MICROBOONE PUBLIC NOTE

STATUS OF THE MICROBOONE INCLUSIVE SINGLE PHOTON SELECTION USING WIRE-CELL RECONSTRUCTION

May 27, 2022

MicroBooNE Collaboration *

MICROBOONE-NOTE-1102-PUB

Abstract

MicroBooNE[1] is a Liquid Argon Time Project Chamber (LArTPC) detector whose primary design goal is to understand the "low-energy-excess" anomaly seen by MiniBooNE[2]. MicroBooNE's currently published results[3] [4] see no excess consistent with the MiniBooNE observation, therefore creating a need for searches in more channels. This note summarizes MicroBooNE's inclusive single photon selection using Wire-Cell reconstruction and pattern recognition, which will be used to search for a low-energy-excess (LEE) anomaly in the inclusive single photon channel. The selection is similar to the Wire-Cell inclusive electron neutrino selection[5], but with a different signal definition and some modifications and additions to the pattern recognition tools. A selection with 15.2% efficiency and 18.5% purity is achieved for simulated Standard Model single photon events.

*MICROBOONE_INFO@fnal.gov

Contents

1 Introduction 3

2 Signal Definition 3

2.1 Signal Categories . . . 3

3 Event Selection 5

3.1 Preselection . . . 6

3.2 BDTs . . . 8

3.2.1 ν_μ CC Background BDT . . . 8

3.2.2 Other (NC, Dirt, Not in FV, and Cosmic) Background BDT . . . 9

3.2.3 NCπ^0 Background BDT . . . 10

3.2.4 ν_e CC Background BDT . . . 11

3.3 Efficiency and Purity . . . 12

4 BNB Open Data 14

4.1 BDT Score Distributions . . . 14

4.2 Energy . . . 16

4.3 Shower Angle . . . 16

4.4 Particle Multiplicity . . . 17

4.5 Shower Energy Deposition (dE/dx) . . . 17

5 Summary 19

1 INTRODUCTION

This analysis aims to probe the MiniBooNE low energy excess anomaly with an inclusive single photon channel hypothesis. This note documents the selection for this analysis, at the time of writing, which uses Boosted Decision Trees (BDTs) trained on Wire-Cell reconstruction and pattern recognition variables. The signal definition for the currently published MicroBooNE single photon analysis[4] is exclusively NC Delta radiative events with either exactly 1 photon and 1 proton or 1 photon and no other particles in the final state. This leaves a large swath of the single photon (or single-photon-like e+e-) event phase-space unexplored, particularly for events with more than one track, or non-proton-like tracks. With no currently obvious MiniBooNE-like anomaly seen in the currently published electron-like LEE[3] or single photon-like LEE[4] analyses, a more broad analysis of the single photon channel is needed. Therefore, the goal of this selection is to select a more broad and inclusive set of single-photon-like events. These events can then be further analyzed and fit to evaluate if any significant excess or anomaly is seen. Additionally, this analysis can provide a high-statistics sample of irreducible single photon events from Standard Model processes, which are important to further study in order to validate both Standard Model predictions and photon shower reconstruction. This note outlines the details of the developed selection, further analysis is still underway and will be covered separately in a future publication.

2 SIGNAL DEFINITION

The signal for this analysis was chosen to be as inclusive as possible within the limits of current MicroBooNE reconstruction, while also targeting events that could reasonably have contributed to the MiniBooNE low energy excess[6]. Toward this goal, a truth study was performed to determine what currently modeled (i.e. Standard Model-based) processes should be used as signal events for this analysis. The results of this study led to the signal definition and categorization outlined below.

2.1 Signal Categories

To ensure consistency with the MiniBooNE selection, as well as ensuring the photon can be seen and reconstructed by MicroBooNE, a "single photon event" is defined as an event with no true primary electrons and with exactly one true photon with true energy above 20 MeV, that is not the secondary product of an electromagnetic shower, and is both produced and begins to shower within the space-charge boundary (i.e. whose true production point and true shower start is within the space-charge boundary). The "space-charge boundary" (SCB) represents the effective active TPC volume. We note that events with a true Dalitz decay are excluded as signal, as this would produce three showers from the same decay, which would not pass the single shower requirement in MiniBooNE's selection. For analysis purposes the signal events have been placed into four categories:

- NC π^0 1γ
  - NC events with a π^0 that has decayed into two photons, where one photon does not fit the requirements above, i.e. it does not shower inside the TPC or is too low energy to reconstruct.
  - Before any selection, 65 events are expected in the 3.4e19 POT of data shown here. This category makes up 73% of the expected SM signal events before selection.
- NC Δ 1γ
  - An NC $\Delta \rightarrow N\gamma$ event where the photon fits the requirements above.
  - Before any selection, 8 events are expected in the 3.4e19 POT of data shown here. This category makes up 10% of the expected SM signal events before selection.
- NC Other 1γ
  - Other NC events with a photon fitting the requirements above, but whose true parent is something other than π^0 or Δ, such as ones from higher resonant state particles like η or ρ_0.
  - Before any selection, 2 events are expected in the 3.4e19 POT of data shown here. This category makes up 2% of the expected SM signal events before selection.
- ν_μ CC 1γ, $\mu < 100 MeV$
  - ν_μ CC events with exactly one photon fitting the requirements above and where the true muon has true kinetic energy less than 100 MeV.
    * 100 MeV was chosen based on MiniBooNE's muon reconstruction threshold. The Cerenkov threshold for a muon in mineral oil is about 40 MeV. However, considering hardware and reconstruction inefficiencies and limitations, it was determined that MiniBooNE was not likely to reliably reconstruct a muon below about 100 MeV in their LEE anomaly result [7].

– Before any selection, 14 events are expected in the 3.4e19 POT of data shown here. This category makes up 15% of the expected SM signal events before selection.

Note that the“within the active volume" requirement mentioned above means that “out of fiducial volume" and “dirt" events, where the neutrino interaction happens inside the detector but outside the active volume or outside the detector entirely, respectively, are not considered signal. In the case where the interaction point is outside the active volume but a single photon travels inside the volume and is detected, the event is not considered signal but does exactly resemble a signal event and so is currently an irreducible background.

3 EVENT SELECTION

The selection is split into two main sections: preselection and BDTs.

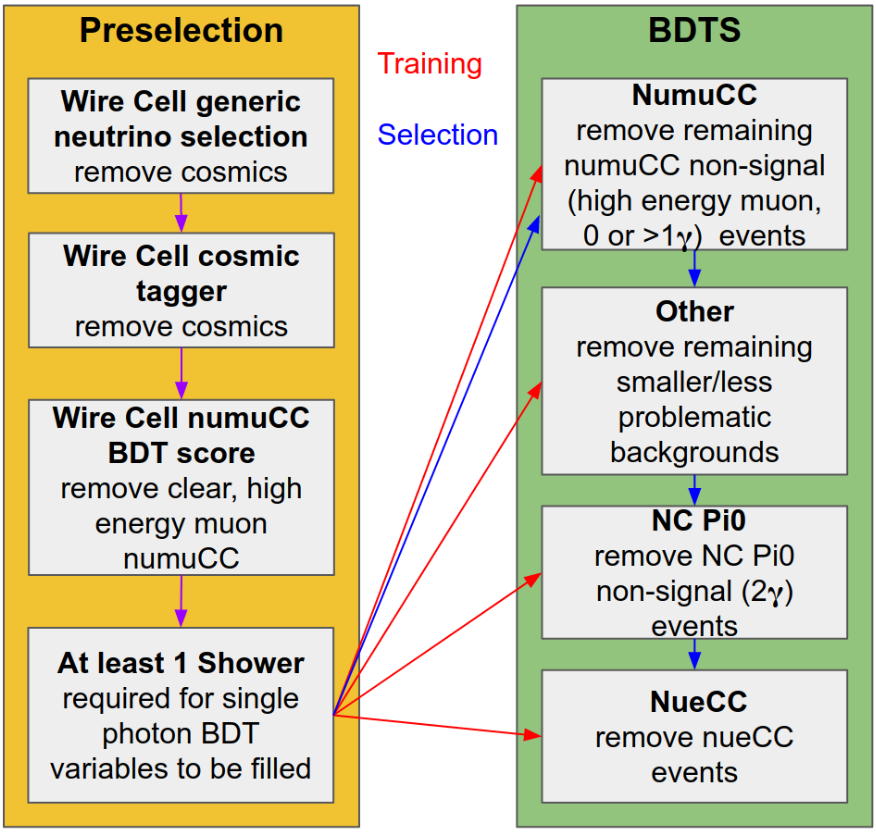


Figure 1: A diagram showing the structure and order of the preselection and BDT training and cuts. The arrows indicate the order of training (red) and selection (blue).

3.1 Preselection

Before training or selecting with the BDTs, events go through a preselection to remove large unsignal-like backgrounds and events with all or almost all BDT variables not filled.

A plot of the reconstructed neutrino energy, with a breakdown of event categories for this analysis, immediately after generic neutrino selection[8] is shown at the top of Figure 2. Plots of the other three Wire-Cell particle flow [9] variables used for preselection are also shown in Figure 2. These variables are all shown immediately after generic neutrino selection, since they are not computed or saved for events that do not pass generic neutrino selection. The efficiency and purity of the preselection cuts can be found in Table 1.

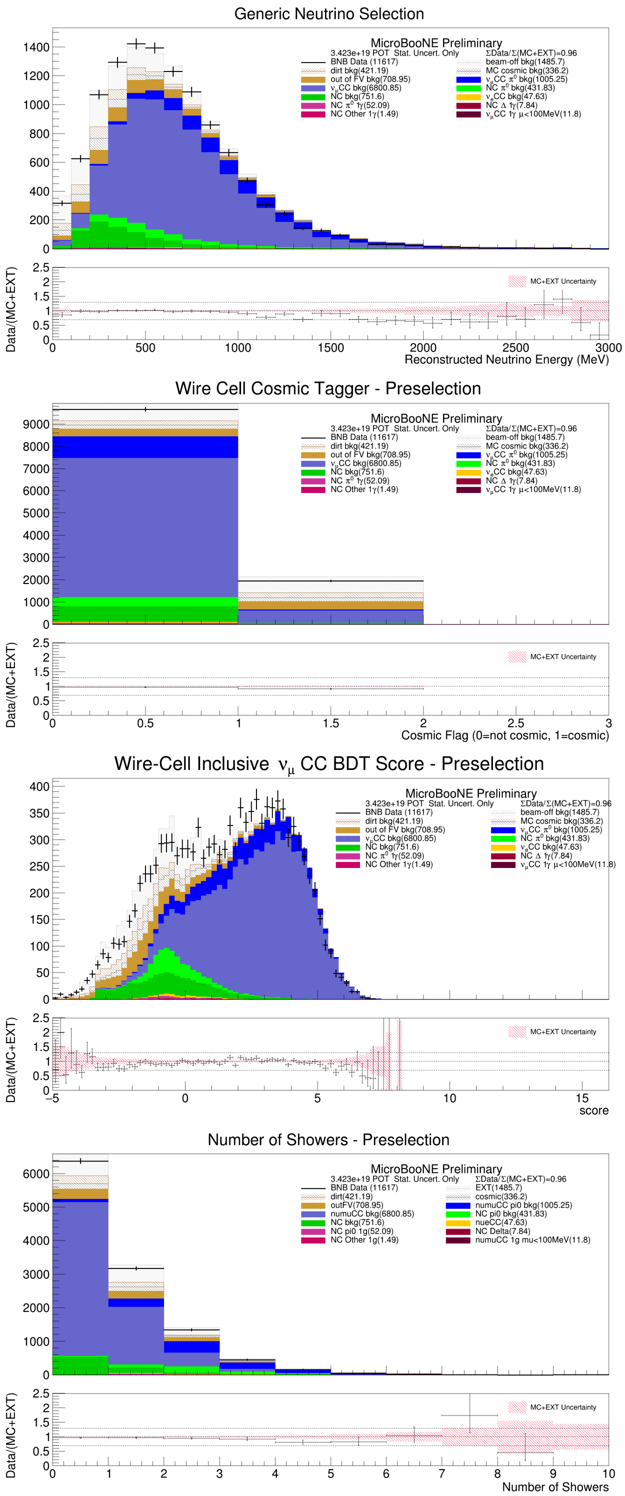


Figure 2: The reconstructed neutrino energy, cosmic tagger flag, Wire-Cell ν_μ CC BDT score, and number of showers for events immediately after generic neutrino selection. “beam-off" events are cosmic ray events from off-beam data and “MC cosmic” events are simulated neutrino events overlaid on cosmic data where a cosmic induced flash/cluster was selected instead of the neutrino induced one. The data is run 1 open data only, which corresponds to 3.423e19 POT. The middle dashed line on the ratio plots is at $\frac{Data}{MC+EXT} = 1$, which indicates exact central value agreement between data and prediction.

3.2 BDTs

All BDTs were trained using XGBoost [10] on 26,048 and tested on 26,477 signal events (events that have exactly one contained photon above 20 MeV and fit criteria from Section 2.1) from samples that were filtered through the preselection described in Section 3.1. Additionally, a total of 123,650 background events for training and 88,106 background test events, split up by type and used only on the relevant BDT were used also after the preselection. Figure 1 shows a diagram of the preselection, training, and selection workflow.

3.2.1 ν_μ CC Background BDT

The first BDT in the chain is made to remove the remaining ν_μCC background."ν_μCC background" is any ν_μCC event that does not fit the requirements for the ν_μCC 1γ, $\mu < 100 MeV$ signal category ("ν_μCC signal"). It uses 73 variables. The preselection removes the majority of well-reconstructed ν_μCC background events, meaning the remaining background is largely from events where the muon has not been reconstructed well. Handscans have shown that this is usually because it is short, low energy, or exiting, which would be reflected in the Wire-Cell ν_μCC inclusive BDT score and length variables, or it has been mis-reconstructed as a proton, pion, or shower, in which case proton track dQ/dx, shower dQ/dx, number of vertices, and various "bad reconstruction" variables are useful. Figure 3 shows the resulting score distribution for the test sample.

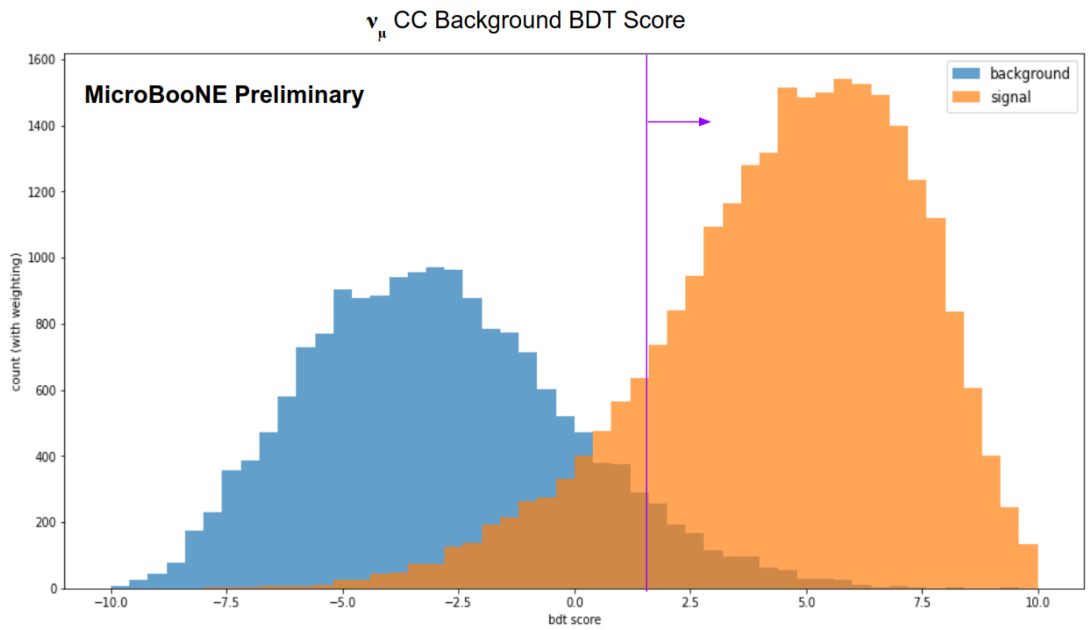


Figure 3: ν_μCC background removal BDT scores. Background includes only ν_μCC background events. The purple line indicate the cut value for the selection.

3.2.2 Other (NC, Dirt, Not in FV, and Cosmic) Background BDT

The second BDT in the chain is made to remove four different, but similar, backgrounds: neutral current events with no π^0 in the final state (NC background), dirt events, which are events where the neutrino interaction occurs outside the TPC, not in fudicial volume ("out of FV") events, which are events where the neutrino interaction position is outside Wire-Cell's 3cm fiducial volume but still in the TPC, and cosmic ray events, which includes both off-beam data ("beam-off") and simulated beam neutrino events that are overlaid on cosmic data events where a cosmic-induced flash/cluster was selected by Wire-Cell in the early reconstruction stage instead of the neutrino induced one ("MC cosmic").

The BDT uses 86 variables. Between the Wire-Cell generic neutrino selection, cosmic tagger, and at least one shower cuts, the preselection removes the majority of these events, meaning the remaining background is largely from events with a cosmic shower, low energy activity, proton stubs that may be reconstructed as a low-energy shower, or partially contained tracks and showers that are likely to not be well-reconstructed. For this reason, a wide range of both track and shower variables, mostly to do with length or identifying a bad reconstruction, are used in this BDT. Figure 4 shows the resulting score distribution for the test sample.

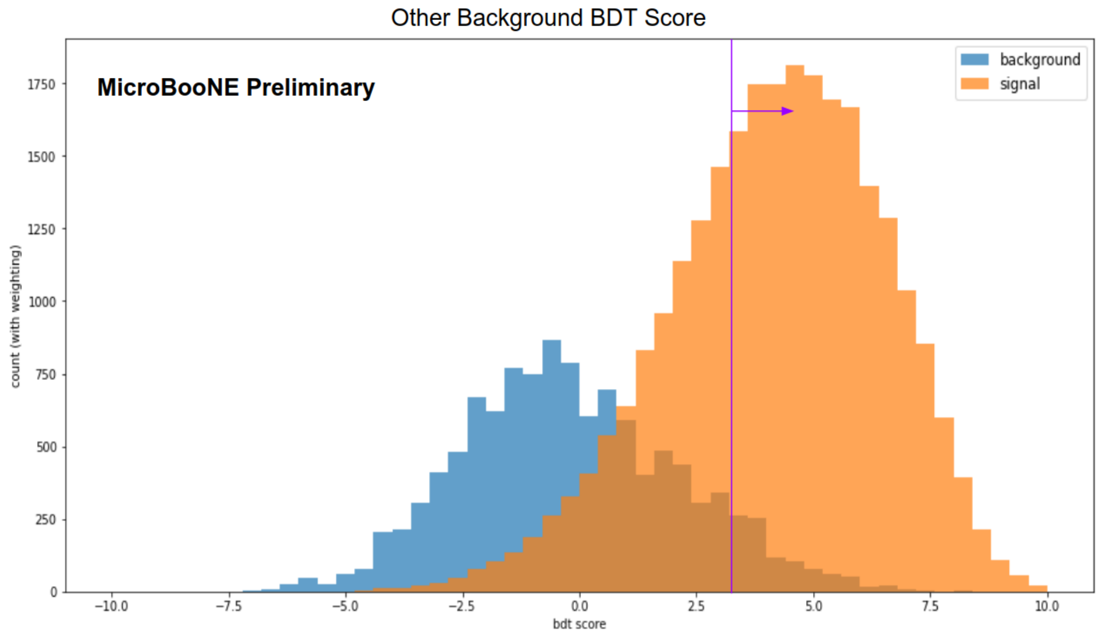


Figure 4: Other background removal BDT scores. Background includes only NC, out of FV, dirt, MC cosmic, and beam-off backgrounds. The purple line indicate the cut value for the selection.

3.2.3 NCπ^0 Background BDT

The third BDT in the chain is made to remove NC π^0 background. It uses 11 variables. This background is very similar to the signal events, especially the NC π^0 1γ category of signal, which is different from this background only by the number of theoretically reconstruct-able showers. As a result, very few variables show any separation, and any reconstruction failure or difficult topology (e.g. showers overlapping) that leads to a different number of showers being reconstructed than are truly there becomes a large issue. Currently, only a few variables have been found to be useful:

- Shower indirect max energy: the maximum energy of the shower subclusters, which are the groups of continuous wire hits that make up the shower. Helpful to identify if two showers have been combined into one.
- Number of showers: numbers of showers passing various criteria (i.e. how many have reco energy greater than 20 MeV, how many have passed one or more of the “bad reconstruction" functions, etc.)
- Number of segments: number of segments or subclusters in the shower.

- Pi0 flag: flag to determine if two shower-like objects that can be used to reconstruct π^0 variables (mass, decay vertex, etc.) exist.
- High energy overlap minimum length and angles: minimum length and angles with the primary shower of either tracks or other showers in the event.

Figure 5 shows the resulting score distribution for the test sample.

In Figure 5, a large amount of build-up in one bin around score=2.5 can be seen for the signal events. This is because the BDT is using very few variables and of these variables: 1 is a flag (only has 0 or 1); 5 are numbers of showers passing some various criteria, where the signal events usually pile up at one or two while the NC π^0 can have many more if both showers were misreconstructed in some way; and the length and angle variables have certain values where events that have some escaping part (like some of the π^0 signal events) are likely to pile up. This means many signal events end up with very similar or the exact same distribution in most or all of the BDT variables and many events end up in this bin.

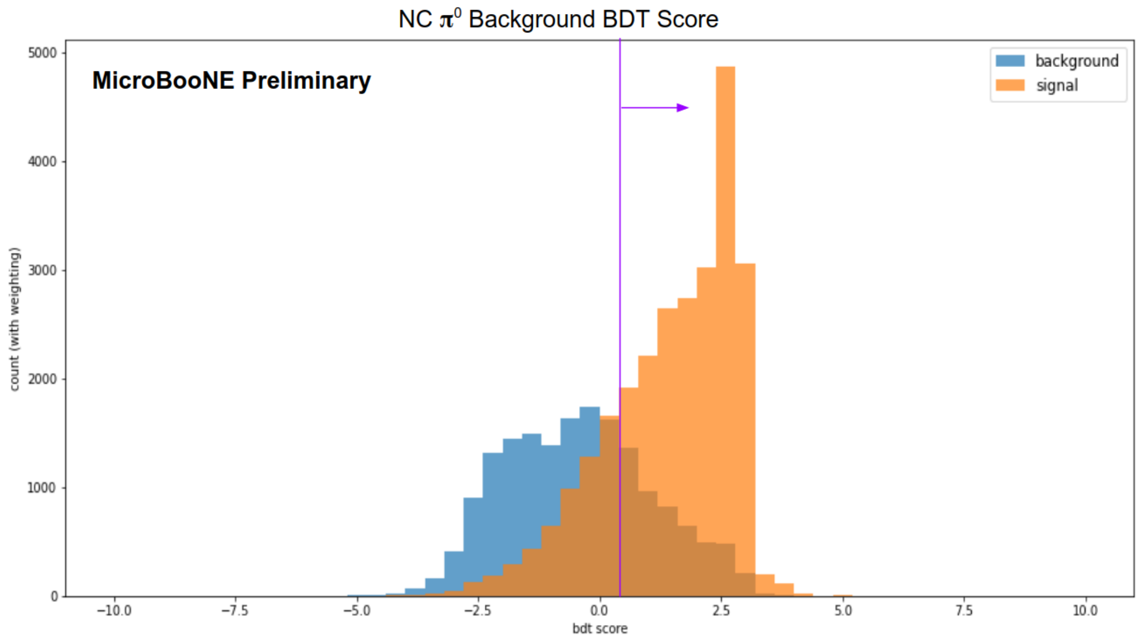


Figure 5: NC π^0 background removal BDT scores. Background includes only NC π^0 events from BNB overlay and NC π^0 overlay. The purple line indicate the cut value for the selection.

3.2.4 ν_e CC Background BDT

The final BDT in the chain is made to remove ν_eCC background. It uses 56 variables. ν_eCC events make up only ~2% of all events after preselection, and therefore their removal does not impact the overall selection purity much. However, to distinguish this analysis from the electron LEE analyses and determine if the MiniBooNE LEE was caused by photon events, it is important to remove as much of the ν_eCC events from this selection as possible. Two main kinds of variables are used to train this BDT: variables used to determine if the shower stem is a minimum ionizing particle (MIP) (dQ/dx chunks, length of the shower stem and full shower, etc.), and the shower conversion distance, which also uses some track (mostly proton) information when available to determine information about the neutrino vertex. Figure 6 shows the resulting score distribution for the test sample.

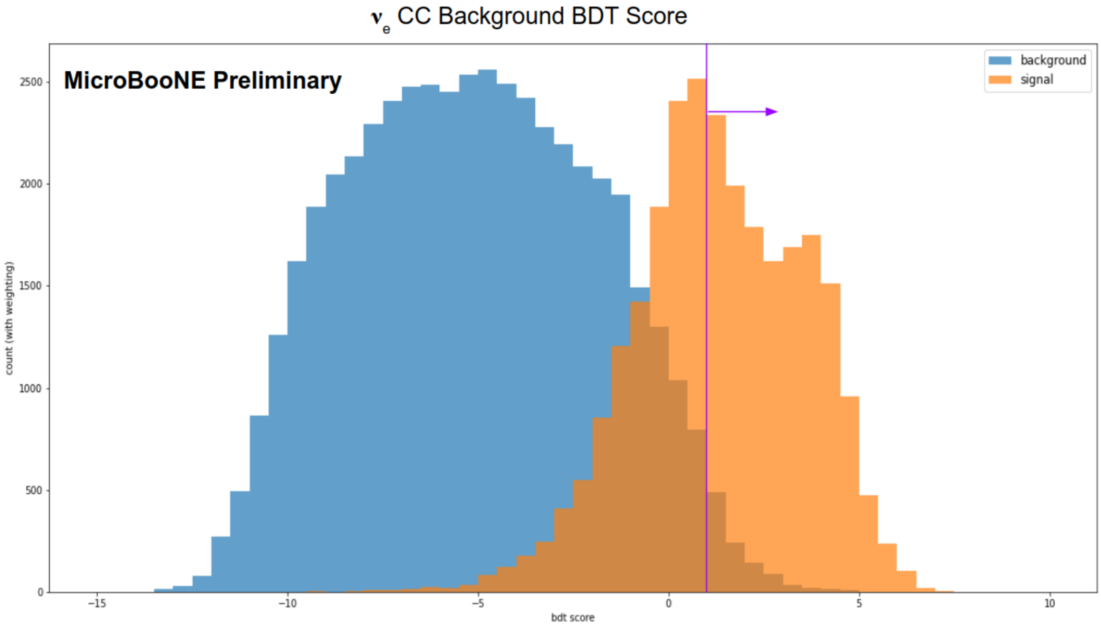


Figure 6: ν_eCC background removal BDT scores. Background includes only ν_eCC events from BNB overlay and ν_eCC overlay. The purple line indicate the cut value for the selection.

3.3 Efficiency and Purity

	Absolute Efficiency	Purity
Generic Neutrino Selection	81.9%	0.6%
Cosmic Tagger	77.7%	0.7%
Wire-Cell ν_μCC Inclusive BDT	70.0%	1.4%
≥ 1 shower	62.1%	1.9%

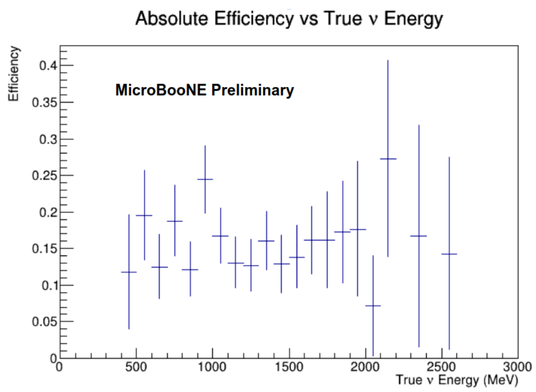
Table 1: Efficiency and purity summary for preselection. The efficiency shown is the number of expected signal events after the cut over the total number of expected signal events with no cuts.

	Efficiency Relative to Preselection	Absolute Efficiency	Purity
Preselection	100%	62.1%	1.9%
ν_μ CC Background BDT	66.6%	41.4%	5.6%
Other Background BDT	44.7%	27.8%	8.0%
NC π^0 Background BDT	32.2%	20.0%	13.3%
ν_e CC Background BDT	24.4%	15.2%	18.5%

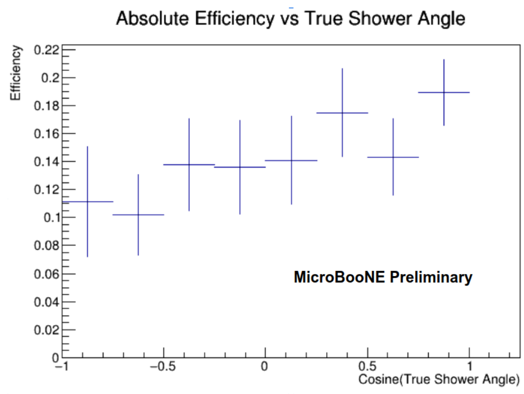
Table 2: Efficiency and purity summary for BDTs. The absolute efficiency shown is the number of expected signal events after the cut over the total number of expected signal events with no cuts. “Efficiency Relative to Preselection" is the same but with the total number of expected signal events after preselection in the denominator.

Of the selected π^0 signal events, 79% are signal because they have one photon outside the TPC and the remaining 21% are signal because they have one photon below reconstruction energy threshold (20 MeV).

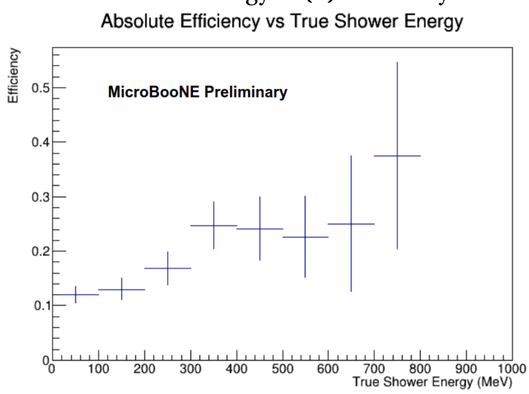
In Figures 7a,7b, and 7c, it can be seen that the efficiency is relatively flat (within error) across a significant range of true neutrino energies, shower angles, and shower energies. This shows no obvious biases in the shower or neutrino kinematics are introduced by the selection.



(a) Efficiency as a function of true neutrino energy.



(b) Efficiency as a function of true shower angle.



(c) Efficiency as a function of true shower energy.

Figure 7: Efficiencies as a function of true kinematic variables. Absolute efficiency means the denominator is all signal events before any cuts or selection.

4 BNB OPEN DATA

Below are a number of data/MC comparison plots, with data and prediction statistical uncertainties. No systematic uncertainties are included currently. The data is run 1 open data only, which corresponds to 3.423e19 POT after some reductions from the beam quality filter and failures in the reconstruction chain. In all figures, the four signal categories are shown in the pink shades.

4.1 BDT Score Distributions

Figure 8 show the BDT score distributions for events immediately before the relevant BDT cut.

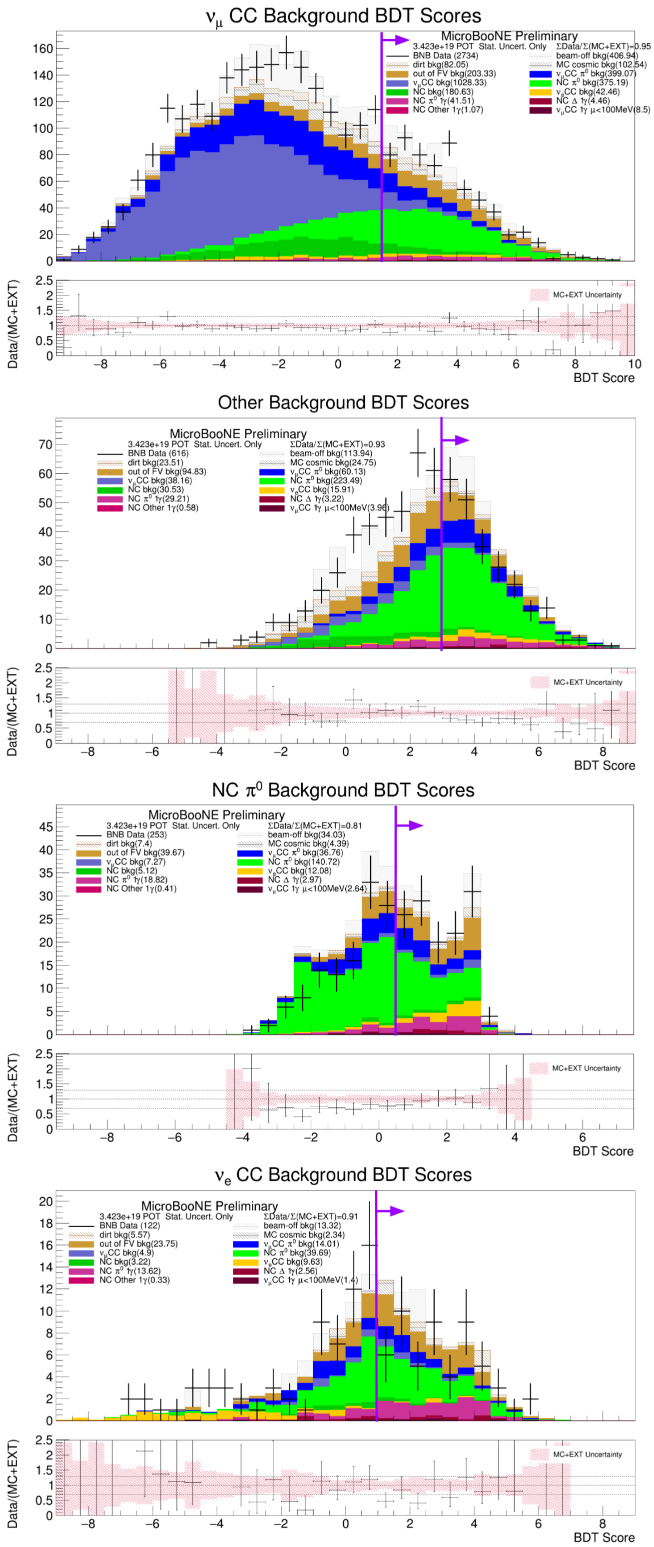


Figure 8: BDT score for each of the 4 BDTs, after preselection and relevant earlier BDT cuts. The purple line indicates the cut value.

4.2 Energy

Figure 9 shows the reconstructed energy of the primary shower for selected events.

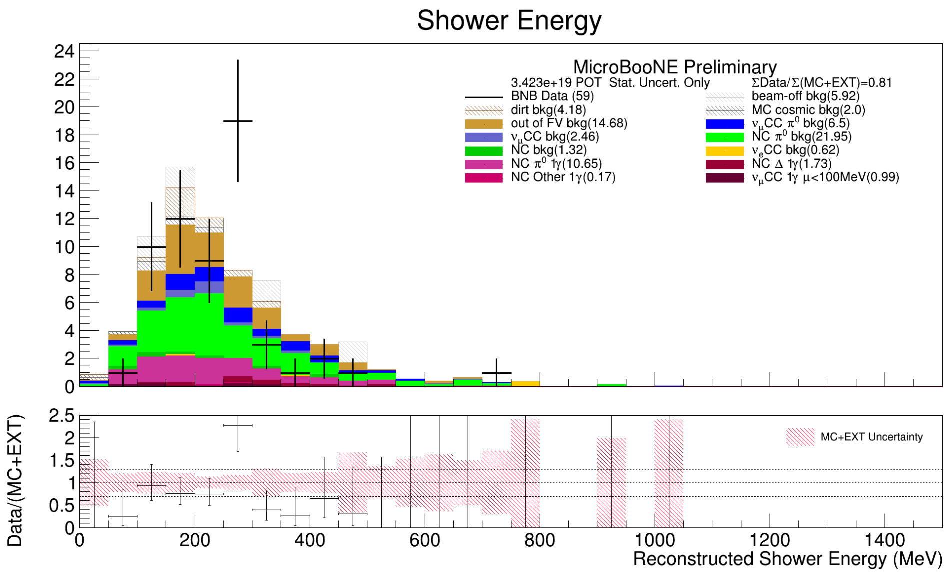


Figure 9: Reconstructed shower energy for selected events.

4.3 Shower Angle

Figure 10 shows the reconstructed angle of the primary shower, with respect to the beam direction, for selected events.

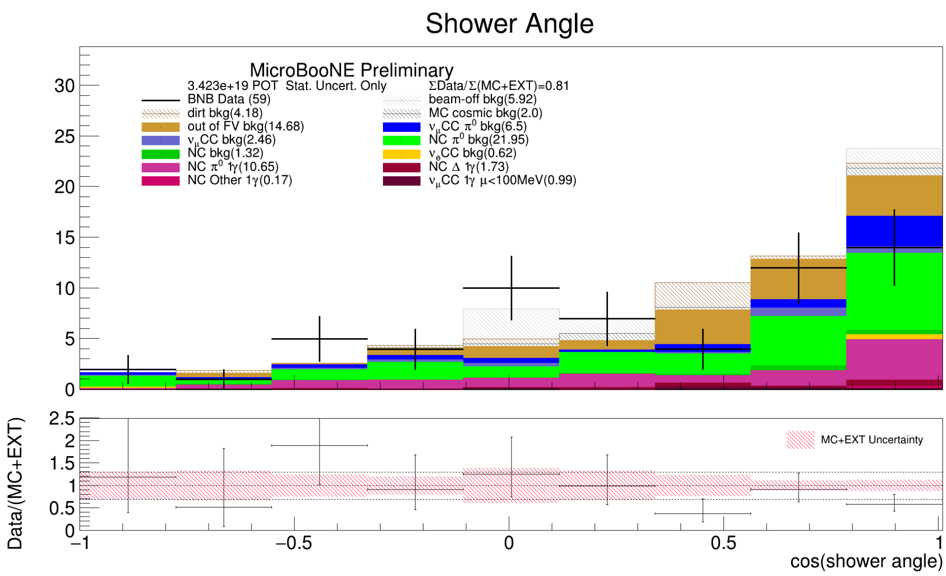


Figure 10: Cosine of reconstructed shower angle with respect to the beam for selected events.

4.4 Particle Multiplicity

Figure 11 shows the number of reconstructed tracks, and their breakdown by particle type as determined by the Wire-Cell particle ID tools [11] for selected events. Selected signal events include events with various numbers of hadronic particles in the final state, which shows that the selection does result in an inclusive sample.

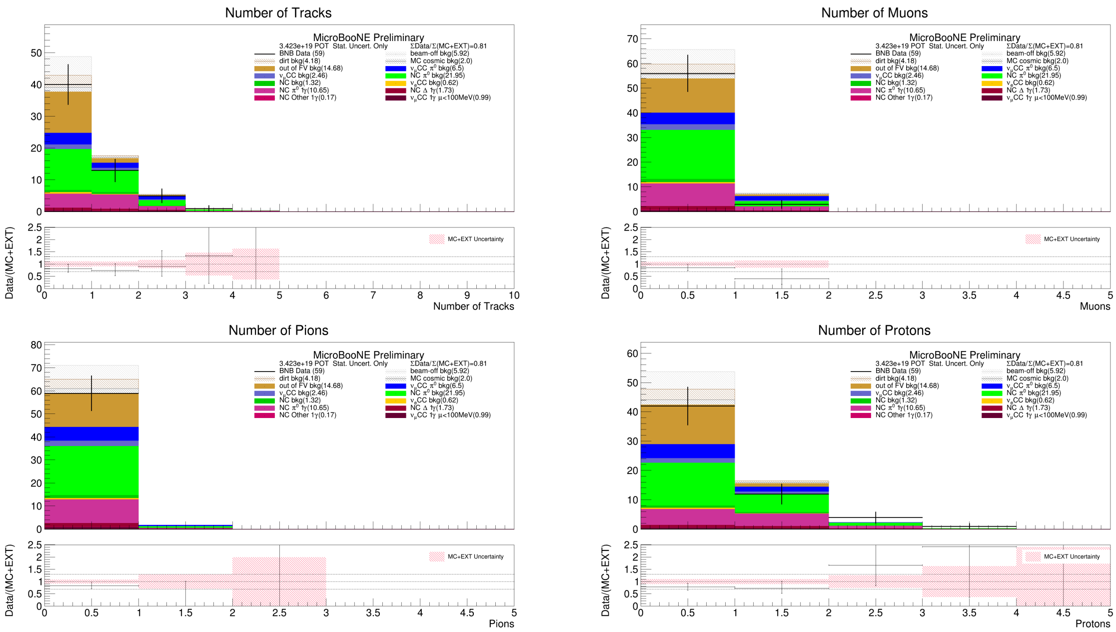
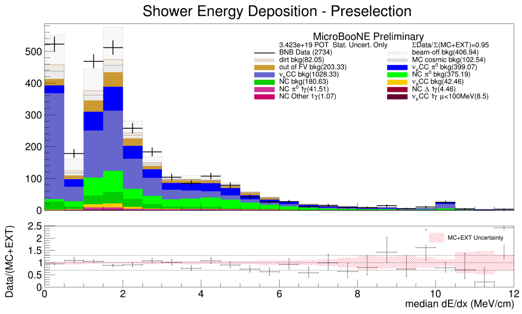


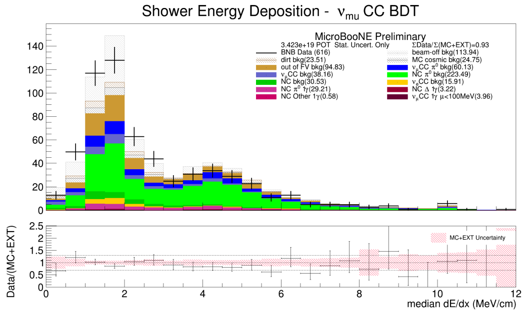
Figure 11: Number of reconstructed tracks and the breakdown of muon, charged pion, and proton tracks, determined by Wire-Cell's particle flow and identification algorithms for selected events.

4.5 Shower Energy Deposition (dE/dx)

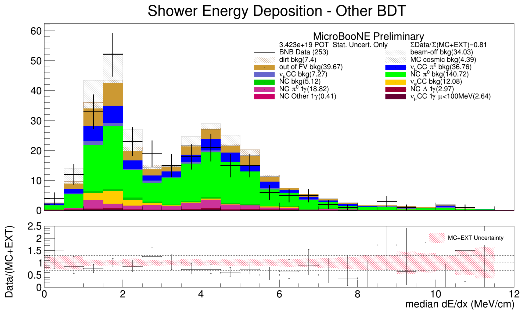
Figure 12 shows the median $\frac{dE}{dx}$ for the first 4cm of the primary shower for events at all stages of the selection. Events with an electron-like shower pile up around 1-2 MeV/cm, which is typical for MIP particles like an electron. Events with a photon-like shower, such as π_0 events, will usually have a $\frac{dE}{dx}$ twice that of an electron at its start, since the shower start should contain the e^+e^- pair from the photon. However, a photon with an incorrectly reconstructed shower start or a very low energy, short shower will often be reconstructed with a lower median $\frac{dE}{dx}$. Badly reconstructed particles, such as a muon track that has been reconstructed as a shower by mistake, often show very low, near-zero $\frac{dE}{dx}$ values, as can be seen by the ν_μCC events in the top left plot.



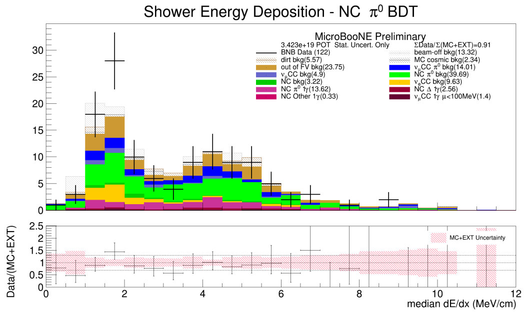
(a) The shower stem dE/dx immediately after preselection.



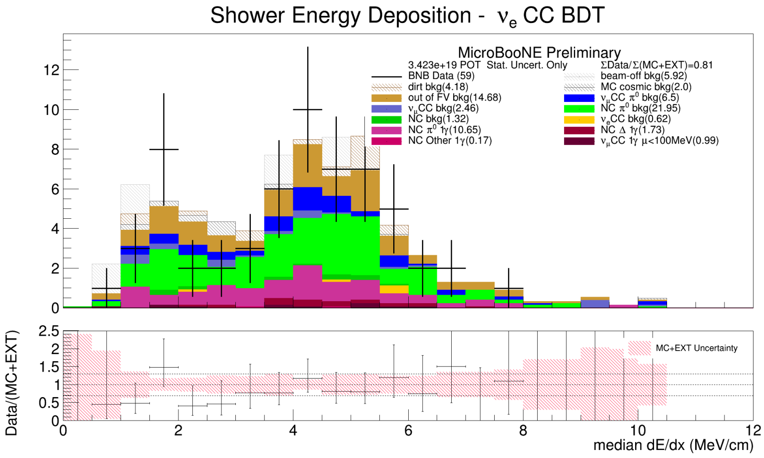
(b) The shower stem dE/dx after preselection and the ν_μCC background BDT cut.



(c) The shower stem dE/dx after preselection and the ν_μCC and other background BDT cuts.



(d) The shower stem dE/dx after preselection and the ν_μCC, other, and NC π^0 background BDT cuts.



(e) The shower stem dE/dx after preselection and the ν_μCC, other, NC π^0, and ν_eCC background BDT cuts.

Figure 12: Median dE/dx for the first 4cm of the shower stem after each of the preselection and BDT cuts.

5 SUMMARY

This document presented a selection for true single visible photon events with an inclusive final state topology, defined as any event with exactly one true photon with true production and shower start inside the TPC and true energy above 20 MeV, any number of hadrons, and no muons with true energy above 100 MeV. The selection uses well-documented Wire-Cell tools from the inclusive electron neutrino LEE analysis that have been expanded to work on photons as well as electrons. Four BDTs, each targeting a specific type of background, have been developed for this selection. The absolute efficiency and purity for the selection after the complete preselection and BDT chain is 15.2% and 18.5%, respectively, which corresponds to 13.54 signal events from Standard Model processes in 3.4e19 POT and a signal-to-background ratio of 1:4. After the selection, the signal events are made up of 78.7% NC π^0 1γ, 12.8% NC Δ, 1.3% NC Other 1γ, and 7.3% ν_μCC 1 γ events. The largest remaining background is NC π^0 background (0 or 2 visible photons) events. With the selection in place further work on the analysis, such as including systematics, evaluating sensitivities, and performing goodness-of-fit tests with the full available MicroBooNE dataset, can begin.

References

[1] R. Acciarri *et al.* (MicroBooNE Collaboration). Design and construction of the MicroBooNE detector. Journal of Instrumentation, 12(02):P02017–P02017, Feb 2017.

[2] A.A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration). Updated MiniBooNE neutrino oscillation results with increased data and new background studies. Phys. Rev. D, 103(5):052002, 2021.

[3] P. Abratenko *et al.* (MicroBooNE Collaboration). Search for an Excess of Electron Neutrino Interactions in MicroBooNE Using Multiple Final State Topologies. arXiv:2110.14054[hep-ex], Oct 2021.

[4] P. Abratenko *et al.* (MicroBooNE Collaboration). Search for neutrino-induced neutral-current Δ radiative decay in microboone and a first test of the miniboone low energy excess under a single-photon hypothesis. Phys. Rev. Lett., 128:111801, Mar 2022.

[5] P. Abratenko *et al.* (MicroBooNE Collaboration). Search for an anomalous excess of inclusive charged-current ν_e interactions in the microboone experiment using wire-cell reconstruction. arXiv:2110.13978[hep-ex], Oct 2021.

[6] A. A. Aguilar-Arevalo et al. Significant Excess of ElectronLike Events in the MiniBooNE Short-Baseline Neutrino Experiment. Phys. Rev. Lett., 121(22):221801, 2018.

[7] Alexis Armando Aguilar-Arevalo. An improved Neutrino Oscillations Analysis of the MiniBooNE Data. PhD thesis, Jan 2008.

[8] P. Abratenko *et al.* (MicroBooNE Collaboration). Cosmic Ray Background Rejection with Wire-Cell LArTPC Event Reconstruction in the MicroBooNE Detector. arXiv:2101.05076[physics.ins-det], Jan 2021.

[9] P. Abratenko et al. First measurement of energy-dependent inclusive muon neutrino charged-current cross sections on argon with the microboone detector. Phys. Rev. Lett., 128:151801, Apr 2022.

[10] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. arXiv:1603.02754[cs], Mar 2016.

[11] P. Abratenko et al. Wire-cell 3d pattern recognition techniques for neutrino event reconstruction in large LArTPCs: algorithm description and quantitative evaluation with MicroBooNE simulation. Journal of Instrumentation, 17(01):P01037, Jan 2022.