Simple Corefitting with Outriggers

Tyce DeYoung

January 29, 2003

Abstract

This note describes the core fit algorithm CoreFit_ORCOM (OutRigger Center Of Mass), available in CVS. The algorithm is designed to be a simple and fast replacement for the current CoreFit_OFF routine, and does not represent the maximum possible improvement from the outrigger information. The algorithm decides whether pond (top-layer) or outrigger information is most useful, and then uses the center of mass of the chosen layer to fit the core. This algorithm is compared to other fast core fitters presently available. Implications for gamma/hadron separation and sensitivity to amplitude calibration errors are investigated.

1 Introduction

The core fitter described in this note is intended to use the outrigger information in a simple and fast manner. The solution is calculated analytically, with no minimization required. No timing information is used, only amplitudes. A simple decision is made whether to use outrigger or pond information to find the core. If the outriggers are used, the center of mass is returned for the core location, and if the pond is used an existing pond core fitter is called. Although crude, this algorithm provides significant improvement over the current fitter.

Unless otherwise noted, all Monte Carlo results in this note are from gamma version 3.2, reweighted to account for core distance biasing, isotropically incident, and on a primary spectrum of $E^{-2.4}$ from 100 GeV to 100 TeV. The standard VME trigger conditions were imposed. The normalization is arbitrary. A total of 43,127 showers are included in this set.

2 Description of the Algorithm

The algorithm in this note is as simple as possible. The first stage is to decide whether to use the outriggers or the pond (the air shower layer) to fit the core. A more sophisticated algorithm would of course attempt to combine this information for the large fraction of events in which both sets of tubes contribute significantly to the event. Doing so would probably improve reconstruction accuracy, especially for cores at the edge of the pond, around 40–70 m from the center (see Fig. 5). For this reconstruction, however, we simply use one or the other.

One complication is that the outriggers have much higher relative noise rates than the pond tubes. The outrigger edge finder window is 600 ns wide, and a typical outrigger noise rate is 40 kHz. This gives an expectation of 4.2 random noise hits in the 175 outriggers. This should be compared to a median 9.6 outriggers hit in a gamma shower (and a most probable value of only 1 hit!), as shown in Figure 1. For sub-TeV gammas the median is only 2.9 hits.

Inspection of data events in Tony's event display indicates that most random hits (i.e., hits which in the combined judgement of Andy, Greg, and Ty were isolated from the real event, especially lone outriggers far from the pond in small events) are single-PE hits. We therefore impose a threshold of 1.5 PE on outrigger hits for consideration in the pond/outrigger decision algorithm. However, it is possible that noise produced by light leaks in the outrigger tanks may have a much higher amplitude spectrum. David Williams has developed an online tube veto to reject high-rate (> 100 kHz) tubes, and this will be implemented into the core fitter. Some additional comparison of day- and night-time data files is in order, though, to better characterize the amplitude spectrum produced in leaky outriggers.

If no outriggers are hit, the choice of algorithm is clear; otherwise, a decision must be made. The ratio of the number of outriggers with multi-PE hits to the number of air shower tubes hit is found to be a fairly good indicator of core distance, as shown in Fig. 2, and of course is extremely easy to calculate. As implemented, events below a threshold ratio of 0.05 are sent to the pond fitter, and the outrigger center of mass is used for events with ratios above 0.05. This value is chosen by inspection of Fig. 3, which shows the accuracy of the two methods as a function of the Nhit ratio.

For events determined to be "off-pond," we treat each outrigger as a point mass equal to the measured amplitude and calculate the center of gravity of the event. This approach has previously been shown to be a reasonably good estimate of the core location by both David Noyes and

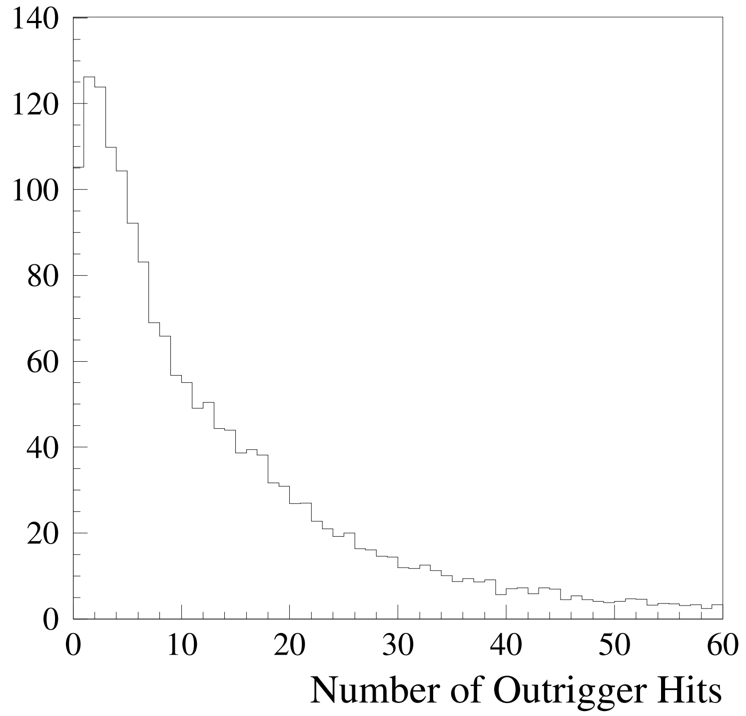


Figure 1: The number of outriggers hit in gamma showers. The median is 9.6 hits, and the mean is 16.0.

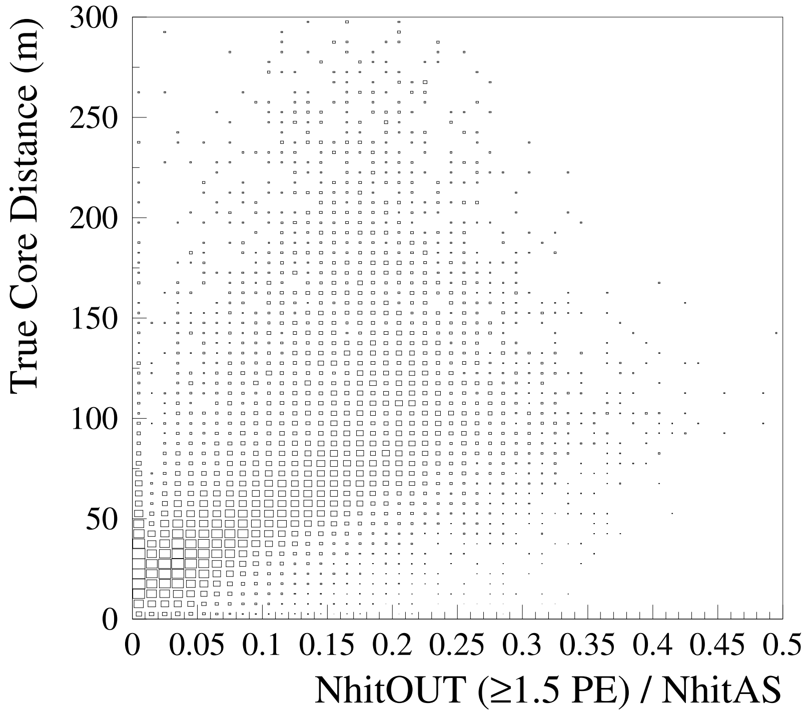


Figure 2: The distribution of the ratio of multi-PE outrigger hits to air shower tubes hit, as a function of true core distance from the center of the pond. Noise is not included in the simulation. With the exception of a population of events without outrigger hits, extending surprisingly far from the pond center (the leftmost column of the plot), the ratio is a good indicator of the true distance. For those events, of course, the point is moot: there is no choice but to use the pond for reconstruction.

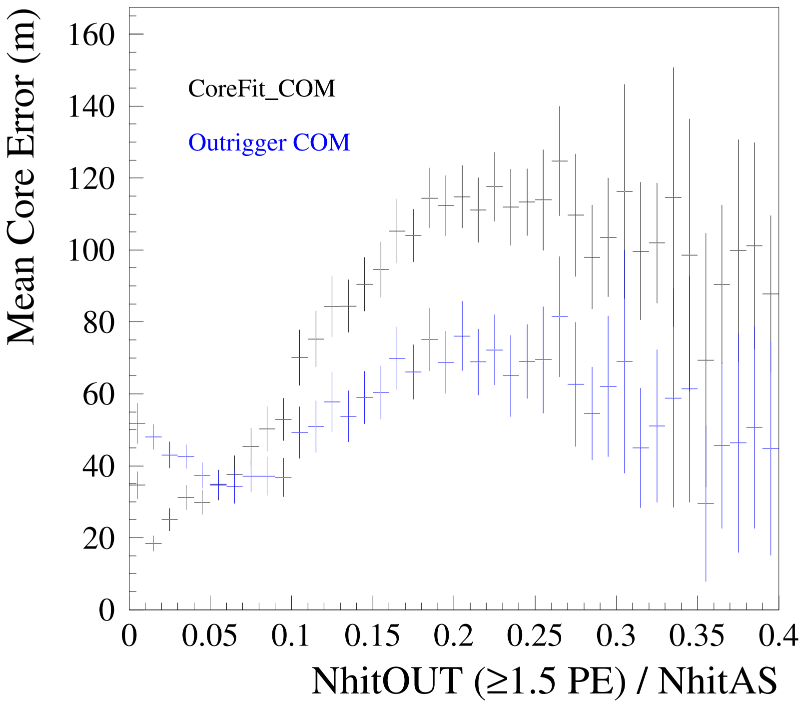


Figure 3: The accuracy of CoreFit_COM and the outrigger center of mass as a function of the ratio of the numbers of multi-PE outrigger hits and air shower tubes hit. CoreFit_COM is used below 0.05 and the outrigger COM above. The plot of median errors shows crossover at the same point.

myself. Note that for many of these "off-pond" events, the center of mass is actually on the pond. This finesses the issue of 'border' events, but one would expect a bias away from the pond center due to the lack of couterweighting hits on the pond. (In fact this seems not to be the case; see Fig. 6.) On the other hand, many showers in this region are quite small (they are close enough to the pond to trigger even with relatively low energy deposition). This means that few outriggers may be hit, and so the event is sent to the pond core fitter even though the core is slightly off the pond; this produces a bias in the opposite direction (towards the pond). At any rate, the poorer performance in this border region is clearly visible in Fig. 5, and is a clear avenue for future improvement.

For "on-pond" events, we call the old CoreFit_COM routine, which was used prior to the installation of Greg's CoreFit_OFF. This routine attempts to find a cluster of hits near the center of mass of the hits in the air shower layer. Although better on-pond algorithms are available (e.g., CoreFit_FFT and CoreFit_GAUSS), CoreFit_COM is faster and probably near the point of diminishing returns for events on the pond (the advantage of 8 m resolution vs. 12 m resolution is not obvious).

3 Performance

The overall performance of three core fitters is shown in Figure 4. This is for the default Monte Carlo set; i.e., gammas on $E^{-2.4}$ from 100 GeV to 100 TeV. The Gaussian fitter is slightly better, mainly in sharpening up good fits. That is, both fitters are equally likely to be completely wrong (the tails are almost identical), but when the Gaussian fitter is right it is likely to be a few meters closer to the core. This is also seen in Fig. 5, where on-pond showers are better reconstructed by the Gaussian. For showers from primaries below 1 TeV (indicative of a GRB, say), the overall difference is even smaller: median error of 18.9 m for the Gaussian, vs. 19.6 m for this fitter. These compare to 26.0 m for the current fitter. In terms of speed, the OR COM fitter is extremely fast; a full reconstruction (i.e., core and angle fitting) based on OR COM runs at some 1430 Hz on a 1.4 GHz Athlon processor. The Gaussian fitter is rather slower, at 220 Hz, but still several times faster than a likelihood fitter. These speeds were determined by processing 10000 MC gamma events with and without the call to Recon2, and then subtracting the times. They are therefore approximate, but repeated trials on different machines at different times seem to give consistent results within 10–15%.

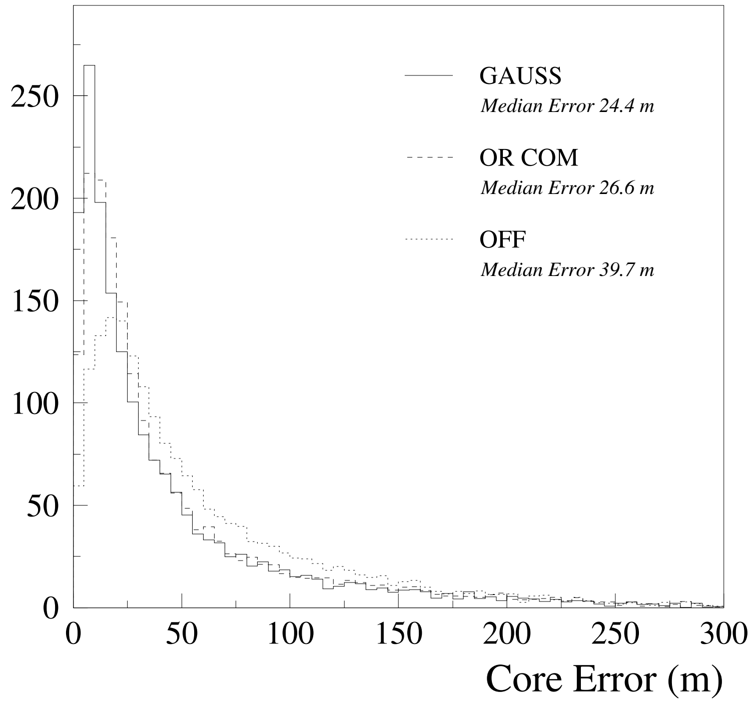


Figure 4: The accuracy of the Gaussian core fitter (Tony's), the outrigger center of mass fitter (this work), and the current fitter (OFF, by Greg). Either of the outrigger-based core fitters is a significant improvement over the current fitter.

The accuracy vs. (true) distance is shown in Fig. 5. The current fitter (OFF) suffers for cores on the pond because of its poor on/off discrimination. It does quite well for cores at 50 m distance, of course, because at that distance it is only searching in one dimension. However, either the Gaussian or the OR COM are clearly superior to the present fitter. As noted above, the Gaussian is slightly better than OR COM on the pond, and also for cores outside the outrigger field (beyond about 100 m). At those distances the Gaussian is able to extrapolate some distance beyond the outriggers, while the COM is strictly limited to be within the outrigger field. Interestingly, in the pond/outrigger border region, from about 40–70 m, the COM fitter is actually slightly better than the Gaussian (the plot shows mean errors, but the result also holds for median errors). This is because the Gaussian fitter is having trouble somewhere in its decision tree; see Fig. 6. Interestingly, the OR COM fitter tends to underestimate the distance to the core; this is true for both the pond fitter (naturally) and for the outrigger center of mass, where one would expect a bias *a*way from the 'hole' in the outrigger array. It may be that the trigger selects events in which the shower is asymmetrically dense on the pondward side of the true core, or it may be that the COM is biased toward the edge of the pond by the dense line of outriggers at the edge.

Another question of interest is whether the two fitters get the *direction* to the core correct, independent of the *distance* to the core. As shown in Fig. 7, the two fitters are very nearly equal in this regard, with the OR COM being just slightly better. This is probably connected with the distribution shown in Fig. 8 of fit core directions, where both fitters display a marked structure, presumably related to the corners of the pond. The structure is worse for the Gaussian fitter, which probably causes the slightly worse directional resolution.

The accuracy of on/off pond classification is also similar for the Gaussian and OR COM fitters. In each case, cores fit as being within 40 m of the pond center are roughly equally likely to be within or without that radius (54.7% within for GAUSS, 51.4% within for OR COM, compared to 51.1% within for the current fitter). The improvement over the current fitter comes in the accuracy of off-pond ($>$ 40 m) classification, which is 97.0% for the Gaussian and 95.5% for the OR COM. With the current fitter, only 78.2% of cores fit as off the pond were in fact off the pond. With either of the new fitters, a core fit on the pond may or may not actually be so (and bear in mind, having the geometric core off the pond does not preclude significant particle density on the pond), but a core fit as off-pond may be reliably taken to be distant.

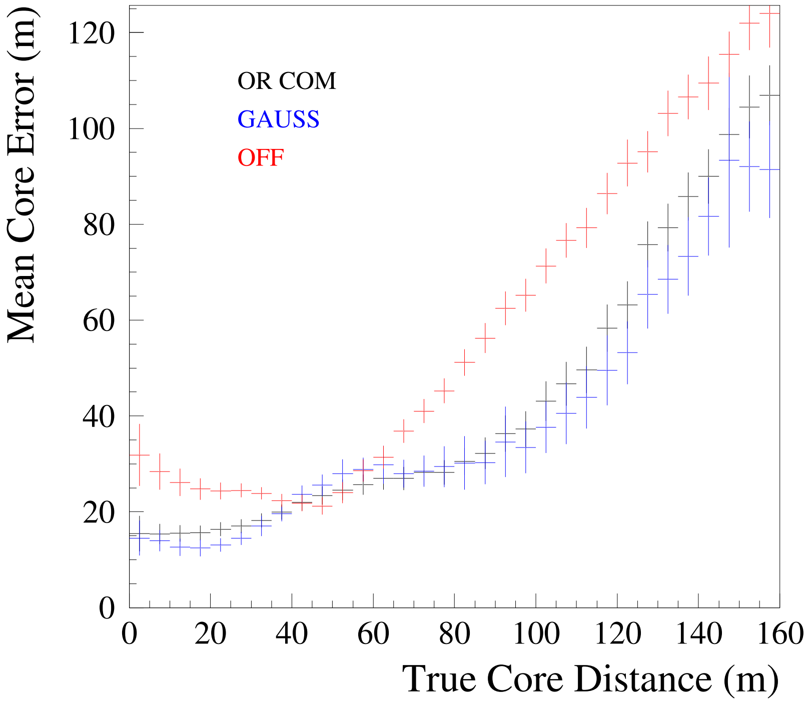


Figure 5: The accuracy of the Gaussian core fitter (Tony's), the outrigger center of mass fitter (this work), and the current fitter (OFF, by Greg), as a function of the true distance of the shower core from the center of the pond.

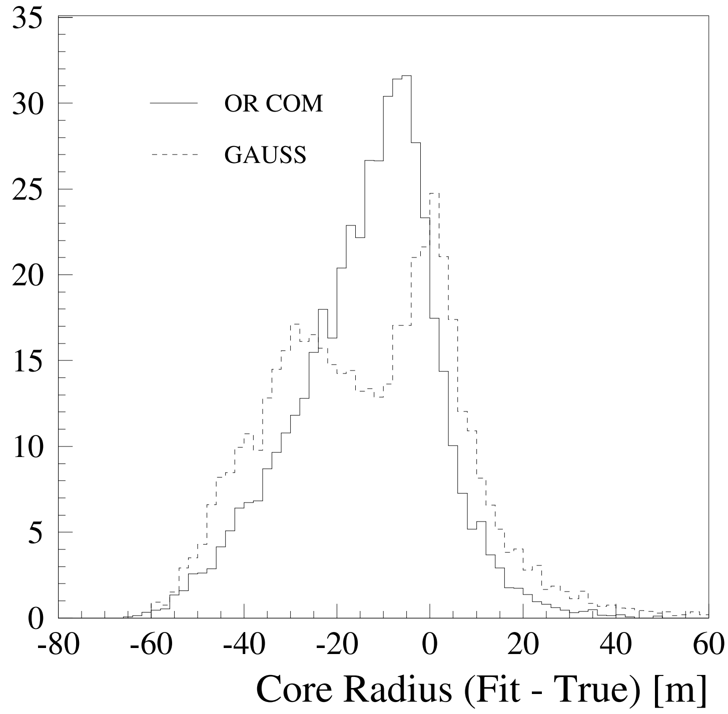


Figure 6: The accuracy of the Gaussian core fitter (Tony's) and the outrigger center of mass fitter (this work) for cores at 40–70 m distance from the pond center. The Gaussian fitter either does quite well or underestimates the distance by about 25 m; presumably this relates to an incorrect on/off decision. See also Fig. 10. The OR COM is more consistently somewhat wrong.

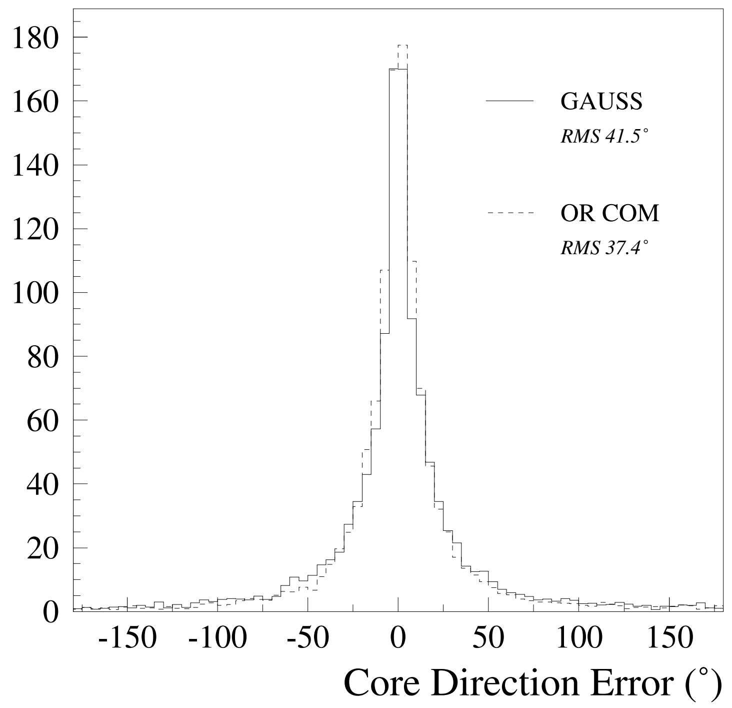


Figure 7: The accuracy with which the Gaussian core fitter (Tony's) and the outrigger center of mass determine the direction to the core, for cores more than 50 m from the pond center.

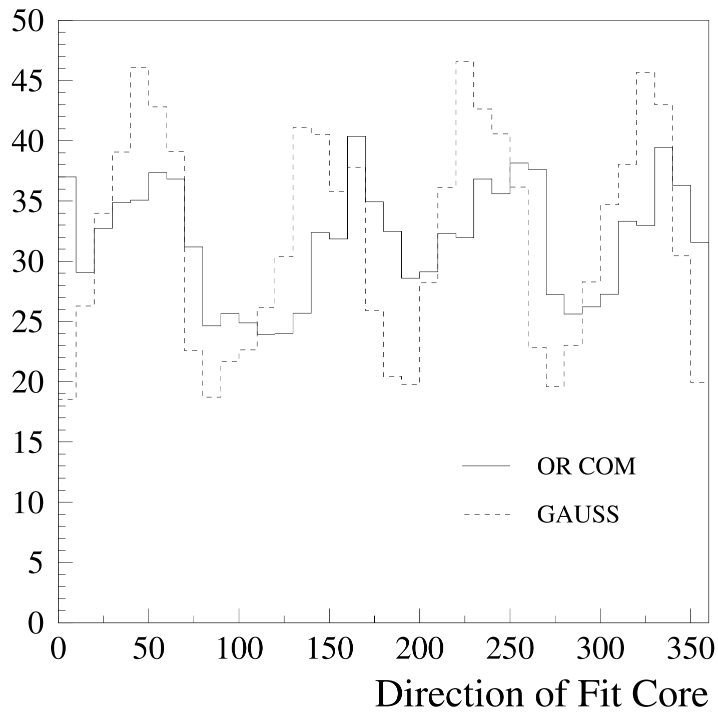


Figure 8: The reconstructed direction to the shower core, for cores more than 50 m from the pond center. The structure seems to be related to the corners of the pond. The true distribution, of course, is flat.

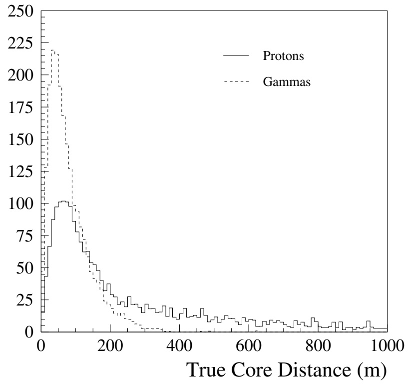
4 Protons and Gamma/Hadron Separation

The algorithm described here has been developed to give good results for gamma ray showers, without regard to its performance on hadronic showers. Nevertheless, it is important to know how it will react to hadrons in order to best incorporate the new fitter into background rejection algorithms.

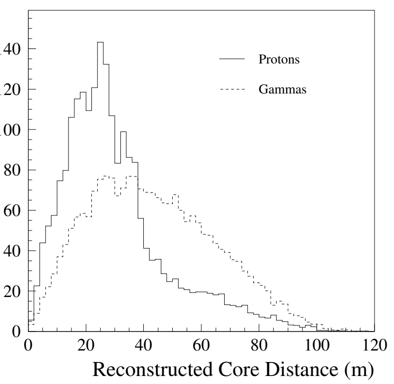
The most salient feature is that although in reality protons trigger Milagro over a much greater area than gammas (Fig. 9(a)), the reconstruction fits the proton cores almost exclusively on the pond, in contrast to gamma cores (Fig. 9(b)). This is almost certainly because proton showers triggering Milagro will have a sub-shower on the pond, with a high $p_\perp$ shower possible even if the core is a kilometer away. This bias is not exclusive to the center of mass fitter; the Gaussian fitter (Fig. 10) shows even a slightly greater bias toward putting protons on the pond (81.4% within 40 m vs. 78.5% for the OR COM).

The unfortunate consequence is that gamma/hadron separation will continue to be hardest for gammas on the pond, i.e., for the gammas that promise the best angular resolution. On the other hand, there will now be more gammas placed off the pond (about half the total on the default spectrum, compared to only 18% with the current fitter), and these should be much easier to retain now that 80% of protons are placed on the pond, as opposed to 31% with the current fitter. In addition, we can now accurately locate the core for these showers, which will allow better curvature corrections and presumably improve the angular resolution for off-pond cores, although this effect will probably be smaller than the improvement in gamma/hadron separation.

As a preliminary attempt to integrate the new core fitter into our present background rejection scheme, I looked at optimizing the X_2 cut jointly in two regions: cores on the pond (less than 45 m from the center) and off the pond (more than 45 m away). With the current scheme, a single X_2 cut for all events, the optimal Q factor of 1.94 is obtained at $X_2 > 4.6$, as shown in Fig. 11. This corresponds to a pass rate for gammas of 15%, and for protons of 0.6%, and is a considerably harder cut than is currently imposed. Dividing the cut into two regions, I find an optimal Q of 2.17 by cutting at 4.5 for events on the pond, and 3.4 for events off the pond. This gives a total pass rate of 22.7% of gammas and 1.1% of protons.



(a) True distances of cores from the center of the pond.



(b) Reconstructed distances from the center of the pond.

Figure 9: True and reconstructed distance of cores triggering Milagro, from the center of the pond; note the change in scale between the plots. Although proton cores actually produce triggers at large distances, they are almost all reconstructed as being on the pond. This probably reflects small subshowers which land on the pond, with the remainder of the shower either outside the array, producing little signal in the outriggers, or not reaching the ground at all.

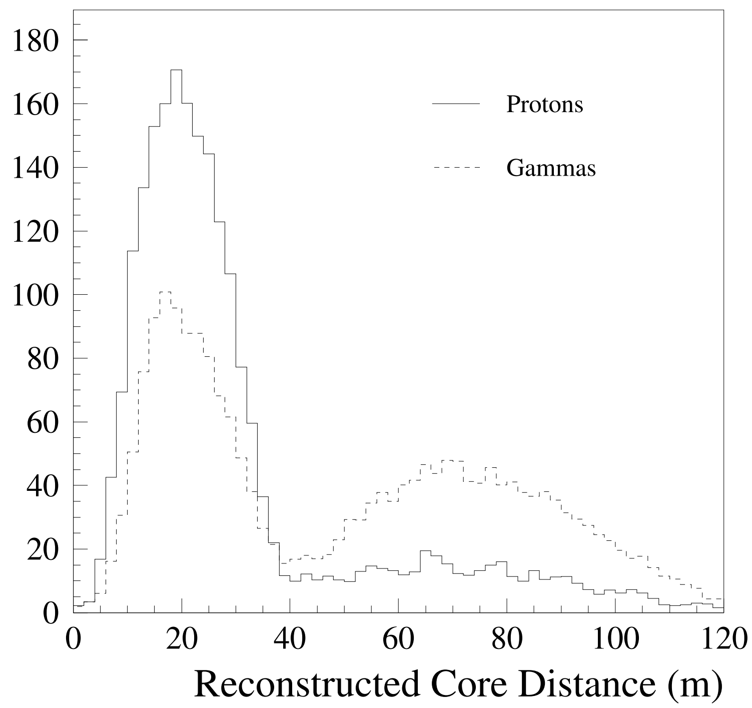


Figure 10: The distance to the shower core as reconstructed by the Gaussian core fitter.

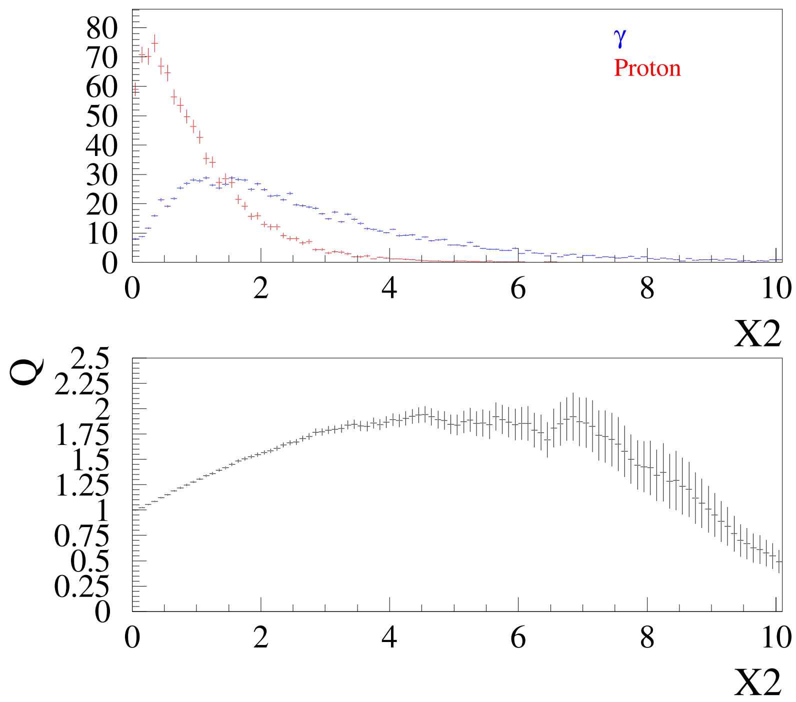


Figure 11: The X_2 distributions of gammas and protons with the VME trigger (top), and the Q factor obtained by cutting at the given X_2 value (bottom). The gamma curve includes all triggers reconstructed within 1.2° of the correct direction, but does not include primaries below 100 GeV, which might change the results slightly. The optimal Q value of 1.94 is found by cutting at 4.6, much harder than currently done. However, this leads to a gamma efficiency of only 15%. Cutting separately on and off the pond can give a Q of 2.17 with an efficiency of 22.7%.

5 Amplitude Dependence

One area of concern in integrating the outriggers into the online reconstruction scheme is the accuracy of outrigger calibrations, and the sensitivity of the reconstruction to miscalibrations. We have attempted to address this concern in two different ways: by Monte Carlo studies of reconstruction accuracy with bad amplitude information, and by investigating reconstruction performance without amplitude information.

In the Monte Carlo studies of amplitude miscalibration, several models of calibration errors were used, none of them particularly realistic. The intention is to give a rough estimate of the amount of degradation that can be expected rather than to quanitatively account for calibration errors. In each case, the simulated amplitudes of all outrigger tubes were modified before being sent to the reconstruction algorithm, and the results compared with reconstructions given the actually simulated amplitudes.

Five models were tested. In the first model, each outrigger amplitude was multiplied by a factor $1 + \alpha_i \cdot \beta \cdot X$, where α_i is the normalized amplitude (PEs)/800, β is a scale factor, and X is a uniform random variate on $[-0.5, 0.5]$. In other words, each amplitude is smeared by an amount which increases with the (true) amplitude of the hit. This is intended to represent calibration uncertainties that are more severe at higher amplitudes but which are centered around the true value. For reference, the amplitude at which there is $\pm$ 50% smearing is at $\beta/800$ PEs. The second model is like the first, save that the X term is dropped, so that each hit is shifted in the same direction and by the same proportionate amount. This is intended to mimic a systematic calibration error which is again worse at higher amplitudes. The third and fourth models are like the first and second, but without the amplitude dependent α_i term. The fifth model uses random smearing with an amplitude dependent magnitude, but instead of being linear in amplitude the magnitude is smallest for high amplitudes, and largest for sub-PE amplitudes and those near the high-threshold/low-threshold turnover, between 2 and 4 PEs. This is loosely based on the Milagrito ADC/TOT calibration comparison curve. The effects of these miscalibrations on the OR COM and Gaussian core fitter median errors are shown in Table 1.

Another means of investigating our dependence on amplitude information is to attempt to use the COM algorithm without amplitude information, or with weaker amplitude dependence. For those events for which the outrigger center of mass is used to tag the core, the median error is 31.7 m (this is higher than the overall median of 26.6 m because the more accurately reconstructed on-pond events are excluded). In this case, the amplitude en-

	β	OR COM	GAUSS
Model One	0	26.6	24.4
	5	27.4	25.2
	10	27.8	25.9
	25	28.3	26.8
	50	28.7	27.8
Model Two	-25	38.1	34.3
	-10	34.5	31.0
	-5	31.9	28.9
	-2	29.1	26.7
	0	26.6	24.4
	2	26.5	24.4
	5	26.3	24.5
	10	26.1	24.6
	25	26.0	25.2
Model Three	1	26.9	24.8
	0.8	26.8	24.7
	0.6	26.7	24.5
	0.4	26.6	24.6
	0.2	26.6	24.5
	0	26.6	24.4
Model Four	-0.5	24.3	20.9
	-0.25	26.2	23.3
	0	26.6	24.4
	0.25	26.9	24.8
	0.5	27.0	25.1
	1.0	27.1	25.2
Model Five	2	26.5	24.5
	5	26.7	23.9
	10	26.9	24.2
	20	27.4	24.8

Table 1: Median reconstruction errors, in meters, for various types of outrigger amplitude miscalibrations. The models are described in the text, β is a scale factor. With the exception of negative β in model two (i.e., correct calibration of small hits, but with increasing amplitude underestimation at higher amplitudes), none of the miscalibration models produce a significant degradation of accuracy.

ters linearly. If instead we weight each hit by the logarithm of the amplitude (offset by one to avoid negative weights), the median error increases to 38.7 m, and if the amplitudes are not used at all, the error is 43.5 m. These results are consistent with the results for model 2 in Table 1, where increasing underestimation of high amplitudes (and thus decreasing influence of high amplitude hits) caused a rapid degradation of accuracy in both the OR COM and GAUSS reconstructions. It should be noted that amplitude-independent methods do exist, for instance a likelihood fitter that attained a median error of 25 m on a somewhat more difficult set of events, but they are much more complicated, time-consuming, and (probably) sensitive to data/Monte Carlo discrepancies.

6 Conclusions

The outrigger center of mass (OR COM) core fitter offers a significant improvement in our core reconstruction accuracy, especially in our ability to identify cores landing off the pond. Tony's Gaussian core fitter offers slightly better accuracy, at a lower speed. The OR COM could be run on the hardware presently available, while the Gaussian would require a few additional CPUs. Both core fitters seem reasonably robust against amplitude miscalibrations, although of course good amplitude information improves performance. However, the higher relative rate of noise hits in the outriggers will require careful identification of spurious hits. Further investigation of noise in the outriggers, and a careful comparison of Monte Carlo with data, should be undertaken before the algorithm is installed online.