Study of π^0 Events in MicroBooNE and Applications to the Deep Learning LEE Search
MICROBOONE-NOTE-1090-PUB

The MicroBooNE Collaboration
MicroBooNE_Info@fnal.gov

June 15, 2020

Abstract

We present a selection of events containing π^0 particles in the MicroBooNE detector. These events are used to understand the important pi0 background in the deep learning (DL) based electron low energy excess (e-LEE) analysis. The selection uses a new shower reconstruction algorithm based on tools in the DL reconstruction chain. We present the efficiency and energy resolution of the selection. We also present the planned use of these events in energy calibration using the reconstructed π^0 mass.

1 Introduction

The MicroBooNE experiment at Fermilab is a liquid argon time projection chamber (LArTPC) detector built to understand the nature of the Low Energy Excess (LEE) observed by the MiniBooNE experiment [1]. One of the analysis techniques employed by The MicroBooNE collaboration uses deep learning (DL) tools to reconstruct low-energy electron-neutrino events with one electron and one proton connected at the vertex. This topology can be mimicked by neutral current π^0 events with a proton and with one gamma close to the event vertex. Similarly, charged current muon-neutrino events with one gamma close to the vertex can be mistakenly interpreted as an electron neutrino event if the muon is mis-reconstructed as a proton. This makes π^0 events an important background to the MicroBooNE LEE analysis.

In this note we discuss a new shower reconstruction algorithm, how this new reconstruction is used to calculate π^0 mass, and how the π^0 events are selected. We will then present the results of this selection including the efficiency and energy resolution, how this π^0 mass will be used for energy calibration, and further discuss the π^0 background to the DL low-energy excess (LEE) Analysis and it's possible use as a side-band.

2 2D Shower Reconstruction

The first step in π^0 reconstruction is reconstruction of the decay gammas. In the MicroBooNE detector contains three wire-readout planes, two induction plane and one collection plane. Images are created for each plane that show the amount of charge deposited on each wire in the plane over readout time window. The wire and time information is transformed into row and column number where the value of each pixel corresponds to the amount of charge deposited or ADC value. The decay gammas from the π^0 create shower-like patterns in the detector images.

To reconstruct showers, the results from Sparse SSNet [2] and the reconstructed vertices[6] are utilitized. Sparse SSNet gives each pixel a score of how track-like and how shower-like it is. The first step is to mask the image using the Sparse SSNet shower score. Pixels with an SSNet Shower score > 0.5 (with 1.0 being a perfect shower score) and ADC value > 10 are kept, all other pixels are masked out. In order to speed up the algorithm, the 2D image is then turned into a sparse matrix. For each non-zero remaining pixel a row, column, and ADC value is saved. This greatly speeds up parts of the algorithms which involve looping over every pixel.

A template triangle is then placed at the reconstructed vertex. Vertices are reconstructed using another algorithm [6]. The vertex finder looks for "vee" shapes of ADC in 2D images and then finds 3D consistent candidates. These 3d vertex positions are the starting point for the template triangle. Initial values for direction, length, and opening angle are used. These values were chosen to capture a significant fraction of the charge in the first 30 cm of a typical EM shower. Future plans for improving the algorithm involve fine-tuning these values. The triangle is then spun around the vertex to choose the shower direction. The direction is chosen is the triangle which contains the most pixels with non-zero ADC.

Once the direction is chosen, the shower start is then moved away from the vertex. The best shower start is chosen again based on the number of enclosed pixels with non-zero ADC. This allows a gap between the shower start and the neutrino vertex, a necessary step to look for the gammas from π^0 decay. Once the direction and gap size have been chosen, length and opening angle are similarly calculated.

In order to get the energy of the shower, all of the ADC enclosed in the triangle is added up. A conversion equation is found by running the shower reconstruction code on "good" reconstructed vertices and comparing the enclosed ADC to true electron energy. A good reconstructed vertex is a reconstructed vertex within 2 cm of a the true vertex. The result is shown in Fig. 1 for a sample of monte-carlo ν_e overlay with cosmic ray data. The black points on this plot are the Gaussian mean of fits to the vertical bins. A best fit line is calculated for these points. The resulting equation is:

$$E = ADCsum * 0.0130 + 31.6 \tag{1}$$
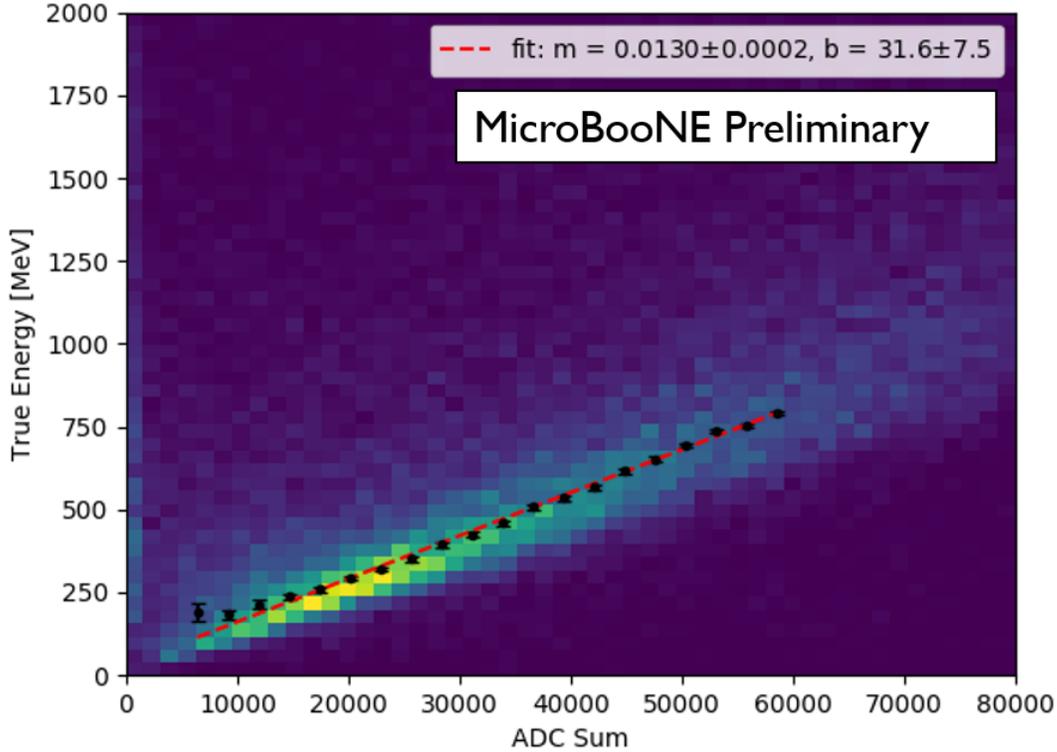
Figure 1: Shows the comparison between true electron energy and the total ADC enclosed in the reconstructed shower. The best fit line is used to calculate the conversion from ADC to energy in MeV.

There is a non zero offset that results. If the fit is rerun without the lowest energy points, the non zero offset remains. Future plans for calibrating the shower energy calculation are presented in Section 6.

Once a shower has been found by the reconstruction algorithm, the pixels found in the shower are masked out. If the total amount of ADC remaining is > 5000, the second shower algorithm is run. This follows the same process as the first shower, just with the first shower masked out. Additionally, the initial length is extended to 60 cm to capture showers that start further from the decay vertex. Later in the selection part of the analysis, showers with < 35 MeV are cut. This is useful here in ensuring the second shower algorithm did not just find noise and is discussed further in Section 4. The resulting energy and angle resolutions are shown in Section 5.

3 π^0 Reconstruction

In order for the π^0 mass to be reconstructed for an event, two reconstructed showers in the collection plane are required. The collection plane is used to calculate the energy of the shower.

Calculating the 3d opening angle is a more complex process. As the shower reconstruction is performed in 2d, the first step is to match the 2d showers from the collection plane to the other planes. Each shower in the collection plane is compared to the up to 4 other possible showers in the other planes and an overlap fraction is calculated. Each pixel in the collection plane shower is

analyzed to see if the time (row) coordinate overlaps with a pixel in the shower it is being compared to. The overlap fraction is calculated as:

$$OverlapFrac = \frac{NumMatch}{N} \tag{2}$$

where N is the total number of shower pixels in the collection plane shower. A threshold of $>$ 0.5 overlap is used to consider two showers as having good overlap. Once the overlap fraction has been calculated, if both collection plane showers have been matched to a showers in another plane, the 3d direction of the showers can be determined. Here it is important to note that in order for the π^0 to be reconstructed, a number of cuts have been necessary.

1. 2 collection plane showers >35 MeV

2. Both collection plane showers overlap fraction with another shower > 0.5

3. If a collection plane shower matches with showers in both u and v, the one with the highest overlap fraction is chosen.

4. The collection plane showers cannot match to the same shower in another plane

The 3D direction is then calculated for each shower. All matching pixels are converted into 3d points using time and the 2 wire coordinates. Principal component analysis is used to determine the center of the 3d points. The 3d line between the vertex and the center is calculated. Once these lines are calculated, the angle between the lines (θ) can be calculated.

At this point the π^0 mass can be found by the equation:

$$M^2 = 4E_1 E_2 sin^2(\frac{\theta}{2}) \tag{3}$$

where E_1 and E_2 are the reconstructed energies of the photons.

4 Selection of π^0 Events

Using the calculation in Section 3, the π^0 mass peak can be plotted. In order to make this plot, pre-selection cuts are applied. These are:

1. PMT Precuts used in the general DL Analysis [3]

2. The reco vertex is in the fiducial volume.

3. Number of reco tracks = 2

4. π^0 mass has been calculated (includes all cuts described in Section 3.)

These pre-selection cuts are applied (except for π^0 mass) have been added to match with the DL e-LEE selection [3]. This was done for the purpose of understanding the π^0 events in the context of a background to the main analysis. For example, the boosted decision tree (BDT) used in the DL e-LEE only produces a score for events passing this pre-selection. In future work, the π^0 selected events' BDT scores will be analyzed [3]. Further, the cut on the number of reconstructed tracks(N) is not expected to have a big impact on this analysis. N is unlikely to be 0 or 1 because of how the deep learning vertex finding algorithm works [6]. N >2 will similarly be a small contribution.

Another cut used in the reconstruction, described in Section 3, is the requirement that reconstructed shower energy > 35 MeV. This value was also chosen to line up with the DL e-LEE selection.

In the future we will analyze π^0 events independently of the LEE search. At this stage we will reanalyze the cuts used in the pre-selection and optimize them for events with a π^0 in the final state instead of for the DL e-LEE search.

For a first simple selection, multiple variables from the 1e1p analysis and shower reconstruction have been investigated. Those that show some separation have been used to make simple box cuts to further refine the selection are:

1. Reconstructed energy of leading shower > 80 MeV

2. The ADC sum near the vertex is > 250 ADC.

3. Electron reconstructed theta < 1.5 radians(angle of leading shower w.r.t beam direction, used in the DL-eLEE BDT)

4. The angle between the two showers is < 2.5 radians.

The values of these cuts have been chosen by eye to try to retain efficiency while improving purity. The plots of these variables can be seen in Appendix A. The result of adding these box cuts is shown in Fig. 2. This plot compares the monte-carlo with the open 5e19 POT of MicroBooNE data. Both have been scaled up to 7e20 to match the expected POT for the first future open data sample in order to illustrate expected event counts in this data set. The monte-carlo has been split into multiple categories. "NC π^0" are neutral current π^0 events with a good reconstructed vertex, where a good reconstructed vertex is within 5 cm of the true vertex. "CC π^0" are defined similarly for charged current π^0. "Offvtx π^0" are true π^0 events with poorly reconstructed vertices. "Non π^0" events are broken into good and bad vertices. ν_e events are all events that originated from a ν_e.

The simulated events have been scaled to match the 4.4e19 data POT. This is the open run1 data set. Systematic uncertainties are found using the same procedure as the whole DL analysis and are described further in [4]. Flux and cross section systematics are included on Fig. 2 and Fig. 6. Detector systematics have not yet been incorporated into the π^0 study. This is also true of Fig. 6.
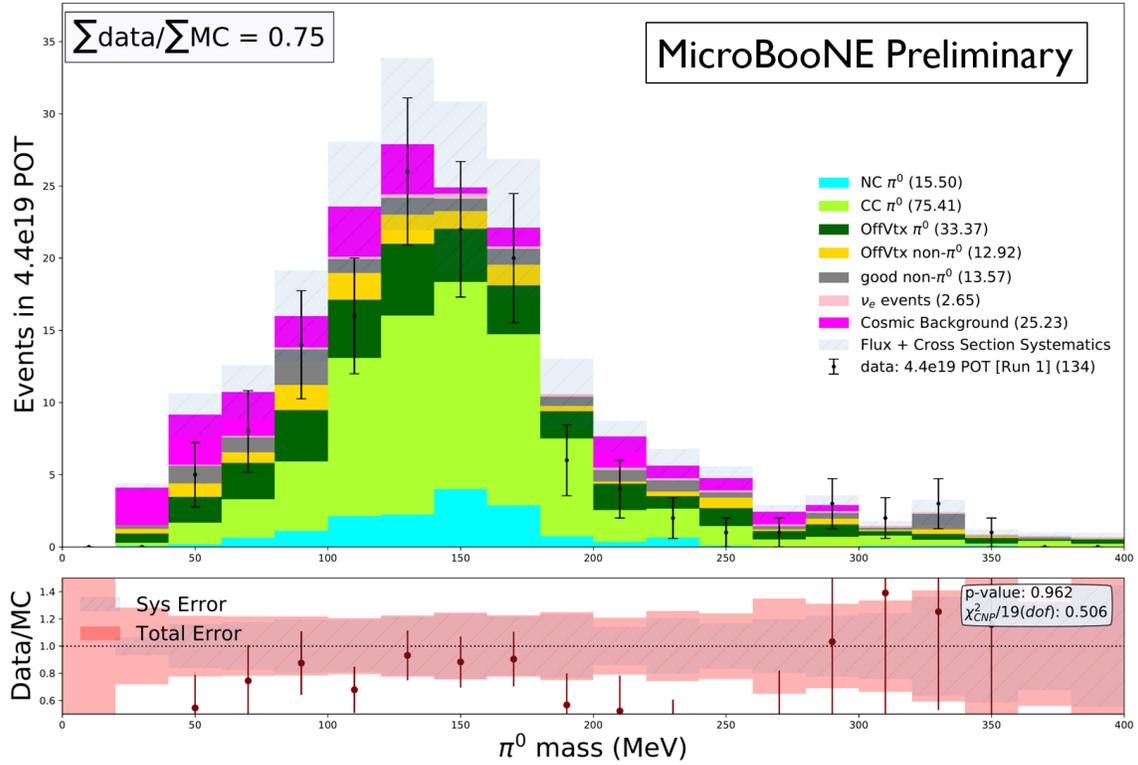
Figure 2: Shows the calculated π^0 mass for events passing simple box cuts. The simulated events have been scaled to match the 4.4e19 data POT. The expected invariant mass is 135MeV.

A gaussian fit to the data distribution gives a mean of $145.28 +/- 4.96$ and a width of $43.60 +/- 4.96$. A gaussian fit to the MC distribution gives a mean of $147.12 +/- 7.95$ and as width of $50.57 +/- 7.96$. The expected invariant mass is 135 MeV, neither distribution matches this within uncertainty. The mean of the two distributions match either other. The energy calibration plan described in Section 6 may shift these distributions.

The normalization of the data is lower than that of the MC simulation by about 25%. This is seen across the MicroBooNE collaboration in π^0 selections and is under investigation. Appendix B provides plots that show the data/MC agreement at various steps of the reconstruction and selection. This is a first step at diagnosing the cause of the disagreement. From these plots, the largest change in agreement appears at the second shower reconstruction, so that step will be one of the first parts to improve in the future.

5   Efficiency and Energy Resolution

There are many steps that go into reconstructing and selecting π^0's. The first of these, not discussed here, is the vertex finding stage [6]. In order to reconstruct a π^0, a reconstructed vertex must be placed. After this step, showers can be reconstructed. The efficiency of various stages of the 3d reconstruction and the selection is shown in 3. Efficiency is defined as:

$$Efficiency = \frac{num \ of \ events \ passing \ cuts}{num \ of \ true \ pi^0 \ events \ with \ good \ reco \ vtx} \tag{4}$$

6

Note that the denominator contains only events with a good reconstructed vertex. This is to closely analyze the π^0 specific reconstruction steps and cuts.
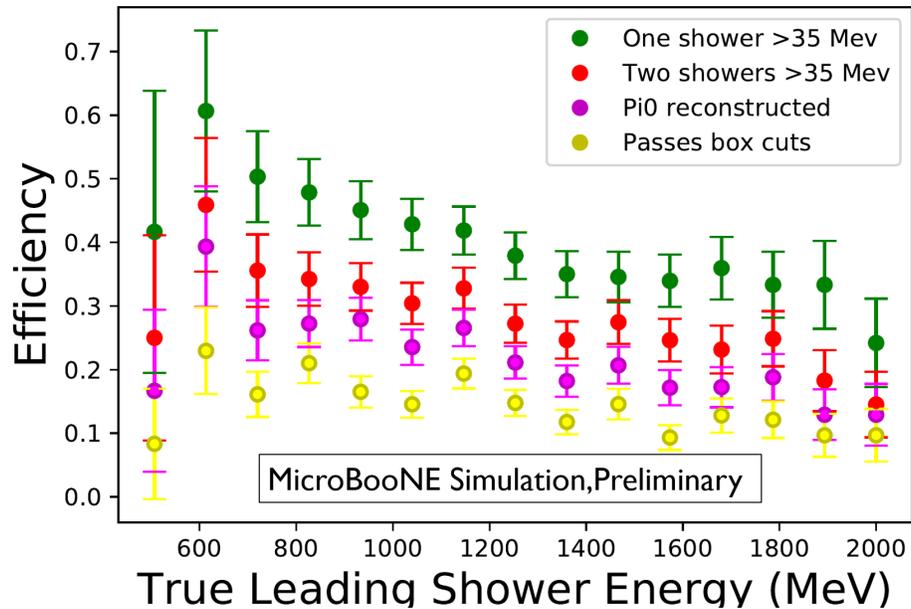

Figure 3: Shows the efficiency of various cuts leading to the π^0 selection

The final selection of box cuts is looped into one category of passes box cuts. The biggest drop in efficiency appears to come from the shower reconstruction. So that step of the reconstruction will be one of the first parts to look at improving in the future. The distribution has a peak at a true neutrino energy of 600 MeV, and efficiency drops off at higher energies. This is currently under investigation.

Shower energy resolution is analyzed to test the reconstruction in Fig. 4. This plot is made using a sample of true π^0 events with good reconstructed vertices. The first shower found is often closer to the true energy than the second shower. This is under investigation.
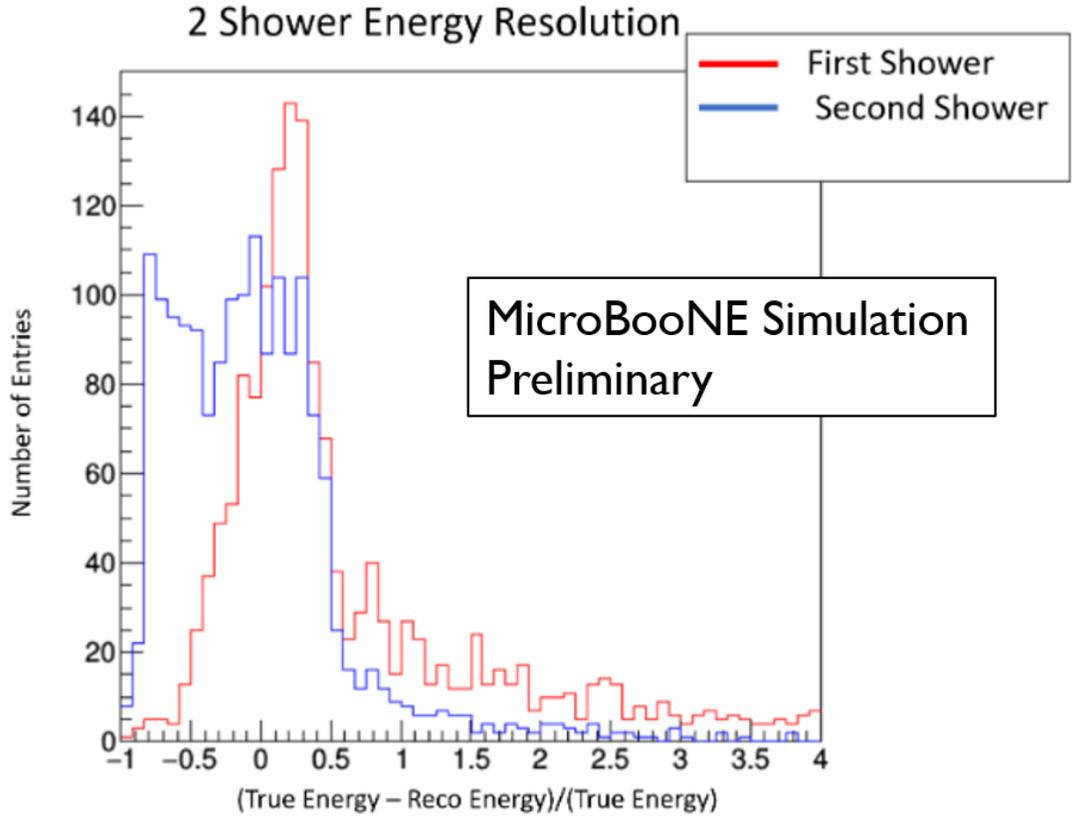
Figure 4: This plot shows the energy resolution for each of the two reconstructed showers for all vertices which has a π^0 mass calculated.

Another important metric is the 3d shower direction resolution. This is shown in Figure 5. This plot shows the cos of the angle between the true direction and reconstructed direction. There is a peak at one where the angle is within 15 degrees of the true value (the last bin in the plot). For the first shower 24.05% of the showers fall in the last bin and 20.71% of the second showers. All vertices with a reconstructed π^0 are shown (no box cuts applied). In this plot, the labels on the simulated events have the same definitions as those shown in Fig 2.
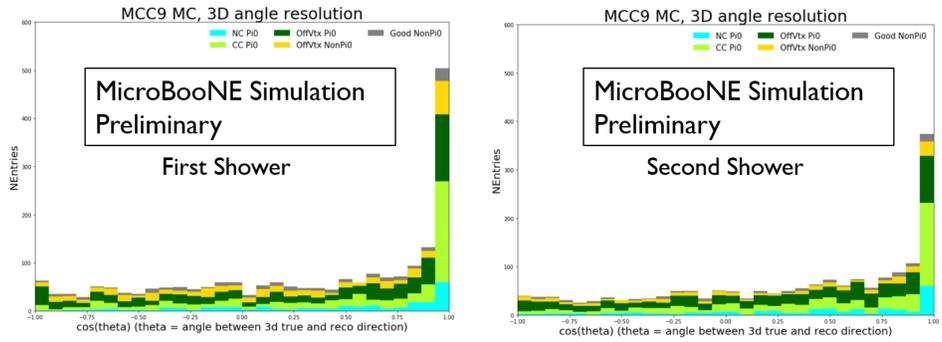
Figure 5: The cosine of the angle between the true and reconstructed showers for events with a reconstructed pi0 particles.

Another check one can do to test the reconstruction is to reconstruct the delta invariant mass.

This is done by reconstructing the 4-vector of the π^0 and the proton. Then the two vectors are used to reconstruct the delta particle. The proton reconstruction is detailed in [3]. The reconstructed delta invariant mass is shown in Fig.6.The expected invariant mass is 1232 MeV.
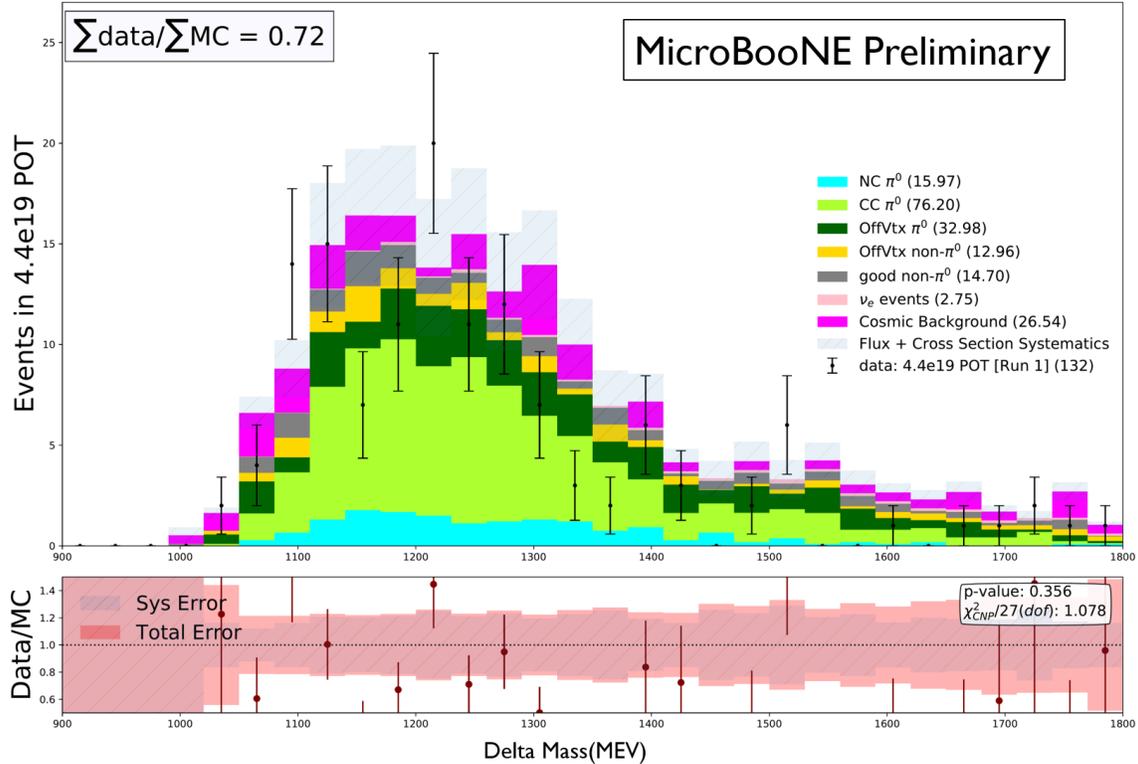

Figure 6: This plot shows the reconstructed delta mass, created using the π^0 reconstruction. Data and MC are both shown for comparison. MC simulation events are scaled to the 4.4e19 data POT.

6 Energy Calibration

Currently a simple equation described in section 2 is used to calculate energy in a shower. Using the reconstructed Michel stopping point and π^0 mass peak, further calibration can be achieved. Both of these values are well measured and can therefore be used to refine the ADC to energy conversion for showers. In order to calibrate the energy to maximize data/MC/truth agreement,the following process will be done for each data run separately (note that this is a general plan and will be refined as work progresses):

1. From the true energy vs. adc distributions, scan different values of the slope and bias, treating data and MC events separately.

2. Recalculate energy of showers using new values –leading to new π^0 mass and Michel stopping point

3. Optimize the chi2 of π^0 (and Michel) distribution.

This work is currently underway and is expected to aid in the general e-LEE analysis as this is very dependent on the shower (electron) energy reconstruction.

9

Further calibration may be added to just the π^0 reconstruction to account for the observed differences in data/MC. The MiniBooNE experiment re-weighted the MC events based on π^0 momentum [5]. A similar process will be attempted.

7 π^0 Background to the DL LEE Analysis

Events with π^0s are an important background to the DL e-LEE search. This search looks for events involving an electron neutrino which results in an electron and a proton. In the language of the deep-learning analysis, we are looking for events with one track-like proton and one shower-like electron (1e1p). This search is described in detail in [3]. Neutral current π^0 events can mimic this topology if a proton is produced and if one of the two decay gammas is located very close to the vertex. The second decay gamma may be further away or also close and just missed by the reconstruction. Charge current π^0's can also be a background due to the possibility of the muon either being missed in the reconstruction or mis-reconstructed as a proton. The π^0 selection can be used to help understand this background.

Events are selected for the DL e-LEE using a BDT as described in [3]. A quick way to help remove the π^0 background is to cut events which have a π^0 mass reconstructed. This is implemented in the 1e1p selection.

A sideband will be made of the π^0 events that have mid-range 1e1p BDT scores. This will involve looking at events with a mid-range BDT score that do not pass the DL-LEE cut (i.e. .5 to .8). These are events which the BDT has determined appear similar to the signal, but have differences. It is expected that there are a large number of π^0 events that meet this criteria. The π^0 selection will be applied to these events and then further analysis will be done to understand the 1e1p π^0 background.

8 Conclusion

In this note we have presented the new DL shower reconstruction for both one and two showers. Using this reconstruction, we have calculated the π^0 mass for candidates passing a pre-selection and simple box cuts. We observe a difference in the measured rate to that predicted by MC simulation, which is seen in other MicroBooNE analyses. Further work is required to understand this difference. The performance of the shower reconstruction in π^0 events has been analyzed using energy and angle resolution in simulated data. The efficiency of various π^0 selection cuts has been analyzed as well and shows that the shower reconstruction is a dominant source of inefficiency. Future work involves using the π^0 mass and Michel spectrum to further calibrate reconstructed shower energy. Currently this reconstruction is used to veto π^0 events that passed the 1e1p selection. A sideband will also be used to further understand the π^0 background to the 1e1p DL e-LEE analysis.

A   Simple Box Cuts

In this Appendix, the variables used for the simple box cuts in Section 5 are shown. These cuts are:

1. Reconstructed energy of leading shower >80MeV

2. The ADC sum near the vertex is > 250 ADC.

3. Electron reconstructed theta < 1.5 radians(angle of leading shower w.r.t beam direction, from the 1e1p BDT)

4. The angle between the two showers is < 2.5 radians.

The cut values are chosen by eye for each variable and aimed at maximizing both efficiency and purity. Other variables were looked at, but only those with some separation are shown. In these plots, only pre-selection and pi0 mass reconstruction cuts have been applied. The arrows indicate the direction of events that pass these cuts.
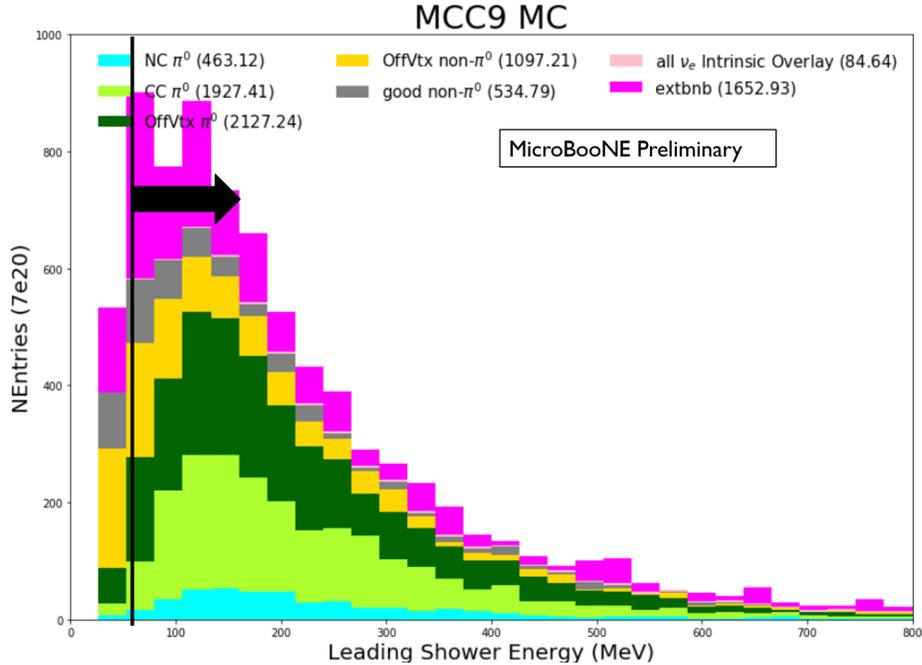
Figure 7: This figure shows the reconstructed leading shower energy. Anything < 80MeV is cut in the box cuts. Note that this cut is used with the purpose of greatly reducing the off-vtx non-pi0 events.
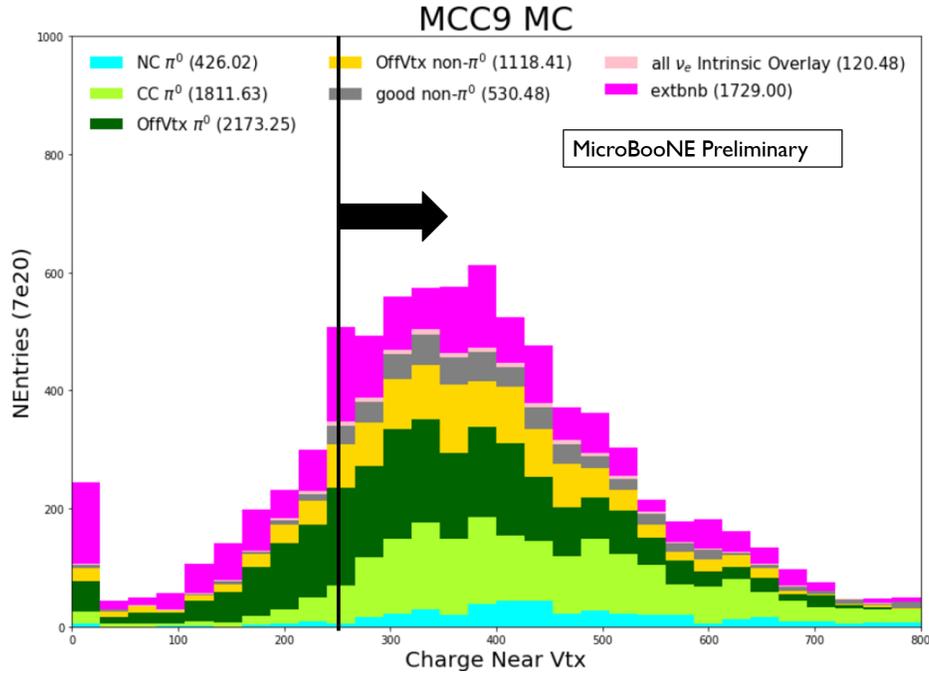
Figure 8: This figure shows the adc sum near the vertex. Anything <250 ADC is cut in the box cuts.
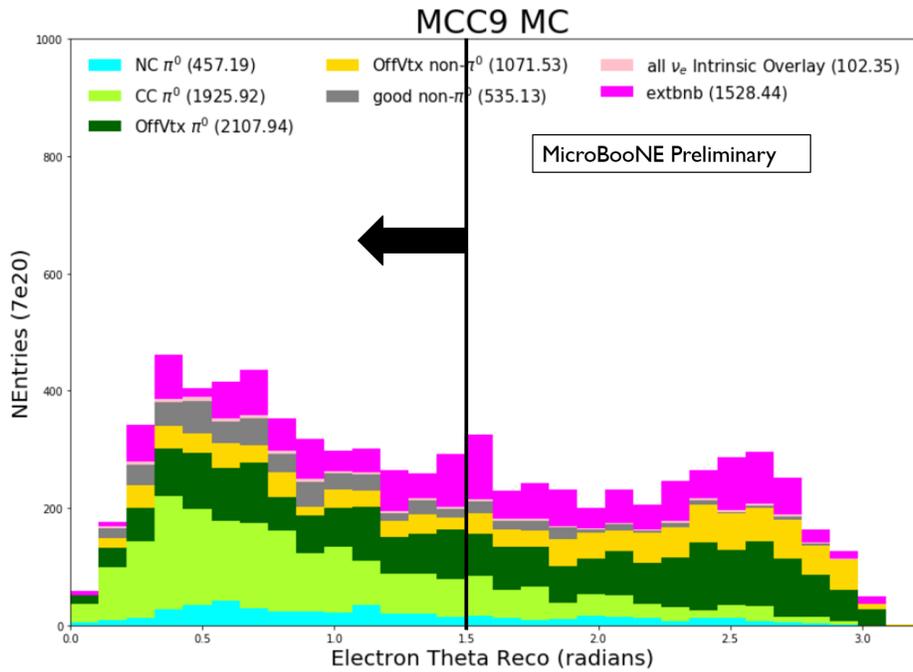
Figure 9: This figure shows the angle of the leading shower w.r.t the beam direction (from the 1e1p BDT). Anything >1.5 radians is cut.
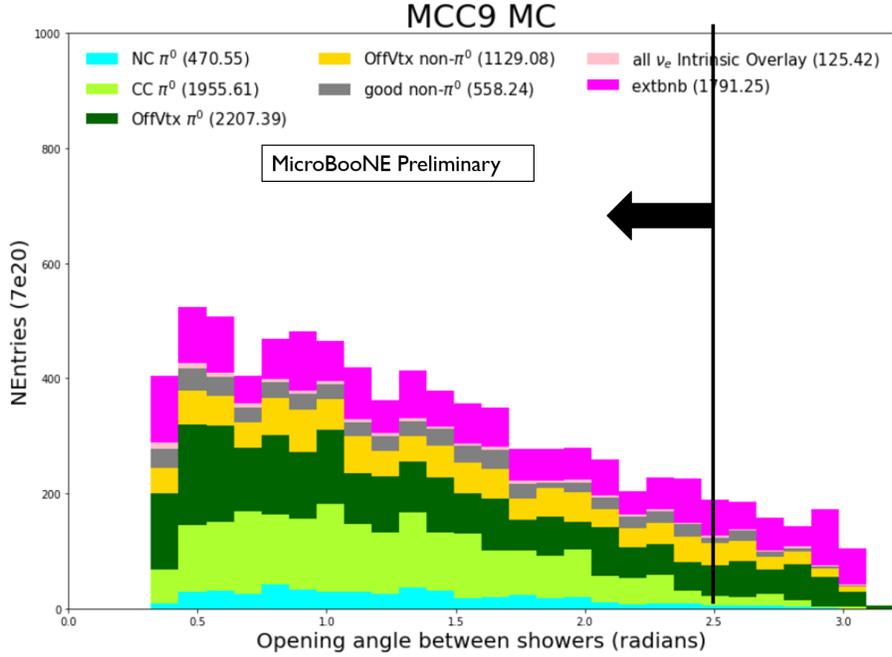
Figure 10: This figure shows the angle between the two reconstructed showers. Anything >2.5 radians is cut.

B Data/MC Agreement

To investigate the data/MC agreement, reconstructed neutrino energy was analyzed for various steps of the selection. The plots are shown here. The only pre-selection applied was, in the fiducial volume, pmt precuts, beam quality cuts, and a good runs cut. π^0 selection plots are added individually to look for changes in agreement. No conclusion about the cause of the disagreement have yet been drawn, though it is noted that the largest change in agreement appears in the second shower reconstruction.
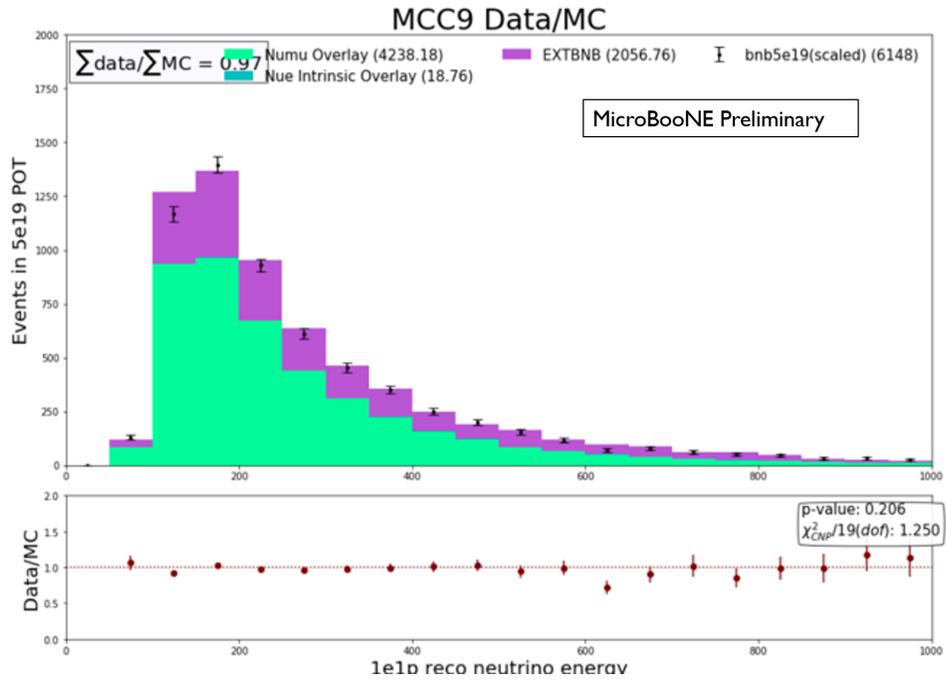
Figure 11: This figure shows reco neutrino energy. Cuts = preselection.
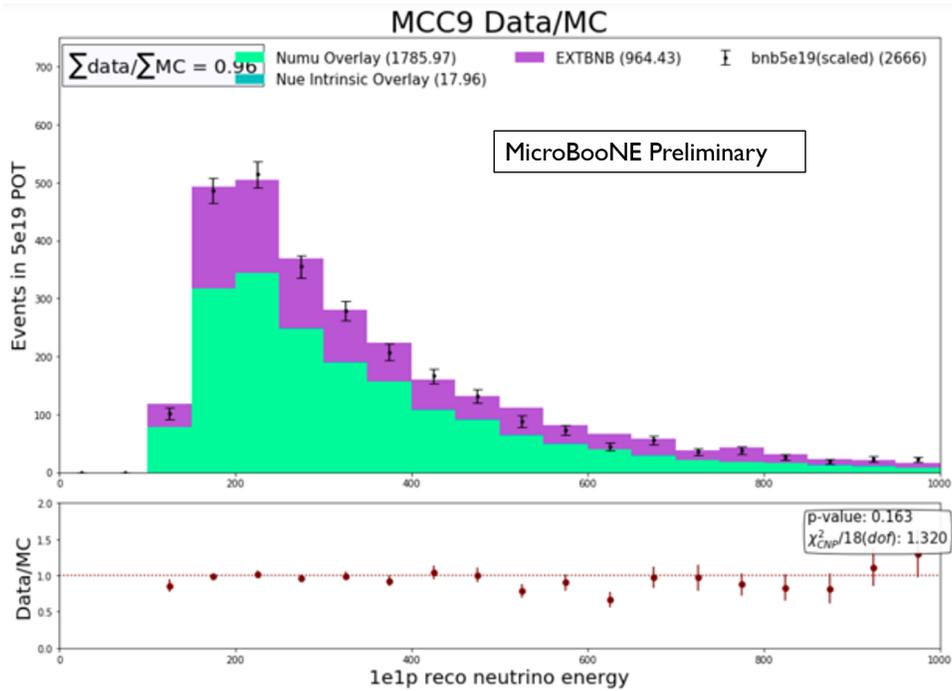

Figure 12: This figure shows reco neutrino energy. Cuts = preselection +one shower found.
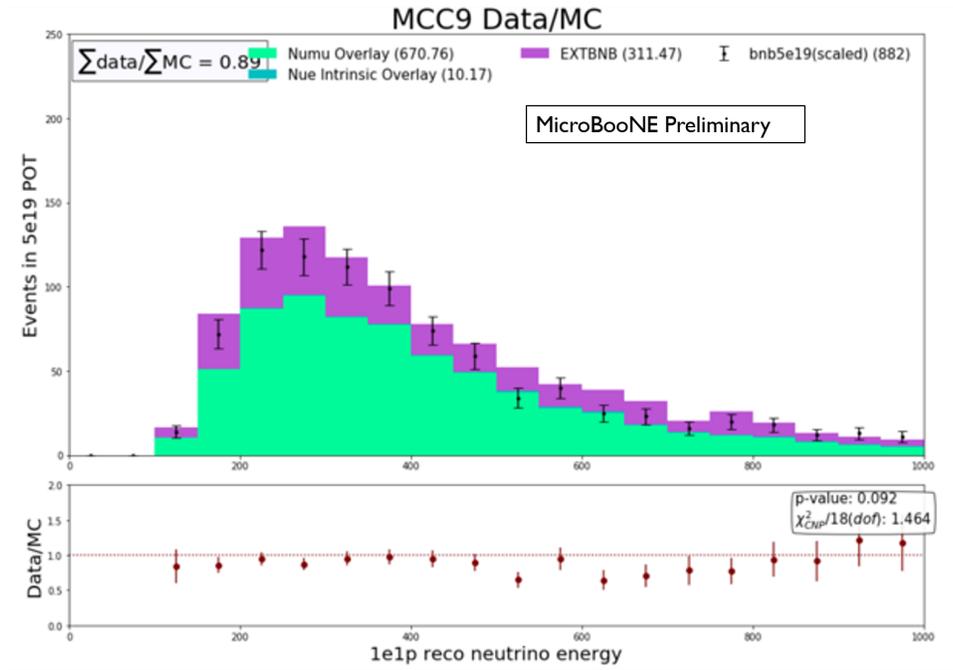
Figure 13: This figure shows reco neutrino energy. Cuts = preselection+ two showers found.
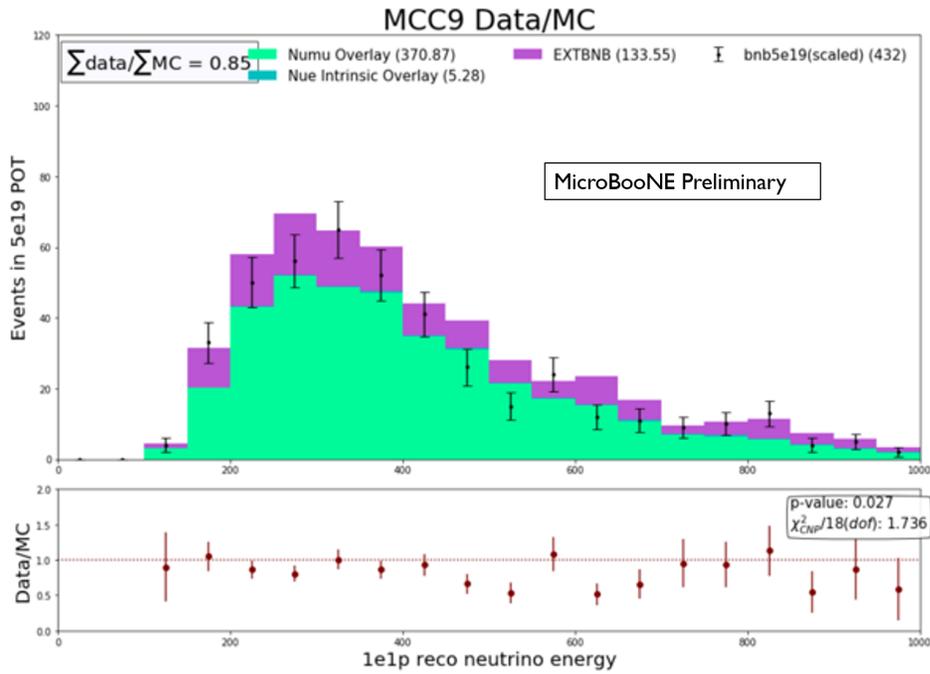
Figure 14: This figure shows reco neutrino energy. Cuts = preselection +pi0 reconstructed.
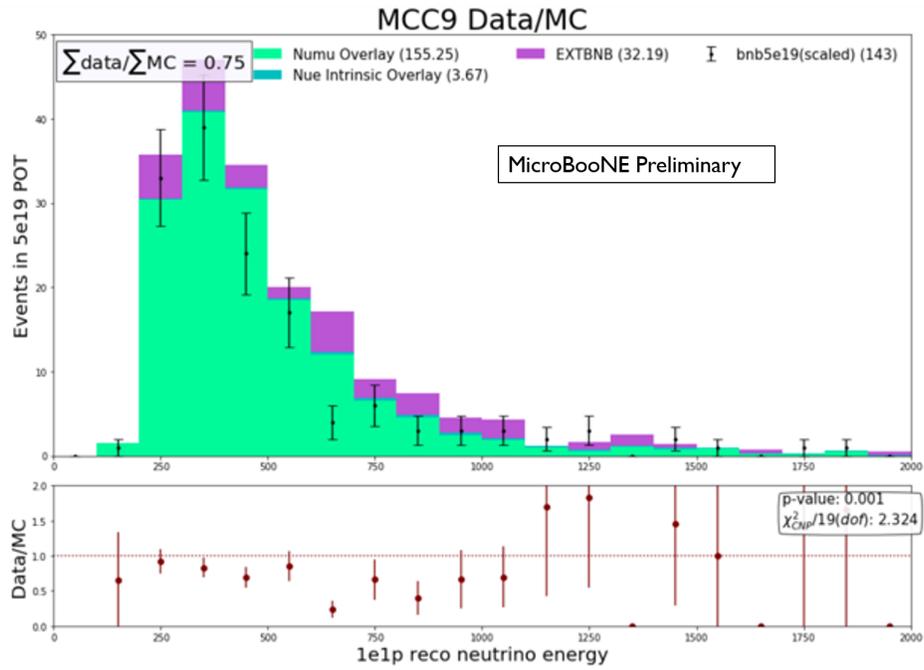
Figure 15: This figure shows reco neutrino energy. Cuts = preselection + pi0 reconstructed +pi0 selected.

References

[1] A. A. Aguilar-Arevalo et al., MiniBooNE Collaboration, Significant Excess of Electronlike Events in the MiniBooNE Short-Baseline Neutrino Experiment, Phys. Rev. Lett.,121,221801,2018

[2] Semantic Segmentation using SparseSSNet, Ran Itay, MicroBooNE docdb 30676-v1

[3] 1e1p public note

[4] systematic note

[5] J.M. Link,*Neutral Current π^0 Production in MiniBooNE*, arXiv:0709.3213v1,2007

[6] The MicroBooNE Collaboration, Vertex-Finding and Reconstruction of Contained Two-track Neutrino Events in the MicroBooNE Detector, https://arxiv.org/abs/2002.09375