Fitter Resolution and Systematic Errors

X.Chen
Lawrence Berkeley National Laboratory
SNO-STR-2000-026

October 17, 2000

Abstract

We use both the laser and the ^{16}N data to estimate systematic error of the grid fitter. Depending on definition, the grid fitter may have a vertex resolution of 20cm and a direction resolution of 26 degrees. The systematic error for the vertex reconstruction may be 10cm while the systematic error for the direction reconstruction may be 6 degrees.

1 Introduction

Fitters are essential for correctly interpreting data from the SNO experiment as they enable us to determine where an event occurred and its direction. Not only resolution of the fitters but also, more importantly, the associated systematic errors have to be understood. Normally study of a fitter can be carried out in two steps:

1. Define some numeric parameters which describe characters of the fitter. Some of the most commonly used parameters are: *Rdev* which is defined as the distance between the fitted vertex and the true vertex; θ which is defined as the angle between the fitted event direction and the actual direction.

Table 1: Reconstruction efficiency of the grid fitter for events taken during several ^{16}N data cleaning runs. A Nhits cut of Nhits $\geq$ 30 was imposed.

Run name	Source position	No. of events successfully fitted	No. of events failed	Failure rate
10353	0,0.3,100.1	147136	7	4.76×10^{-5}
10355	-0.17,0.29,200.1	154514	4	2.59×10^{-5}
10357	0.15,0.23,300.2	154257	2	1.30×10^{-5}
10359	0,0.42,400.5	148684	4	2.69×10^{-5}
10361	-1.42,14.12,498.23	98455	0	0
10517	-10.56,2.2,600.	21484	0	0
10564	-586,208,-100	78707	2	2.54×10^{-5}
10566	-586,208,-200	83346	2	2.40×10^{-5}
10568	-586,208,-300	80501	3	3.73×10^{-5}
10573	-586,208,-500	26891	7	1.49×10^{-4}

2. Tune the Monte Carlo so that the MC prediction of these characteristic distributions agree with the calibration data as much as possible. Then the tuned Monte Carlo can be used to predict/extrapolate behavior of the fitter under the scenarios that the calibration source can not simulate.

3. No matter how we tune the Monte Carlo, there are always going to be some discrepancies between the prediction of the tuned MC and the reality indicated by the calibration data. We can simply take these discrepancies as systematic errors of the fitter. An alternative way is to systematically vary certain parameters in the Monte Carlo within some reasonable ranges and take variations among the MC predictions as the systematic errors.

As far as the fitter study is concerned, we have not made any attempts to tune the SNOMAN to match the calibration data. We simply take prediction of the Monte Carlo as a measure of the fitter resolution and use calibration data to estimate the systematic errors. As a start, we used the optic title files produced by the Queen's university as it correctly reproduced the Nhits spectra of the ^{16}N data. Only the grid fitter with its maximum likelihood direction fitting is considered in the study. And we only considered events with high Nhits (Nhits $\geq$ 45) as the solar neutrino analyses is carried out above such threshold.

2 Fitter Characters

Generally speaking, a good fitter has to not only fit events as accurately as possible but also have a high success rate on reconstruction. The reconstruction efficiency of a fitter can be estimated by studying the ^{16}N data. As table 1 shows, the grid fitter reconstructs almost all of the ^{16}N events no matter where the source is located. Its reconstruction efficiency is so high that any associated systematic errors can be safely ignored.

Table 2: Quantitive evaluation of performance of the grid fitter on the CC events.

Criterion	No. of events	Fraction
Rdev < 200cm	11882	0.9946
Rdev < 100cm	11882	0.9878
Rdev < 60cm	11882	0.9269
Rdev < 40cm	11882	0.7751
Rdev < 20cm	11882	0.2920
Rdev < 10cm	11882	0.0516

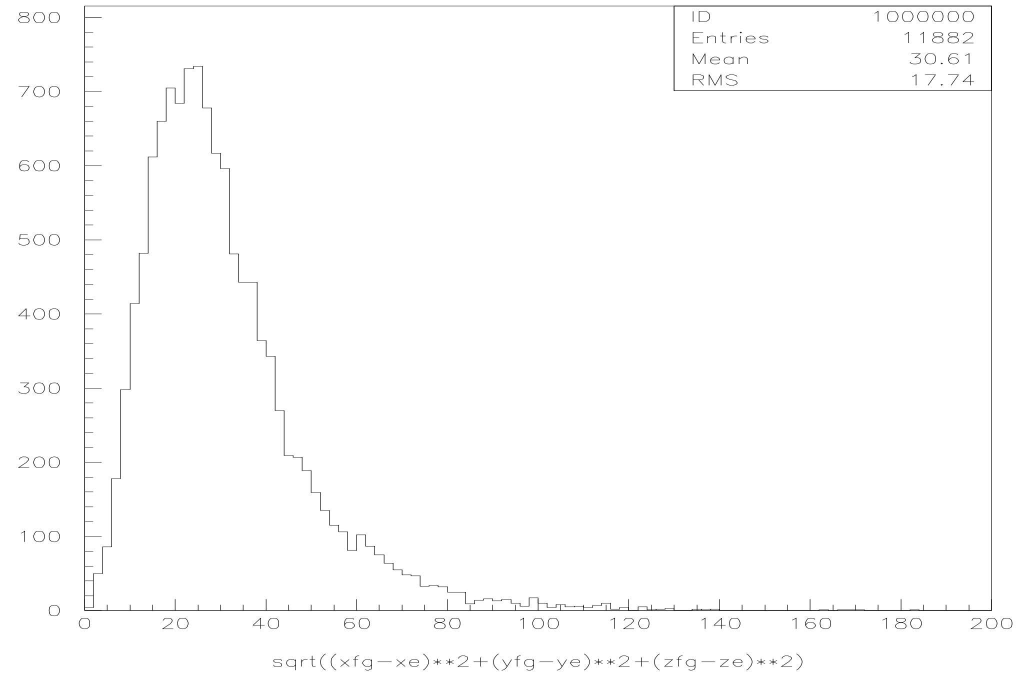
Criterion	No. of events	Fraction
cosθ > 0.99	1980	0.1666
cosθ > 0.95	5938	0.4997
cosθ > 0.90	8012	0.6742
cosθ > 0.85	9119	0.7674
cosθ > 0.80	9761	0.8214
cosθ > 0.75	10175	0.8563
cosθ > 0.70	10508	0.8843
cosθ > 0.60	10745	0.9042
cosθ > 0.50	11133	0.9369

For MC events, the actual event position and event direction are known, thus performance of the fitter can be easily evaluated. Figure 1 shows the Rdev and the cosθ distributions for 15000 simulated CC events. Table 2 is a quantitive description of the fitter performance. If we define the peak of the Rdev distribution as the fitter vertex resolution, then for the CC events, the grid fitter has a resolution of roughly 20cm. If we define the angular resolution as the value above which 67.42% of events have larger cosθ, then the angular resolution is roughly 26 degrees. The grid fitter in SNOMAN uses 22.355 cm/ns as the speed of the light in reconstruction. As shown later, a more optimized value averaged over the group speed may increase the reconstruction accuracy.

3 Fitter Systematic

3.1 Systematic on Vertex Reconstruction

As stated before, we simply take the difference between the MC prediction and the calibration data as the systematic. Up to now, data have been taken from four calibration source: the LED tube, the laser ball, the ^{16}N source and the neutron source. The LED data is not very useful because it is located at the PSUP, far away from the fiducial volume. For vertex study, the neutron source is inferior to the ^{16}N source because it is a more diffused source due to neutron propagation. As far as the fitter study is concerned, the laser ball and the ^{16}N decay chamber are the most relevant source, they are also the most frequently runned source. Main product of the ^{16}N source is a 6Mev or 7Mev gamma, which then produces a scattered electron,



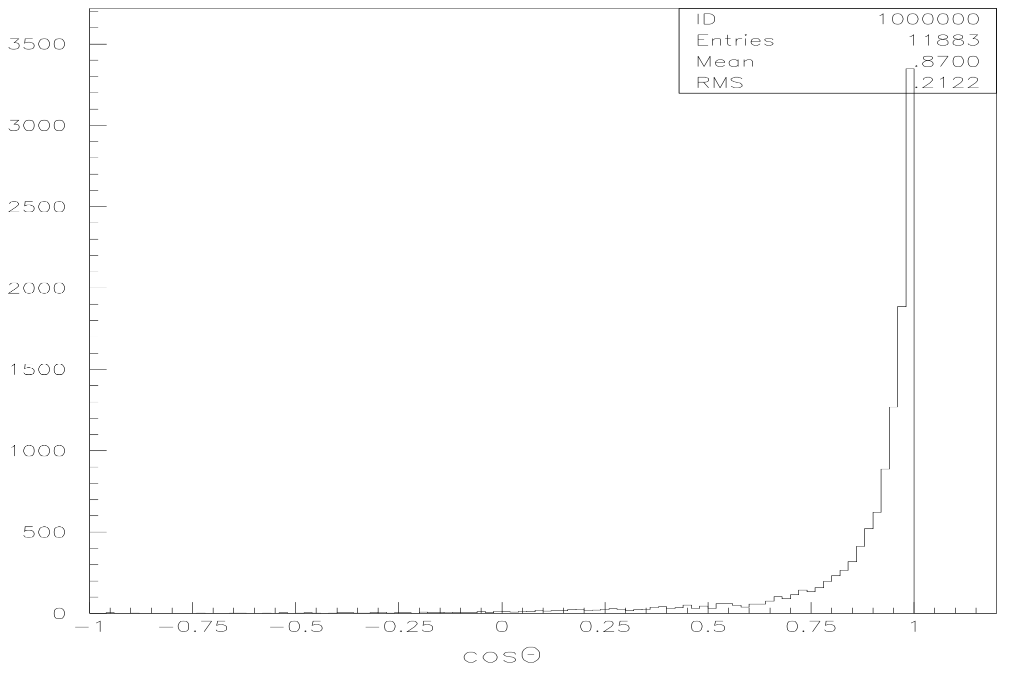


Figure 1: Rdev (upper plot) and cosθ (bottom plot) distributions of CC events generated uniformly throughout the D_2O volume. Rdev is the distance between the fitted vertex and the true vertex. θ is the angle between the fitted event direction and the electron direction. A cut of Rdev $<$ 200cm and Nhits $\geq$ 45 is imposed on the events in the upper plot while a cut of Rfit $<$ 700cm and Nhits $\geq$ 45 is imposed on the events in the bottom plot.

Table 3: Laser runs used in the analyses.

Run number	Source position(cm)	wave length
10988	-0.3,0.8,0.0	500nm
10996	0.1,1.4,-552.2	500nm
11023	-7.6,12.9,549.9	500nm

thus data of the ^{16}N source directly mimic the neutrino signals. However, because the gamma ray can travel a certain distance before it is converted to an electron, source of the photons detected by the PMTs could be far away from the decay chamber and is not known on an event by event basis. This makes study of the fitter based on the ^{16}N data somewhat difficult, especially if the systematic is small. The laser source is a point source, but its light production is isotropic and differs from that of a scattered electron. We used data from both sources in the study.

One thing worth mentioning with reconstruction of the laser data is that we used a different light speed (22.082 cm/ns) in the reconstruction in contrast with the reconstruction of the ^{16}N data where the default light speed in SNOMAN was used. It is because laser events produce photons with the same wavelength instead of a wide spectrum produced by the ^{16}N events. Difference between the group speed and the phase speed may have a big effect for the laser data. Table 3 lists laser runs used in the analyses. As laser events normally have several hundred hits, we randomly sampled the hits down to the solar neutrino region. Figure 2 shows Rdev vs. Nhits profile histogramms of three laser runs, superimposed with those of MC data. Apparently there is a 3cm difference between the data and the MC prediction. The difference is consistent with laser at different location, thus systematic on nhits and radius dependence may be small. However, isotropy of the laser light normally make it much easier to reconstruct than a electron. Thus 3cm can only be considered as the lower limit of the fitter systematic on neutrino events.

The ^{16}N data used in the study were triggered only by the Nhits trigger so that it ensures the same data taking condition as the solar neutrino data. Although the actual event vertex is not known on the event by event basis, the distance (between the fitted vertex and the source position) distribution can still be used as a rough measurement of the fitter systematic. Figure 3 shows such distribution of the ^{16}N data superimposed with the distribution produced by the Monte Carlo. It is apparent that there is an excellent agreement between the data and the MC prediction, which indicates that the systematic error of the grid fitter must be much less than attenuation length of a 6 Mev gamma. As figure 4 shows, centroid of the fitted vertex distribution agrees well with the nominated source position which indicates there are no tangible systematic shift by the grid fitter when the source was located at the center of the detector.

However, the systematic shift became tangible when source was moved away from the detector center, as shown in figure 5. One thing worth mentioning is that we define the systematic error as the difference between the data and the MC prediction. For example, the fitter pull is NOT a systematic error if magnitude of the pull is accurately predicted by the Monte Carlo. This is evident in the bottom

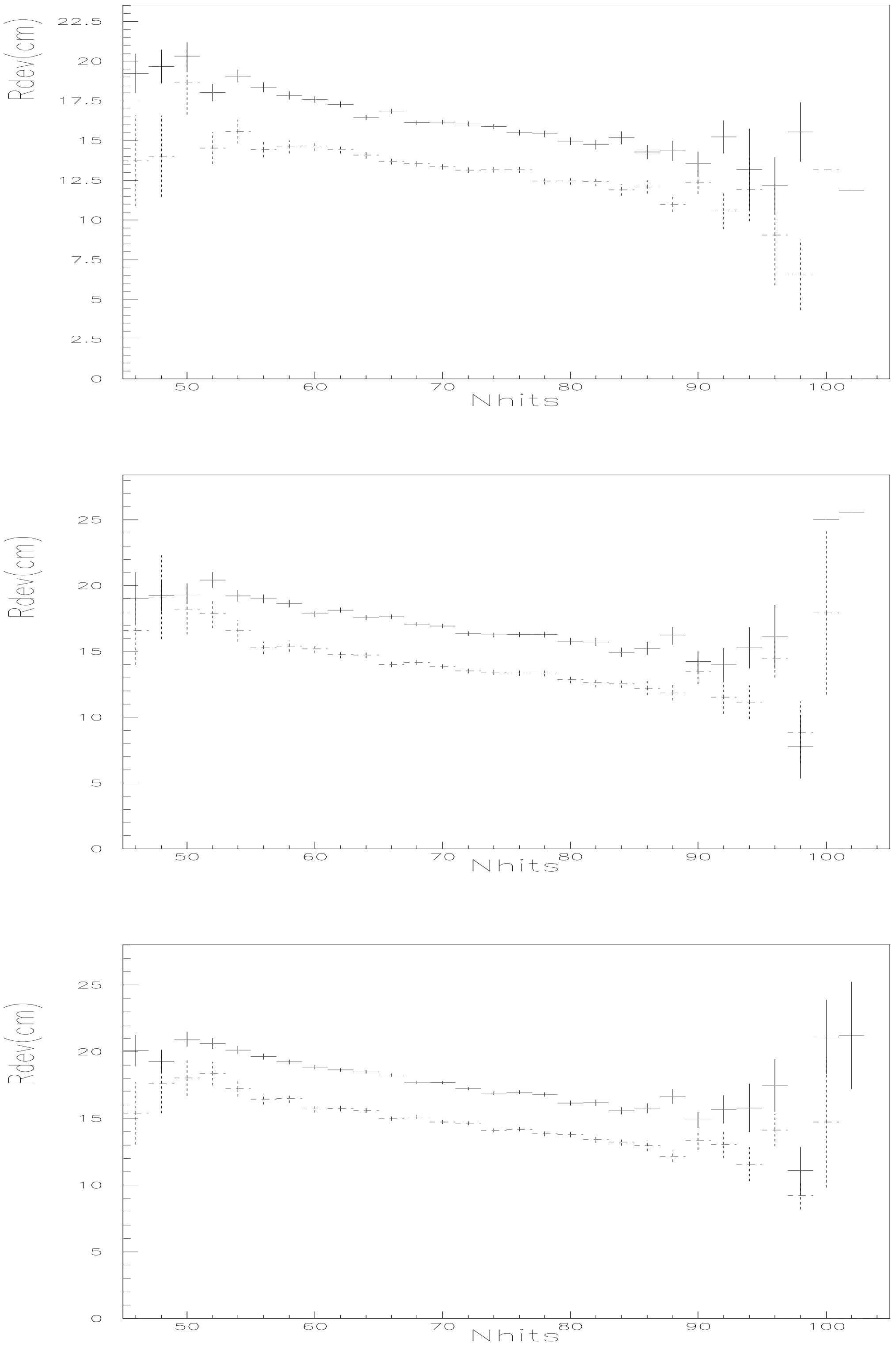


Figure 2: Rdev vs. Nhits profile histograms for laser run 10988(upper plot), 10996(middle plot) and 11023(bottom plot). Rdev is the distance between the source position and the fitted position. Solid lines are real data while dashed line are MC data.

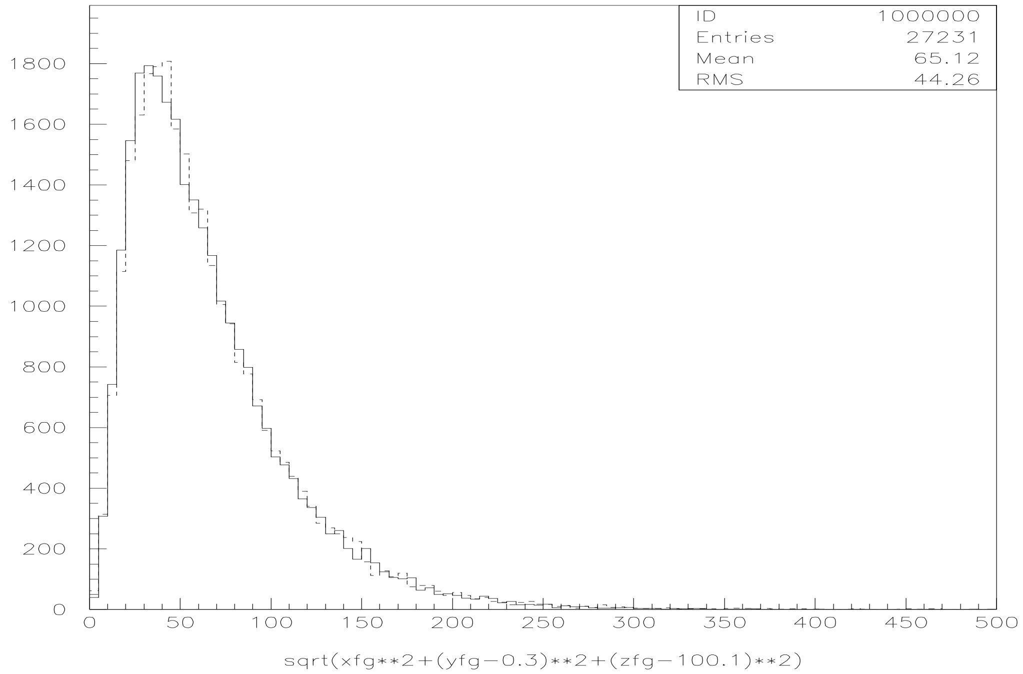


Figure 3: Distance between the fitted vertex and the source position for the ^{16}N source. Data was taken during run10353, the source position is (0, 0.3, 100.1). A Nhits cut of Nhits $\geq$ 45 is imposed.

plot of figure 5, where peak of the z_{fit} distribution is around 510cm, 10cm above the nominated source position (z = 498.23cm). The tangible systematic shift is at the y direction. In order to investigate dependence of the systematic shift on position, we also studied data from several other ^{16}N data cleaning runs, the run numbers are shown in table 4. Figure 6, figure 7 and figure 8 show the fitted x, y and z distributions respectively for various ^{16}N runs, compared with the MC prediction. Several observations can be obtained from the plots:

- The shape of the data distribution and that of the MC distribution is fairly similar along 3 axis for almost all of the runs. But there is shift between the two distributions in some plots.
- There are fairly good agreement between the data and the MC prediction along the x axis.
- It seems there may be a systematic error associated with the fitter pull along the z axis for runs with sources at lower half of the detector.
- There are systematic shifts along the y axis for runs with source at z=400cm, z=-200cm, z=-400cm and z=-500cm.
- The agreement between the data and the MC prediction is poor when source is located at the neck region.

Many factors could lead to such systematic difference:

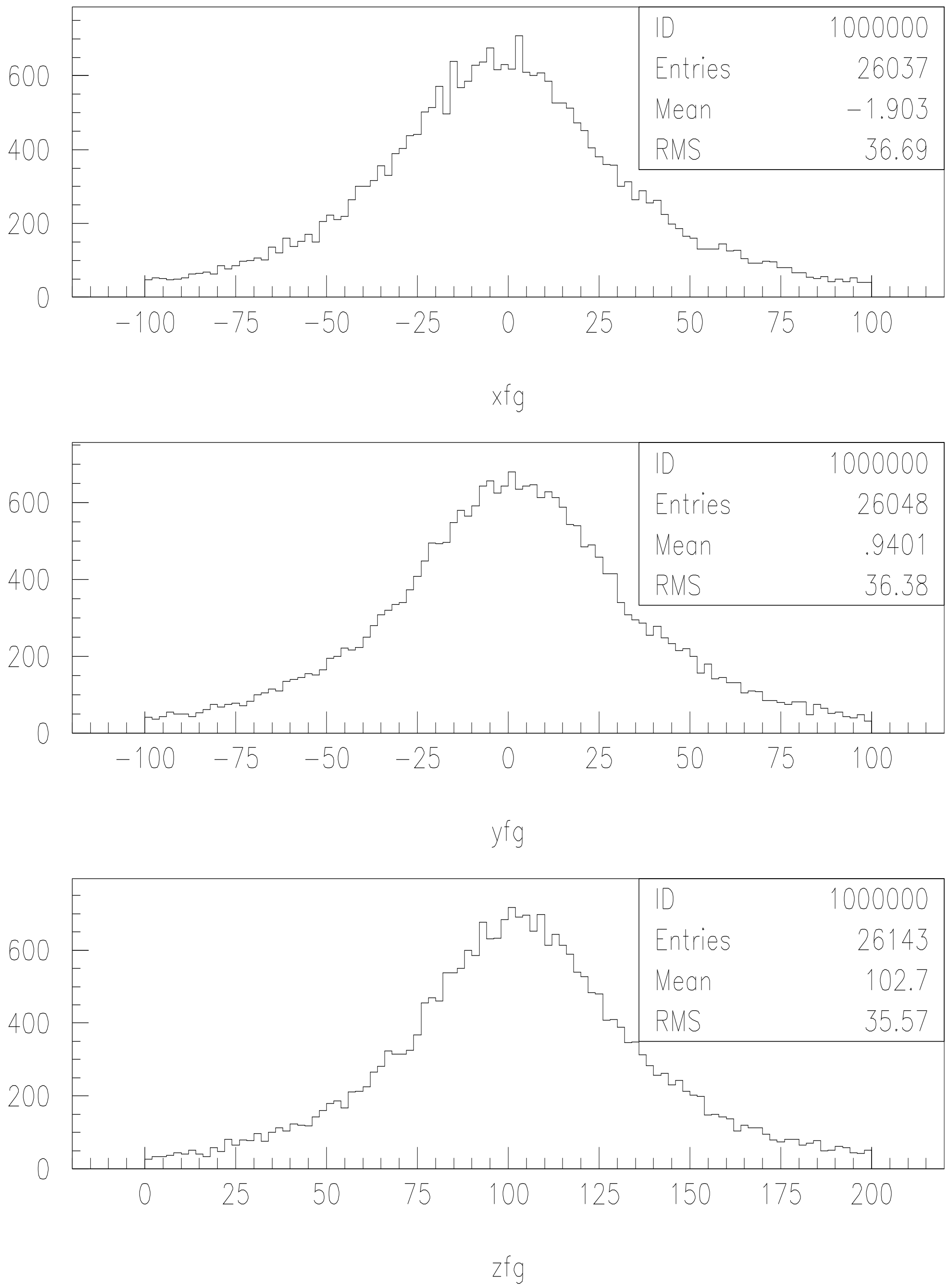


Figure 4: Fitted x, y, z distributions of the ^{16}N data. Data was taken during run10353, the source position is (0, 0.3, 100.1). A Nhits cut of Nhits $\geq$ 45 is imposed.

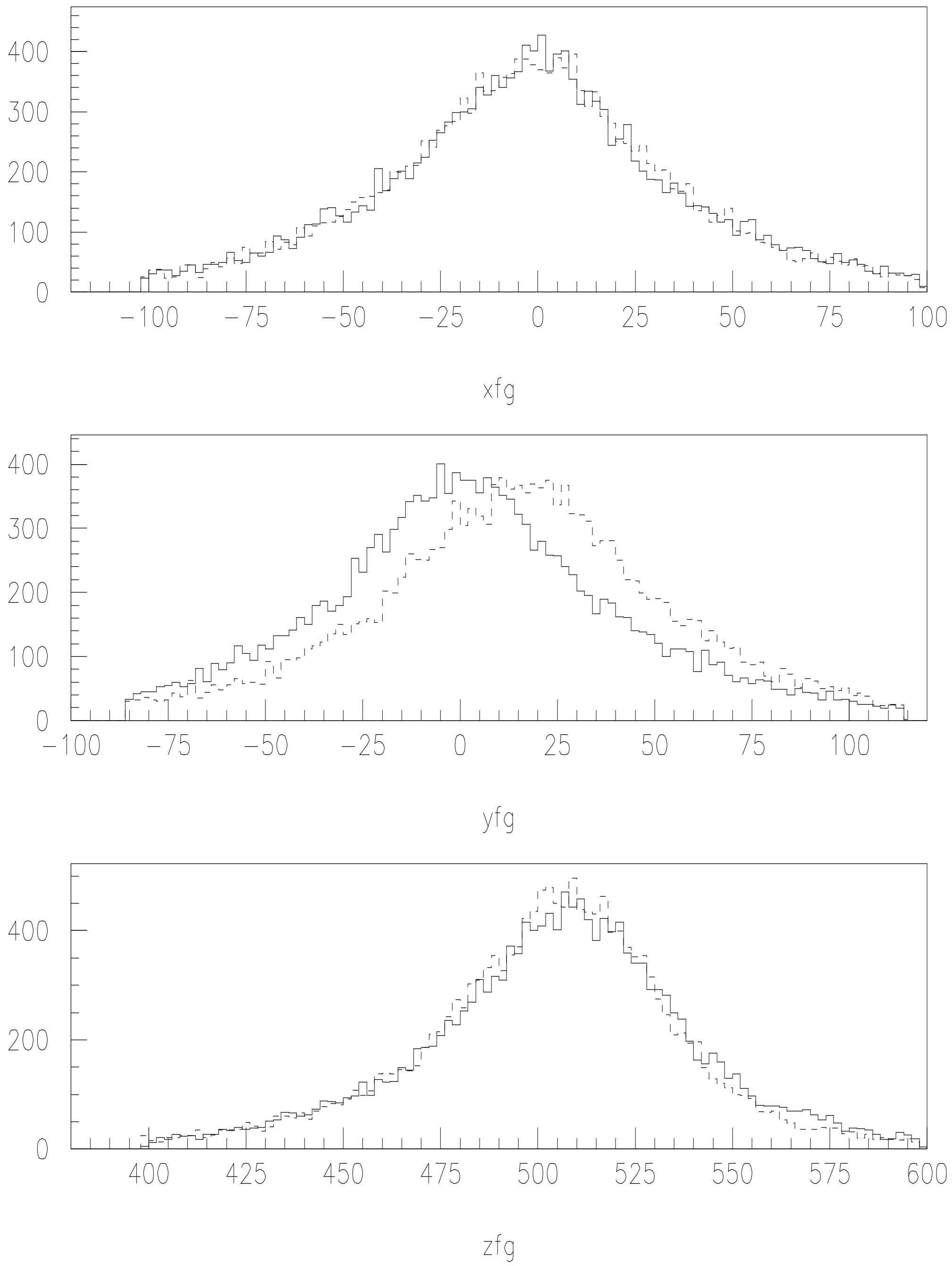


Figure 5: Fitted x, y, z distributions of the ^{16}N data. Data was taken during run10361, the source position is (-1.42, 14.12, 498.23). A Nhits cut of Nhits ≥ 45 is imposed. Solid line is the data while dashed line is the MC prediction.

Table 4: List of the ^{16}N data cleaning runs in study of the fitter systematic.

Run name	Source position
10353	0,0.3,100.1
10355	-0.17,0.29,200.1
10359	0,0.42,400.5
10361	-1.42,14.12,498.23
10482	0., 0., -200.
10486	0., 0., -400.
10488	0., 0., -500.
10517	-10.56, 2.2, 600.

- Systematic errors of the grid fitter.
- Inaccuracy on simulation of the ^{16}N source.
- Errors on the source position obtained from the manipulator.

We do not know which is the main reason. We suspect there may be more than one factor here as the shifts seen with various runs are not symmetric although the detector is symmetric along the x and the y axis.. Conservatively, we attribute such systematic difference entirely to the fitter. Thus there may be a 10cm systematic errors on vertex reconstruction for the grid fitter.

Another thing need to consider is the speed of the light used in the event reconstruction. Strictly speaking, this may not be a systematic error as long as the same light speed is used in reconstruction of both the real data and the MC data. The standard grid fitter in SNOMAN uses 22.355 cm/ns as the light speed. This may not be correct as it is based on the average of the phase speed of the light in the D_2O and the H_2O region. It was pointed out [2] the group speed should be used in the reconstruction. As the Cerenkov light has a wide spectrum and the photon could traverse through different medias, it may be difficult to determine the optimized speed, a detailed ray-tracing fitting may be far too complex. In order to estimate systematic caused by different light speed, we naively use 21.9 cm/ns as the light speed in the reconstruction and compare the fit results with that obtained with c=22.355 cm/ns. Such comparison is shown in figure 9. The improvement with a optimized light speed is tangible, which could be as large as 7cm. It may be an evidence that we should use a more optimized value for the light speed in the event reconstruction. We are studying the effect with real data. Report will be soon available.

Two important topics are not covered in this paper:

1. Accuracy of the source position reported by the manipulator system. PCA relies on correct information of the laser position to calibrate out time delays due to difference on cable length among each PMTs. It is essential to use an independent method to verify the coordinates obtained by the manipulator. Such study is under progress.

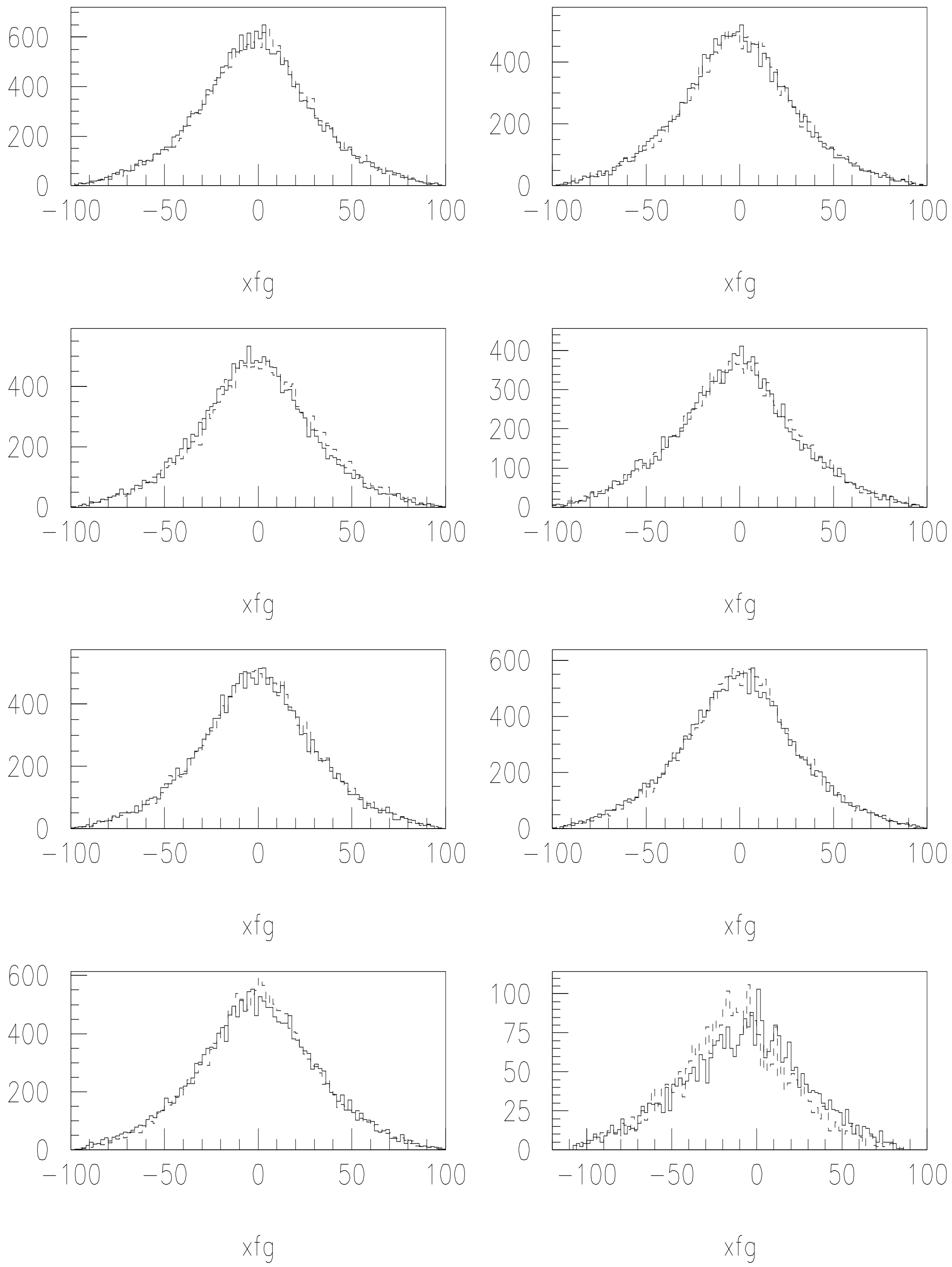


Figure 6: Fitted x distributions of the ^{16}N data. From top to bottom and from left to right, the plots represent data from run10353, run10355, run10359, run10361, run10482, run10486 and run10517. A Nhits cut of Nhits ≥ 45 is imposed. Solid line is the data while dashed line is the MC prediction.

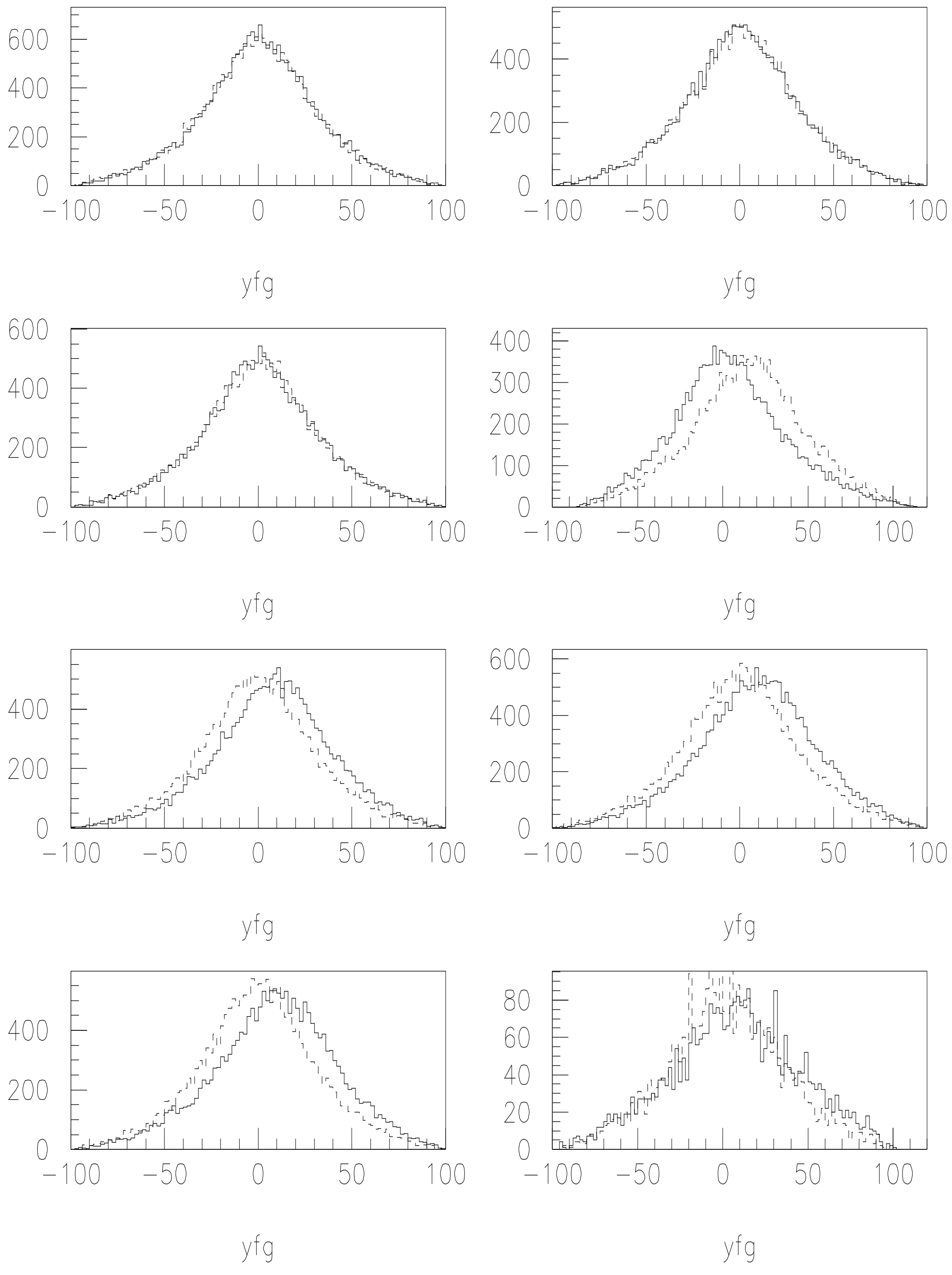


Figure 7: Fitted y distributions of the ^{16}N data. From top to bottom and from left to right, the plots represent data from run10353, run10355, run10359, run10361, run10482, run10486 and run10517. A Nhits cut of Nhits ≥ 45 is imposed. Solid line is the data while dashed line is the MC prediction.

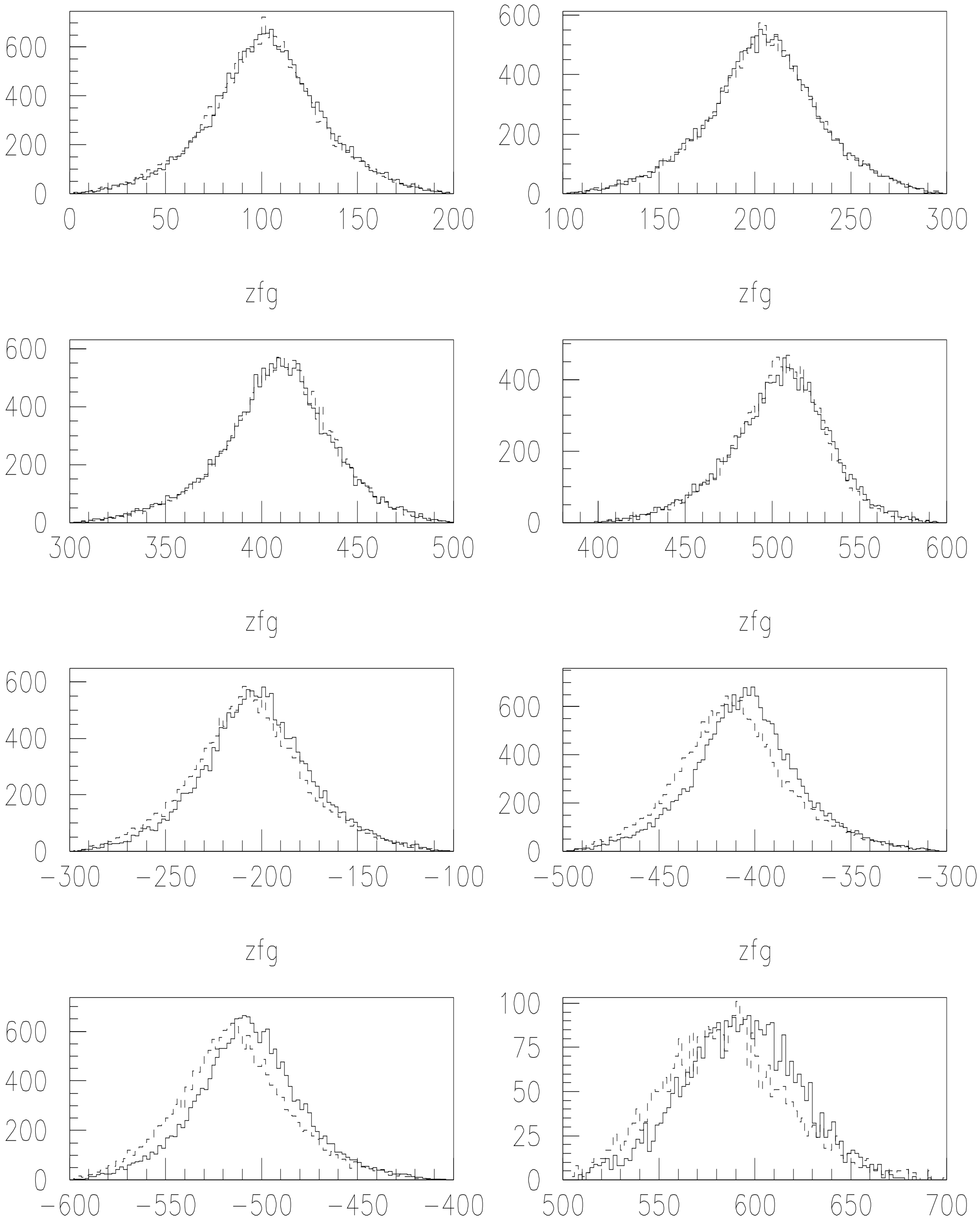


Figure 8: Fitted z distributions of the ^{16}N data. From top to bottom and from left to right, the plots represent data from run10353, run10355, run10359, run10361, run10482, run10486 and run10517. A Nhits cut of Nhits ≥ 45 is imposed. Solid line is the data while dashed line is the MC prediction.

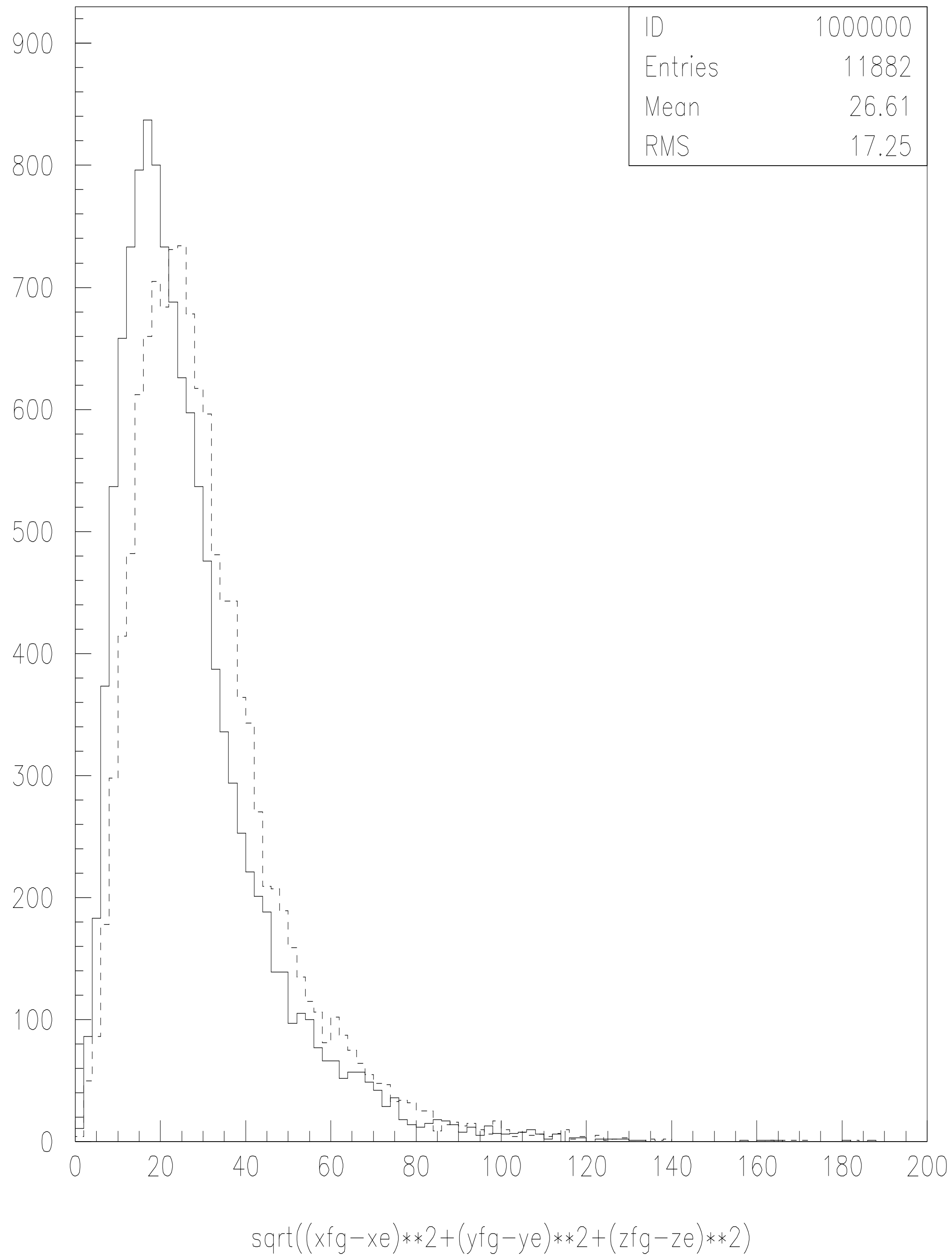


Figure 9: Rdev distributions of the CC events fitted with c=21.9cm/ns (solid line) and with c=22.355cm/ns (dashed line). A cut of Nhits ≥ 45 and Rdev < 200cm is imposed.

2. Effect of the chi2 problem. We still do not fully understand the effect at the moment, but the preliminary study with low rate laser and ^{16}N data indicates the effect could be fairly large. Study of this effect is also under progress.

3.2 Systematic on Direction Reconstruction

A widely used method [1] to evaluate fitter performance on direction reconstruction is to look at the ^{16}N events reconstructed far away from the source. Most of the gammas emitted from the source produce electrons by Compton scattering, the scattered electron normally retain the direction of the gamma. Thus the vector from the source to the electron vertex is aligned with the direction of the electron. Thus the angle between the vector and the fitted direction could be used to measure fitter's systematic on direction reconstruction. Figure 10 shows such distributions. The difference between the upper plot of the figure, which represents the data, and the bottom plot, which represents the MC direction, could be due to either or both of two factors:

- Fitter's systematic errors on direction reconstruction
- Fitter's systematic errors on vertex reconstruction

Using only the events reconstructed far away from the source reduce impact of the second factor, but we do not know how to quantify it at the moment. Study is under progress to identify the mis-reconstructed events statistically. We assume the first factor is dominant, thus the difference between the upper and the bottom plot is an manifestation of the fitter systematic on direction reconstruction. One way to estimate effect of the mis-reconstructed events is to vary the distance cut. Figure 11 and figure 12 shows the $\cos\theta$ distributions with different distance cut. Apparently it is difficult to describe the difference between two distributions with a single number. We applied the Kolmogorov-Smirnov test to quantify the difference. As table 5 shows, the difference increases as the cut on distance between the fitted vertex and the source increases. It may be due to the increasing effect of the mis-reconstructed events as we look further away from the source. Nevertheless, the probability that two distributions come from different underlying distributions is not statistically significant. If we simply define the difference between mean value of the data distribution and that of the MC distribution as the systematic error, conservatively the systematic may be 6 degrees.

We also studied the ^{16}N data with the source at other locations. Figure 13 shows the comparison between the data and the MC prediction when the source is located at (-1.42, 14.12, 498.23) while figure 14 shows the comparison when the source is located at (0, 0, -500). As the source is located fairly close the AV, some of the gammas from the source could be converted fairly deep into the light water region. Angular resolution could be skewed by unknown optics and acceptance in the light water region, thus we applied a cut of rfit $<$ 700cm to minimize such effect. Generally speaking, the agreement between the data and the MC prediction is also good for these two runs. It seems the systematic difference may be smaller than the conservative number given above.

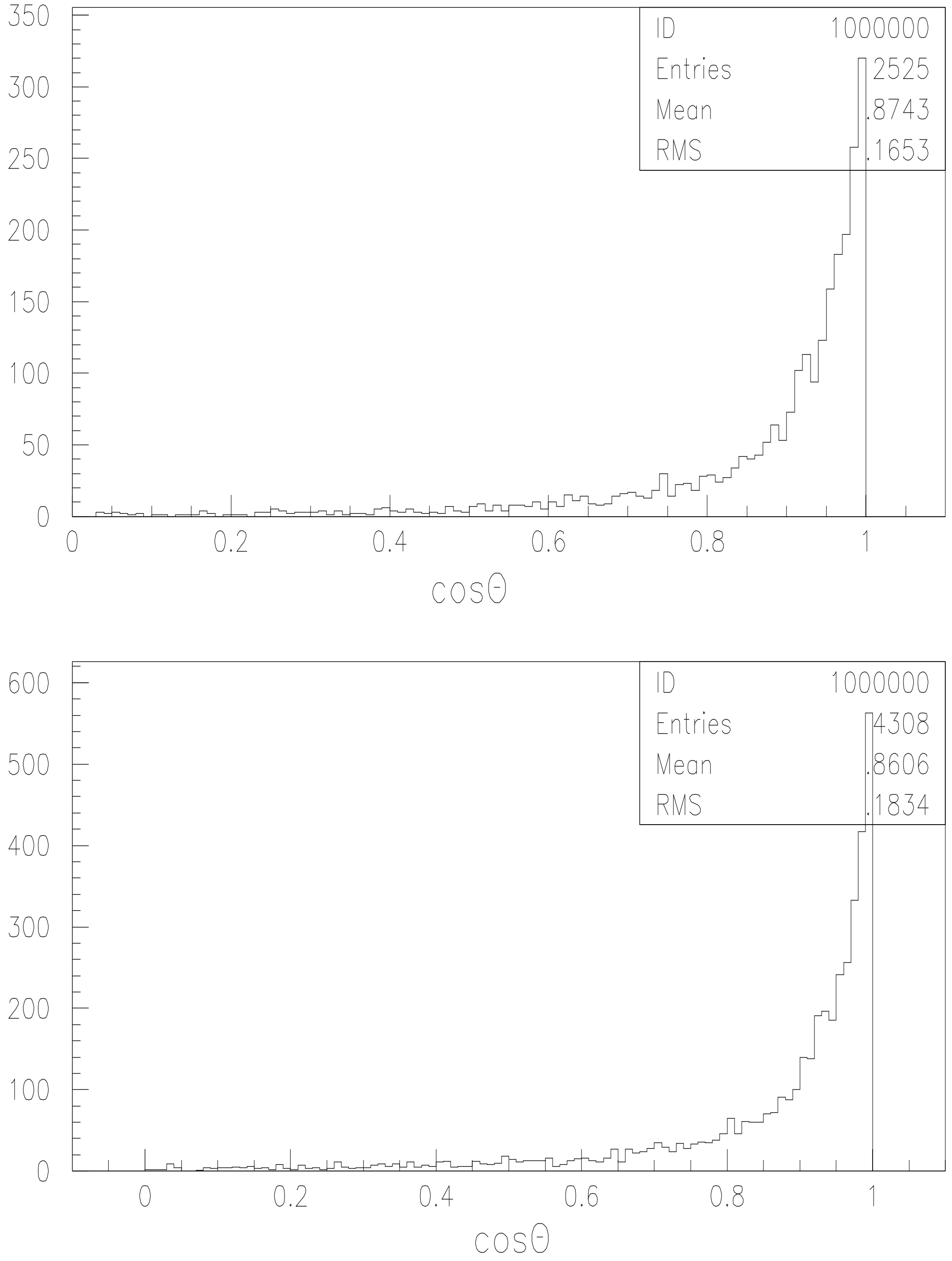


Figure 10: Cosθ distributions of ^{16}N run10353(upper plot) and MC data(bottom plot). θ is the angle between the fitted direction and the vector from the source to the fitted vertex. A cut of Nhits geq 45, 100 $<$ dis $<$ 500 and cosθ $>$ 0 is applied. dis is the distance between the source and the fitted position.

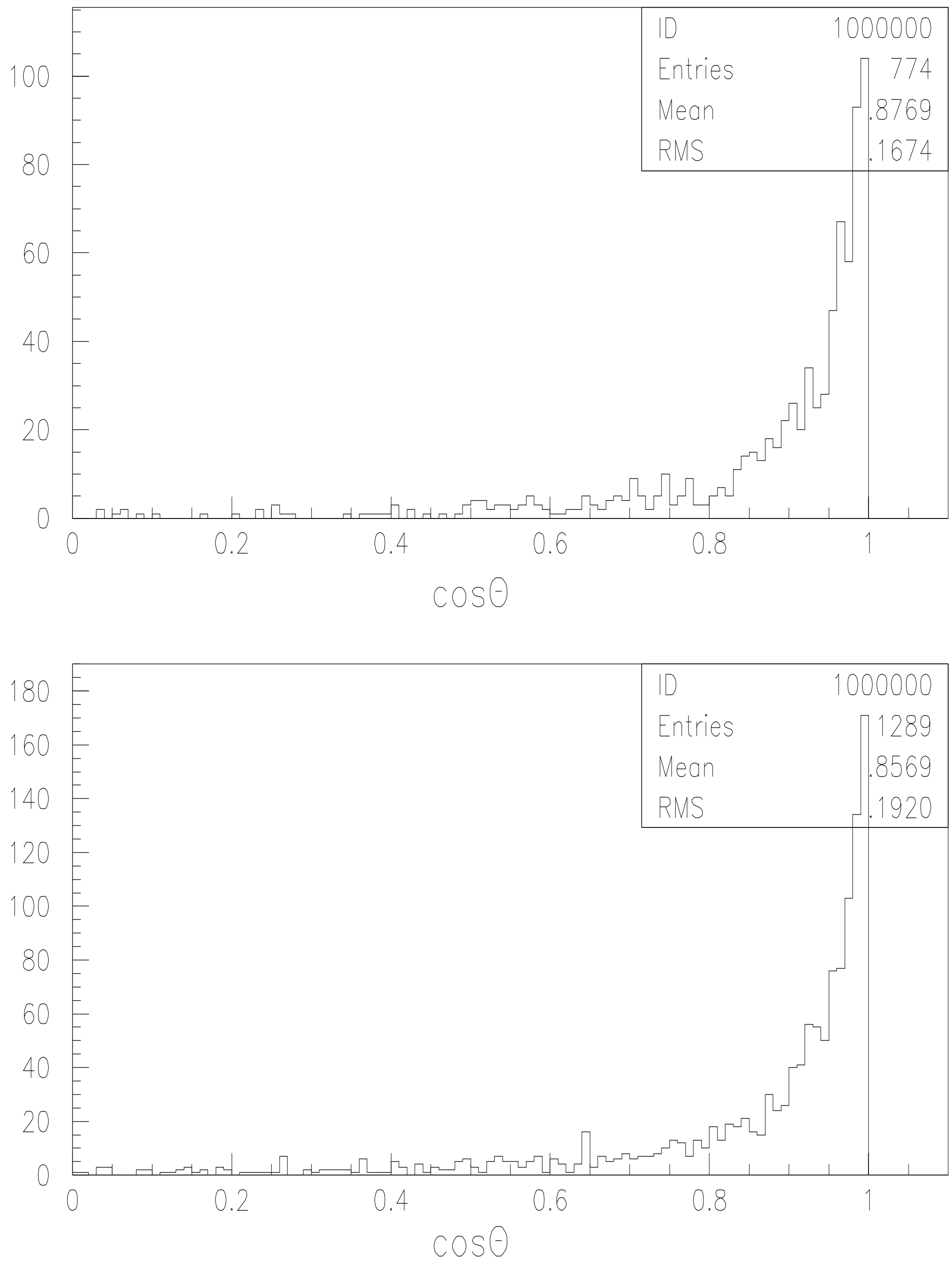


Figure 11: Cosθ distributions of ^{16}N run10353(upper plot) and MC data(bottom plot). θ is the angle between the fitted direction and the vector from the source to the fitted vertex. A cut of Nhits *geq* 45, 150 < dis < 500 and cosθ > 0 is applied. dis is the distance between the source and the fitted position.

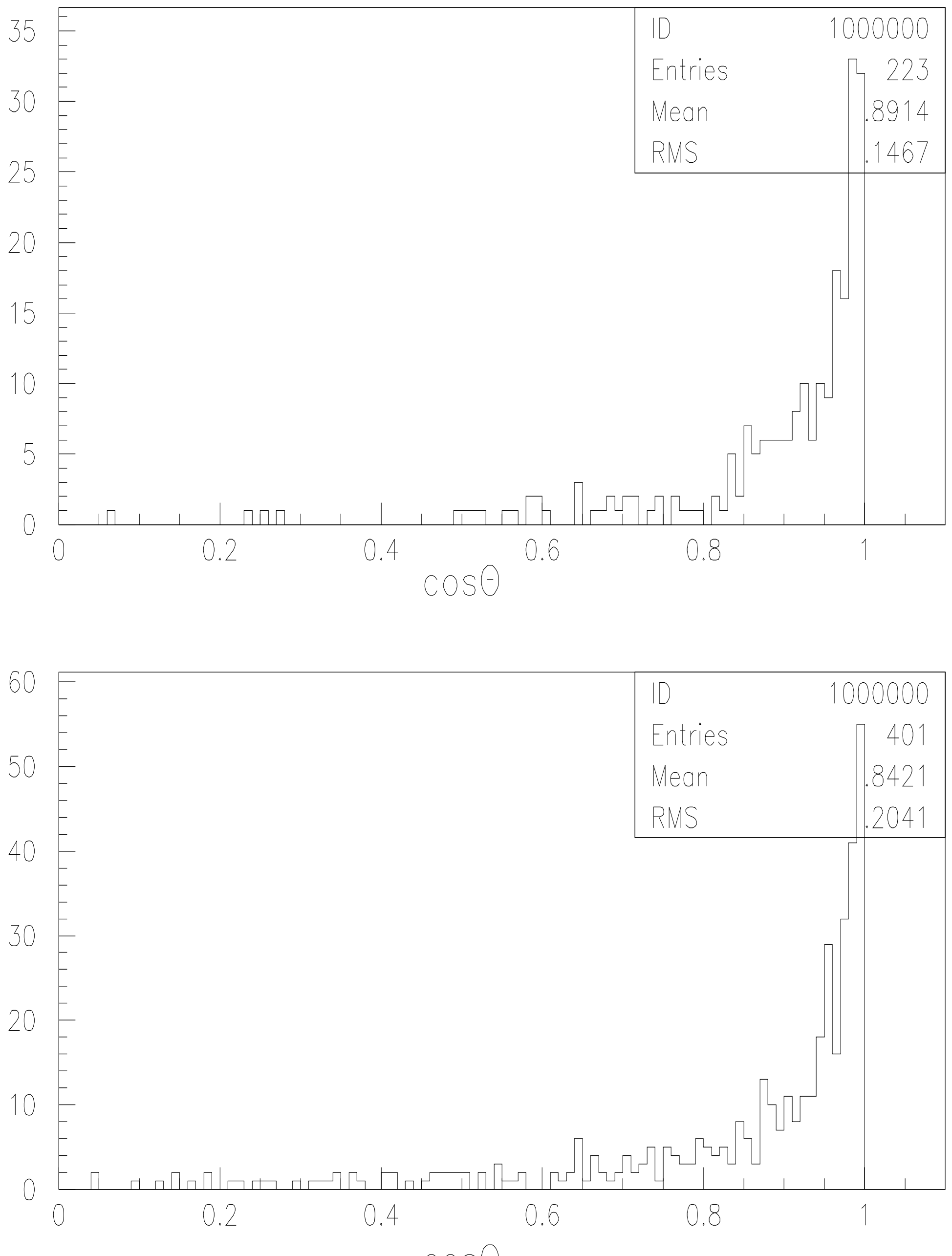


Figure 12: Cosθ distributions of ^{16}N run10353(upper plot) and MC data(bottom plot). θ is the angle between the fitted direction and the vector from the source to the fitted vertex. A cut of Nhits *geq* 45, 200 < dis < 500 and cosθ > 0 is applied. dis is the distance between the source and the fitted position.

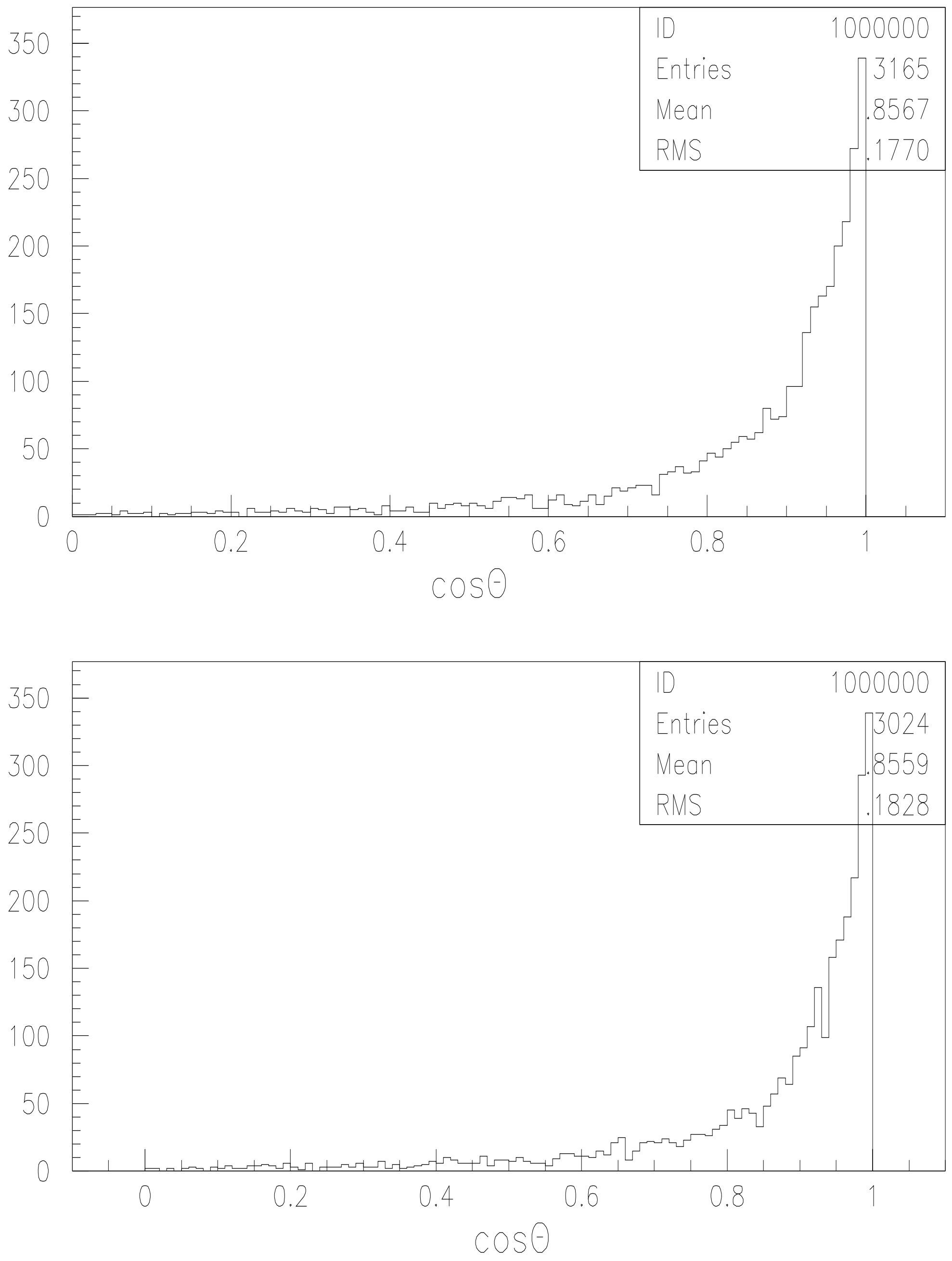


Figure 13: Cosθ distributions of ^{16}N run10361(upper plot) and MC data(bottom plot). θ is the angle between the fitted direction and the vector from the source to the fitted vertex. A cut of Nhits *geq* 45, $100 < \text{dis} < 500$, $\cos\theta > 0$ and rfit < 700cm is applied. dis is the distance between the source and the fitted position.

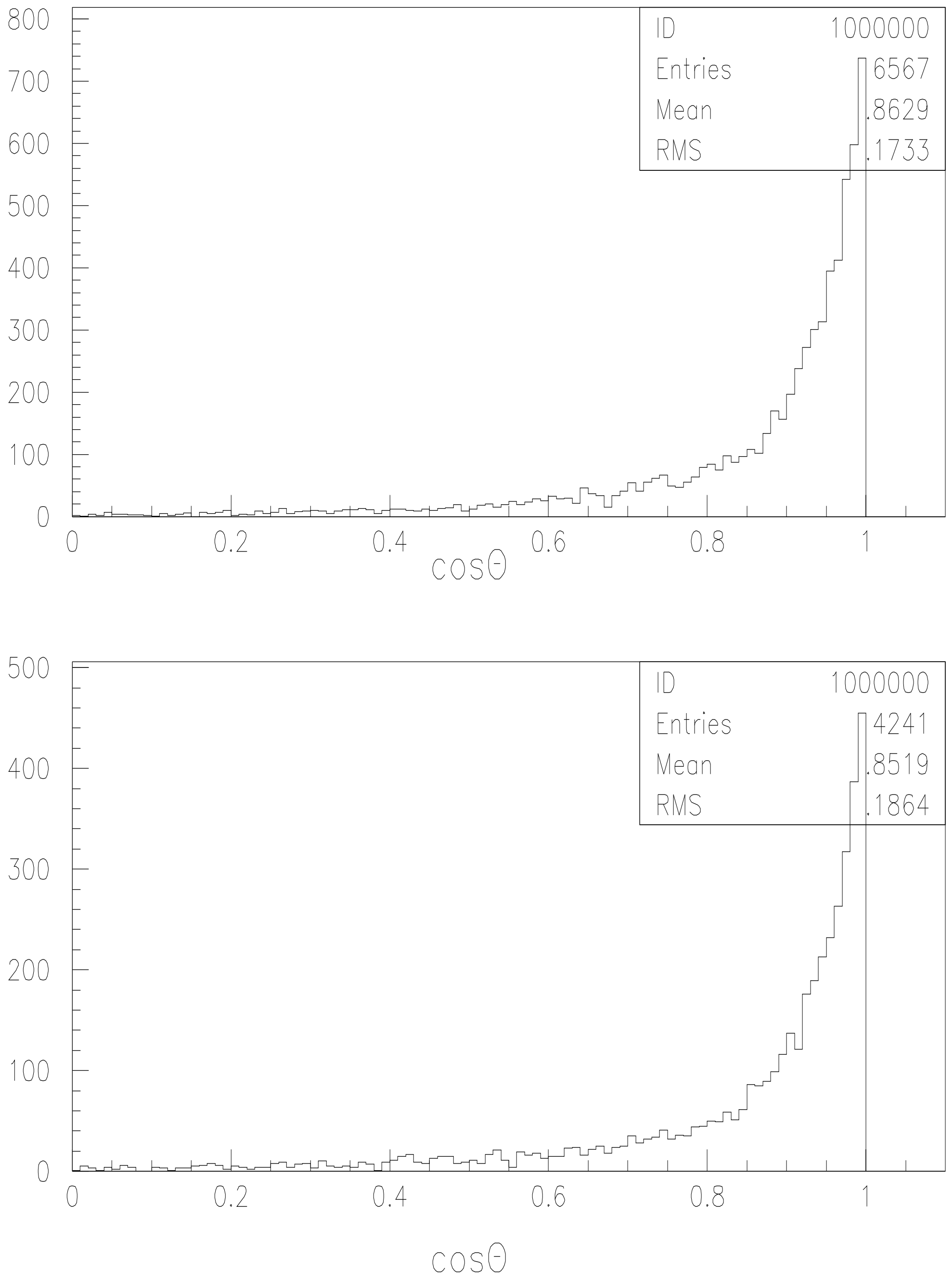


Figure 14: Cosθ distributions of ^{16}N run10488(upper plot) and MC data(bottom plot). θ is the angle between the fitted direction and the vector from the source to the fitted vertex. A cut of Nhits *geq* 45, 150cm < dis < 500cm, cosθ > 0 and rfit < 700cm is applied. dis is the distance between the source and the fitted position.

Table 5: Results of the K-S test on the systematic difference between the angular distribution of the ^{16}N data and that of the MC prediction. Data from run10353 were used.

Distance from the source	Maximum K-S difference	Probability
100cm $<$ dis $<$ 500cm	0.0297	0.1205
150cm $<$ dis $<$ 500cm	0.0501	0.1768
200cm $<$ dis $<$ 500cm	0.1264	0.0206

4 Conclusion

We use Monte Carlo to study performance of the grid fitter on both vertex reconstruction and direction reconstruction. Following definitions of fitter characters described in the paper, for the CC events, the vertex resolution is roughly 20cm while the direction resolution is roughly 26 degrees.

We compare both the laser and the ^{16}N data with the MC prediction to study systematic error of the grid fitter. We find that the vertex systematic is roughly 10cm while the direction systematic is roughly 6 degrees.

We find that a more optimized value for the light speed may tangibly improve the fitter performance.

The paper does not cover the systematic caused by the possible error on the nominated laser position and the chi2 problem. Preliminary result indicates the latter effect could be very large.

References

[1] C.J.Jillings, The Electron-Scattering Reaction in the Sudbury Neutrino Observatory, Ph.D thesis, Queen's University, Canada.

[2] Hamish Robertson, private communication, 2000..