Selection efficiencies in signal only samples for $p \to \mu^+\gamma$ and $p \to e^+\pi^0$ nucleon decay modes in MicroBooNE

William De Rocco, Elena Gramellini

August 24, 2015

Abstract

The study of nucleon decay in liquid argon time projection chambers (LArTPCs) is a subject of much interest since it widens nucleon decay searches to modes hardly detectable with other technologies. Assessing the performances of LArTPCs for complex topologies detection and background rejection is crucial for future massive LArTPCs like DUNE [1].
This note is the first of a series of 3 tech notes and addresses the first step towards assessing background processes to nucleon decay in MicroBooNE. Specifically, a procedure for the identification of the $p \to \mu^+\gamma$ and $p \to e^+\pi^0$ decay modes in MicroBooNE is presented along with preliminary selection efficiencies. Since the reconstruction of 3D objects plays a fundamental role in the identification of the nucleon decay modes, track reconstruction has been optimized for the $p \to \mu^+\gamma$ mode. The study of cosmogenic background will be addressed in a future note.

Contents

1 Introduction

In the Standard Model, baryon number is conserved, causing for the proton and bounded neutrons to be stable. However, many Grand Unified Theories (GUTs) predict the nucleon decay on long time-scales. Both gauge-mediated GUTs, in which new gauge bosons are introduced that allow for the transformation of quarks into leptons and vice versa, as well as supersymmetric GUTs allow for the nucleon decay. The experimental identification of even one proton decay would point to an indirect proof of GUTs and is therefore a subject of great interest [1].

Previous experiments have placed limits on the proton lifetime [2]. The dominant experimental setup for these searches has historically been a water Cherenkov detectors. This technology suffers from an inability to reconstruct events thru decay products that are below the Cherenkov threshold and relays on the gammas from nuclear deexcitation for those modes.

For this reason, an attractive alternate approach to identifying nucleon decay is the use of a liquid argon time projection chamber (LArTPC). LArTPCs have the ability to identify nucleon decay modes that a water Cherenkov detector has less sensitivity to, most notably the $p \to K^+\bar{\nu}$ (see Figure 3 for MC event display using Larsoft, raw digits only). Previous MC studies [3] have claimed a very high selection efficiency of many nucleon decay topologies, making this an important area of study.

Future LArTPC experiments such as DUNE will have active volumes large enough and will run for lengths of time that would allow for the identification of nucleon decay in liquid argon or, alternatively in the case of a null result, the ability to bound the proton lifetime in many channels. Though the active volume of MicroBooNE is too small to be able to place any limits on nucleon decay lifetimes, it provides an excellent source of data on backgrounds, primarily from cosmogenic sources, that will affect future proton decay searches in LArTPCs. Previous work has been done to assess the potential identification efficiencies for different decay modes in a LArTPC [3] but as of yet, no study of selection efficiencies and dominant backgrounds to nucleon decay in LArTPCs has been performed with data, nor taking into account reconstruction efficiencies. This note discusses the first step of an analysis that will make the first investigation into proton decay backgrounds in MicroBooNE for the $p \to \mu^+\gamma$ and $e^+\pi^0$ decay modes.

2 Strategy

The general strategy to identify nucleon decay backgrounds in MicroBooNE consist in the following 5 steps:

1. Study the signal in MC
2. Fixing selections and workflow in MC, calculate the selection efficiency
3. Apply analysis framework to cosmics MC
4. Apply analysis framework to cosmics data
5. Count how many cosmics events pass our selections.

This note concerns exclusively the first 2 steps for the $p \to \mu^+\gamma$ and $e^+\pi^0$ decay modes.
The $p \to \mu^+\gamma$ mode is particularly interesting for background searches in MicroBooNE. Given MicroBooNE surface positioning, we expect an important statistical background. The $e^+\pi^0$ decay mode is a dominant decay mode in many GUT models. This is the "golden mode" for water Chrenkov detectors and can serve as a standard candle.

We developed a series of cuts specific to each decay mode. Firstly, we apply the cut-flow on mcreco data products (in what follows also referred to as "Monte Carlo energy deposition") in a signal only sample. This first pass ensures a good quality of the cuts and serves as a normalization for the reconstruction efficiency. Secondly, we tune our selection parameters on reco data products in the signal only sample. After this second pass, all the parameters of the cut flow are fixed and the reconstruction efficiency are calculated. The ratio between the number of mcreco and reco events that pass our cuts determines the efficiency to detect that decay topology.

3 MC samples

We used two different MC samples for this analysis, which are defined as follows:

1. $p \to \mu^+\gamma$ signal only
2. $p \to e^+\pi^0$ signal only

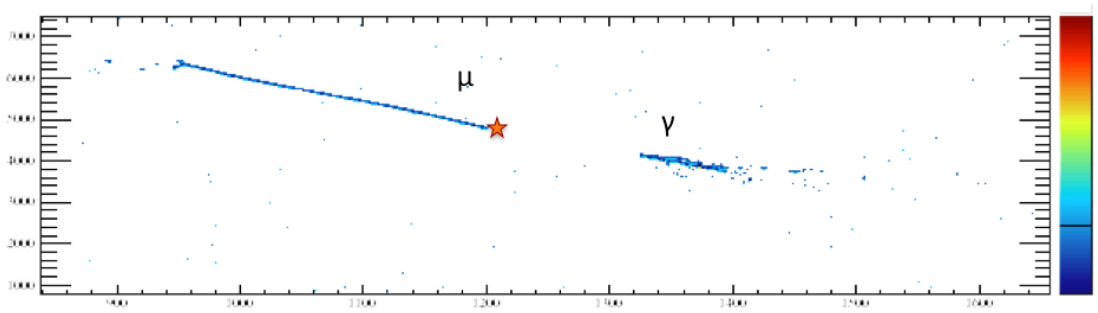

Figure 1: Event display for the $p \to \mu^+\gamma$ mode. Raw digit only.

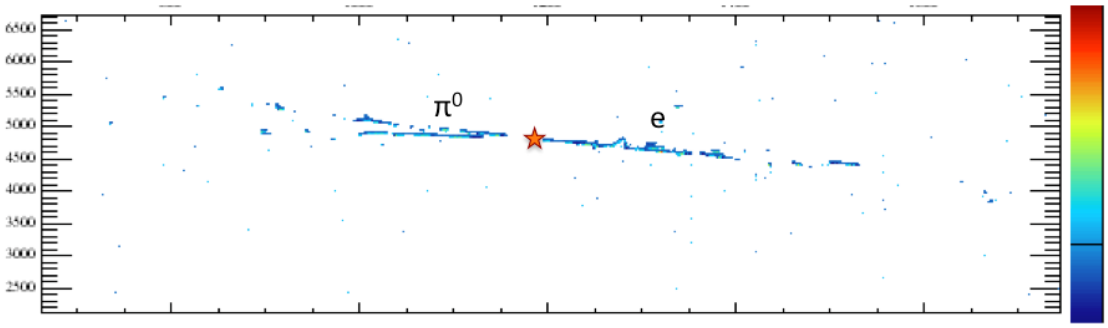

Figure 2: Event display for the $p \to e^+\pi^0$ mode. Raw digit only.

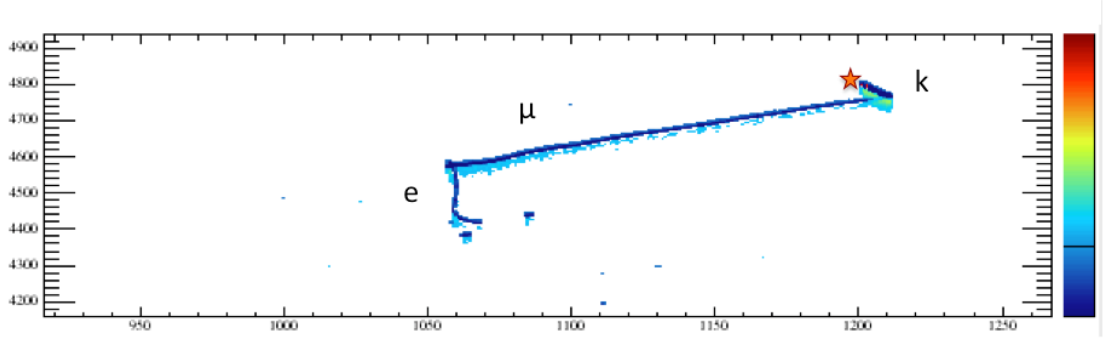

Figure 3: Event display for the $p \to K^+\bar{\nu}$ mode. Raw digit only.

For both the samples, we used the reco data products "showerrecopandora" for shower reconstruction and "kalmanhit" for track reconstruction. Additional studies with different data products are viable, but not yet performed.

1. $p \to \mu^+\gamma$ **signal only** Each event of this sample contains a single proton decay in the $p \to \mu^+\gamma$ mode, generated at the center of the TPC. A visual representation of one event can be found in figure 1, generated with Larsoft event display, raw digits only, 5 ACD threshold. 10000 $p \to \mu^+\gamma$ decays were generated using a slightly modified version of the GENIE 2.8.6 package gevgen_ndcy. We added the $p \to \mu^+\gamma$ decay mode to the standard gevgen_ndcy generator by implementing a muon and a gamma in the two body decay list of modes. The modified code can be found in /uboone/app/users/elena/NewGENIE/altups/genie/my286alt/Linux64bit+2-debug-e7/
GENIE_my286alt/src/NucleonDecay/ .
No nuclear re-interaction occurs for the muon and the gamma, so no other action was taken on the GENIE data set. The GENIE sample can be found in
/uboone/data/uboonerd/nucleonDecay/decayModes/Mode13/Mode13.out .
We used larsoft and uboonecode version v04_12_00 to parse the GENIE output. We used a slightly modified version of NDKGen and its parser, so to be able to run on the grid and facilitate the bookkeeping between GENIE event number and the larsoft event number. The module and its parser are respectively in
/uboone/app/users/elena/LarDev04_12_00/srcs/larsim/EventGenerator/NDKGen_module.cc
and
/uboone/app/users/elena/LarDev04_12_00/LiteProduction/Mode13/v04_12_00/
sim/prod_NdK/prodndkGoldenSim13.fcl.
Standard reconstruction and standard conversion to larlite format were finally applied. 10000 events were passed to the analysis stage. We modified the REco track attributes to correct for reconstruction pathologies during the analysis stage only (see 4.1.1).

2. $p \to e^+\pi^0$ **signal only** Each event of this sample contains a single proton decay in the $p \to e^+\pi^0$ mode, generated at the center of the TPC. A visual representation of one event can be found in figure 2, generated with Larsoft event display, raw digits only, 5 ACD threshold. About 2500 $p \to e^+\pi^0$ decays were generated with the GENIE 2.8.6 package gevgen_ndcy. However, in the $e^+\pi^0$ mode, GENIE simulates hadronic re-interaction in the nucleus. So, the π^0 can cause for the ejection of a secondary nucleon as it escapes. Alternatively, the pion can fail to escape the nucleus and its energy goes instead into the ejection of several additional nucleons. This analysis has no sensitivity to these events, so these events are removed at generator level. Only 48% of signal is kept at this stage. Therefore, in the context of this analysis, we define a signal event as a single proton decay with no additional ejected nucleons. We decided to parse with larsoft a skimmed sample of 10000 GENIE events that can be found in
/uboone/data/uboonerd/nucleonDecay/decayModes/Mode0/Mode0.out .

We used larsoft and uboonecode version v04_12_00 to parse the skimmed sample. We used a slightly modified version of NDKGen and its parser, so to be able to run on the grid and facilitate the bookkeeping between GENIE event number and the larsoft event number. The module and its parser are respectively in
/uboone/app/users/elena/LarDev04_12_00/srcs/larsim/EventGenerator/
NDKGen_module.cc
and
/uboone/app/users/elena/LarDev04_12_00/LiteProduction/Mode0/v04_12_00/
sim/prod_NdK/prodndkGoldenSim0.fcl.
Standard reconstruction and standard conversion to larlite format were finally applied. Due to a grid job failure at the reconstruction stage, we were left with 9990 events for the analysis.

4 $p \to \mu^+\gamma$

The following paragraphs address the analysis of the $p \to \mu^+\gamma$ proton decay mode. The signal topology for this mode is constituted by one electromagnetic shower and one track pointing back to a common origin.

4.1 Signal only analysis

In an attempt to isolate the signal, the following procedure was implemented (for a detailed explanation about running the code, see Appendix A):

1. Tracks refinements

2. Passing thru tracks cuts and loose energy cuts
3. Combinatorics pairing
4. Geometrical cuts
5. PID and Calorimetry cuts

4.1.1 Tracks refinement

Hand scanning reconstructed events and comparing the length of MCReco tracks with the length of Reco tracks highlighted some pathologies of track reconstruction. The following three aspects were particularly noticeable:

1. there are cases where the gamma shower is reconstructed as one or multiple tracks
2. there are cases where the michel electron is attached to the muon track
3. there are cases where the gamma is attached to the muon track

As a first action to eliminate the mis-reconstructed showers from the track pool, only tracks with more than 10 3D points in MCReco data product and tracks with more than 100 3D points in Reco data product are selected[1]. In general, showers reconstructed as tracks have very few 3D points.
Then, we apply a technique to refine the edges of the track. We start assuming that the central part of the track is well reconstructed and that the mis-association of 3D points can occur only at the edges. We divide the track in 5 segments, each containing a fifth of the total number of 3D points. We analyze the first and last segment. In the first segment, we calculate the distance between each 3D point starting from the first 3D point. If the distance is greater that 5 cm for Reco tracks (100 cm for MCReco), we claim the existence of a "gap". All the points before the gap are removed from the track. The same procedure is performed on the last segment of the track, starting from the last 3D point. We define then the edges of the refined track to be the first point of after the gap in the first segment and the first point before the gap in the last segment. We are using here the concept of track edge instead of beginning and end of the track since the track direction is not defined. In order to calculate the energy deposited by the track, we multiply the length of the refined track by 2.3 MeV/cm. No Bragg peak correction is performed. We can now comparing the muon energy deposition calculated with this method with the muon energy deposition from the Reco data product. We find a better accordance between MCReco and Reco using the former method.

4.1.2 Combinatoric pairing

As the first stage of event selection, we perform a combinatoric pairing. Every combination of one shower and one track is kept as a preliminary candidate. This implies that the muon track can be paired to its michel electron at times. The idea of this stage is to be as inclusive as possible, even at the risk of over-identify pairs in the event. Further cuts are applied at a later stage will eliminate mis-id pairs, like muon - michel e pairs.

4.1.3 Passing thru tracks cuts and loose energy cuts

The second stage of event selection is the application of some loose cuts. Every track passing thru the detector is discarded. The track is defined as "passing thru" if one of the following conditions is met:

1. One edge x $>$ 3 cm and the other edge x $<$ 250 cm
2. One edge y $>$ -113 cm and the other edge y $<$ 113 cm
3. One edge z $>$ 3 cm and the other edge z $<$ 1050 cm

It is important to notice that this selection keeps tracks that originates in the detector but are not contained as well as tracks that enter and stop the detector. Shower and tracks with energy deposition greater than 2 GeV are also discarded.

4.1.4 Geometrical cuts

Three geometrical cuts are then applied. The first geometrical cut performed is a cut on the angle between the 3-momenta of the reconstructed μ and γ. Naively, this would be expected to be steeply peaked near π because the proton decays at rest and momentum is conserved. This is not the case. Two factors contribute

[1]The difference in the value of this cut is due to the way the MCReco and Reco data products are stored. The MCReco track data product saves only the 3D points where physics process takes place in G4. On the contrary, the spacing among 3D points in Reco track data products is driven by actual wire separation (3mm), and therefore a count of trajectory points is inherently different from MCReco track data product.

to this. The first is that the proton does not decay at rest, but instead decays with some momentum less than the Fermi momentum of the nucleus, which is on the order of 200 MeV/c. In addition to this effect, the direction of the track is mis-reconstructed at times. This implies the opening angle distribution to be peaked at both 0 and π. For these reasons, the opening angle cut is quite loose, since we keep pairs with angle greater than 1.2 rad and pairs with angle less than 0.5 rad.
The second geometrical cut is a cut on the distance of closest approach between the shower and the track (the impact parameter, IP). This cut assesses the coplanarity of shower and track. The third geometrical cut is a cut on the distance between the start point of the shower and the start point of the track. We define the "start point of the track" to be the track edge closest to the shower under consideration. This cut serves as a proximity cut.

4.1.5 PID and Calorimetry cuts

After geometrical cuts are performed, PID and calorimetric cuts on the decay products and reconstructed proton are performed. Tracks whose PID corresponds to muons and pions are kept, the PID information being retrieved from the MCReco (or Reco) track data product. For the shower ID, we define the interaction point to coincide with the start point of the track (see definition in the previous section). The γ-like nature of a shower is found by comparing the projective log-likelihood of its dE/dx and radiation length corresponding to a photon or electron. Specifically,

$$\mathrm{LL}_\gamma(dE/dx, radlength) = \log \frac{\mathrm{PDF}_\gamma(dE/dx, radlength)}{\mathrm{PDF}_\gamma(dE/dx, radlength) + \mathrm{PDF}_e(dE/dx, radlength)}$$

and

$$\mathrm{LL}_e(dE/dx, radlength) = \log \frac{\mathrm{PDF}_e(dE/dx, radlength)}{\mathrm{PDF}_\gamma(dE/dx, radlength) + \mathrm{PDF}_e(dE/dx, radlength)}$$

are compared and the greater log-likelihood determines the e-like or γ-like nature of the shower. For details about the "PDF" see appendix A.

The calorimetry cut consists of a simple cut on the reconstructed deposited energy of the gamma. Additionally, cuts are applied on the components of the reconstructed proton 3-momentum to seek out decays in which the proton has initial momentum not much larger than the Fermi momentum.

4.2 Results

The preliminary results presented below show the efficiency of this selection methodology at the MCReco and Reco level. Table 1 shows that the cutflow keeps a good percentage of events when applied to both MCReco and Reco level. Calculating the ratio of the number of events that pass the cut flow for Reco over the number of events that pass the cut flow for MCReco (last line in Table 1), we find a overall reconstruction efficiency for this channel of ~ 56 %.

Table 1: Topology-based and calorimetry-based set of cuts to identify proton decay. All samples are signal only.

Parameter	**Cut**	**MCReco**	**Reco**
Combinatorics total	-	18616	14370
Passing track filter	eliminate crossing tracks	18616	14370
Track deposited energy	< 2 GeV	18616	14358
Shower deposited energy	< 2 GeV	18616	14358
Angle between μ and γ	> 1.2 rad or < 0.5 rad	16168	12961
Impact parameter	< 15 cm	15520	8881
Distance between first energy depositions	< 70 cm	15236	7534
Track ID	μ or charge π	15233	7028
Shower ID	γ	9679	5885
Shower deposited energy	[65, 700] MeV	9678	5554
$\lvert\Sigma p_x\rvert$, $\lvert\Sigma p_y\rvert$, $\lvert\Sigma p_z\rvert$	< 400 MeV	9648	5372

We can now evaluate the goodness of the selected candidates by comparing the energy deposition distributions for the muon, the gamma and the total energy (which is simply the sum of the two) at the MCReco and Reco level. Figures 4, 5 and 6 show the energy deposition of the muon, the gamma and the total for MCReco products only. A comparison of the MCReco vs Reco distributions for the muon and the gamma can be found in figures 7, 8. The muon deposited energy is defined as in 4.1.1. From MCReco to Reco, the

deposited energy distributions broaden slightly due to imperfections in clustering and, in the gamma case, to imperfection in shower reconstruction. Despite this, the peaks appear to be in roughly the same position and the distributions are comparably shaped, indicating that the events selected were well-reconstructed. While the reconstruction of the μ is very good, especially after the track refinement, the reconstruction of the γ can be tricky. The primary error made during the reconstruction of these events is to reconstruct the γ as multiple showers. The shower energy is thus lowered, explaining the low energy tail in figure 8.

We can compare the total energy deposited distributions between MCReco and Reco level (figures 9, 10 and 11) and evaluate the power of our cut flow. Figure 9 shows the total deposited energy for all the combinatoric paris in the case of Reco. In the Reco distrubution, two peaks are visible. The lower energy peak is given by the muon-michel electron pair and the mis-reconstructed pieces of the gamma shower. In Figure 10, the geometrical cuts are applied and the low energy peak almost disappears. In Figure 11, geometrical, PID and calorimetry cuts are applied. The Reco distribution builds up to peak roughly the same position as MCReco when all the cuts are applied.

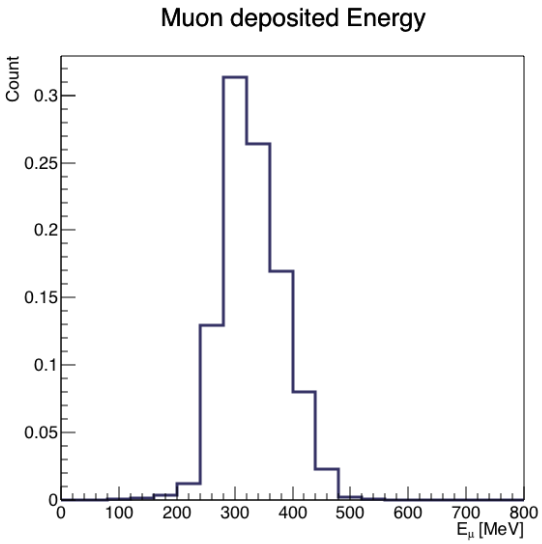

Figure 4: Deposited energy distribution of the muon at MCReco level. All cuts in Table 1 have been applied. The plot is area normalized.

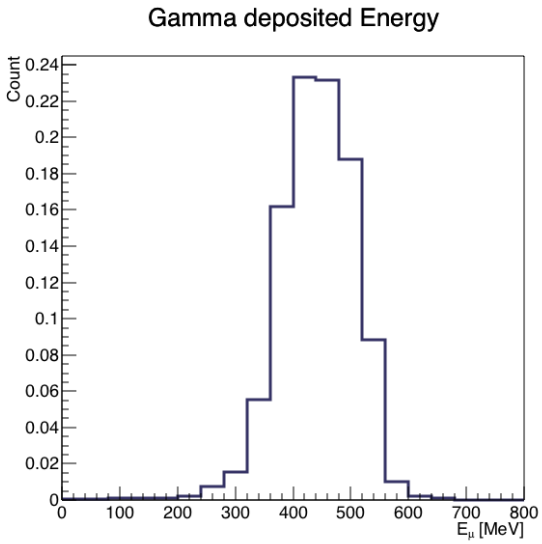

Figure 5: Deposited energy distribution of the gamma at MCReco level. All cuts in Table 1 have been applied. The plot is area normalized.

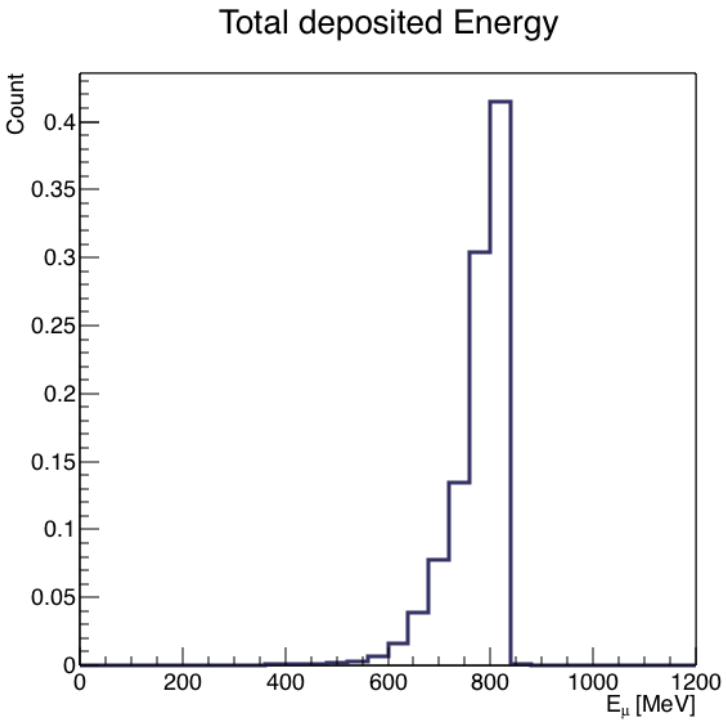

Figure 6: Candidate deposited energy at MCReco level. All cuts in Table 1 have been applied. The plot is area normalized.

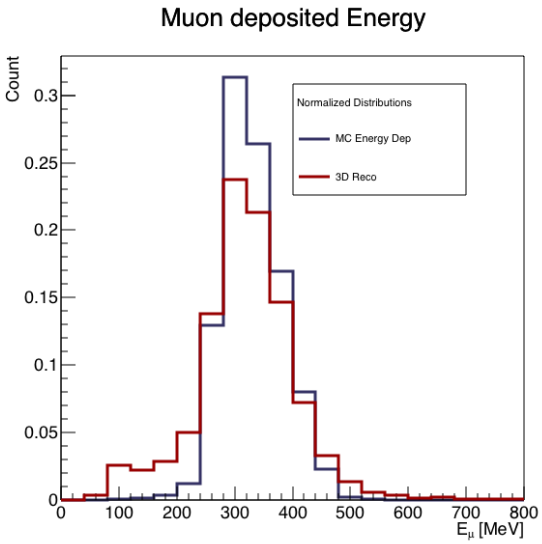

Figure 7: Deposited energy of the muon at MCReco and Reco level. All cuts in Table 1 have been applied. The two plots are area normalized.

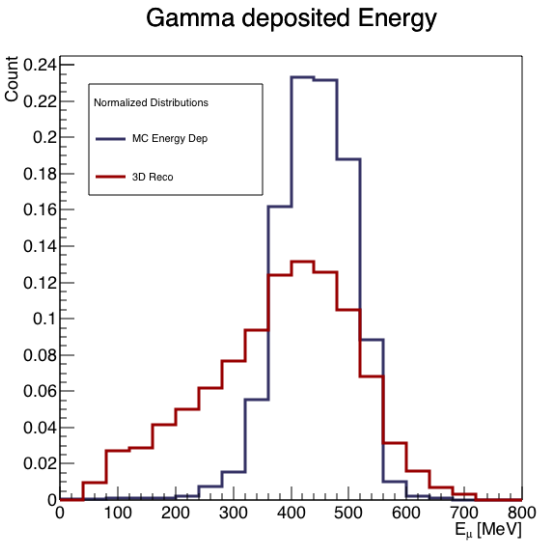

Figure 8: Deposited energy of the gamma at MCReco and Reco level. All cuts in Table 1 have been applied. The two plots are area normalized.

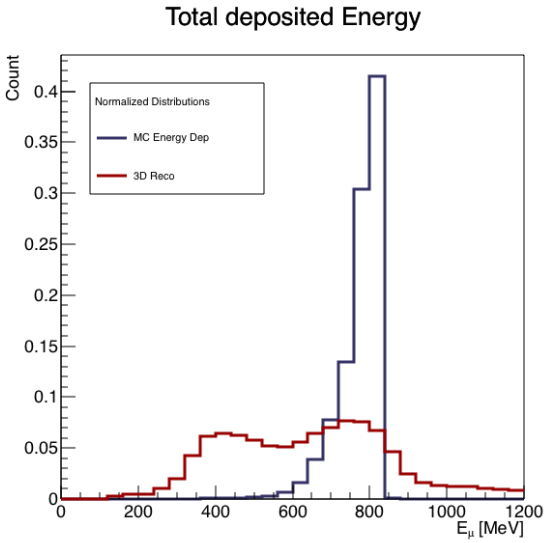

Figure 9: Candidate deposited energy at MCReco and Reco level. No cuts in Table 1 have been applied for Reco. The two plots are area normalized.

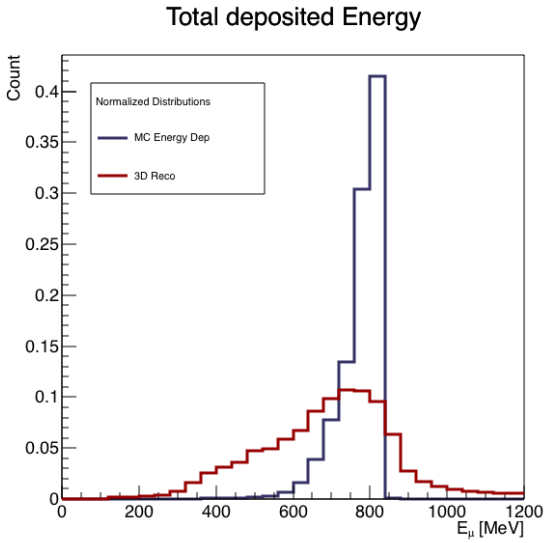

Figure 10: Candidate deposited energy at MCReco and Reco level. Only geometrical cuts in Table 1 have been applied for Reco. The two plots are area normalized.

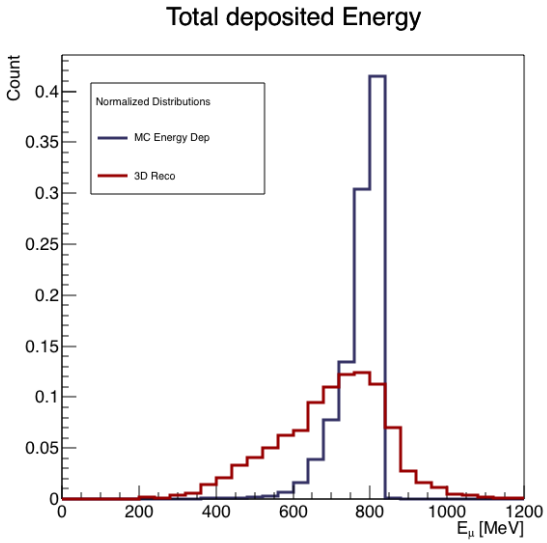

Figure 11: Candidate deposited energy at MCReco and Reco level. All cuts in Table 1 have been applied for Reco. The two plots are area normalized.

5 $p \to e^+\pi^0$

The following paragraphs address the analysis of the $e^+\pi^0$ proton decay mode.
The neutral pion decays immediately into two photons, both of which produce EM showers in the detector. The positron showers as well, leading to a signal topology of three electromagnetic showers emerging from a common origin.

5.1 Signal only analysis

In an attempt to isolate this signal, the following procedure was implemented (for a detailed explanation about running the code, see Appendix A):

1. Shower filter
2. Strict e^+ identification
3. Strict π^0 identification
4. Loose e^+ identification
5. Loose π^0 identification
6. Geometrical cuts
7. Calorimetric cuts

5.1.1 Shower filter

The first stage of event selection is the shower filter. A three shower filter is imposed that discards all events without three reconstructed showers. Though in future analyses, this will be changed to reject all events with fewer than three reconstructed showers, it was beneficial as first step to impose a strict three shower cut on signal sample. This filter allows for the identification efficiency of this event topology to be partially separated from shower reconstruction efficiencies. This analysis depends crucially on accurate shower reconstruction, which is still under development. By limiting the analysis to signal events that have exactly three reconstructed showers, there is a greater likelihood that the showers have been reconstructed reasonably well. Even so, shower reconstruction effects still dominate the factors contributing to the loss of signal events.

5.1.2 Particle identification

After the shower filter, a first attempt at particle identification is performed. The strategy for particle identification is to loop through reconstructed showers and to try to associate the shower with either a positron or photon from neutral pion decay. It is superior at this stage to over-identify pions and positrons pairs in the event: further cuts are applied later to eliminate misreconstructed pairs. If the identification is too tight, there is a risk of losing good events at this stage. Since the identification is performed sequentially, however, only one loose pass cannot be performed because the first identification algorithm applied will misidentify certain showers as its own. For concreteness, take the example of searching for a neutral pion first, using a loose definition. This algorithm may reconstruct pions using the shower actually coming from the positron, eliminating it from the viable pool of showers when the positron identification algorithm is applied next.

To circumvent this, particle identification is performed iteratively. First, positrons and neutral pions are identified according to a tight definition so that only showers that are very likely to be correctly associated with that particular particle are associated with it. Then, a looser iteration is performed on the remaining showers. This methodology helps to avoid the problem of "overzealous identification" for a particular particle.

Both the tight and loose definitions employed for particle identification are displayed in Tables 2 and 3. The e-like nature of a shower is found by comparing the log-likelihood of its dE/dx corresponding to a photon or electron as described in sec 4.1.5.

The definition of "likelihood" for the neutral pion is similar but takes into consideration five different aspects of the reconstruction. The likelihood is the sum of the log-likelihoods of the radiation length for both γ showers, the dE/dx for both showers, and the impact parameter for the two showers. The log-likelihoods are defined identically to the definitions supplied above but different PDFs are used for radiation length and impact parameter.

It should be noted that a particularly strong discriminating variable in π^0 identification is the angle between the two γ showers. For a given π^0 energy, there is a fixed minimum angle between the showers $\alpha_{\min} = 2\arcsin(\frac{m_\pi}{E_\pi}) = 2\arcsin(1/\gamma)$. This can be calculated by boosting a back-to-back decay from the rest frame of the pion into a frame moving perpendicularly to the photons. The distribution of $\theta_{\gamma\gamma}$ drops sharply

to zero below the minimum angle for this energy range, roughly 0.45 radians. This provides a good handle on identifying pions in the appropriate energy range.

Table 2: Cuts applied during e^+ identification for both loose and tight IDs.

Parameter	**Tight Cut**	**Loose Cut**
dE/dx	is e-like	is e-like
Shower energy	> 200 MeV	> 100 MeV

Table 3: Cuts applied during pion identification for both loose and tight IDs.

Parameter	**Tight Cut**	**Loose Cut**
"Likelihood"	> -10	> -10
IP of two showers	< 5 cm	< 10 cm
Pion mass	[100, 200] MeV	[75, 275] MeV
Shower energies	[20, 600] MeV	[15, 1000] MeV
Angle between showers	[0.5, 3.14] radians	[0.3, 3.14] radians
Pion energy	> 100 MeV	> 100 MeV

5.1.3 Geometrical cuts

After particle identification has been performed, two geometrical cuts are applied. The first is a cut on the radius of the smallest sphere that encloses the start points of the three showers. This serves as a proximity cut. This radius is a more robust variable to cut on than the distance between the reconstructed π^0 location and the start point of the positron shower. This is because during pion reconstruction, the position of the neutral pion is determined by finding the point of closest approach between the two half-lines extending from the photon showers in the opposite direction from the reconstructed shower direction. If the shower direction is even slightly misreconstructed, the change in angle can change the point of closest approach significantly. For this reason, using the reconstructed pion position as a parameter to cut on is inadvisable. In contrast, the positions of the initial energy deposits for the showers is a quantity measured directly by the detector and is therefore subject to less uncertainty due to reconstruction effects.

The second geometrical cut performed is a cut on the angle between the 3-momenta of the reconstructed π^0 and e^+. Naively, this would be expected to be steeply peaked near π because the proton decays at rest and momentum is conserved. This is not the case. Two factors contribute to this. The first is that the proton does not decay at rest, but instead decays with some momentum less than the Fermi momentum of the nucleus, which is on the order of 200 MeV/c. In addition to this effect, the angle between the decay products is changed because the pion undergoes intranuclear interactions before escaping the nucleus that can alter its momentum. For these reasons, the cut on angle is very loose.

5.1.4 Calorimetric cuts

After geometrical cuts are performed, calorimetric cuts on the decay products and reconstructed proton are performed. These consist of a simple energy cut on the reconstructed positron and pion as well as an energy cut on the reconstructed proton (found by summing the four-momenta of the electron and pion). The energy of the positron is taken to the be the total energy associated with the shower. The energy of the π^0 is found by reconstructing the momentum and invariant mass of the π^0 according to the following two equations:

$$|P_{\pi^0}| = \sqrt{E_1^2 + E_2^2 + 2E_1E_2\cos(\theta_{\gamma\gamma})} \tag{1}$$

and

$$M_{\gamma\gamma} = \sqrt{4E_1^\gamma E_2^\gamma \sin^2(\theta_{\gamma\gamma}/2)} \tag{2}$$

where E_1 and E_2 are the energies of the γ showers and $\theta_{\gamma\gamma}$ is the opening angle between the showers [4]. Note that substituting these formulas into $E^2 = p^2c^2 + m^2c^4$ returns $E_{\pi^0} = E_1 + E_2$, the sum of the energy depositions of the two photons.

Additionally, cuts are applied on the components of the reconstructed proton 3-momentum to seek out decays in which the proton has initial momentum not much larger than the Fermi momentum.

5.2 Results

The preliminary results presented below show the efficiency of this selection methodology at the MCReco and Reco level.

Table 14 shows that the cutflow keeps a good percentage of events when applied to MCReco level. This is not true at the reconstruction level. The cause of such a low efficiency seems to be due primarily to shower reconstruction effects. For example, two-thirds of events are lost because the number of reconstructed showers is not equal to 3. On top of this, π^0 reconstruction is poor at reco level and requires further work. The overall efficiency for this channel is $\sim 2\%$.

Table 4: Topology-based and calorimetry-based set of cuts to identify proton decay. All samples are signal only. The poor efficiency of the reconstructed events is due predominately to shower reconstruction effects.

Parameter	**Cut**	**MCReco**	**Reco**
Initial total	-	9990	9960
Shower filter	3 showers in event	9867	3222
IDed π^0s in event	> 1	9551	1753
IDed $e^{+/-}$s in event	> 1	9548	932
Radius of sphere enclosing shower starts	< 50 cm	9373	546
Angle between e^+ and π^0	> 2 radian	9337	396
e^+ energy	> 200 MeV	9334	381
π^0 energy	> 250 MeV	9324	347
Proton energy	[800, 1000] MeV	9164	223
$\lvert\Sigma p_x\rvert$, $\lvert\Sigma p_y\rvert$, $\lvert\Sigma p_z\rvert$	< 300 MeV	9039	190

Despite the low efficiency, the events that are selected as proton decay candidates at reco level do seem to be well-reconstructed events, indicating that the selection behaves well. Figures 12 and 13 demonstrate this. Figure 12 shows the positron energy at both the levels of analysis after all cuts have been applied. From MC energy deposition to reco, the distribution broadens slightly due to imperfections in clustering and shower reconstruction. Despite this, the peak appears to be in roughly the same position and the distributions are comparably shaped, indicating that the events selected, though few, were well-reconstructed. Therefore, for well-reconstructed events, our selection appears to work well to identify the three shower topologies associated with this channel of proton decay.

Figure 13 shows the same plot for the π^0 energy. Here, the distribution has been shifted significantly downward. This is due mainly to reconstruction inefficiencies. On a side note, it is interesting to notice that the reconstruction of the boosted π^0 is particularly difficult due to the high collimation of the decay products leading to overlapping showers. The primary error made during the reconstruction of these events is to reconstruct the π^0 as a single, highly-energetic shower. In fact, 4360 signal events out of 9960 total have two reconstructed showers or fewer. By hand-scanning a small portion of them, we noticed that the overlapping showers tend to be reconstructed as some set of showers and tracks. In three shower events specifically, the energy of the π^0 is lower because some of the particle's energy has been reconstructed as tracks. A strategy to improve these reconstruction efficiencies may be to extend the π^0 algorithm by searching for reconstructed showers with $dE/dx \approx 8.5$ MeV/c^2 that correspond to the combination of two γ-showers. This methodology would also allow for the identification of Dalitz decay, a decay mode of π^0 to which this analysis is currently not sensitive.

Figure 14 shows the energy of the reconstructed proton (sum of decay product 4-momenta) at each level. Though there is significant variation in the distribution, it appears to peak slightly below the truth energy, an effect that is the result of the misreconstruction of the pion energy discussed previously.

6 Conclusions

In this note, we calculated the reconstruction efficiencies for the $p \to \mu^+\gamma$ ($\epsilon \sim 56\%$) and $p \to e^+\pi^0$ ($\epsilon \sim 2\%$) mode in signal only samples. Despite the non stellar efficiency, we demonstrated that the events selected were well-reconstructed for both the modes. These results can only improve as reconstruction in MicroBooNE improves. The next step is applying our cut flow for both these modes to cosmics only events.

References

[1] C. Adams et al. The Long-Baseline Neutrino Experiment: Exploring Fundamental Symmetries of the Universe. 2013.

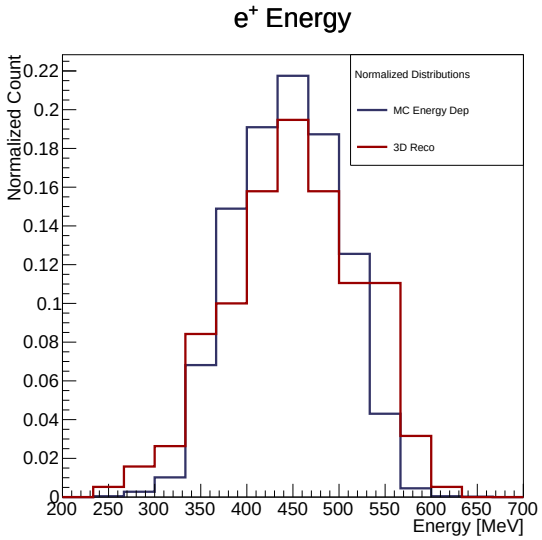

Figure 12: Energy distributions of the positron at MCReco and Reco level. All cuts in Table have been applied. The two plots are area normalized.

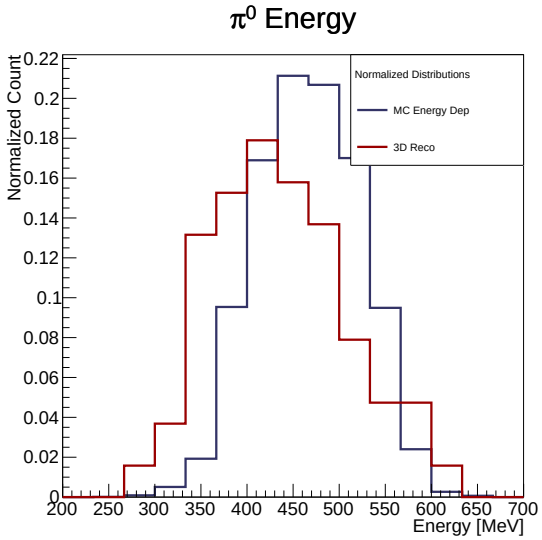

Figure 13: Energy distributions of the pion at MCReco and Reco level. All cuts in Table have been applied. The two plots are area normalized.

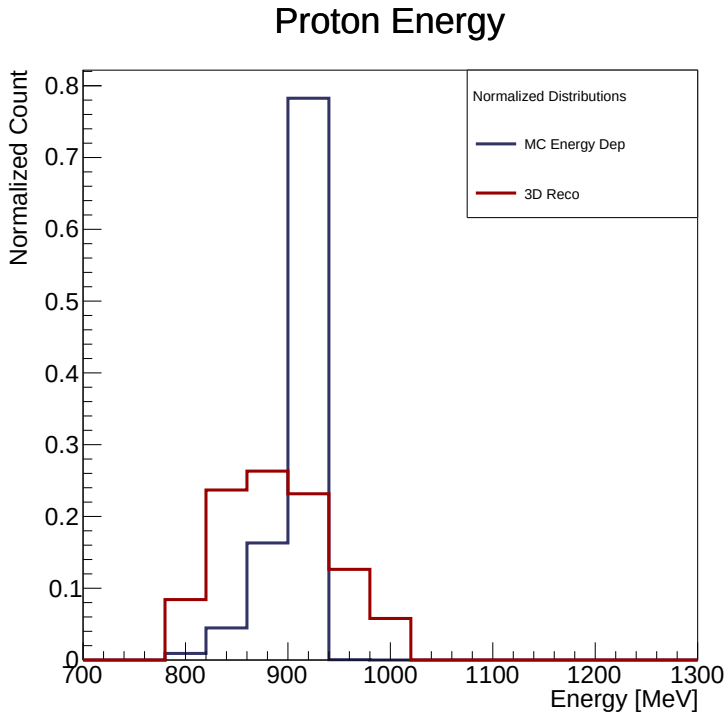

Figure 14: Energy distributions of the reconstructed proton at MCReco and Reco level. All cuts in Table have been applied. The two plots are area normalized.

[2] T. Akiri et al. The 2010 Interim Report of the Long-Baseline Neutrino Experiment Collaboration Physics Working Groups. 2011.

[3] A. Bueno, Z. Dai, Y. Ge, M. Laffranchi, A. J. Melgarejo, A. Meregaglia, Sergio Navas, and A. Rubbia. Nucleon decay searches with large liquid argon TPC detectors at shallow depths: Atmospheric neutrinos and cosmogenic backgrounds. *JHEP*, 04:041, 2007.

[4] R. Acciarri et al. Measurements of Inclusive Muon Neutrino and Antineutrino Charged Current Differential Cross Sections on Argon in the NuMI Antineutrino Beam. *Phys. Rev.*, D89(11):112003, 2014.

A Electron and gamma PDFs

Kazu will write it.

B Manual

This section will explain the machinery used to perform this analysis. The modules for topology identification were all written within ERTool, a lightweight analysis framework for Larlite data products. This assumes that you already have LArLite installed and compiled.

All of the modules used for this are stored in the "Ndk" git repository. To clone this, run

```
git clone git@github.com:ElenaGramellini/Ndk.git
```

For $p \rightarrow \mu + \gamma$,
`cd Mode13`
and
`make`.

Within the directory, multiple modules where used to tune the analysis. To obtain the plots shown, you'll need to run
`ERAnaOneToOneMu.cxx` and `ERAnaMode13.cxx`.
Unless you need to fix a bug, there should be no need to interact directly with these modules once they are compiled. Everything is done instead through the launcher macros, located under `mac/`.

To run the macro that outputs the muon ntuples for MCReco, use
`python mac/Ana121MuMC.py [`*filename1*`] .`
To run the macro that outputs the muon ntuples for Reco, use
`python mac/Ana121MuReco.py [`*filename1*`] [`*filename2*`] .`

To run the macro that outputs the gamma and proton ntuples for MCReco, use
`python mac/WholeMode13.py [`*filename1*`] .`
To run the macro that outputs the gamma and proton ntuples for Reco, use
`python mac/RecoWholeMode13.py [`*filename2*`] [`*filename3*`] .`

This will run the analysis over *filename* (a LArLite file). Running this will output a ROOT files that contain TTree containing the information about all the identified gammas, muons and candidates in the sample provided.

Once this step has been completed, the actual cutflow can be applied to the reconstructed gammas and muons stored in `outfile`. To do this, we use another macro, `mac/SignalPlotsForTechNote.C`. In order to get table 1 run `mac/Table.C`

The files used for this study (at the time of writing) are as follows:

`[`*filename1*`]` : `/pnfs/uboone/scratch/users/elenag/v04_12_00/Mode13/sim/prod_NdK/2598033_*/larlite_mcinfo_*.root`
`[`*filename2*`]` : `/uboone/data/uboonerd/nucleonDecay/LarLiteProduction/Mode13/Mode13/liteMaker2/prod_NdK/3289417_*/larlite_reco3d_kalmanhit.root`
`[`*filename3*`]` : `/uboone/data/uboonerd/nucleonDecay/LarLiteProduction/Mode13/Mode13/liteMaker2/prod_NdK/3289417_*/larlite_reco3d_pandora.root`

For $p \to e^+\pi^0$,

`cd Mode0`

and

`make`.

Within the directory, there are multiple modules that are run during the analysis, primarily `ShowerFilter_NdkModeZero.cxx`, `ERAlgoSingleE_NdkModeZero.cxx`, `ERAlgoPi0_NdkModeZero.cxx`, and `ERAnaNdkModeZero.cxx`.

Unless you need to fix a bug, there should be no need to interact directly with these modules once they are compiled. Everything is done instead through the launcher macro `modezeroselection.py`, located under `mac/`.

To run this, use

`python mac/modezeroselection.py [`*`filename`*`] > log.log &`.

This will run the analysis over *filename* (a LArLite file) in the background and redirect `stdout` to `log.log` so you can read the status using `tail -f log.log`. Running this will output a ROOT file that contains several plots and a TTree containing the information about all the identified pions and electrons in the sample provided.

There are various parts of `modezeroselection.py` that must be changed each time a new sample is analyzed. First, there is the output filename (line 21). Change the name after `outfile =` to be whatever output file you want. If you plan on running over an MCTruth file, you must set `mc` to `True`, otherwise, `False` (line 18). Lines 29 and 30 let you set the minimum and maximum number of showers accepted in a given event.

Lines 36-72 allow you to modify the cuts placed when identifying π^0s and e^+s. Most are self-explanatory. For the electrons, there is only one that matters (`setEThreshold`, which sets the energy value that any candidate electron must be above). The others are deprecated. For the pion, you can specify minimum and maximum shower energy, maximum impact parameter between the two showers, minimum and maximum reconstructed mass, and minimum and maximum angle between the two showers.

Once this step has been completed, the actual cutflow can be applied to the reconstructed pions and electrons stored in `outfile`. To do this, we use another macro, `mac/ModeZero_cutflow_v3.py`.

To run, do

`python mac/ModeZero_cutflow_v3.py [`*`outfile`*`]`

where *outfile* refers to the ROOT file output at the end of the previous step. This will return a new ROOT file with histograms of relevant quantities. The actual yield will be output to `stdout` when the process finishes.

There are also many tunable parameters at this stage. They can be found in lines 11-35. The first thing to note is the ability to specify the output filename (edit `cutflowoutputfile =`). Then, there is the `doTruth` option, which must be `False` for any reco files and can be either `True` or `False` for MCTruth files (this will access the truth branches of the TTree and so will throw an error if there are no truth branches). Then there are the parameters, each of which has an associated comment describing the cut.

The files used for this study (at the time of writing) are as follows:

1. Truth signal: `/pnfs/uboone/scratch/users/elenag/v04_12_00/Mode0/sim/prod_NdK/2597833_*/larlite_mcinfo_*.root`
2. Reco signal: `/pnfs/uboone/scratch/users/elenag/v04_12_00/Mode0/reco2/prod_NdK/*_*/larlite_reco3d.root`