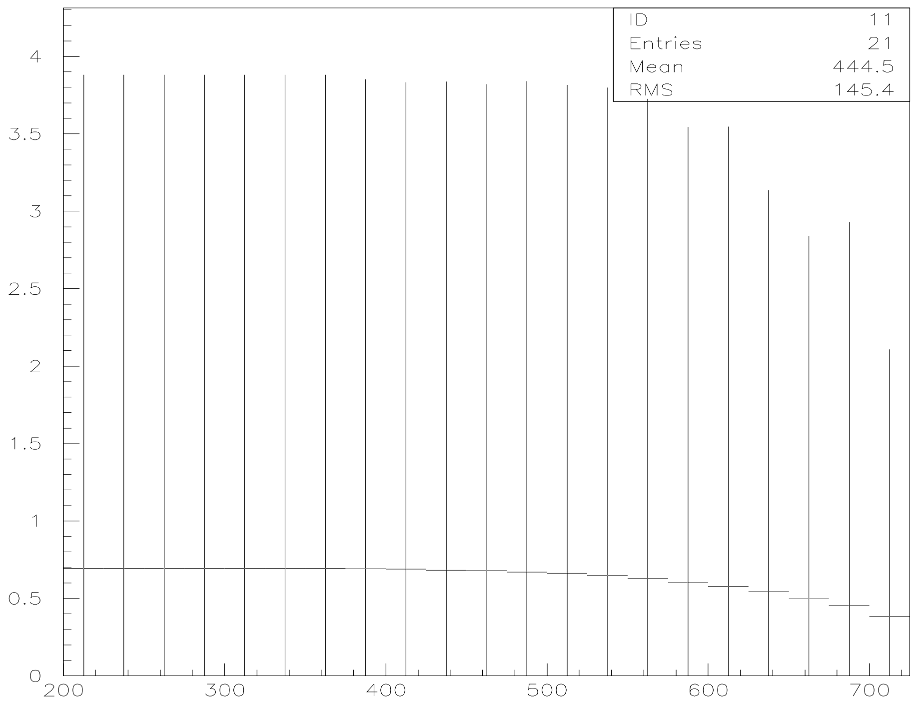

Figure 10: Differential plot of median Δ_{eo} as a function of Nhit band.

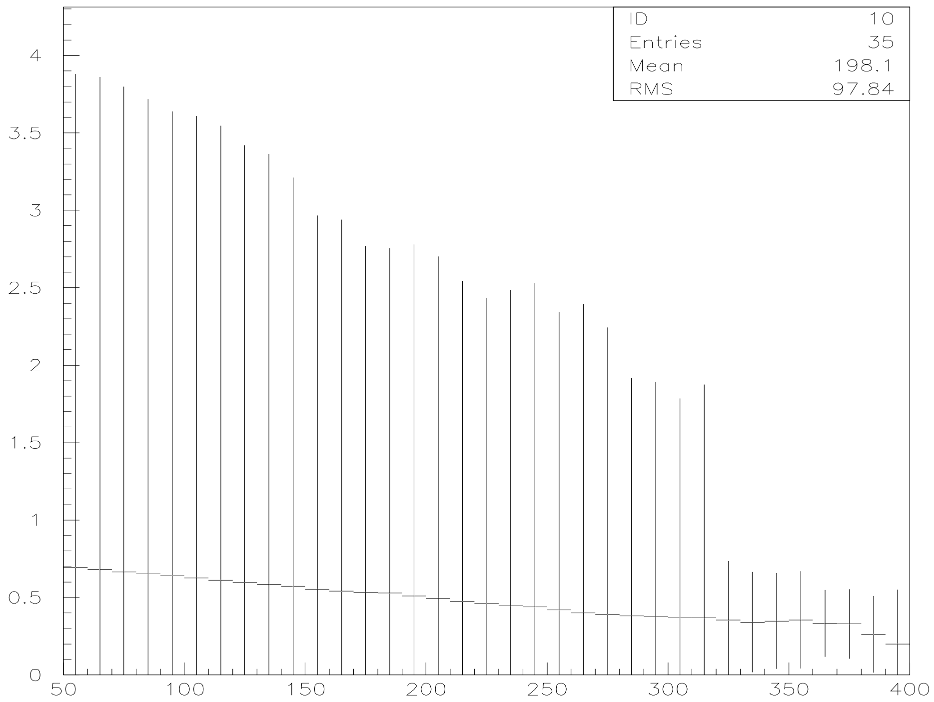

Figure 11: Integral plot of median Δ_{eo} as a function of Nfit cut for SIGcuts(1.5,1.25,1,1,1) and PEcuts(4,3,2,1,1). Note: errorbars should become smaller on a larger dataset.