

Event Selection in the MicroBooNE Deep Learning Based Low Energy Excess Analysis Using Two-Body Scattering Criteria

MicroBooNE Collaboration

May 2020

MICROBOONE-NOTE-1086-PUB

Abstract

The uniquely detailed neutrino event information from liquid argon time projection chambers allows reconstruction of a set of kinematic quantities that over-constrain the expectations for charged current quasielastic scattering (CCQE). MicroBooNE makes use of the CCQE consistency requirements in a deep-learning-based search for the MiniBooNE low energy excess analysis. This requirement rejects backgrounds as well as events with poorly reconstructed neutrino energy due to final state interactions of the outgoing proton.

The results presented here demonstrate the quality of the selection of ν_e and ν_μ events. We show excellent agreement between the data and the simulation across many data sets. This positions us to be ready to unblind the MicroBooNE low energy excess analysis in the very near future.

Contents

1	Introduction	2
2	Variables Used to Identify CCQE Interactions	3
2.1	Kinematic Variables in the Lab Frame	4
2.2	Quasi-elastic Scattering Consistency, Discussed in the Lab Frame	6
2.3	Model Independent Boosting to Reduce Nucleon Motion Effects	7
2.4	Ionization Asymmetry of the Particles	12
2.5	Statistical Analysis Presented on Plots	12
3	Boosted Decision Trees Used in This Analysis	13
4	Simulation and Systematic Uncertainties	13
5	$1\mu 1p$ Selection, With Data-to-MC Comparisons	14
5.1	Preliminary Cuts	14
5.2	Boosted Decision Tree	15
5.3	Selected Data-MC Comparisons	16

6	1e1p Selection, With Limited Data-to-MC Comparisons	16
6.1	Preliminary Cuts	16
6.2	Boosted Decision Trees	19
6.3	Selected Data-MC Comparisons	21
7	Conclusions	25
A	Further 1μ1p Data-to-Monte Carlos Comparisons	25
B	Further 1e1p Data-to-Monte Carlos Comparisons	30
B.1	Run 1+3 Open Data Set	30
B.2	Run 1+2+3 Blindness-Safe Plots	44

1 Introduction

Liquid argon time projection chambers (LArTPCs), such as the MicroBooNE detector, allow for detailed, nearly-bubble-chamber-quality event reconstruction. The detector is composed of two sub-systems, a set of 32 cryogenic Photo-Multiplier Tubes (PMT), and a Time Projection Chamber (TPC). The detector design can be found in great details in [2]. Events consisting of two-body interactions, such as charged current quasi-elastic scattering (CCQE), $\nu_\ell + n \rightarrow \ell + p$, can be particularly well reconstructed for the cases where the lepton, ℓ , and proton p traverse sufficient distance to span multiple wires. Backgrounds can mimic this “1 ℓ 1p” signature. However, one can require the two-body scattering kinematics of the CCQE signal in order to greatly reduce backgrounds, assuming that the paths of the neutrinos in the beam are parallel when they reach the detector. In this analysis, we exploit the fact that, under the two-body scattering assumption, the 1 ℓ 1p CCQE signal is over-constrained. Thus, we can demand two-body scattering consistency between the reconstructed kinematic variables to isolate highly pure 1 μ 1p and 1e1p CCQE samples.

Well-reconstructed two-body scatters are ideal for investigating the MiniBooNE “low energy excess” (LEE) [7] [8]. The model for the LEE investigated here assumes that the excess derives from $CC\nu_e$ -like interactions. In a highly pure analysis, like the one presented here, the background to the LEE signal is dominated, by far, by the intrinsic ν_e in the beam. These produce CCQE interactions that are typically, although not exclusively, at higher energy than the LEE signal. Thus, if excellent energy resolution can be achieved, the intrinsic background can be separated from the signal. Moreover, the analysis will use the ν_μ measured events to constrain the intrinsic ν_e rate. Hence, excellent energy resolution on these events is also desirable. The two-body scattering kinematics of the events in this analysis sample lead to excellent energy resolution.

The goal of this analysis is to isolate neutrino events in the energy range of 200 MeV to 1 GeV that are consistent with the signature of CCQE in the rest frame of the neutron. We will consider 1e1p events, which will potentially include the signal of the LEE and also 1 μ 1p events that will be used to constrain the systematic uncertainties and backgrounds. The largest backgrounds are removed at the start of the analysis through minimum requirements for photo-electrons observed in the photo-multiplier tubes, followed by application of a novel algorithm for removing cosmic rays, called WireCell [3]. This analysis then makes use of Deep Learning techniques for track versus shower identification (see Public Note 1091 [6]), as well as a specialized 3-D reconstruction algorithm described in Ref. [5].

The analysis begins with cosmic ray identification. Reconstructed charge that is not associated with cosmic rays is then passed to the pre-selections. These place requirements on light observed

in the event, two-prong vertex within the fiducial volume, track containment within the active volume, lepton energy > 35 MeV, proton energy > 60 MeV, opening angle > 0.5 radians and agreement between total event energy and momentum, in order to select high-quality candidate events for the next stage. Events are likely to have more than one vertex found. The vertices are separated into those with track-like and shower-like leptons using Deep Learning metrics [4], and then boosted decision trees (BDT) assign a probability to the vertex. The event vertex is the one with the highest probability.

The kinematic constraints are then applied to the identified vertex to determine if this is a CCQE event. We introduce the kinematic variables that are employed below. We use variables reconstructed in the laboratory frame. We can also reconstruct the β and γ of the event, as described below, allowing a boost back to the frame where the target neutron is at rest. This is important for the consistency requirements we will use to reject backgrounds. The kinematic constraints are introduced in the form of BDTs—the same that were used to identify the most likely vertex. An important aspect of this analysis is to establish there is excellent agreement between the BDT data inputs and outputs compared to the simulation across many data samples. This will be shown in the note below and the appendices.

Finally, two Particle Identification Cuts are applied to the analysis to remove the small level of remaining background from event mis-identification. The relevant algorithms are described in Public Note 1080 and 1090. The first makes use of the deep-learning-based MPID algorithm. In this case we cut events where there is a $> 20\%$ probability of a muon appearing in the event. The second makes use of a second-shower-finding algorithm that identifies displaced electromagnetic showers to allow for π^0 identification, requiring π^0 mass of < 50 MeV for an event to be selected.

In the following discussion, we describe how we construct variables to test the CCQE hypothesis. We then describe the $1\mu 1p$ selection and compare with data. Because this is a blind analysis that is not yet complete, we do not have access to the complete data set. However, we can compare our predictions to data from a set of “open boxes.” We also compare our predictions to a set of “blindness-safe plots” of variables that have less than 0.1σ significance for the predicted LEE signal. Table 1 described the various data sets that appear in this report. Additional $1\mu 1p$ plots are available in Appendix A. We demonstrate the consistency of this data set with the MicroBooNE flux and cross section, and discuss modeling issues. Then, we describe the $1e 1p$ selection, providing comparisons with open data sets as a proof of principle. Again, additional plots are provided in Appendix B.

The most important conclusions of this note will be:

- We can isolate a very pure sample of $1e 1p$ and $1\mu 1p$ CCQE events.
- The selected data are in excellent agreement with the simulation.

This gives us confidence that the MicroBooNE Deep Learning Low Energy Excess analysis is nearly ready to be unblinded.

2 Variables Used to Identify CCQE Interactions

The 3D reconstruction provides information on basic kinematic quantities in the lab frame. Below, we describe how higher level kinematic quantities that will be used in the analysis are formed from these basic quantities. We then describe how to boost to the frame where the nucleon momentum is negligible, and reevaluate some of the event variables.

Title of box	Description	Used for testing
4.4×10^{19} POT	An unbiased sample of data from Run 1	$1\mu 1p$ and $1e 1p$
5.3×10^{19} POT	An unbiased sample of data from Run 3 added to the 4.4×10^{19} POT sample	$1\mu 1p$ and $1e 1p$
High Energy Box	$E_\nu > 700$ MeV- open sideband from 4.8×10^{20} POT sampled from Runs 1, 2 and 3)	$1e 1p$
Blindness-safe histograms	Plots derived from from 4.8×10^{20} POT. No correlations between plots are available	$1e 1p$

Table 1: Open boxes employed for studies of data compared to prediction.

2.1 Kinematic Variables in the Lab Frame

Kinematic variables provide the strongest source of signal identification. Insofar as only contained events are considered, once the particles are identified, a LArTPC enables full 4-momentum reconstruction from the energies and angles. From here, virtually any useful kinematic quantity can be reconstructed. The most useful of these and their associated definitions are summarized in Table 2.

Variable Name	Definition
Base Variables	
E_p	Energy of proton determined from range
E_μ	Energy of muon determined from range in detector
E_e	Energy of electron determined from deposited charge
m_ℓ, m_n, m_p	Masses of the lepton, neutron and proton
$\cos \theta_p$	$p_p^z / \vec{p}_p $
$\cos \theta_\ell$	$p_\ell^z / \vec{p}_\ell $
ϕ_p	$\text{atan2}(p_p^y, p_p^x)$
ϕ_ℓ	$\text{atan2}(p_\ell^y, p_\ell^x)$
$P_p = (E_p, \vec{p}_p)$	Reconstructed 4-momentum of the proton
$P_\ell = (E_\ell, \vec{p}_\ell)$	Reconstructed 4-momentum of the lepton
E_b	Binding Energy for argon; the analysis assumes $B = 28.5$ MeV
Definitions Related to Neutrino Energy	
E_ν^{range} *	$E_p + E_\ell - (m_n - E_b)$
E_ν^{QE-p}	$\frac{E_p(m_n - E_b) + \frac{1}{2}(m_\ell^2 - (m_n - E_b)^2 - m_p^2)}{(m_n - E_b) + \vec{p}_p \cos \theta_p - E_p}$
$E_\nu^{QE-\ell}$	$\frac{E_\ell(m_n - E_b) + \frac{1}{2}(m_p^2 - (m_n - E_b)^2 - m_\ell^2)}{(m_n - E_b) + \vec{p}_\ell \cos \theta_\ell - E_\ell}$
Δ^{QE}	$\sqrt{(E_\nu^{QE-p} - E_\nu^{QE-\ell})^2 + (E_\nu^{QE-p} - E_\nu^{range})^2 + (E_\nu^{QE-\ell} - E_\nu^{range})^2}$
* Unless explicitly noted otherwise, any reference to E_ν in this memo or histograms refer to E_ν^{range} as reconstructed in the laboratory frame.	
Event Kinematics	
Q^2	$2E_\nu(E_\ell - P_\ell^z) - m_\ell^2$
Hadronic Mass (m_{had})	$E_\nu - E_\ell$
Björken's Scaling x (x_{Bj})	$Q^2 / 2m_n m_{had}$
Björken's Scaling y (y_{Bj})	m_{had} / E_ν
Opening angle	$\cos^{-1}(\hat{p}_\ell \cdot \hat{p}_p)$
p_T	$\sqrt{(p_\ell^x + p_p^x)^2 + (p_\ell^y + p_p^y)^2}$
p_L	$p_p^z + p_\ell^z$
α_T	$\cos^{-1} \left(-\frac{\vec{p}_T^\ell \cdot \vec{p}_T}{ \vec{p}_T \vec{p}^\ell } \right)$
ϕ_T	$\cos^{-1} \left(-\frac{\vec{p}_T^\ell \cdot \vec{p}_T^p}{ \vec{p}_T^p \vec{p}^\ell } \right)$

Table 2: Kinematic variables derived from the reconstruction code used in this analysis.

2.2 Quasi-elastic Scattering Consistency, Discussed in the Lab Frame

The elastic kinematics associated with CCQE processes not only enables a ready computation of the initial neutrino energy, but it provides several methods by which this can be done. Crucially, this will require the knowledge that the neutrino momentum was aligned with the z-axis. Consider the following:

$$(P_\nu - P_p)^2 = (P_\ell - P_n)^2 \quad (1)$$

$$E_\nu = E_p + E_\ell - m_n \quad (2)$$

$$\vec{p}_\nu = \vec{p}_p + \vec{p}_\ell \quad (3)$$

Where 1 is one of the Mandelstam variables - a Lorentz invariant - and 2 and 3 enforce energy and momentum conservation. From 1

$$\begin{aligned} (p_\ell - p_n)^2 &= (p_\nu - p_p)^2 \\ E_\ell^2 - |\vec{p}_\ell|^2 + E_n^2 - 2E_\ell E_n &= E_\nu^2 - |\vec{p}_\nu|^2 + E_p^2 - 2E_\nu E_p + 2\vec{p}_\nu \cdot \vec{p}_p - |\vec{p}_p|^2 \\ m_\ell^2 + m_n^2 - 2E_\ell m_n &= m_p^2 - 2E_\nu E_p + 2\vec{p}_\nu \cdot \vec{p}_p \\ m_\ell^2 + m_n^2 - 2E_\ell m_n &= m_p^2 - 2E_\nu E_p + 2E_\nu |\vec{p}_p| \cos\theta_p \\ E_\nu &= \frac{E_\ell m_n + \frac{1}{2}(m_p^2 - m_n^2 - m_\ell^2)}{E_p - |\vec{p}_p| \cos\theta_p} \end{aligned}$$

Using the fact that the neutrino momentum was aligned with the z-axis, and that $|\vec{p}_\nu| \approx E_\nu$, Eq.3 becomes $E_\nu = |\vec{p}_p| \cos\theta_p + |\vec{p}_\ell| \cos\theta_\ell$. Subtracting this from Eq.2 yields $E_p - |\vec{p}_p| \cos\theta_p = m_n + |\vec{p}_\ell| \cos\theta_\ell - E_\ell$. Substituting this into the denominator above yields an equation solely in terms of the reconstructed lepton. Furthermore, the above procedure would proceed equally well in terms of either final state particle $p \leftrightarrow \ell$. This yields two equations, either of which can be used to obtain the initial neutrino energy in terms of the 4-momentum of the proton or lepton. We will call these E_ν^{QE-p} and $E_\nu^{QE-\ell}$ hereafter.

$$E_\nu^{QE-\ell} = \frac{E_\ell m_n + \frac{1}{2}(m_p^2 - m_n^2 - m_\ell^2)}{m_n + |\vec{p}_\ell| \cos\theta_\ell - E_\ell} \quad (4)$$

$$E_\nu^{QE-p} = \frac{E_p m_n + \frac{1}{2}(m_\ell^2 - m_n^2 - m_p^2)}{m_n + |\vec{p}_p| \cos\theta_p - E_p} \quad (5)$$

And in fact we have already seen a third equation, 2, which requires no directional information but requires that we have reconstructed both the lepton and proton. Hereafter we will refer to this as E_ν^{range} .

$$E_\nu^{range} = E_p + E_\ell - m_n \quad (6)$$

Because these three formulae in principle reconstruct the same quantity, it will prove useful to consider the following quantity which captures their level of mutual agreement. The smaller the value of this parameter, the better the agreement between the quasi-elastic energies. We will hereafter refer to this as the quasi-elastic scattering consistency.

$$\Delta^{QE} = \sqrt{(E_\nu^{QE-p} - E_\nu^{QE-\ell})^2 + (E_\nu^{QE-p} - E_\nu^{range})^2 + (E_\nu^{QE-\ell} - E_\nu^{range})^2} \quad (7)$$

In principle, these three formulae are exactly equivalent for the quasi-elastic scattering of a neutrino off of a free neutron. In other words, the idealized quasi-elastic scattering consistency is 0. However, we must account for the fact that the neutron exists within a nuclear environment. This introduces several complications.

Firstly, the neutron is not free. A certain amount of energy must be expended to remove it from the nucleus. In reality this is not a fixed value within a given nucleus. However the differences are small, $\mathcal{O}(10 \text{ MeV})$. So at the energy scales of interest to this analysis we approximate it as a fixed value capturing the mean removal energy. A value of $E_b = 28.5 \text{ MeV}$ is used. This acts as a reduction to the neutron's effective scattering mass $m_n \rightarrow m_n - E_b$ in Eqs. 4-6.

Secondly, we have assumed that the momenta of the outgoing lepton and proton which are measured exactly reflect those which emerged from the scatter. Because of the potential for final state interactions within the nucleus this will not always be true. We do not introduce a correction to account for this. Rather we will proceed under the assumption that these final state interactions are small. There will be a population of interactions for which this is false. Our equations will accordingly show tension with each other in situations such as this. The decision to neglect final state interactions is thus a tacit shift in the sought after signal to be quasi-elastic interactions which have small final state interactions.

Finally, we deduced these equations under the assumption that the struck nucleon was at rest. Because of Fermi motion this is also not necessarily true. An approach to help minimize this is discussed in the following section.

2.3 Model Independent Boosting to Reduce Nucleon Motion Effects

Variables associated with quasielastic scattering are generally derived in a picture in which a neutrino scatters off of a bound nucleon at rest. Using these formulae with values reconstructed in the lab frame works to a first approximation, but doing so neglects the initial Fermi motion of the nucleon which was struck and thus introduces tacit approximations. If these formulae could be properly computed in the rest frame of the nucleon then their power and accuracy can be increased.

The Fermi motion is random, and *a priori* cannot be known. However, our signal of interest comprises only quasielastic interactions. If final state interactions are small, we then have sufficient reconstructible information to estimate the Fermi momentum if the momenta of all final state particles are reconstructed and so long as the initial direction of the neutrino beam is known. This follows from a consideration of the kinematics of the scattering.

$$\begin{aligned} \vec{p}_\nu + \vec{p}_{fermi} &= \vec{p}_{final} \\ \vec{p}_\nu + \vec{p}_{fermi} &= \sum_{particles} \vec{p} \\ &\quad 1 \text{ lepton} + 1 \text{ proton final state} \end{aligned} \quad (8)$$

$$\begin{aligned} \langle 0, 0, p_\nu \rangle + \langle p_f^x, p_f^y, p_f^z \rangle &= \langle p_p^x, p_p^y, p_p^z \rangle + \langle p_\ell^x, p_\ell^y, p_\ell^z \rangle \\ \rightarrow \vec{p}_f &= \langle p_p^x + p_\ell^x, p_p^y + p_\ell^y, p_p^z + p_\ell^z - p_\nu \rangle \\ \rightarrow \vec{p}_f &\approx \langle p_p^x + p_\ell^x, p_p^y + p_\ell^y, p_p^z + p_\ell^z - E_\nu \rangle \end{aligned}$$

This allows us to estimate all three components of the Fermi momentum using reconstructible quantities. This, combined with the energy of the struck nucleon E_n , can then be used to define the boost vector $\vec{\beta}$. The value of E_n in this case is the off shell mass given by Eq. 9 in which m_n is the on shell nucleon mass, E_b is the removal energy, and T_f is the final state nuclear recoil kinetic energy.

$$\begin{aligned} E_n &= m_n - E_b - T_f \\ T_f &\ll (m_n - E_b) \\ &\approx m_n - E_b \end{aligned} \tag{9}$$

The boost vector β is then given by Eq. 10

$$\vec{\beta} = \frac{\vec{p}_f}{E_N} \tag{10}$$

It then becomes possible to boost all reconstructed 4-momenta into the struck nucleon rest frame. These boosted vectors can then be used to compute derived kinematic quantities which will be more powerful in this frame.

$$\begin{pmatrix} E' \\ p'_x \\ p'_y \\ p'_z \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta_x & -\gamma\beta_y & -\gamma\beta_z \\ -\gamma\beta_x & 1 + \frac{\gamma-1}{\beta^2}\beta_x^2 & \frac{\gamma-1}{\beta^2}\beta_x\beta_y & \frac{\gamma-1}{\beta^2}\beta_x\beta_z \\ -\gamma\beta_y & \frac{\gamma-1}{\beta^2}\beta_x\beta_y & 1 + \frac{\gamma-1}{\beta^2}\beta_y^2 & \frac{\gamma-1}{\beta^2}\beta_y\beta_z \\ -\gamma\beta_z & \frac{\gamma-1}{\beta^2}\beta_x\beta_z & \frac{\gamma-1}{\beta^2}\beta_y\beta_z & 1 + \frac{\gamma-1}{\beta^2}\beta_z^2 \end{pmatrix} \begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} \tag{11}$$

For example, in the prior section we saw that for CCQE interactions we can compute the initial neutrino energy in three distinct ways. However the formulae assumed we were in the nucleon rest from. Performing this boost minimizes the detrimental impact of that assumption. As illustrated in Figs. 1- 3, for a simulated CCQE sample boosting provides reconstructed energies which are both in tighter agreement with each other and with the true value.

Correlations Between Energy Reconstruction Equations With and Without Boosting

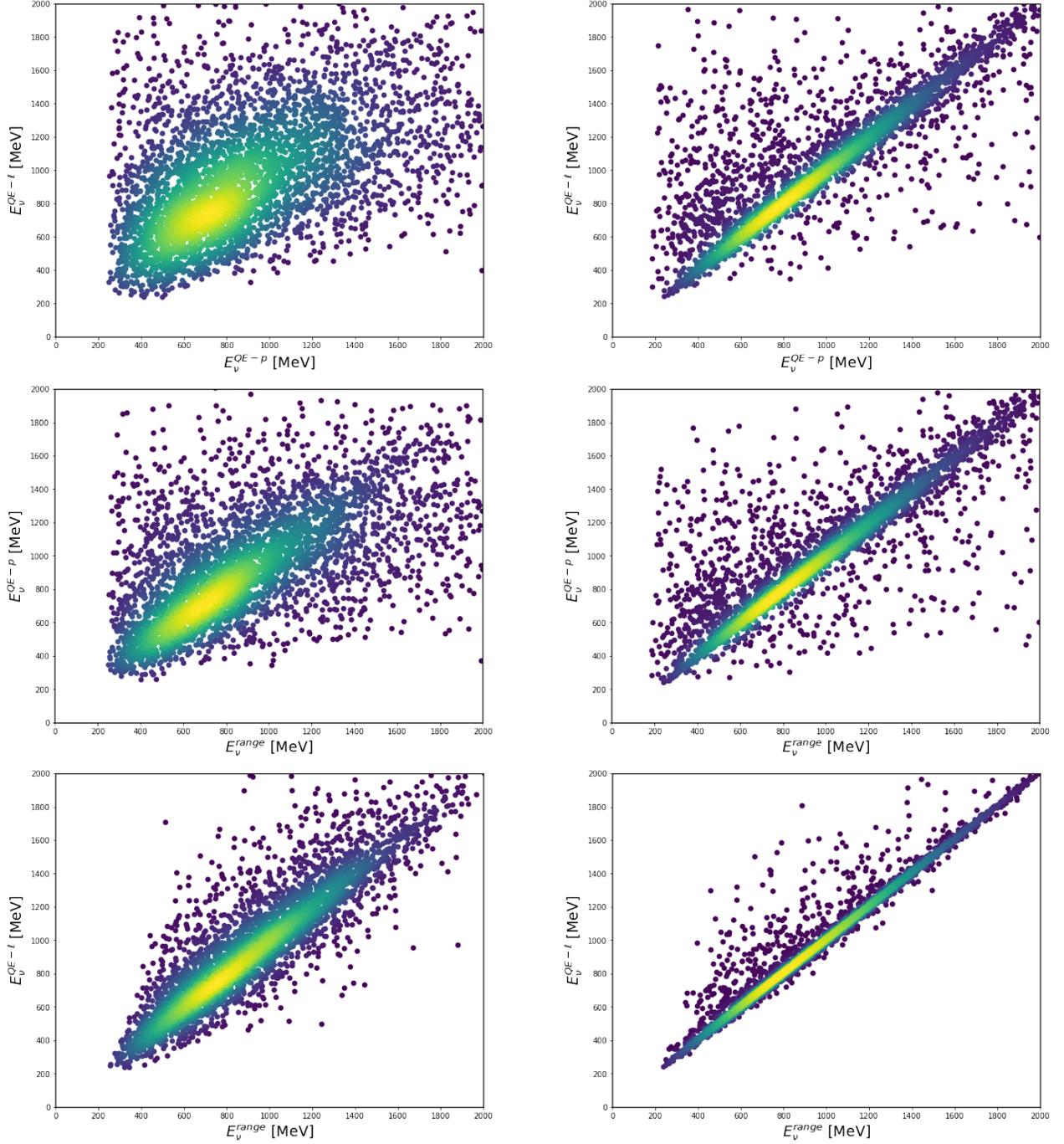


Figure 1: Using $1\mu 1p$ as an example, correlations between the three different equations used to compute the reconstructed neutrino energy, E_{ν}^{range} , E_{ν}^{QE-p} , & E_{ν}^{QE-l} are shown. These are computed using truth level kinematic values for a simulated sample of CCQE interactions with 1 proton and 1 muon in the final state. These are illustrated prior to (left) and after (right) boosting into the frame in which the struck nucleon was at rest.

Comparison of True and Reconstructed Energy Definition Equations With and Without Boosting

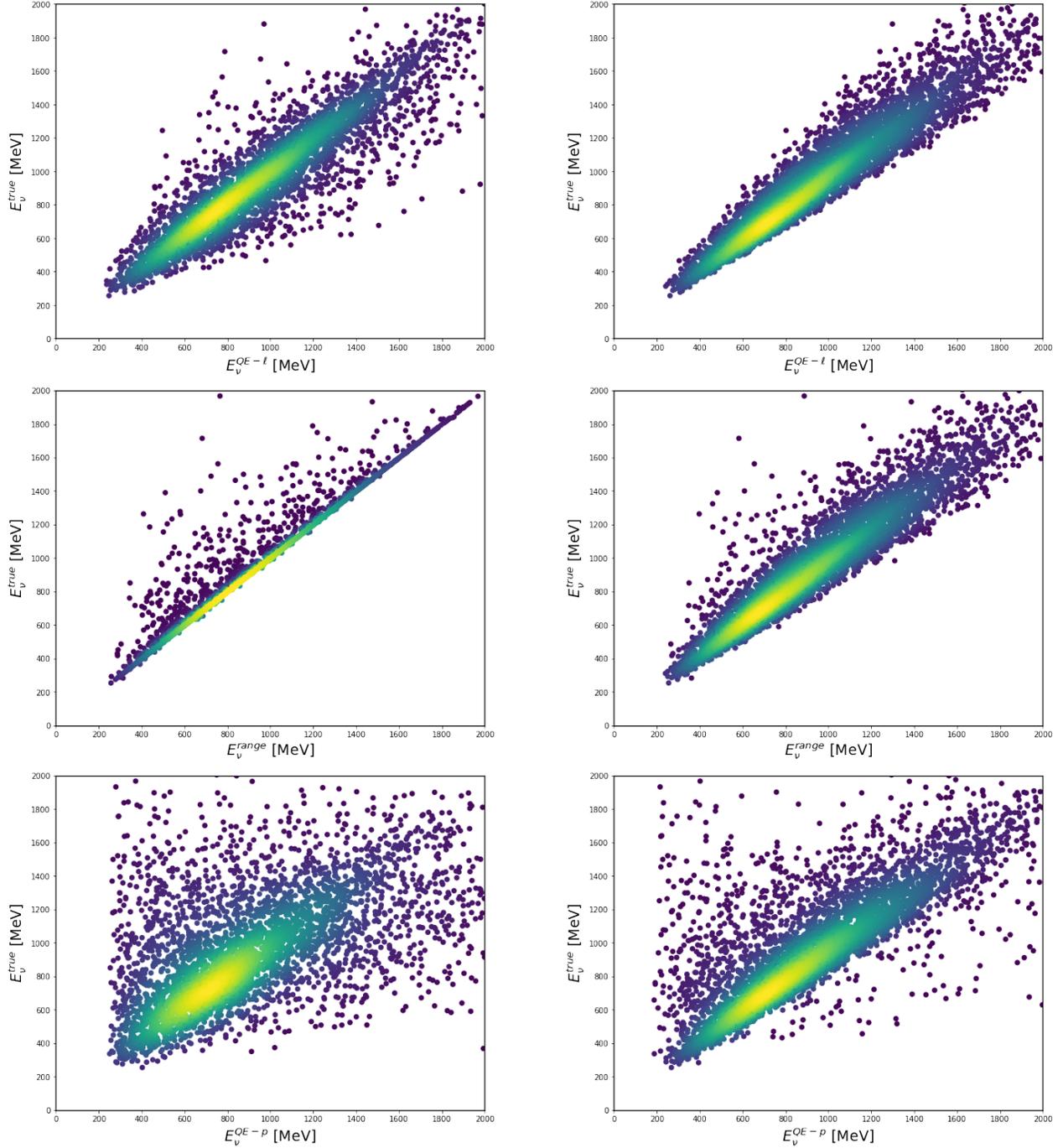


Figure 2: Using $1\mu 1p$ as an example, the degree to which three neutrino energy reconstruction equations, E_ν^{range} , E_ν^{QE-p} , & E_ν^{QE-l} , correspond with the true energy are shown. These are computed using truth level kinematic values for a simulated sample of CCQE interactions with 1 proton and 1 muon in the final state. These are illustrated prior to (left) and after (right) boosting into the frame in which the struck nucleon was at rest.

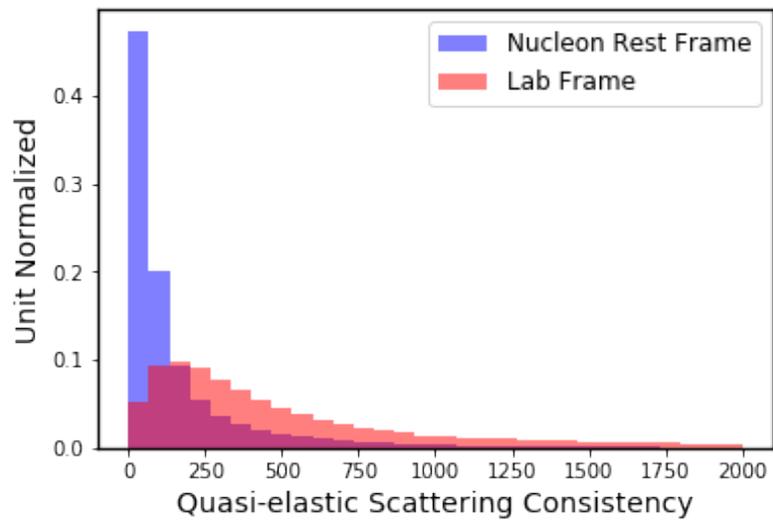


Figure 3: Improvement produced in Δ^{QE} by boosting for a simulated sample of CCQE interactions with 1 proton and 1 muon in the final state.

2.4 Ionization Asymmetry of the Particles

Along with the kinematic variables, in the $1e1p$ analysis a variable that describes the ionization asymmetry of the particles is used in the BDT. For completeness, we explain this variable here.

In principle, the asymmetry of the ionization of the two particles in the event can be a useful discriminator against backgrounds. At MicroBooNE energies, the muon track and the trunk of the electron shower are expected to be minimum ionizing. If the electromagnetic particle in the event is a converted photon rather than an electron, the resulting e^+e^- pair will produce twice the ionization in the portion of the shower trunk near the vertex. Protons are heavily ionizing, typically depositing more than twice that of minimum ionizing particles, with increasing rate of deposit increasing with distance and ending in a Bragg peak. This information can be used to evaluate an asymmetry variable that can discriminate between $1\ell 1p$ events and backgrounds.

In practice, this analysis does not directly use the ionized charge per unit length, called “dQ/dx”; to form the asymmetry, but instead uses an average pixel intensity per track. Averaging is found to reduce sensitivity to variations in the deposited charge along the track due to stochastic processes related to ionization, δ rays, and small-scale detector defects like dead and noisy wires.

In this analysis, the event images provide information about the ionization density of the objects we have reconstructed, as discussed in Ref. [5]. The ADC counts are first mapped to pixels in an image for each wire plane, and this information on “pixel intensity” is then used for the 3D event reconstruction. The 3D tracking produces a series of points along a track, as it performs the track finding. The average intensity within ± 2 pixels (± 6 mm) of each projected track point is found. Then, across the length of each reconstructed particle, the mean pixel intensity is calculate μ_{PI}

The magnitude of μ_{PI} provides a simple yet effective particle ID within an event. We have limited ourselves to 1 lepton-1 proton events, so the proton is identified by choosing the track with the largest μ_{PI} . Because this quantity can be sensitive to the detector response at any given location, we make us of the asymmetry of μ_{PI} between the particles, divided by the sum:

$$\eta = \left| \frac{\mu_{PI}^p - \mu_{PI}^\ell}{\mu_{PI}^p + \mu_{PI}^\ell} \right| \quad (12)$$

In the case of signal interactions, which are $1\ell 1p$, the minimum ionizing lepton will have an average pixel intensity that is roughly half of the intensity of a proton. On the other hand, in the case of a track that is misreconstructed to have a false vertex, the values of μ_{PI} for the two vertex particles will be approximately equal.

2.5 Statistical Analysis Presented on Plots

The primary goal of this note is to benchmark the agreement between data and simulation. In order to do this, we must address the problem of low data statistics and finite Monte Carlo statistics. In this section, we briefly describe the method used to present the uncertainties and to characterize the agreement between a predicted spectrum and observation.

In general, the plots will be presented in two frames. The top frame shows the data overlaid on the Monte Carlo prediction. The bottom frame will present the ratio of data to Monte Carlo. Several uncertainties will be shown. The uncertainty bars on data points in the top frame are 68% Poisson confidence intervals. The ratio plots in the lower frame will contain shaded bands which indicate the flux, cross-section, and detector systematics. The points on the ratio plot illustrate the 68% credible interval for the data-to-prediction ratio with a flat prior. This credible interval

accounts for the statistical uncertainty in both data and in the Monte Carlo used to produce the prediction. In the case of empty bins, the upper bar illustrates the same 34% upper coverage as other points, but because the lower bar extends fully to 0 these are technically 84% total coverage intervals.

The computation of the χ^2/NDF and associated P -values are decoupled from the uncertainties shown graphically. The χ^2 which is reported on all plots is a combined Neumann-Pearson (CNP) χ^2 , which all MicroBooNE analyses will ultimately use in LEE sensitivity calculations. Where μ_i and M_i are the number of predicted and observed events in a given bin, the χ^2_{CNP} is defined as:

$$\chi^2_{\text{CNP}} = \sum_i \begin{cases} \frac{(\mu_i - M_i)^2}{\frac{3}{1/M_i + 2/\mu_i}} & M_i \neq 0 \\ \frac{(\mu_i - M_i)^2}{\frac{\mu_i}{2}} & M_i = 0 \end{cases}$$

3 Boosted Decision Trees Used in This Analysis

In this analysis, the BDT training samples make use of “well-reconstructed CCQE” simulated events. This is defined as a simulated CCQE event where the reconstructed energy is reconstructed within 20% of the true neutrino energy. By training on well-reconstructed CCQE events, the BDT learns to reject MEC events and events with large final state interactions, leaving reconstructed events with a reliable energy estimation.

The $1e1p$ analysis uses a single BDT, while the $1\mu1p$ analysis uses two BDTs. The $1\mu1p$ analysis uses one BDT trained to select well-reconstructed CCQE neutrino events and reject cosmics (called the “Cosmic BDT”), and another to select well-reconstructed CCQE events and reject other types of neutrino events (called the “Neutrino BDT”). The BDTs are regularized to reduce the potential for spurious results.

To reduce the effect of systematic uncertainties in the LEE analysis, the variables used in the $1\mu1p$ BDTs and the $1e1p$ BDTs have a great deal of overlap. The input variables are listed in Table 3.

4 Simulation and Systematic Uncertainties

The simulation used for this analysis uses a GENIE Tune for the cross sections described in Ref. [9]. This tune is based on the preliminary GENIE release G18_10a_02_11a and is based on fits to the T2K “CC0pi” ν_μ data. The ν_e cross section assumes the same underlying cross section parameters; hence it differs from the ν_μ cross section mainly due to lepton-mass threshold effects.

The data below were acquired over three running periods. Between Run 1 and Runs 2 and 3, substantial improvements were made to the electronics. The TPC response for Run 1 is therefore different from Runs 2 and 3, which are very similar. The detector experienced a degradation in light collection during Run2, and this is simulated.

Systematic uncertainties are included in the analysis below. These fall into three categories: Flux, Cross Section, and Detector Response. The methods for determining these uncertainties are described in Ref. [10]. In the case of detector systematic uncertainties, the total variation was found to be 7% and this is directly applied.

Variable	Used in $1\mu 1p$ BDTs	Used in $1e1p$ BDTs
Opening Angle	Yes	Yes
$ \phi_p - \phi_\ell $	Yes	Yes
Charge within 5 cm of vertex	Yes	Yes
Neutrino Energy	Yes	Yes
α_T	Yes	Yes
Event p_T	Yes	Yes
Event p_T/p	Yes	Yes
Δ^{QE} (QE Consistency)	Yes	Yes *
Q_0	Yes	Yes
Q_3	Yes	Yes
Proton ϕ	Yes	Yes
Proton θ	Yes	Yes
Lepton ϕ	Yes	Yes
Lepton length	Yes	Yes
ϕ_T	Yes	No
Bjorken's x	Yes	No
Bjorken's y	Yes	No
Proton length	No	Yes
Energy of electromagnetic shower	No	Yes
η	No	Yes
$p_z - E_\nu$	No	Yes
$\theta_p + \theta_e$	No	Yes
proton shower fraction	No	Yes
Electron shower fraction	No	Yes
Shower charge in event image / shower charge clustered as electron	No	Yes

Table 3: Variables used in the $1\mu 1p$ BDTs and the $1e1p$ BDT, with overlap noted. If a * appears, the variable is used in the boosted frame. Note that more variables are used for validation than are used in the BDTs.

5 $1\mu 1p$ Selection, With Data-to-MC Comparisons

Selection of $1\mu 1p$ events is done in two steps: (1) A set of preliminary cuts is applied to both clean out events more likely to be misreconstructed and to ensure that this selection is entirely independent from the $1e1p$ signal selection; and (2) A final cut is made on BDT scores.

5.1 Preliminary Cuts

These preliminary cuts serve to cut out events that are safely out of the realm of our desired sideband signal. It is further intended to limit the set of candidate events to those for which all variables the BDT requires are well defined. Specifically, these cuts are:

1. **Optical Precuts:** Require a signal of more than 20 photo-electrons to have been observed within a ~ 100 ns coincident window during the time frame coincident with the accelerator

the beam spill. Such a flash is typically from a neutrino, but could be due to a cosmic Michel decay. To veto this latter case, no > 20 photo-electron signal in the $2 \mu s$ prior to the beam spill is permitted. Finally, no more than 60% of the total light in an event can be found in any one PMT. These cuts retain more than 97% of the neutrino events (all types) while rejecting more than 75% of neutrino-empty background events.

2. **Vertex Requirements:** Require a 2 pronged vertex reconstructed within the fiducial volume, defined as more than 10 cm from the edge of the active detector volume.
3. **Containment Requirements:** Require containment, *i.e.* that no particle can be tracked within 5 cm of an edge for the $1e1p$ and $1\mu1p$ analyses.
4. **Electron Shower Cut** Using the MPID neural network score, an event which has a 20% or greater probability of having an electron shower present is rejected.
5. **Basic Quality Cuts:** Require opening angle greater than 0.5 radians and that the reconstructed momenta yield a physical $\beta < 1$ value so that boosting is defined (described in Sec. 2.3). If an event contains multiple vertices, select the vertex with the highest cosmic BDT score within that event.

5.2 Boosted Decision Tree

A final selection is applied via a cut on two boosted decision tree (BDT) predictions, made using the XGBoost Python library. BDT training requires careful definition of both the signal and any background one wishes to discriminate against. To achieve a pure selection with high enough stats to provide a meaningful future constraint on $1e1p$ systematic errors, two separate BDTs are trained: one to weed out cosmic events and one to weed out other, non-CCQE events.

In response to individual variable studies, it was found that our reconstruction is particularly sensitive to events for which we either fail to reconstruct the full energy or we mistake a non- $1l1p$ event for having only two prongs. To amend this, we can create a very specific signal definition which will respond well to our reconstruction capabilities.

Three new qualifiers are introduced for Monte Carlo: **1L1P**, **On Vertex** and **Good Reco**. An MC event is 1L1P if it contains **exactly one** lepton with true kinetic energy greater than 35 MeV and **exactly one** proton with true kinetic energy greater than 60 MeV. An event is On Vertex if its reconstructed neutrino vertex position is within 5cm of the true interaction vertex. An event which has Good Reco has total neutrino energy reconstructed within 20% of true neutrino energy. The signal definitions used to train the BDTs are therefore as follows:

- **Signal: ν_μ CCQE** — Our signal definition, for the purpose of training the BDT network is ν_μ MC events which are 1L1P, On Vertex, have Good Reco and feature an interaction mode corresponding to CCQE.
- **Background One: Cosmic** — Our cosmic background definition comprises off-beam events from our ExtBNB sample, dirt events (interactions simulated outside the detector’s active volume) as well as ν_μ MC events which are 1L1P but **not** On Vertex.
- **Background Two: ν_μ non-CCQE** — Our non-CCQE definition is ν_μ MC events that either: a) are 1L1P, On Vertex and with Good Reco but have an interaction mode corresponding to anything besides CCQE; b) Are 1L1P and On Vertex but not Good Reco, or c) Are not 1L1P.

These background definitions are motivated by the different capabilities of our reconstruction. In Background 1 events, a vertex has been reconstructed somewhere far from the true neutrino vertex, regardless of what kind of interaction that neutrino underwent. In Background 2 events, a vertex has been successfully placed on the true neutrino vertex, but either the interaction is not our desired $\nu_{mu}u$ CCQE signal or the energy was not reconstructed properly.

A score is then assigned to each event, with a number closer to one implying a more background-like event and one closer to zero implying more signal-like.

A cut is chosen to eliminate events with a Cosmic Background BDT probability of more than 0.5 and a ν_{mu} non-CCQE Background BDT probability of more than 0.5, as illustrated in Fig 4.

5.3 Selected Data-MC Comparisons

At this point, the full analysis has only been run over a small subset of recorded data corresponding to approximately $5e19$ protons on target. A plot of reconstructed neutrino energy after the full suite of selection cuts have been made can be seen in Fig 5.

The result of these cuts is a selection with almost 70% well-reconstructed ν_{mu} CCQE purity and which is estimated to identify over 4k ν_{mu} events when run over the full data set. Plots for a large handful of different variables after the full selection can be found in Appendix A.

6 $1e1p$ Selection, With Limited Data-to-MC Comparisons

Just like in the $1\mu1p$ sideband, identification of $1e1p$ signal events is done in two steps: (1) A set of preliminary cuts is applied (2) A cut on a BDT score.

6.1 Preliminary Cuts

The preliminary cuts will identify and remove populations of backgrounds that are very distinctly not signal and despite having different requirements, serve much the same purpose as those for the $1\mu1p$ selection. Specifically, these cuts are:

1. **Optical Precuts:** Require a signal of more than 20 photo-electrons to have been observed within a ~ 100 ns coincident window during the time frame coincident with the accelerator the beam spill. Such a flash is typically from a neutrino, but could be due to a cosmic Michel decay. To veto this latter case, no > 20 photo-electron signal in the $2 \mu s$ prior to the beam spill is permitted. Finally, no more than 60% of the total light in an event can be found in any one PMT. These cuts retain more than 97% of the neutrino events (all types) while rejecting more than 75% of neutrino-empty background events.
2. **Vertex Requirements:** Require a 2 pronged vertex reconstructed within the fiducial volume, defined as more than 10 cm from the edge of the active volume, and omitting a region in z between 700-740 cm corresponding to dead collection plane wires.
3. **Containment Requirements:** Require containment, *i.e.* that no particle can be tracked within 5 cm of an edge for the $1e1p$ and $1\mu1p$ analyses. The $1e1p$ adds an additional requirement that prevents the shower from being tracked into a region of poor efficiency, defined by $y = \frac{z}{\sqrt{3}} - 117$ and $y = \frac{z}{\sqrt{3}} - 80$ cm.

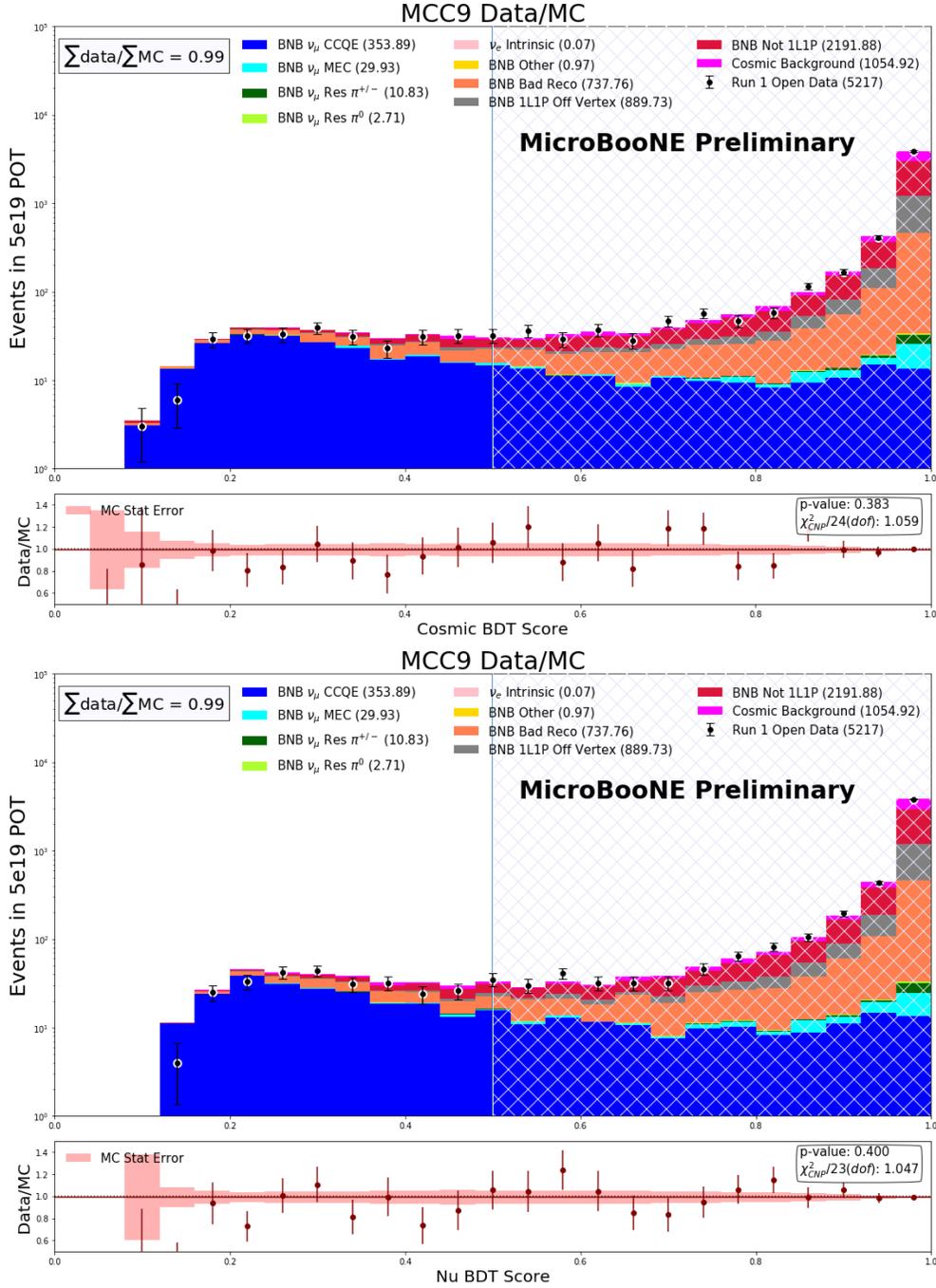


Figure 4: Spectra of each BDT score in all Run 1 $1\mu 1p$ samples. Note the log scale for readability, and that the ratio of data to Monte Carlo is presented with a suppressed zero. Although the point in the lowest bin is low, overall the data are in good agreement with the simulation, as indicated by the goodness of fit and also as seen in the plots in Appendix A.

4. **Particle Energy Cuts:** Require the reconstructed deposited lepton energy to be above 35 MeV and proton energy above 60 MeV. These values correspond to somewhat conservative lower bounds on how large a feature need be to regularly reconstruct well.
5. **Analysis Orthogonality Cut:** Require the particle which has the most shower like char-

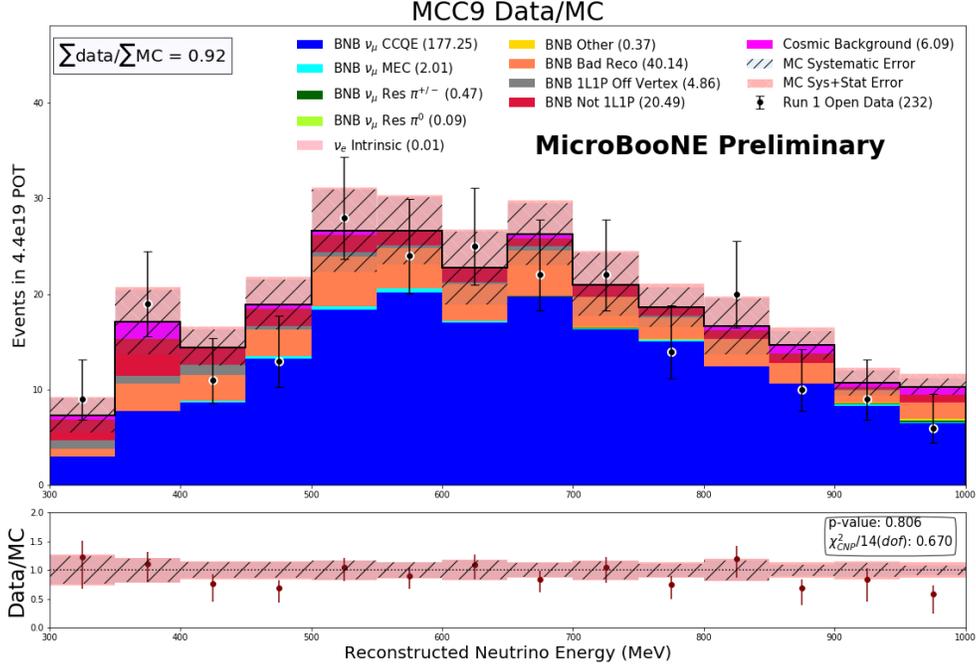


Figure 5: Comparison of Reconstructed ν_μ Energy for $1\mu 1p$ events between Run 1 Open Data (5e19 POT) and Montecarlo after full selection.

acter to comprise at minimum 20% shower like pixels. The shower fraction for each track is the fraction of of pixels which is identified as shower-like by *SparseSSNet*. The $1\mu 1p$ selection demands this value to be lower than 20%. This leads to a natural orthogonality in the selected samples.

6. **Muon Presence Cut** Using the MPID neural network score, an event which has a 20% or greater probability of having a muon present is rejected.
7. **π^0 Rejection** If the second shower finding algorithm yields a second shower which permits the reconstruction of a π^0 invariant mass, and this mass is > 50 MeV then the event is rejected.
8. **Basic Quality Cuts:** Require opening angle greater than 0.5 radians and that the reconstructed momenta yield a physical $\beta < 1$ value so that boosting is defined (described in Sec. 2.3). If an event contains multiple vertices, select the vertex with the highest BDT score within that event.

6.2 Boosted Decision Trees

The BDT is implemented using the XGBoost framework. Training is accomplished using three samples:

1. Data taken at a time external to the event readout window time, this provides a cosmic background sample.
2. Monte Carlo ν_μ with data overlay: Provides ν_μ sample with data driven cosmic background.
3. Monte Carlo ν_e intrinsic: Provides signal sample

The specific variables used during the training are specified in Table 3. Training is performed as conservatively as possible with 40% of the available simulation while 60% is reserved for validation. After reconstruction and initial selection cuts, this provides $\mathcal{O}(10^3)$ signal training events and $\mathcal{O}(10^4)$ ν_μ and cosmic background examples. The result of this training is a single score assigned to any given vertex. This score ranges from 0 to 1 and can be loosely interpreted as a probability that a given event is 1e1p signal like. The final selection requires that the BDT signal probability score is > 0.9 . This cut is preliminary and was chosen based on a preliminary sensitivity optimization accounting only for ν_e statistics. The ultimate value used will be chosen by optimization accounting for systematic uncertainties and the ν_μ constraint.

As one validation, we perform a comparison between a POT normalized predicted probability spectrum and an open data set corresponding to 4.4×10^{19} POT. This is illustrated in Fig. 6. It is also useful to verify that at any particular BDT cut chosen that the number of selected events anticipated matches the number we observe. This is also characterized using the same open data set and is illustrated in Fig. 7

For all plots shown in this note, the BDTs were trained on Run 1 simulation. Electronics improvements were implemented between Run 1 and the later runs, which substantially reduced noise. This may lead to the choice of employing BDTs tuned for separate runs in the future. However, as shown below, the goodness of fit measurements for the analysis using only the Run-1-trained BDT is acceptable on all plots. Thus, problematic effects of using a BDT trained only on Run 1 for the subsequent runs are modest.

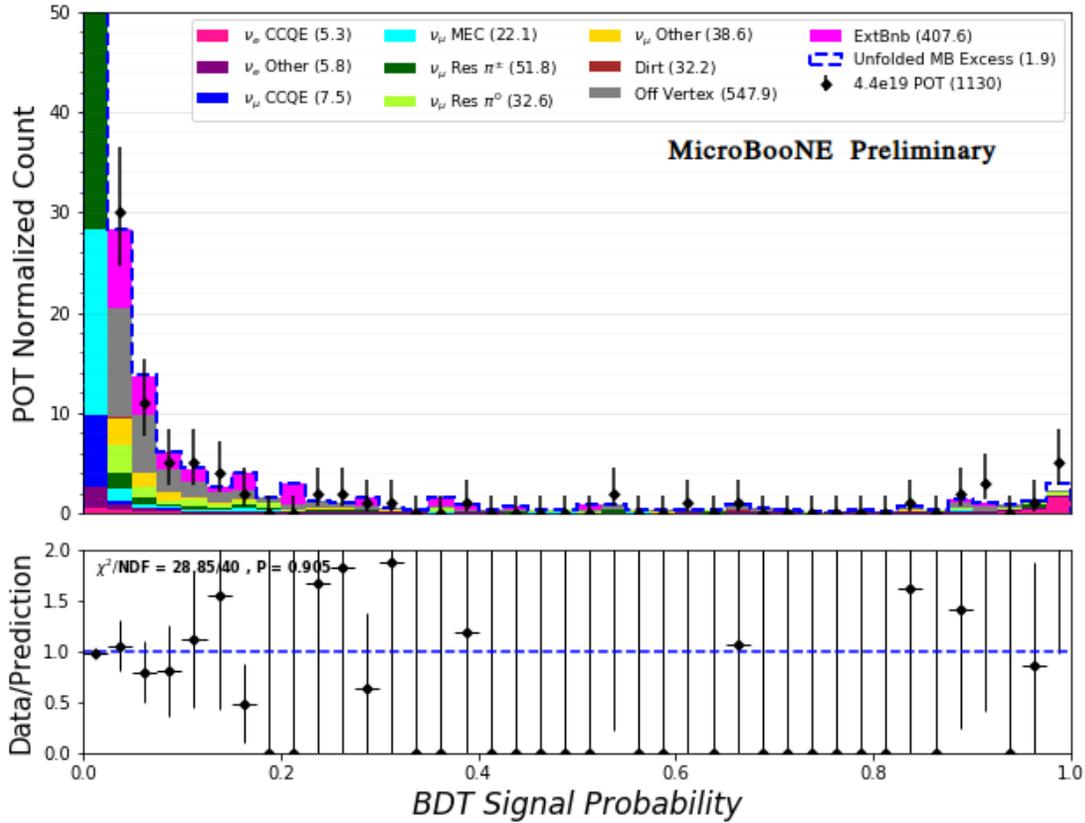


Figure 6: Left, Top: Comparison between a stacked predicted BDT score for the $1e1p$ selection normalized to 4.4×10^{19} POT and an open data set; Bottom: Ratio of data to prediction. Note that the first bin extends off the top of the upper plot, but is demonstrated to be in good agreement by the ratio shown on the lower plot. Right, shape normalized comparisons of a low energy signal shape as a function of BDT score compared with various backgrounds.

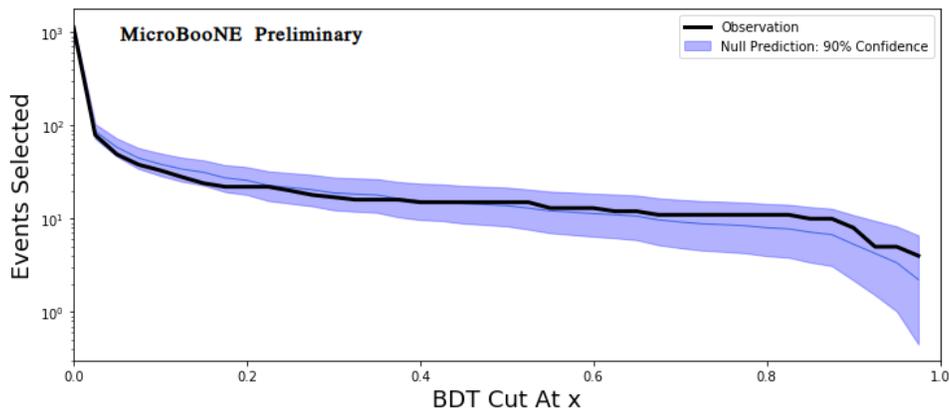


Figure 7: Comparison between the number of $1e1p$ events selected in data and prediction if the BDT cut is placed at $>x$. Moving further to the left indicates a cut at a lower BDT score value and permits more events into the selection. We compare the observed value with 90% confidence bands on the prediction. This allows us to see that the agreement between simulation and data is not an artifact of any particular cut value, but rather would persist for any chosen BDT score value.

6.3 Selected Data-MC Comparisons

While this analysis has not yet been performed over a large quantity of unblinded data, we nonetheless have performed validations using several useful data sets, described in Table 1. We highlight selected plots here. All are available in Appendix B. The specific sample that was used for a given plot will be specified.

The unbiased 5.3×10^{19} POT is a relatively small data set, but is completely unrestricted and thus allows any validation, albeit with limited statistics. For example, Fig. 8, a plot of the neutrino energy for the $1e1p$ selection, is derived from this data set with the full selection applied. A larger set of data to Monte Carlo comparisons for the unbiased 5.3×10^{19} POT is provided in Appendix B.1

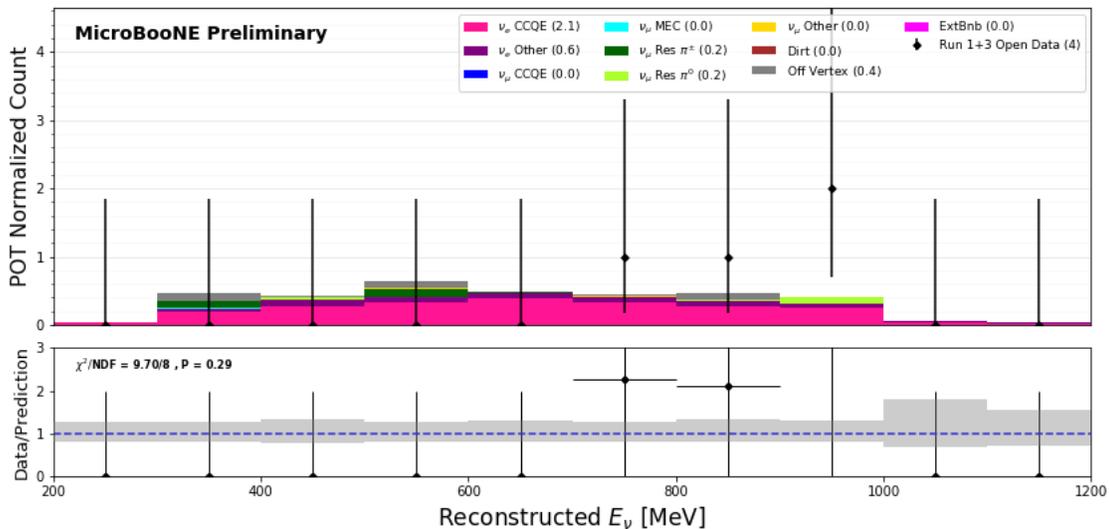


Figure 8: Stacked comparisons between $1e1p$ prediction and data for reconstructed neutrino energy using the open 5.3×10^{19} POT data.

Because the LEE signal is primarily at $E_\nu < 400$ MeV, we can also verify our energy reconstruction & selection with higher statistics while remaining blind, using the High Energy box, but only at energies >700 MeV. The neutrino energy for $1e1p$ events from this sideband is presented in Fig. 9. This and other plots we have studied show excellent agreement with expectation in the higher energy regime.

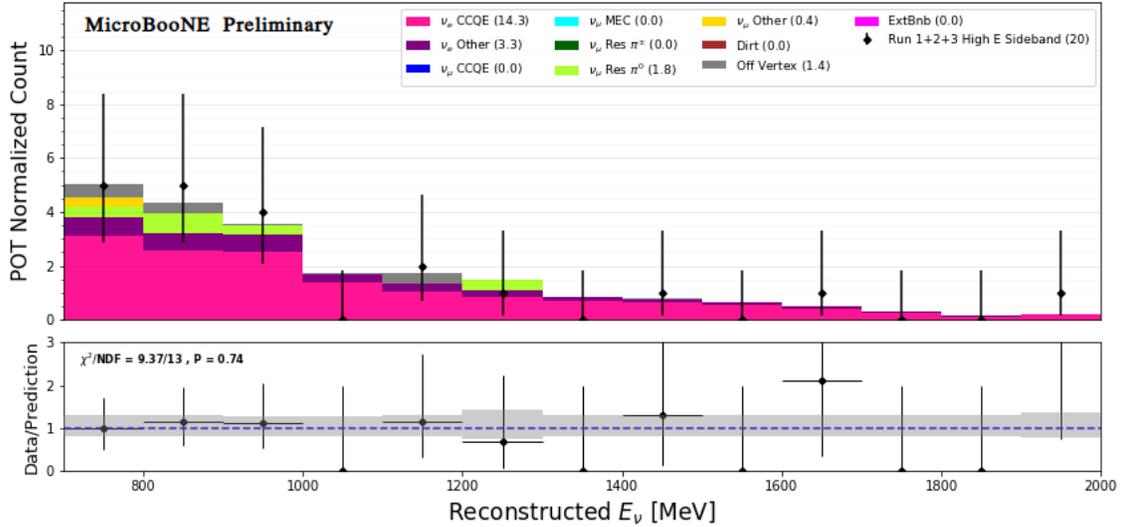


Figure 9: Combined runs 1+2+3 prediction for the high energy sideband with full $1e1p$ selection applied compared to observed selected data events in this sideband.

The blindness-safe histogram set opens histograms but not raw reconstructed data for about half of the total data set. By opening histograms, we can gain information without exploiting correlations between the variables. The variables are chosen such that shape information in any given histogram would give less than 0.1σ sensitivity to the LEE signal. All of the $1e1p$ selection cuts described above are applied. In order to remain blind to normalization, the total number of events was set to the predicted total minus 1.6σ (90% CL). This allows relatively high statistics: 38 events.

The Blindness-safe histograms provide invaluable comparisons between predicted and observed shapes with much higher statistics than available with open data, all of which show excellent agreement. Fig. 10 and 11 provide two examples of the excellent agreement observed: the 2-Body QE consistency and Bjorken's scaling x (which are both evaluated in the boosted reference frame. The boost provides better separation between the CCQE candidates, and the misreconstructed backgrounds. Both plots show good agreement with data and illustrate that our selected data events exhibit CCQE-like behavior. The good agreement on Bjorken's x is also notable because this is not a variable used in the BDT. Also interesting is $p_z - E_\nu$ (Fig. 12), the goodness of agreement between the final state longitudinal momentum and the reconstructed neutrino energy. For a mass-less neutrino and a $1e1p$ final state, this quantity should be zero. This provides a powerful kinematic check against backgrounds and is thus a strong BDT variable. Another interesting variable is α_T (Fig. 13, the transverse momentum relative to the lepton axis. The shape of this distribution can reveal the magnitude of final state interaction in the selected sample. The relative flatness of the distribution suggests that only small FSI events are being selected as anticipated. Appendix B.2 provides further examples.

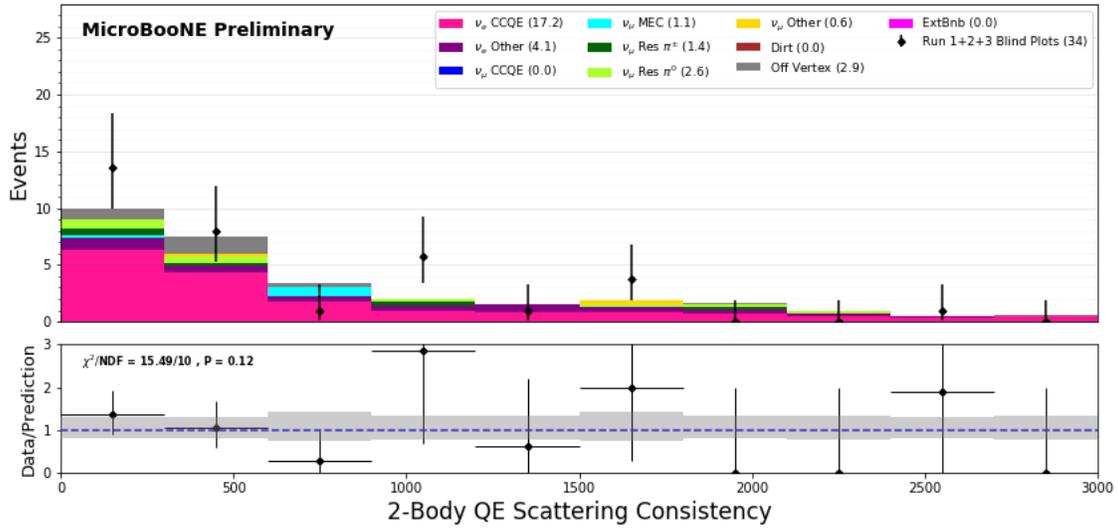


Figure 10: Blindness-safe plot from Run 1 + 2 + 3 : Consistency between 3 QE reconstruction formulae. Evaluated in the boosted, nucleon at rest reference frame

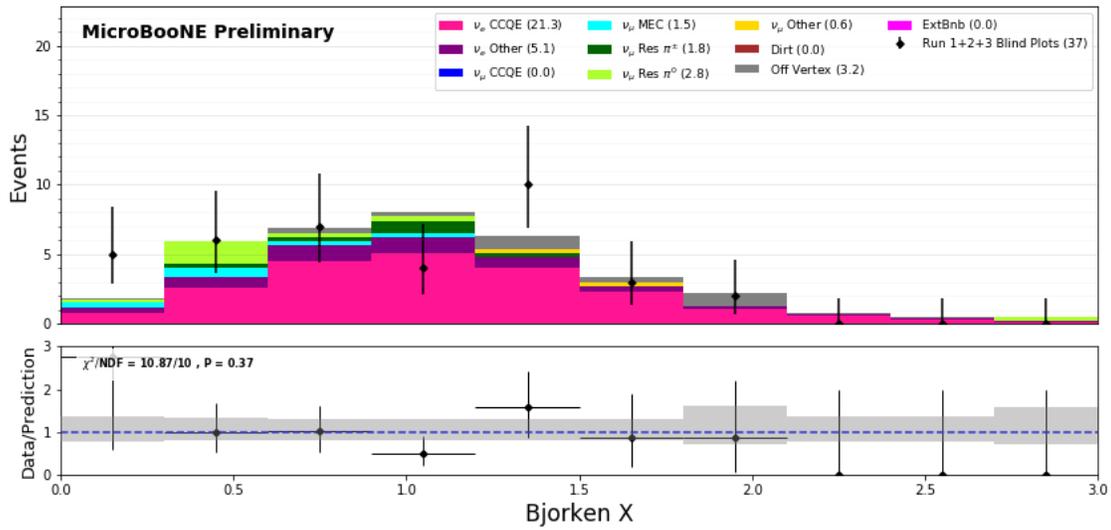


Figure 11: Blindness-safe plot from Run 1 + 2 + 3 : Bjorken X. Evaluated in the boosted, nucleon at rest reference frame.

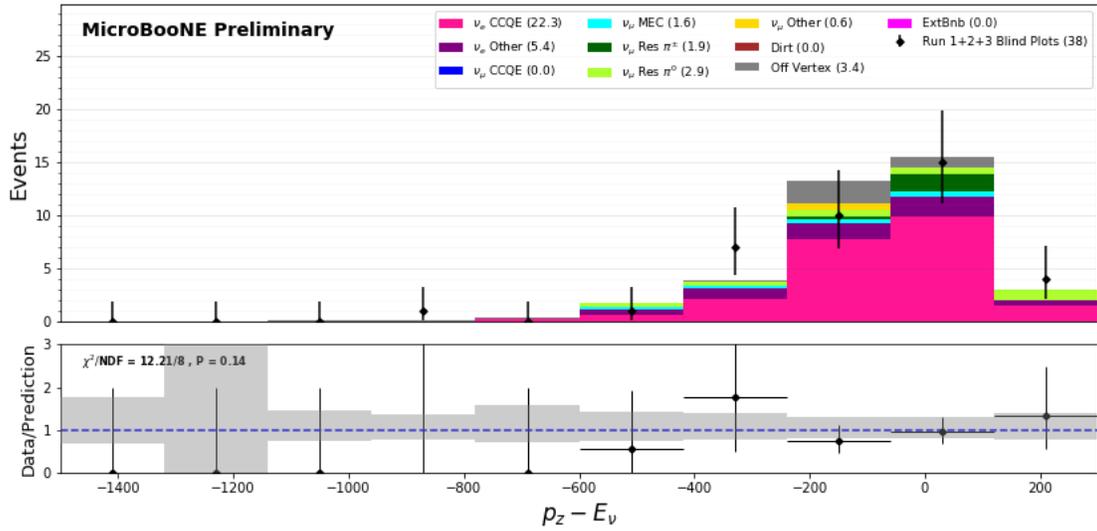


Figure 12: Blindness-safe plot from Run 1 + 2 + 3 : $p_z - E_\nu$

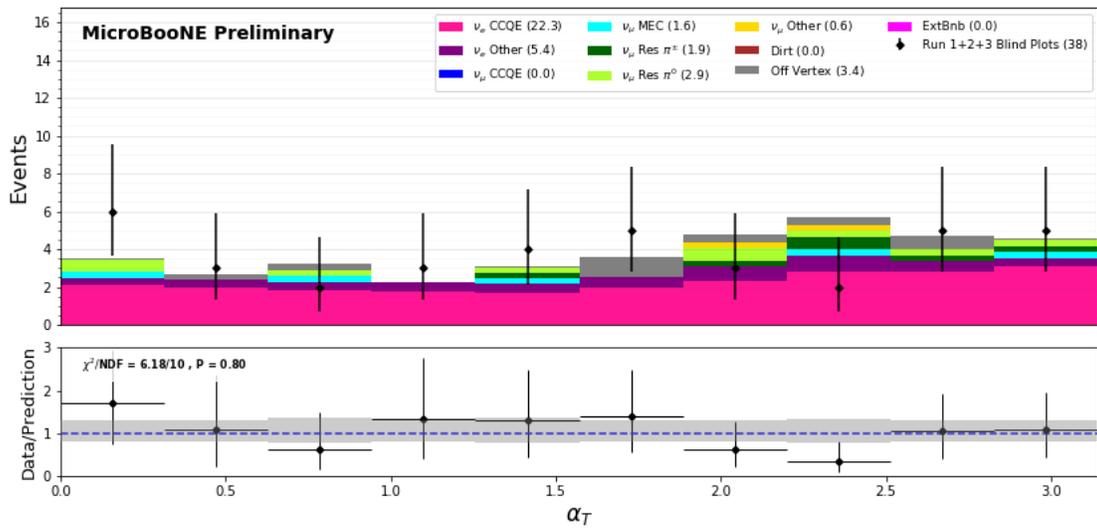


Figure 13: Blindness-safe plot from Run 1 + 2 + 3 : α_T

7 Conclusions

Using the unique reconstruction strengths afforded by the MicroBooNE LArTPC, we have developed an end to end reconstruction and event selection capable of identifying CCQE ν_μ ν_e interactions with high purity. This is possible because of the very complete kinematic topological information reconstructible in MicroBooNE as well as due to the application of powerful machine learning techniques. These reconstructions and selections will be pivotal in MicroBooNE's ultimate checks on the MiniBooNE low energy excess.

Appendices:

A Further $1\mu 1p$ Data-to-Monte Carlos Comparisons

These plots are derived from Run 1 open data corresponding to 4.4×10^{19} POT. Systematic errors comprise those from neutrino simulation flux and cross section, as well as major detector systematic effects. This represents the complete collection of plots used to check agreement, and includes some plots also in the main text, above.

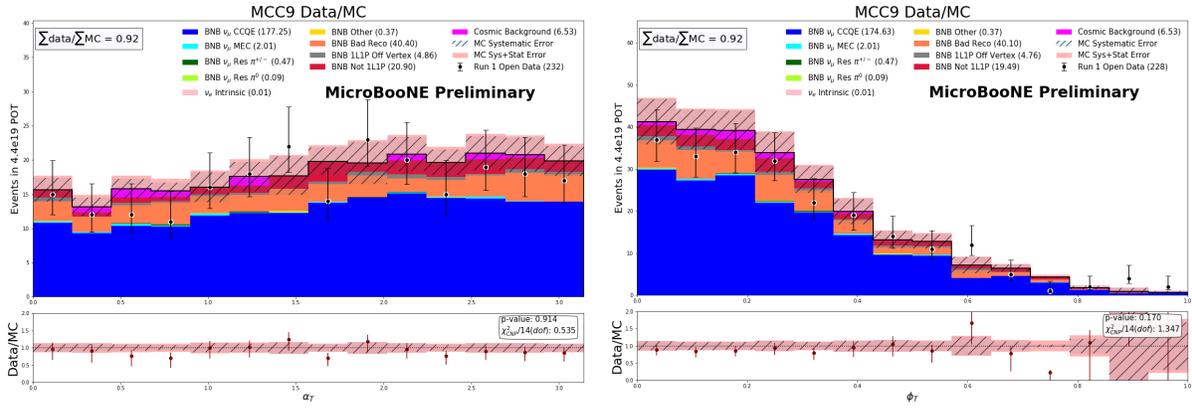


Figure 14: $1\mu 1p$ plots of α_T and ϕ_T

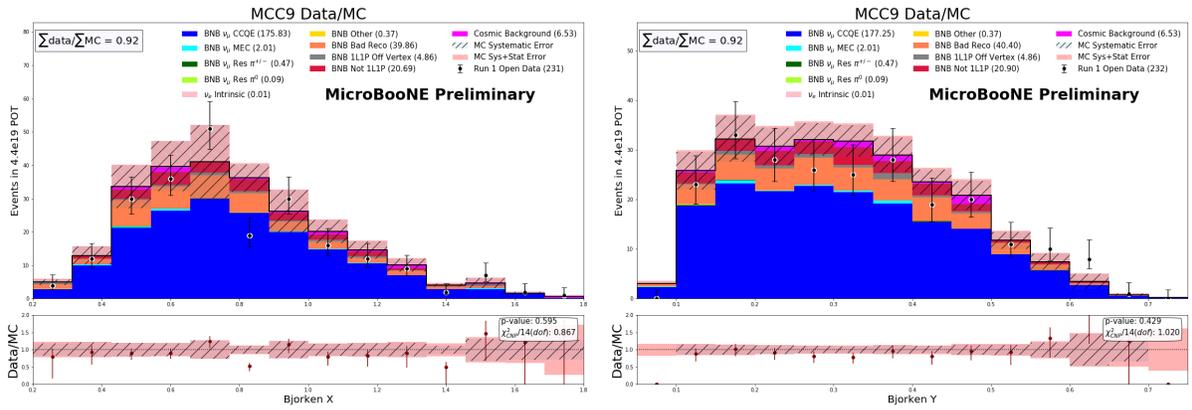


Figure 15: $1\mu 1p$ plots of x_{Bj} and y_{Bj}

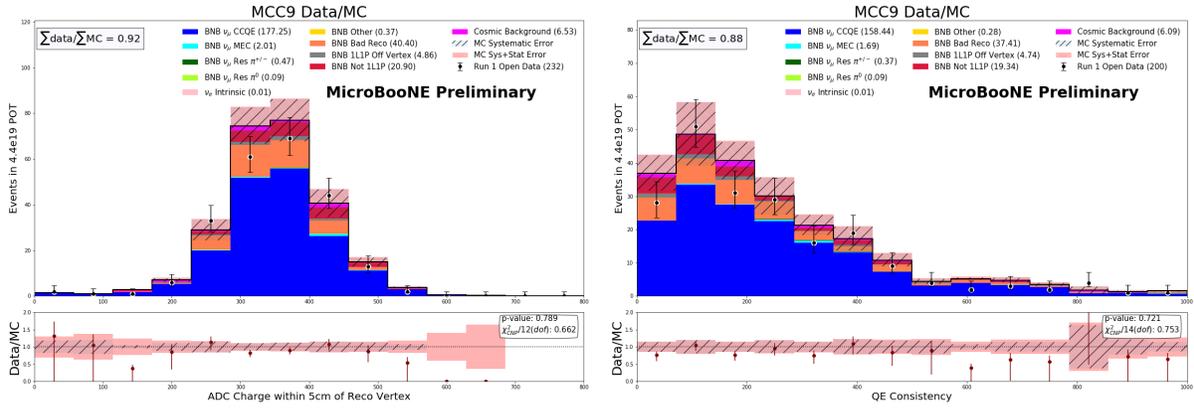


Figure 16: $1\mu 1p$ plots of charge near the vertex and quasi-elastic consistency

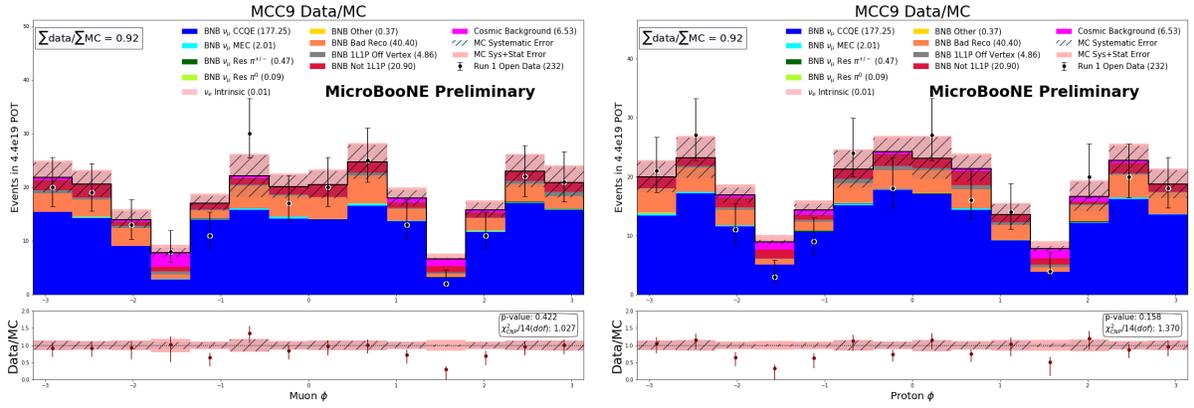


Figure 17: $1\mu 1p$ plots of muon and proton ϕ

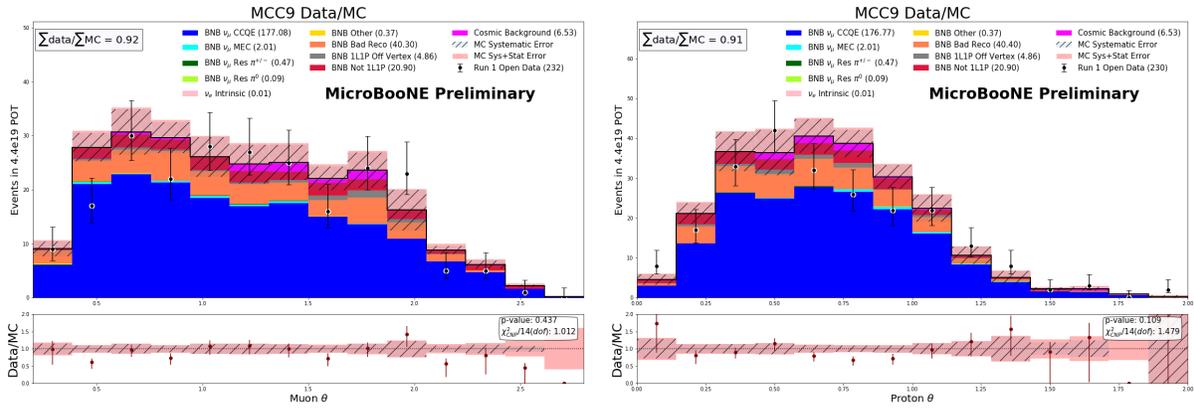


Figure 18: $1\mu 1p$ plots of lepton and proton

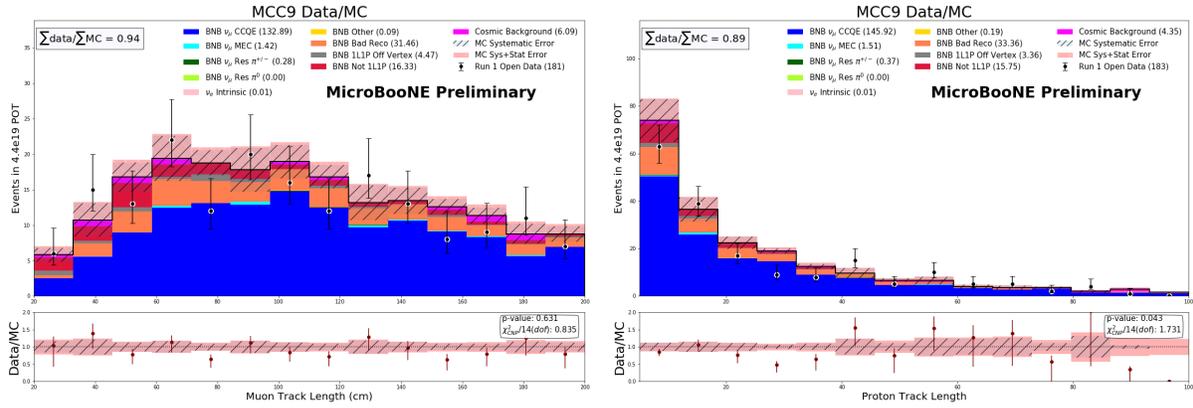


Figure 19: $1\mu 1p$ plots of lepton and proton track length

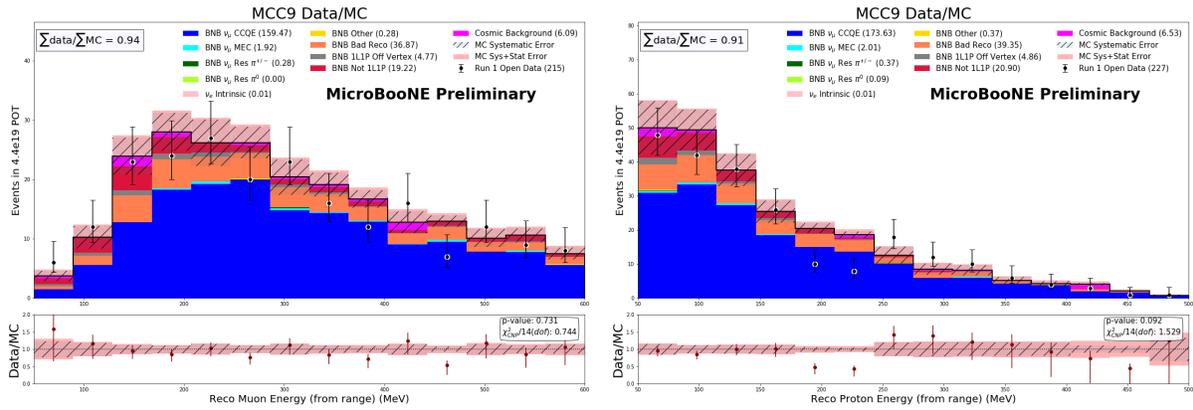


Figure 20: $1\mu 1p$ plots of lepton and proton energies

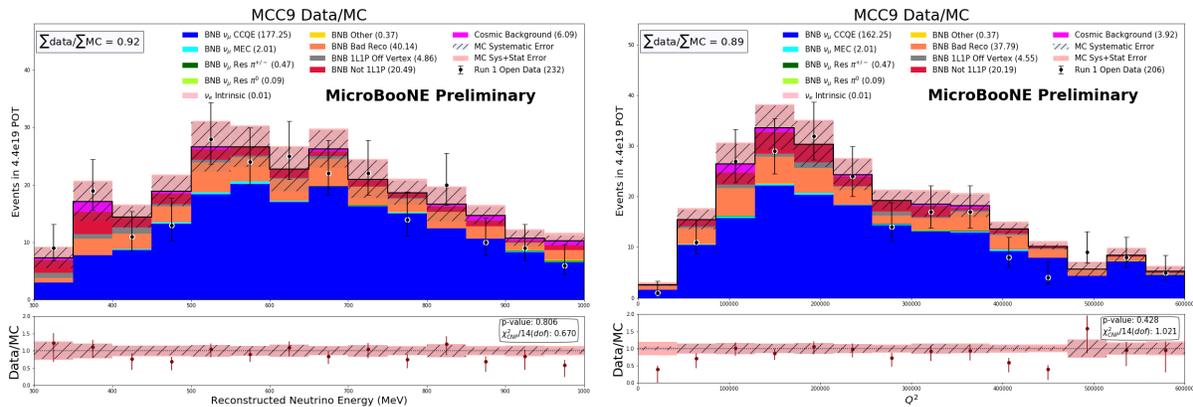


Figure 21: $1\mu 1p$ plots of neutrino energy and Q^2

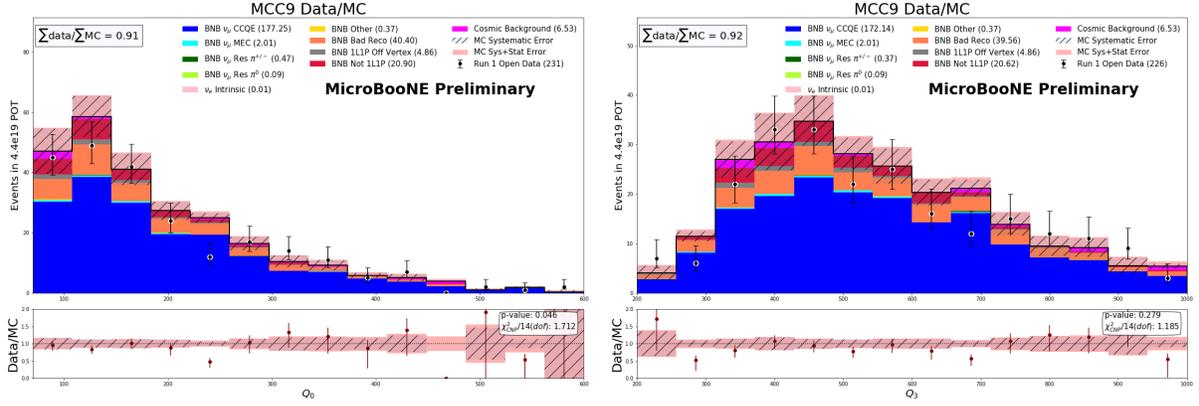


Figure 22: $1\mu 1p$ plots of the energy and momentum magnitude components of the exchange boson vector

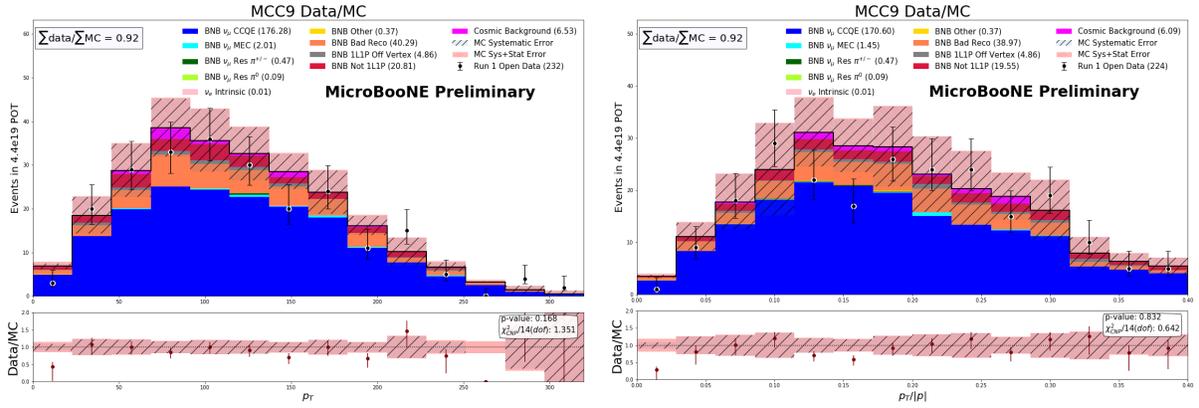


Figure 23: $1\mu 1p$ plots of p_T and the ratio of p_T to total momentum magnitude, $p_T/|p|$

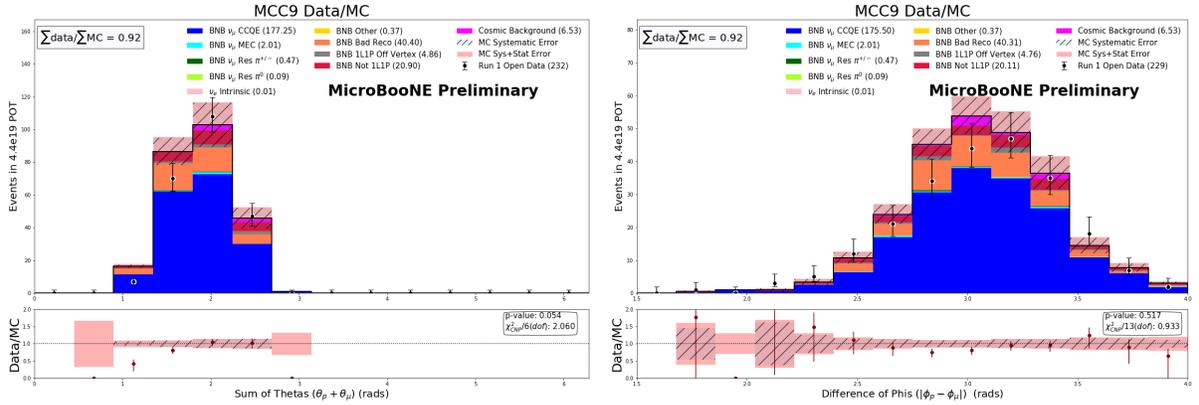


Figure 24: $1\mu 1p$ plots of $\theta_p + \theta_\mu$ and $|\phi_p - \phi_\mu|$

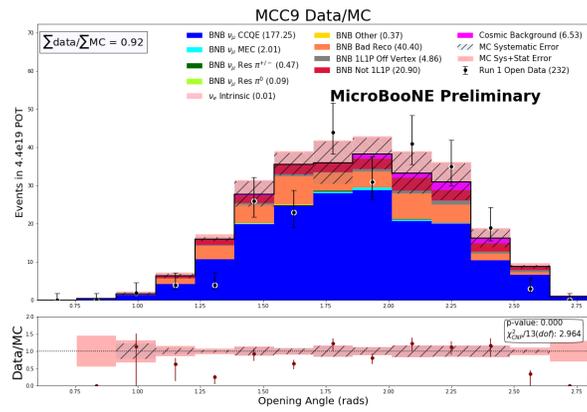


Figure 25: $1\mu 1p$ plot of opening angle of between muon and proton

B Further $1e1p$ Data-to-Monte Carlos Comparisons

B.1 Run 1+3 Open Data Set

These $1e1p$ histograms are derived from the open data sets from Run 1 + Run 3. This corresponds to a total POT of 5.3E19 from each run. This represents the complete collection of plots used to check agreement, and includes some plots also in the main text, above.

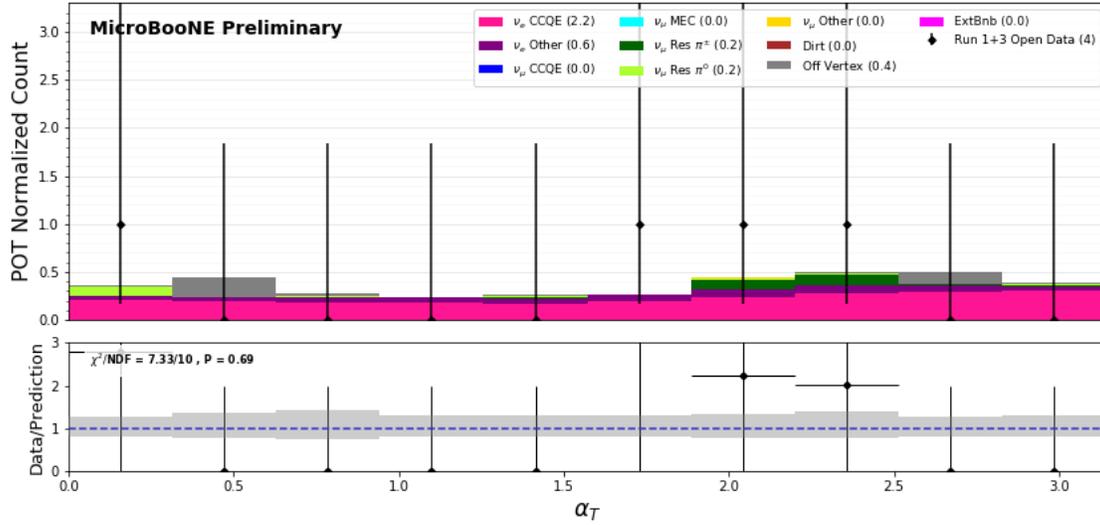


Figure 26: Stacked comparisons between $1e1p$ prediction and data for α_T .

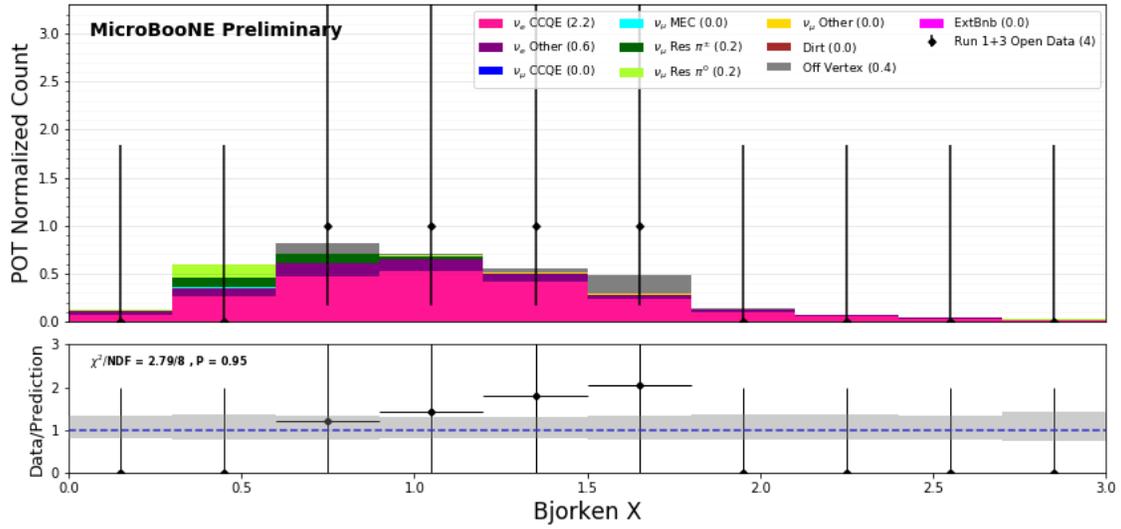


Figure 27: Stacked comparisons between $1e1p$ prediction and data for Bjorken X.

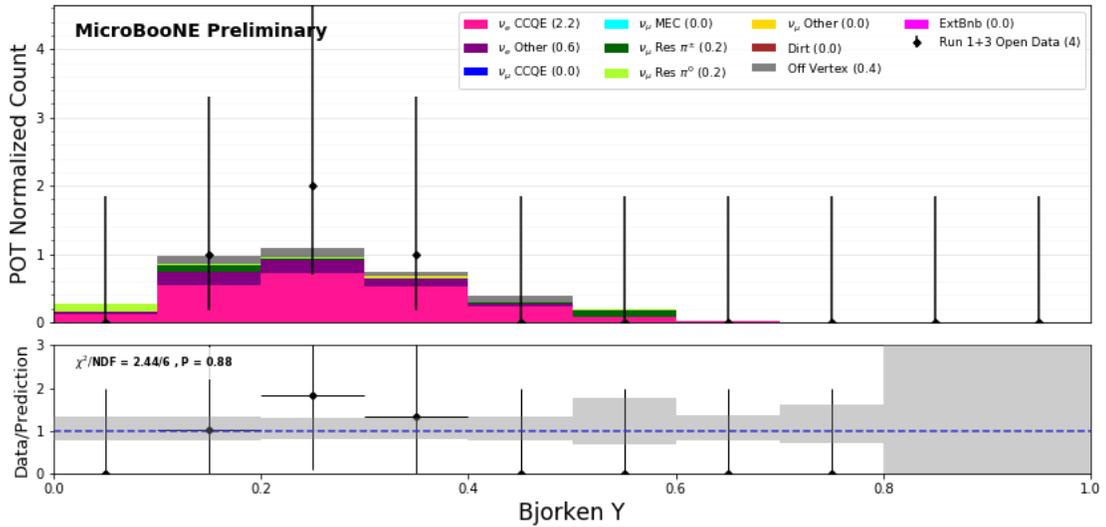


Figure 28: Stacked comparisons between $1e1p$ prediction and data for Bjorken Y.

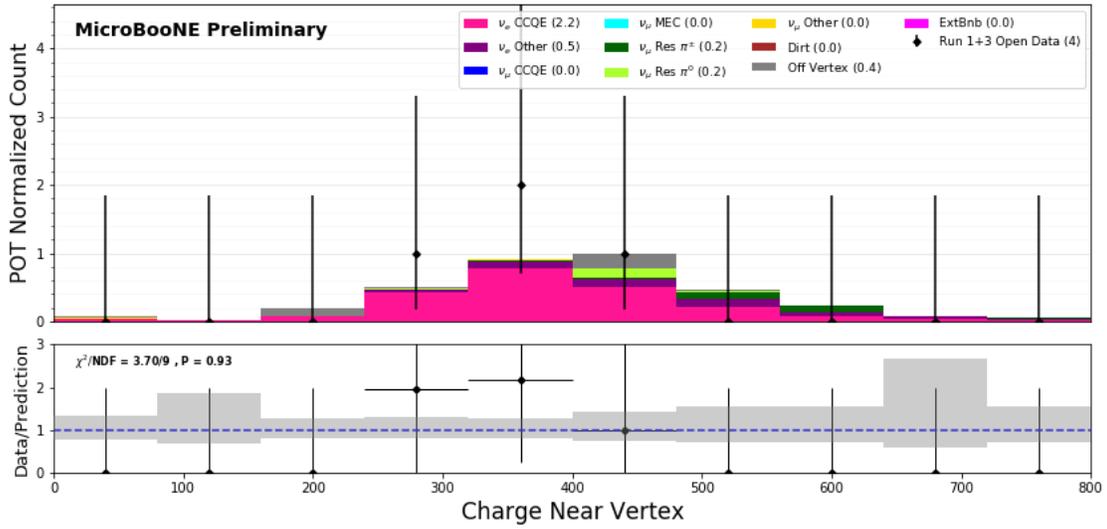


Figure 29: Stacked comparisons between $1e1p$ prediction and data for Charge Near Vertex.

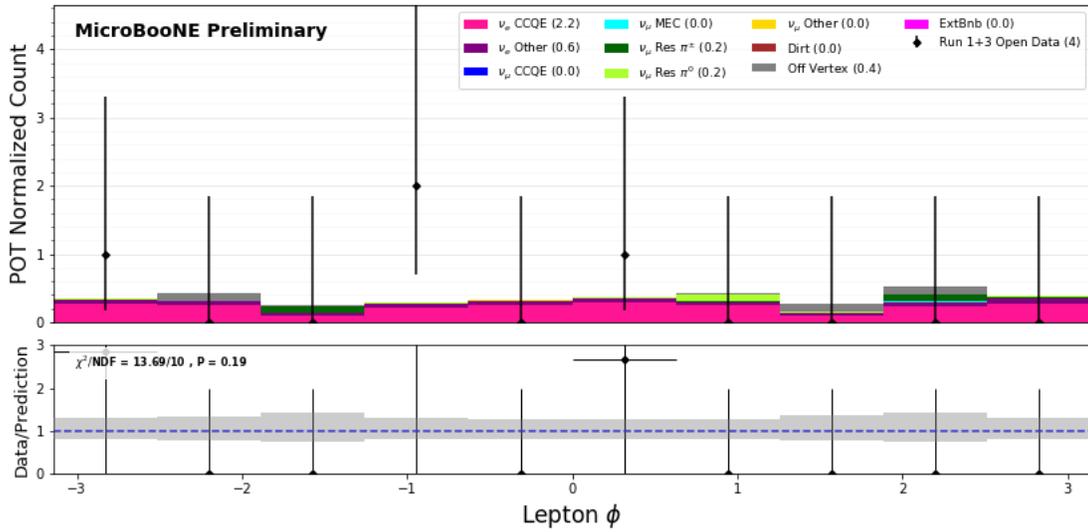


Figure 30: Stacked comparisons between $1e1p$ prediction and data for Electron ϕ .

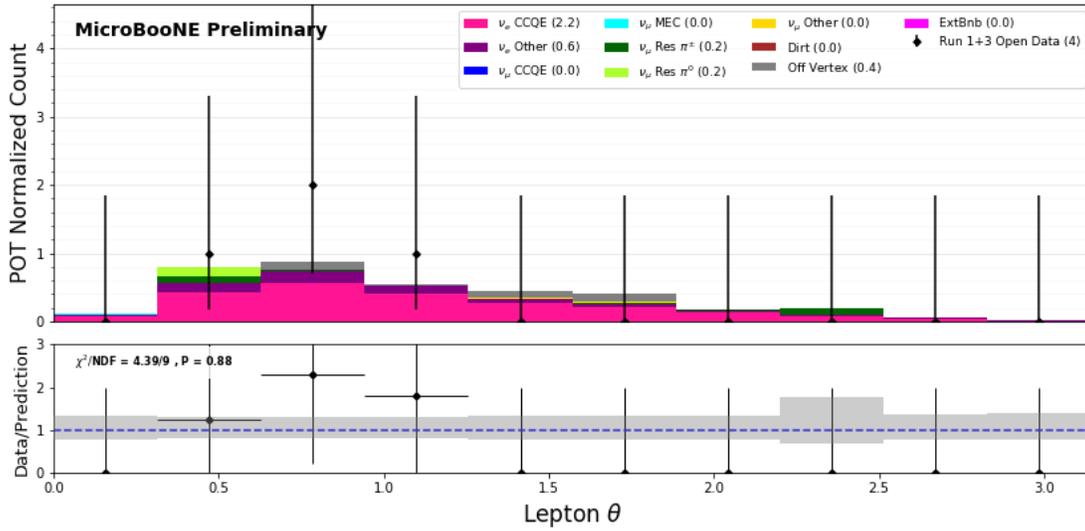


Figure 31: Stacked comparisons between $1e1p$ prediction and data for Electron θ .

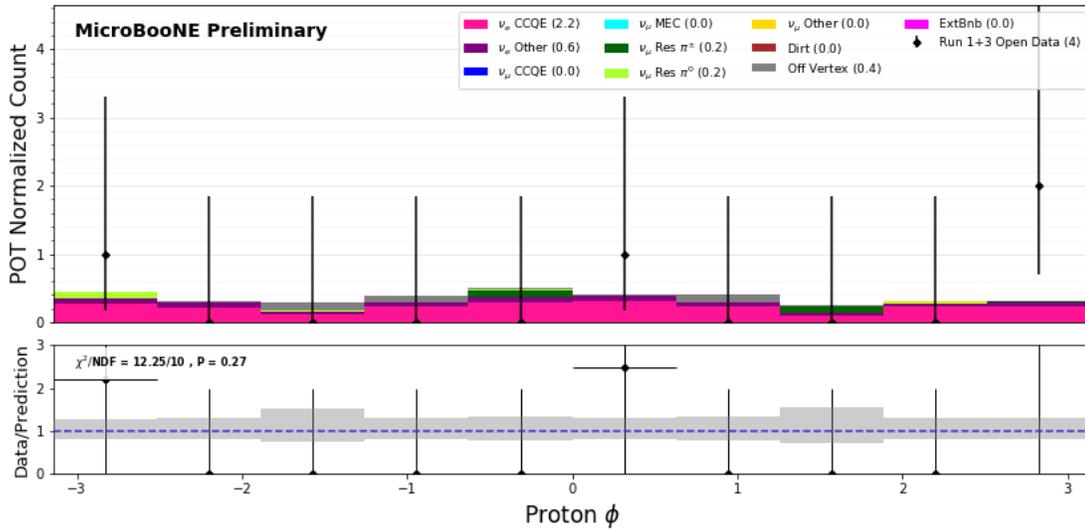


Figure 32: Stacked comparisons between $1e1p$ prediction and data for Proton ϕ .

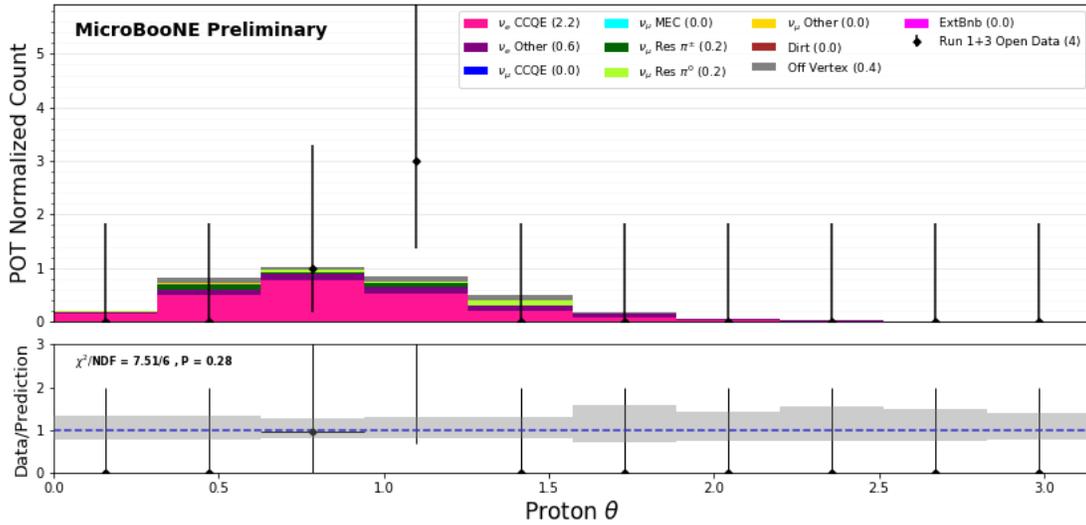


Figure 33: Stacked comparisons between $1e1p$ prediction and data for Proton θ .

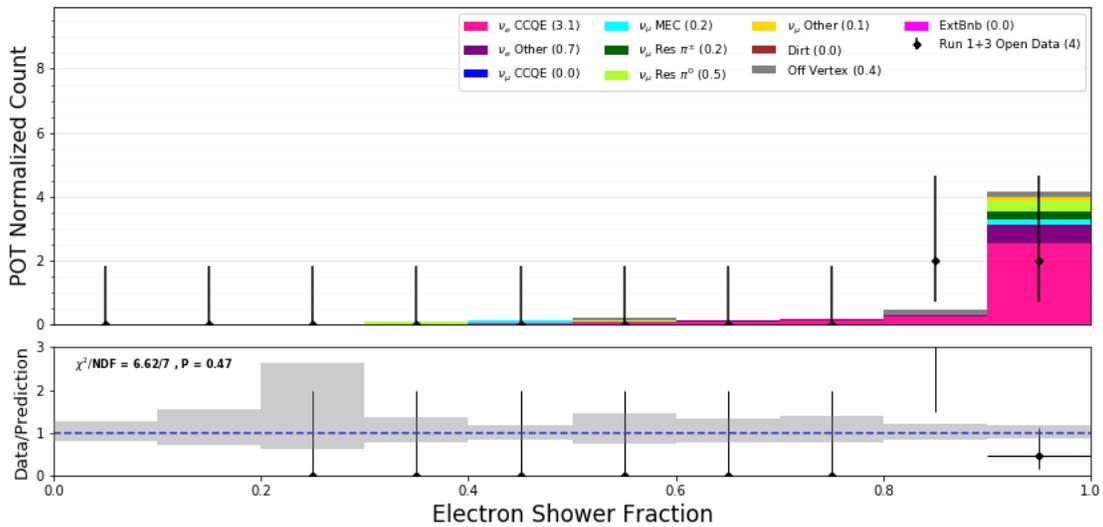


Figure 34: Stacked comparisons between $1e1p$ prediction and data for electron SSNet shower fraction.

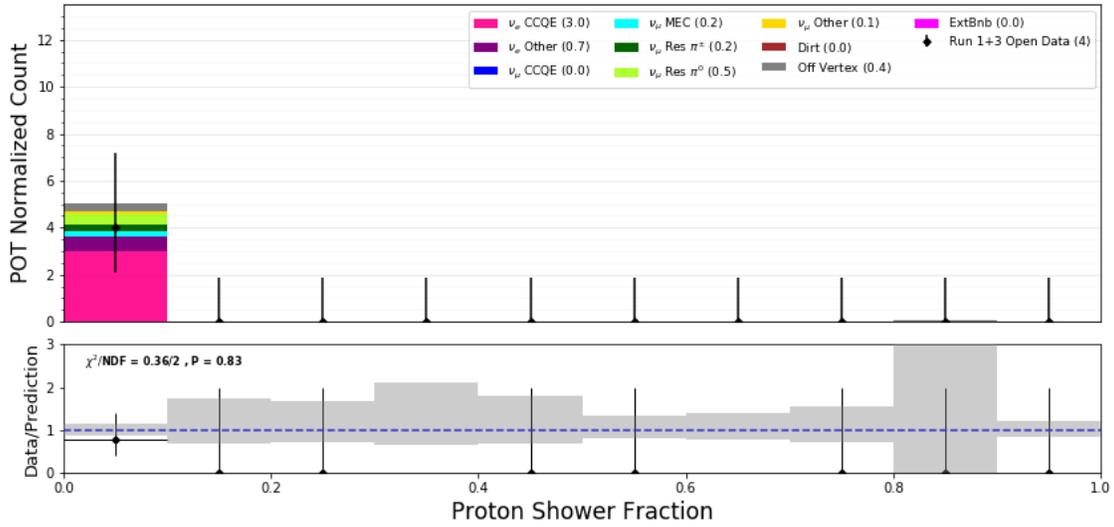


Figure 35: Stacked comparisons between $1e1p$ prediction and data for proton SSNet shower fraction.

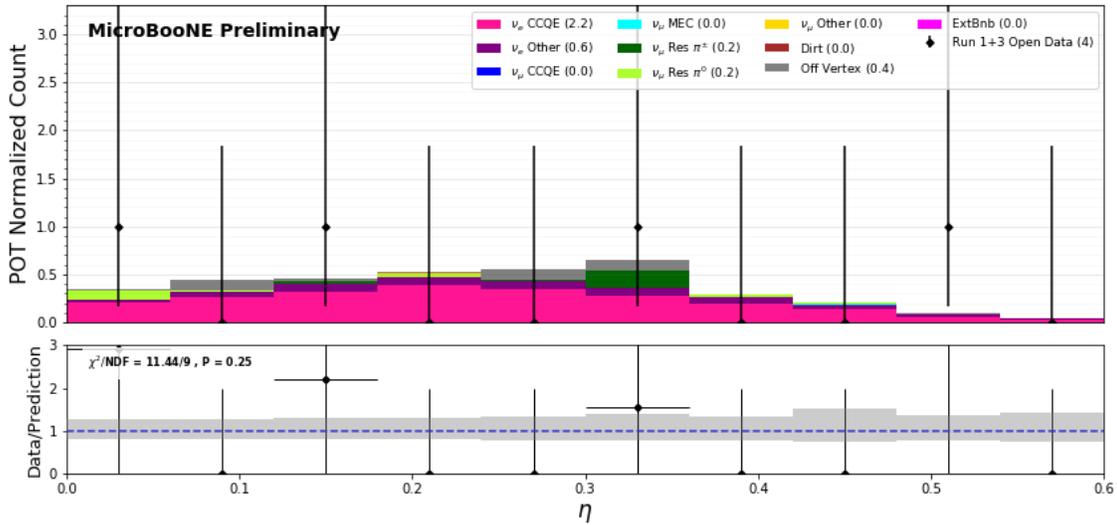


Figure 36: Stacked comparisons between $1e1p$ prediction and data for η .

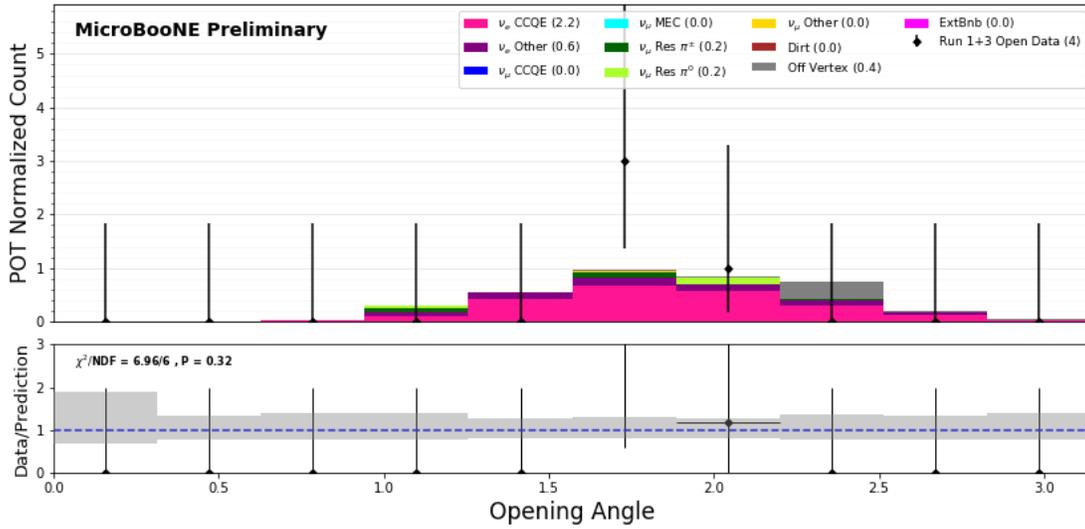


Figure 37: Stacked comparisons between $1e1p$ prediction and data for opening angle.

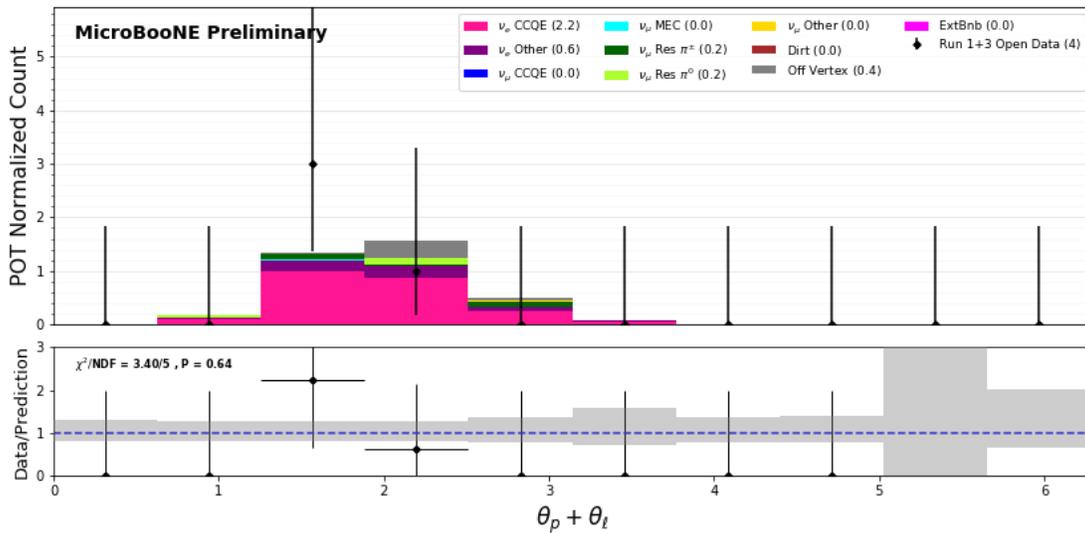


Figure 38: Stacked comparisons between $1e1p$ prediction and data for $\theta_p + \theta_e$.

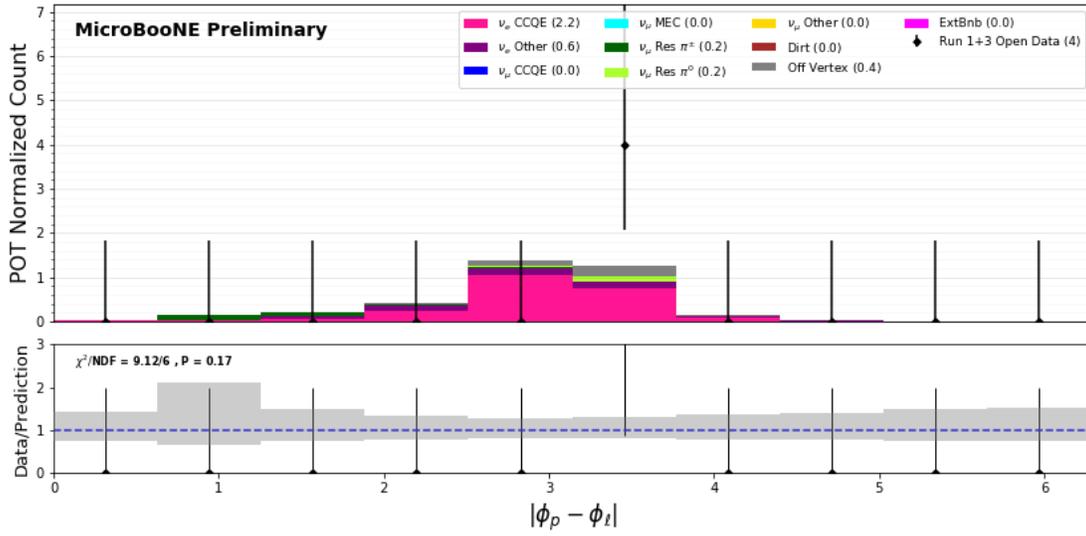


Figure 39: Stacked comparisons between $1e1p$ prediction and data for $|\phi_p - \phi_e$.

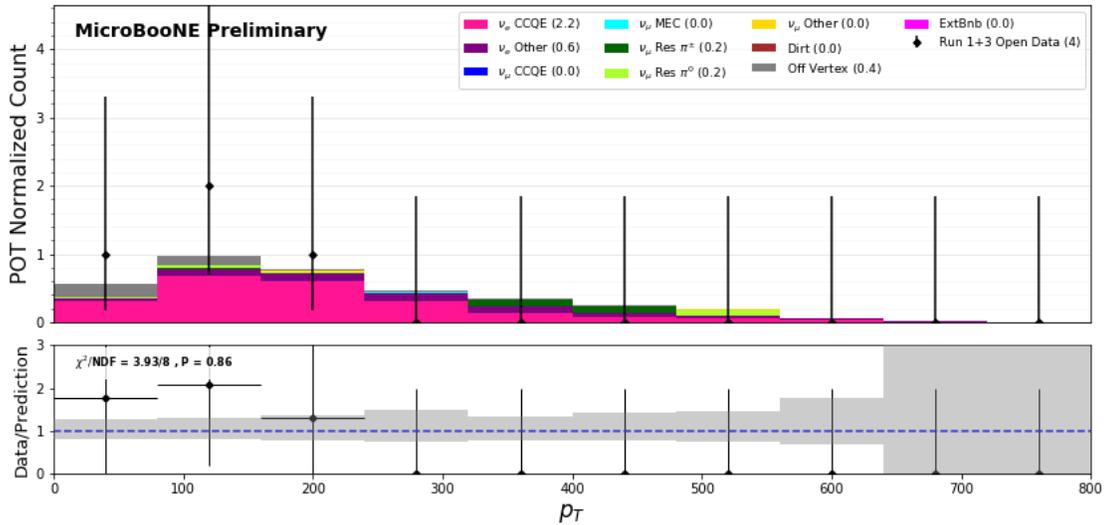


Figure 40: Stacked comparisons between $1e1p$ prediction and data for p_T .

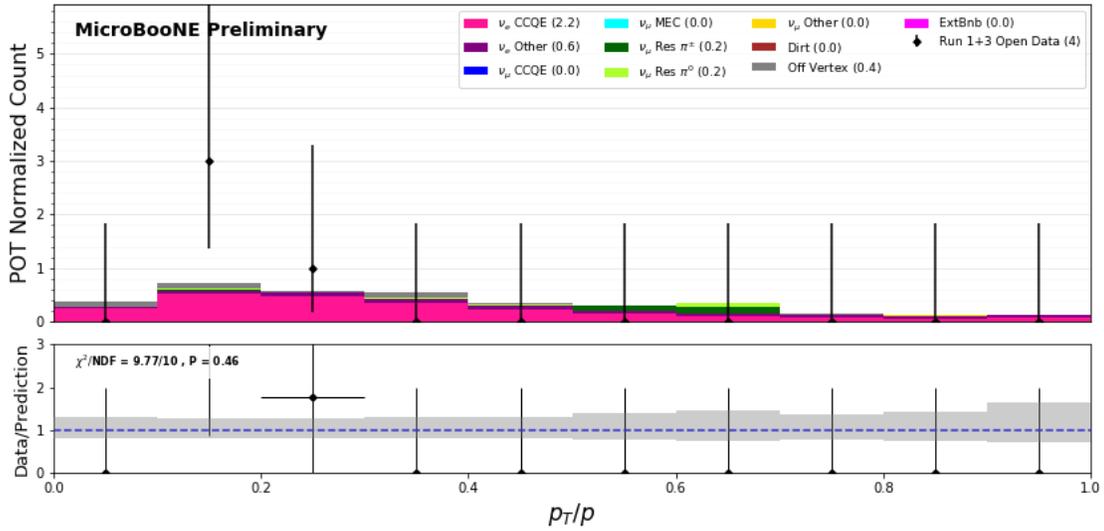


Figure 41: Stacked comparisons between $1e1p$ prediction and data for p_T/p .

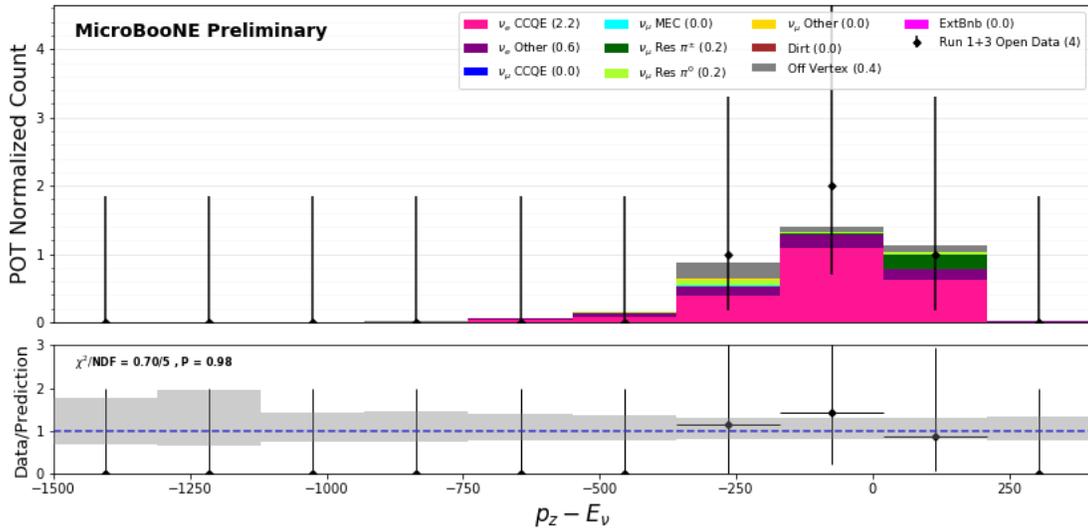


Figure 42: Stacked comparisons between prediction and data for $p_z - E_\nu$.

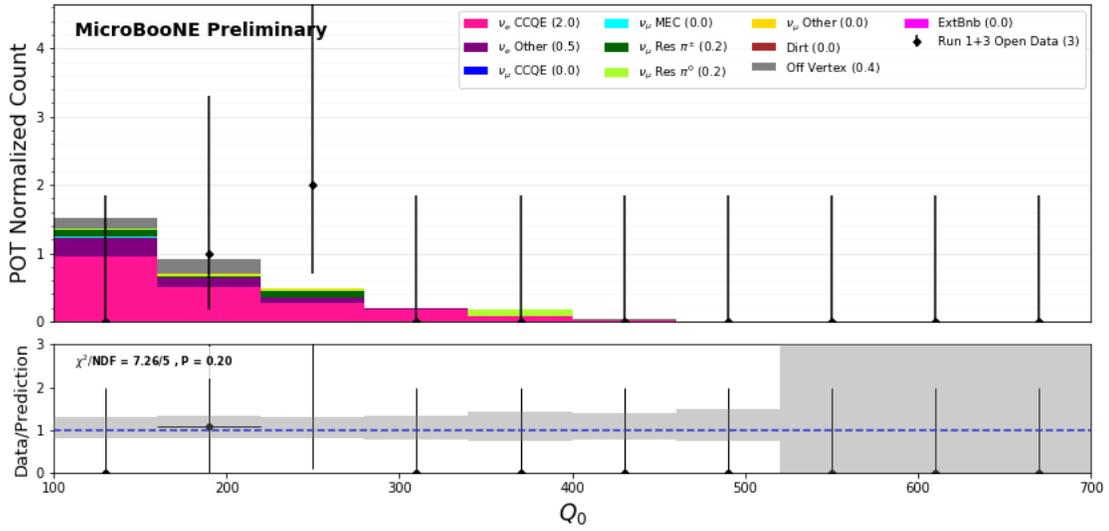


Figure 43: Stacked comparisons between $1e1p$ prediction and data for Q_0 .

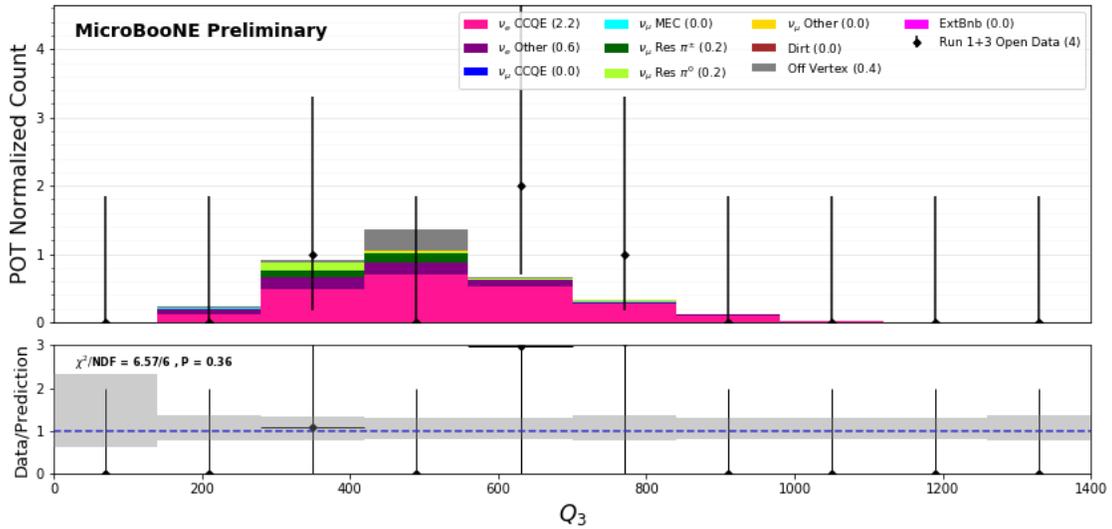


Figure 44: Stacked comparisons between $1e1p$ prediction and data for Q_3 .

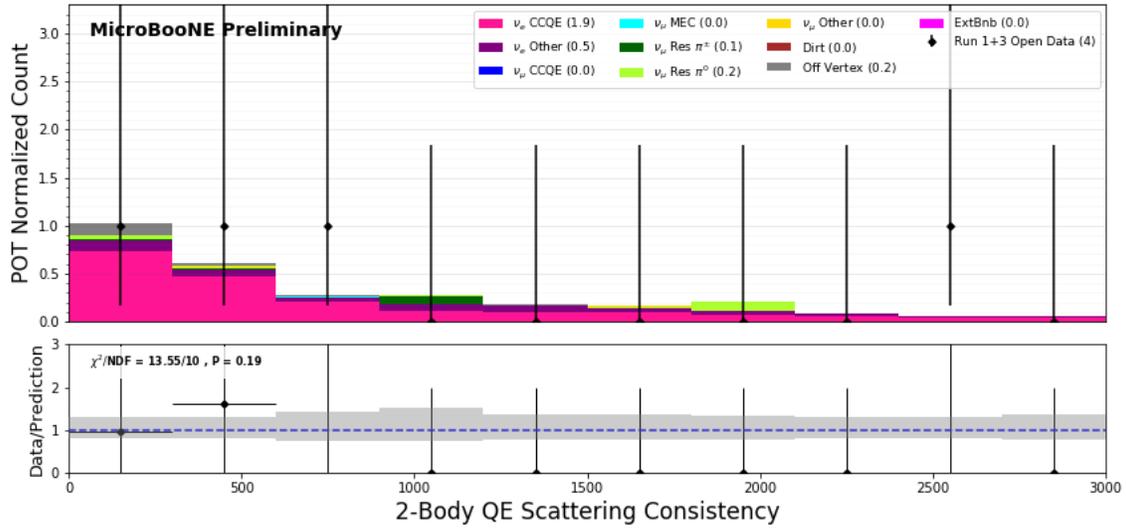


Figure 45: Stacked comparisons between prediction and data for 2-body elastic scattering consistency.

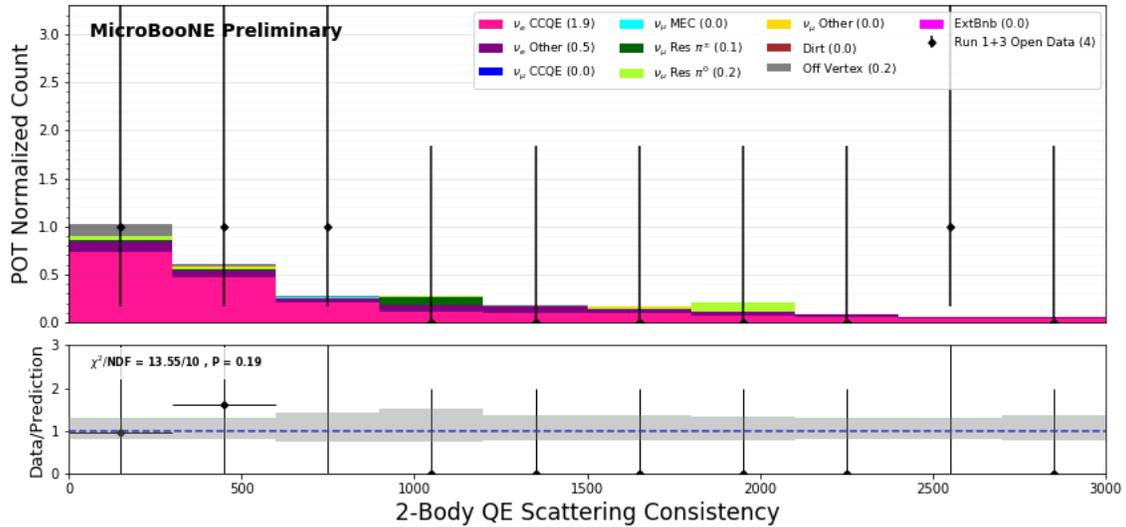


Figure 46: Stacked comparisons between $1e1p$ prediction and data for 2-body elastic scattering consistency.

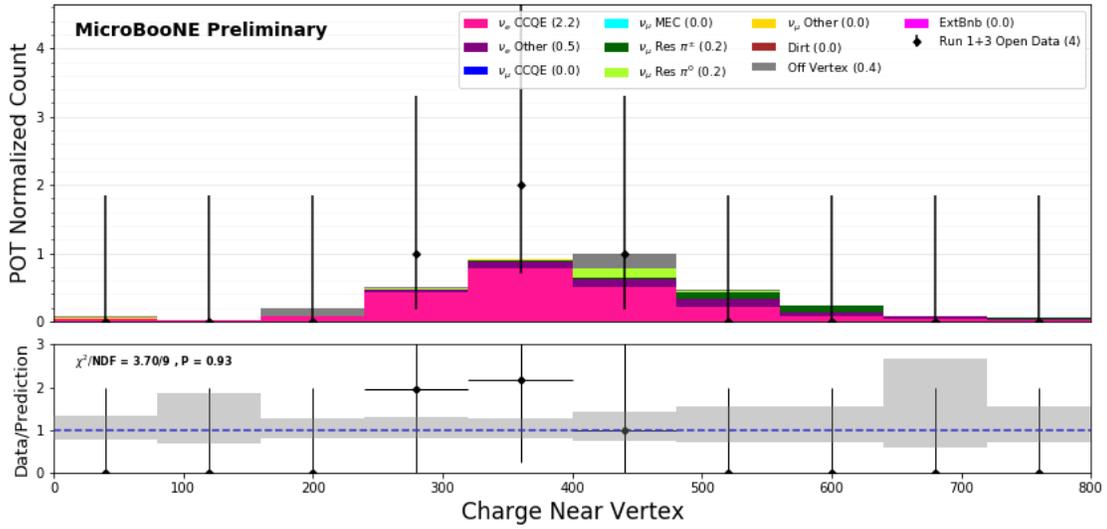


Figure 47: Stacked comparisons between $1e1p$ prediction and data for total ADC charge near the vertex.

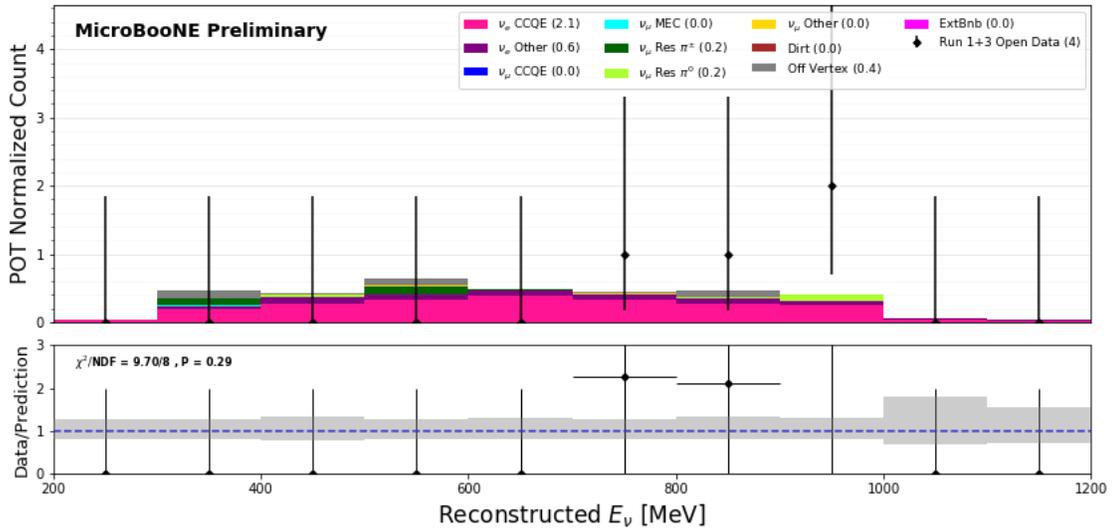


Figure 48: Stacked comparisons between $1e1p$ prediction and data for reconstructed neutrino energy.

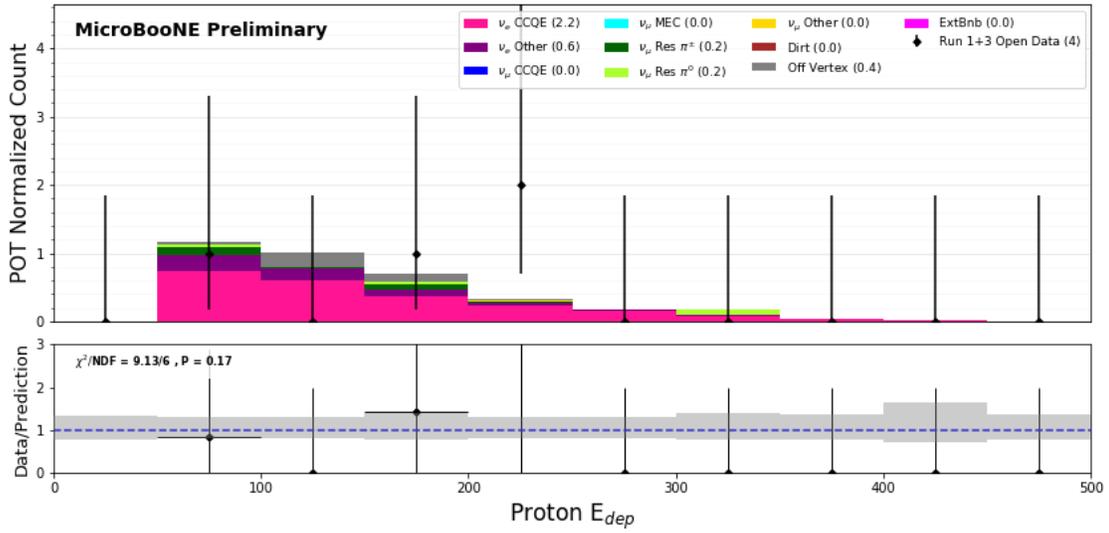


Figure 49: Stacked comparisons between $1e1p$ prediction and data for reconstructed proton energy.

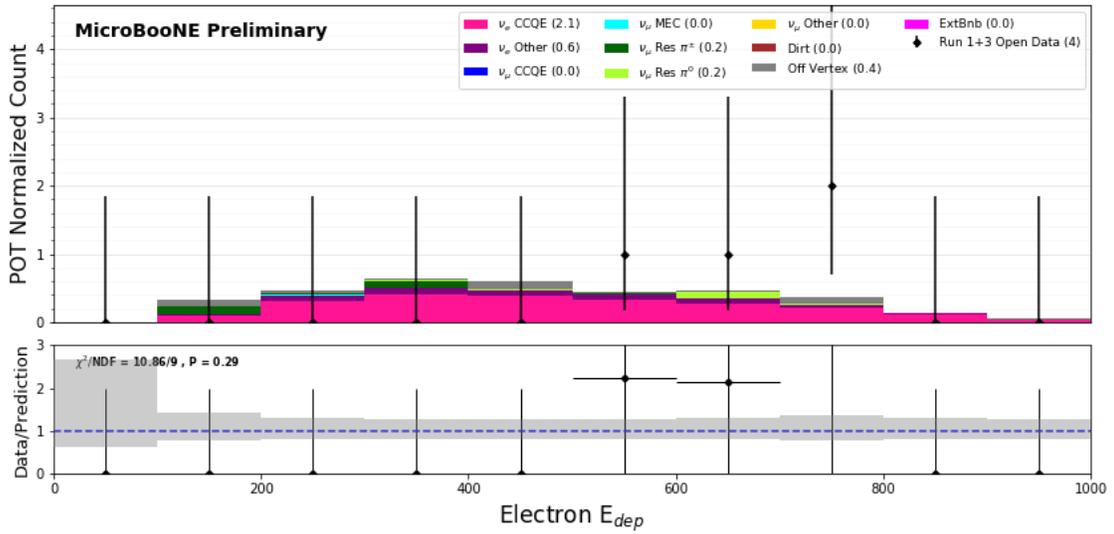


Figure 50: Stacked comparisons between $1e1p$ prediction and data for reconstructed electron energy.

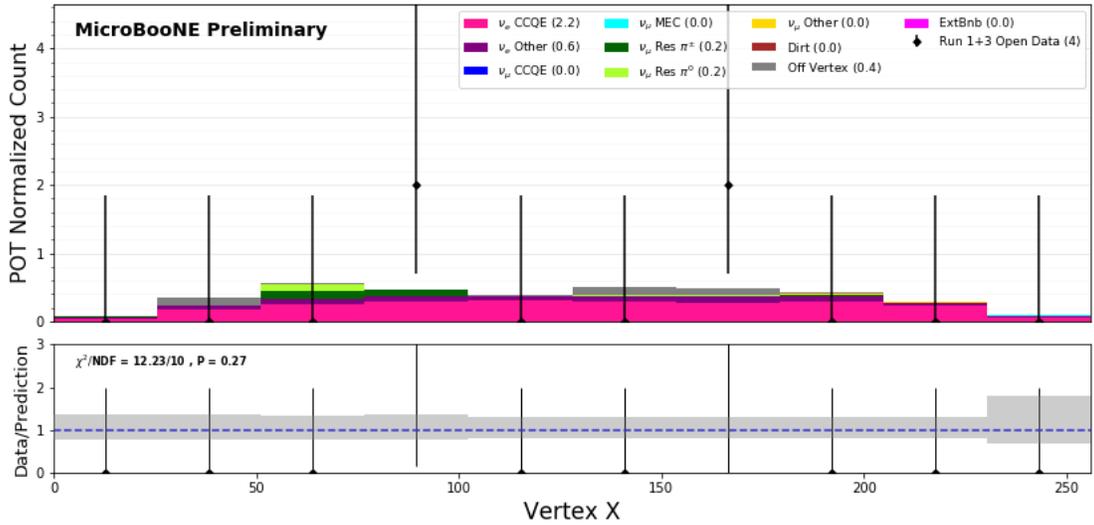


Figure 51: Stacked comparisons between $1e1p$ prediction and data for reconstructed vertex X position.

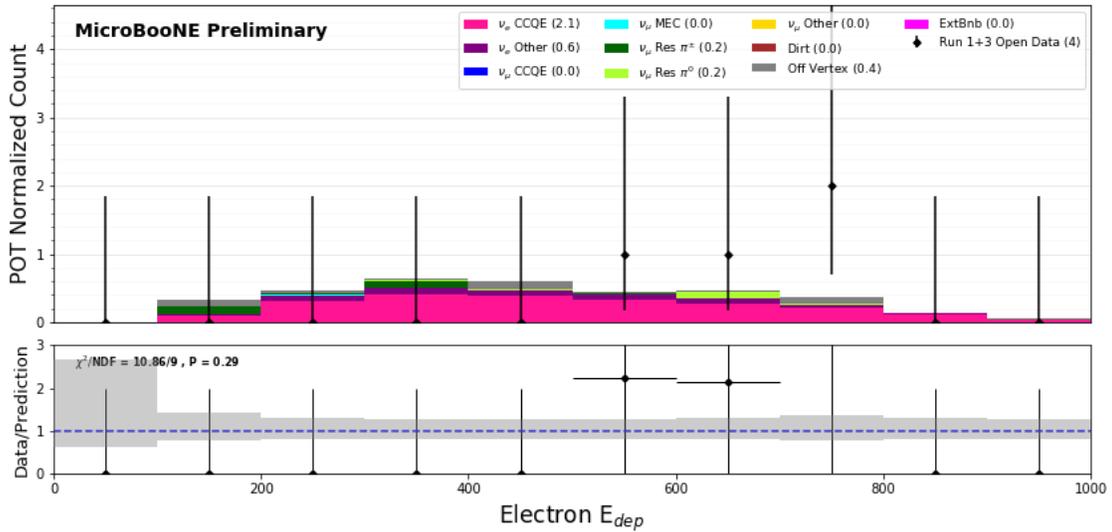


Figure 52: Stacked comparisons between $1e1p$ prediction and data for reconstructed vertex Y position.

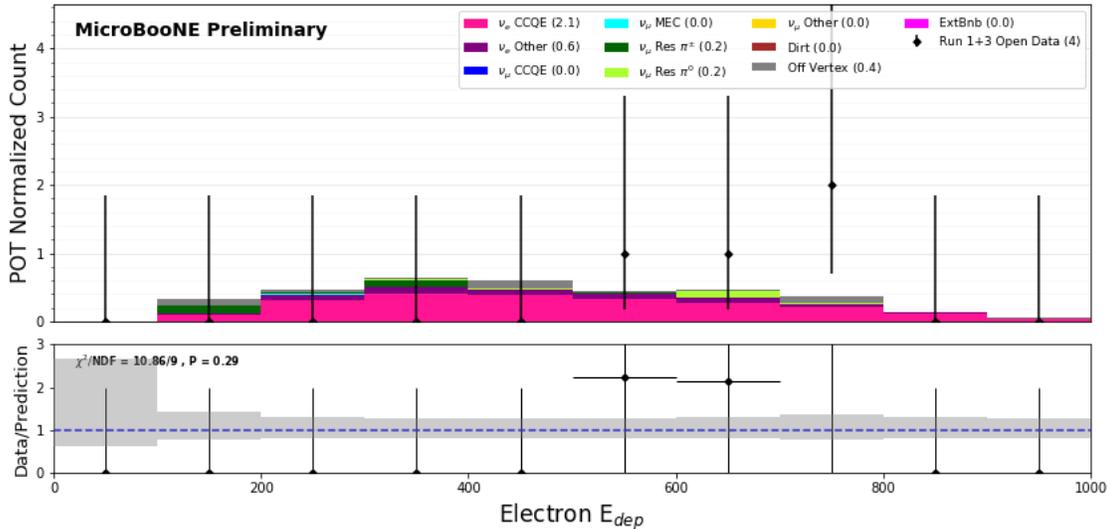


Figure 53: Stacked comparisons between $1e1p$ prediction and data for reconstructed vertex Z position.

B.2 Run 1+2+3 Blindness-Safe Plots

These plots reflect comparisons for different variables between a stacked MC predicted spectrum and the observed spectrum from 38 selected events drawn from runs 1,2, & 3. The stacked prediction is normalized such that the number of events matches the number of data events. In certain situations there will be fewer than 38 events in a given histogram. This occurs when the value is outside of the blind histogram range. This may occur either because the variable has a long tail, or because the variable is not defined in all situations. This represents the complete collection of histograms used to check agreement, and includes some plots that are also in the main text, above. The illustrated bands correspond to flux, cross section, and detector systematic uncertainties. Bins which contain no simulated prediction may be illustrated, but are omitted from the χ^2 computation.

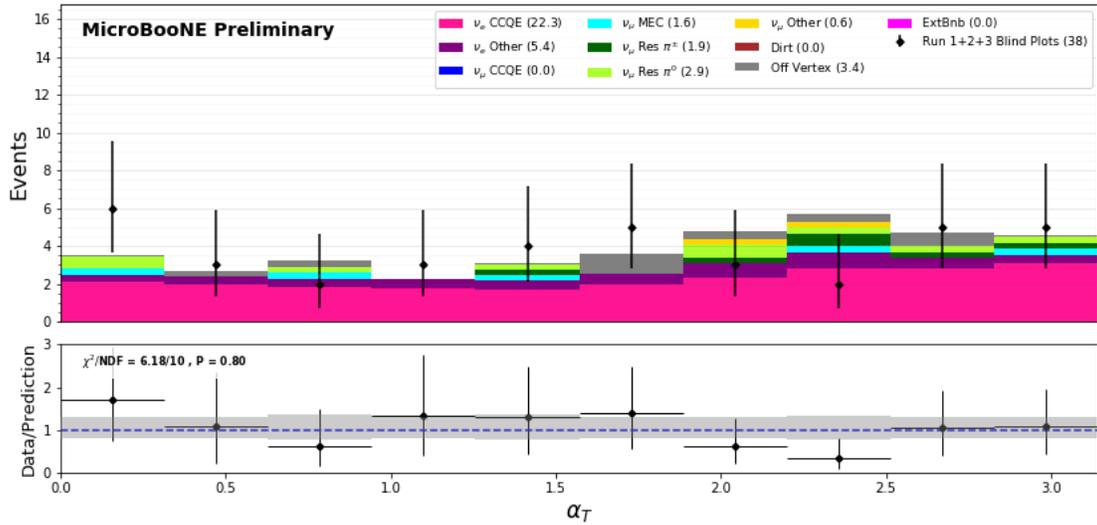


Figure 54: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : α_T .

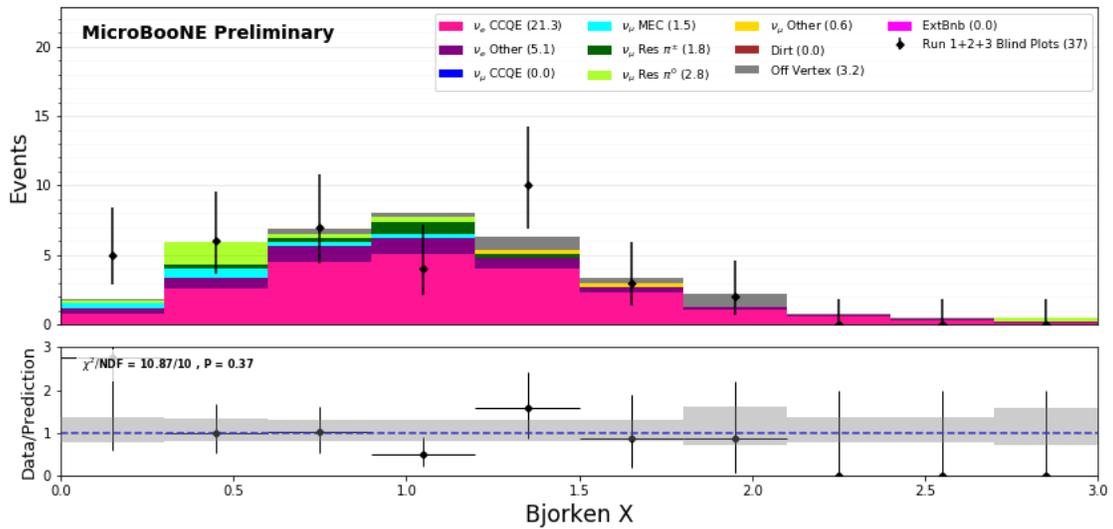


Figure 55: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Bjorken X

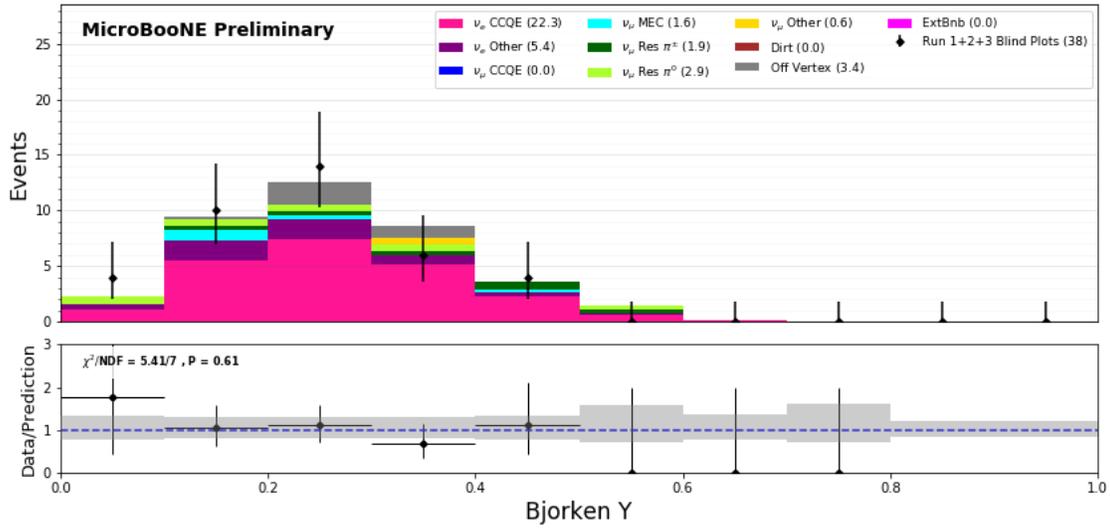


Figure 56: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Bjorken Y

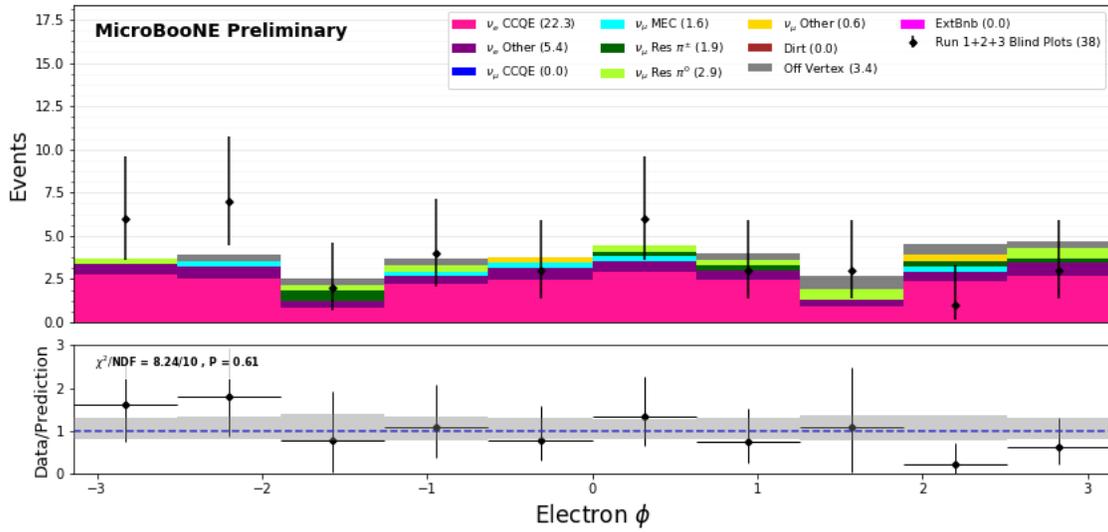


Figure 57: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Electron ϕ

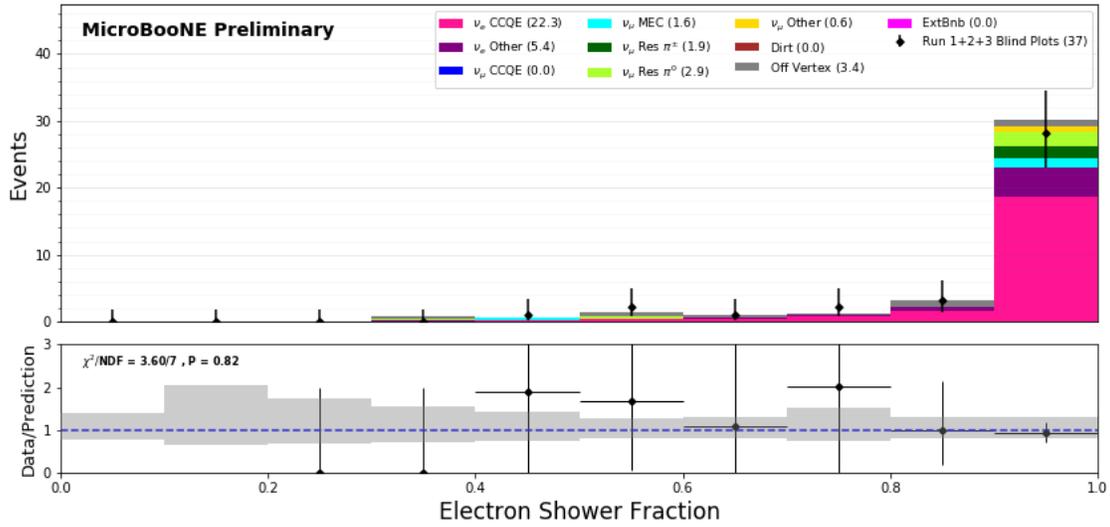


Figure 58: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Electron Shower Fraction

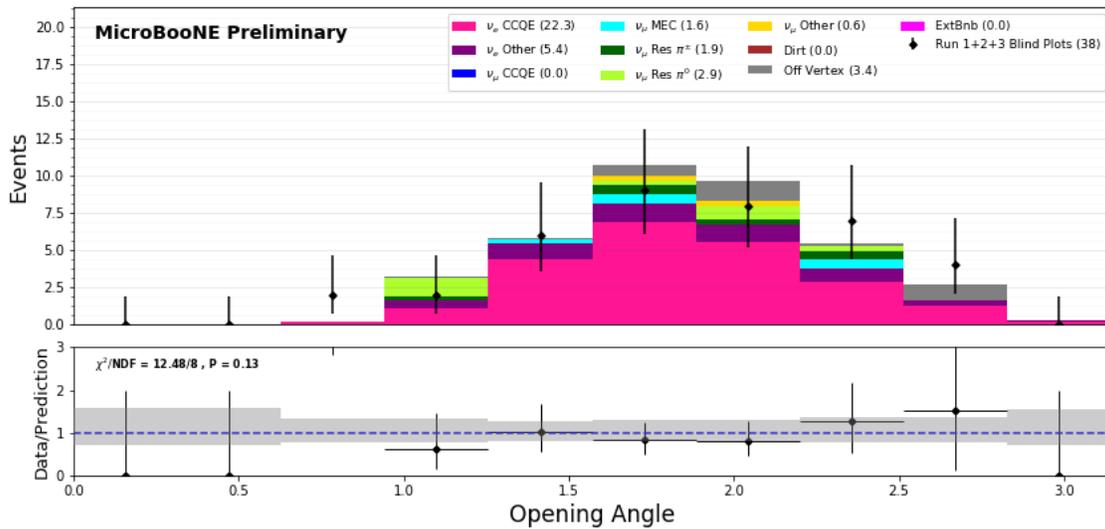


Figure 59: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : 3D Opening Angle

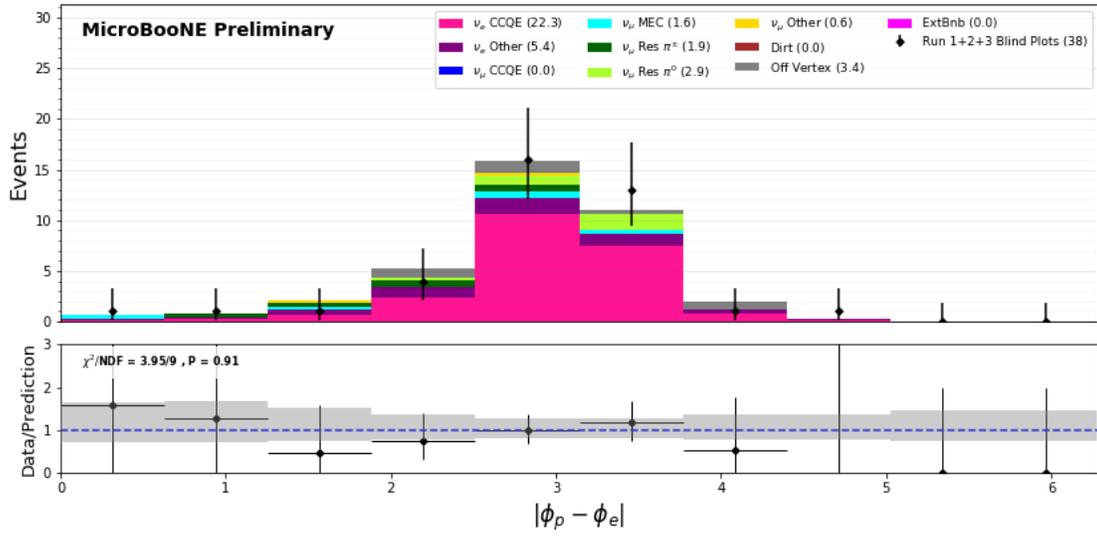


Figure 60: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Absolute difference in ϕ , $-\phi_p - \phi_e$

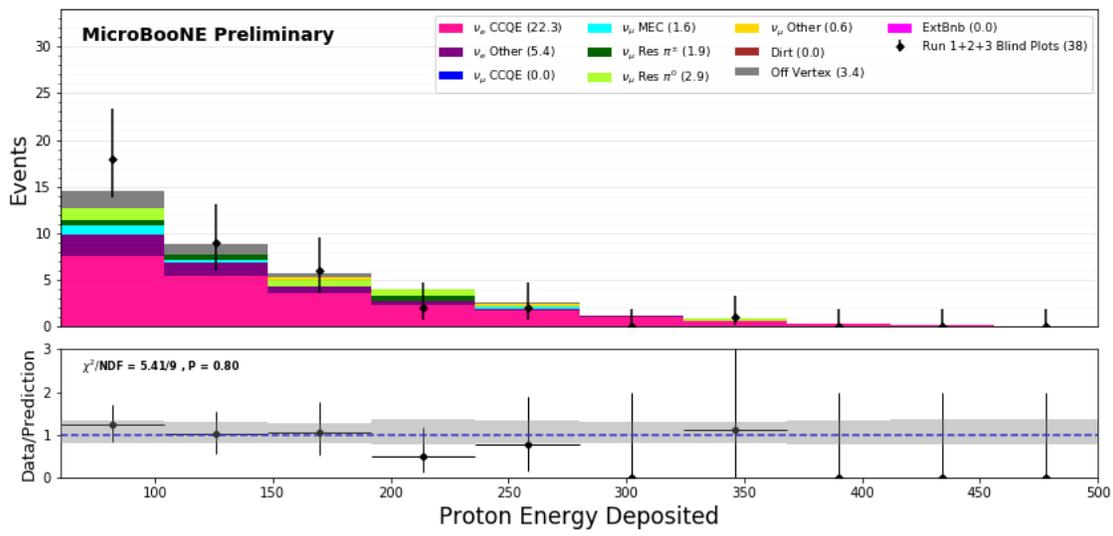


Figure 61: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Proton deposited energy

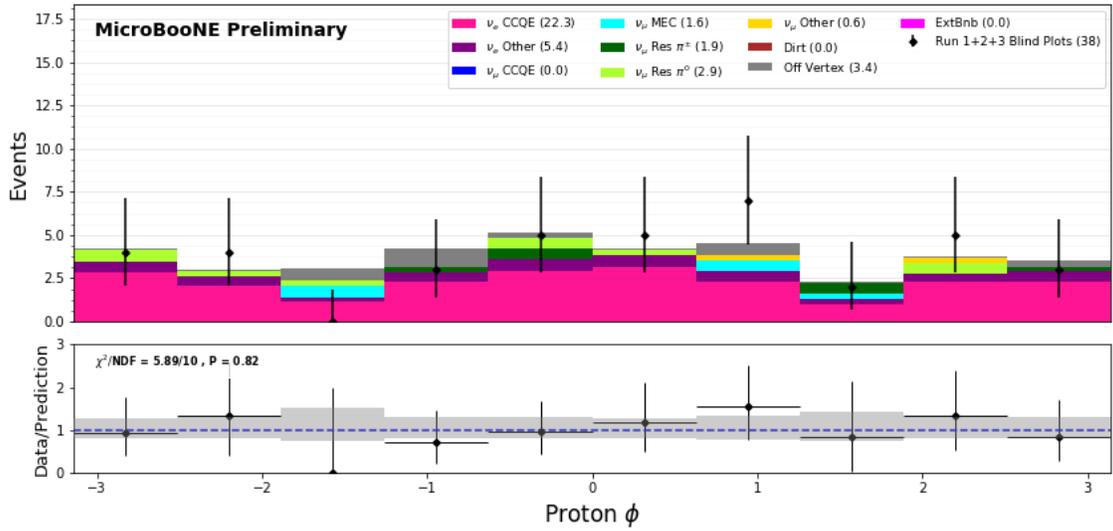


Figure 62: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Proton ϕ

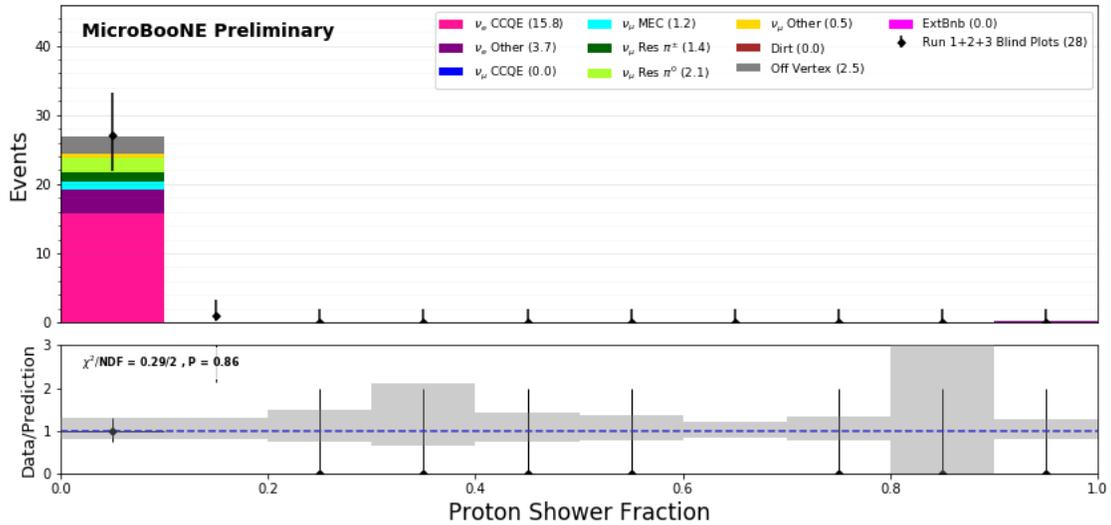


Figure 63: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Proton Shower Fraction

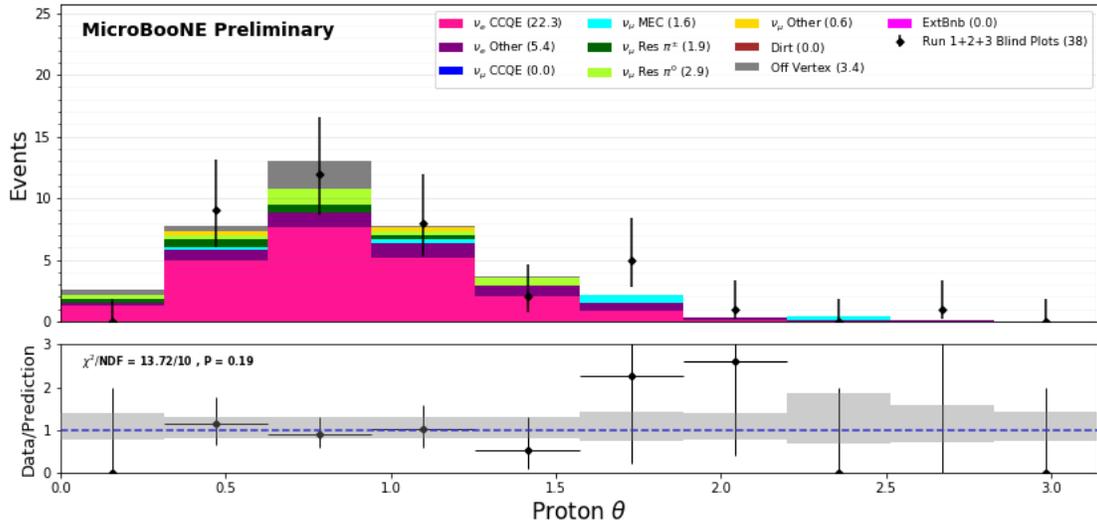


Figure 64: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Proton θ

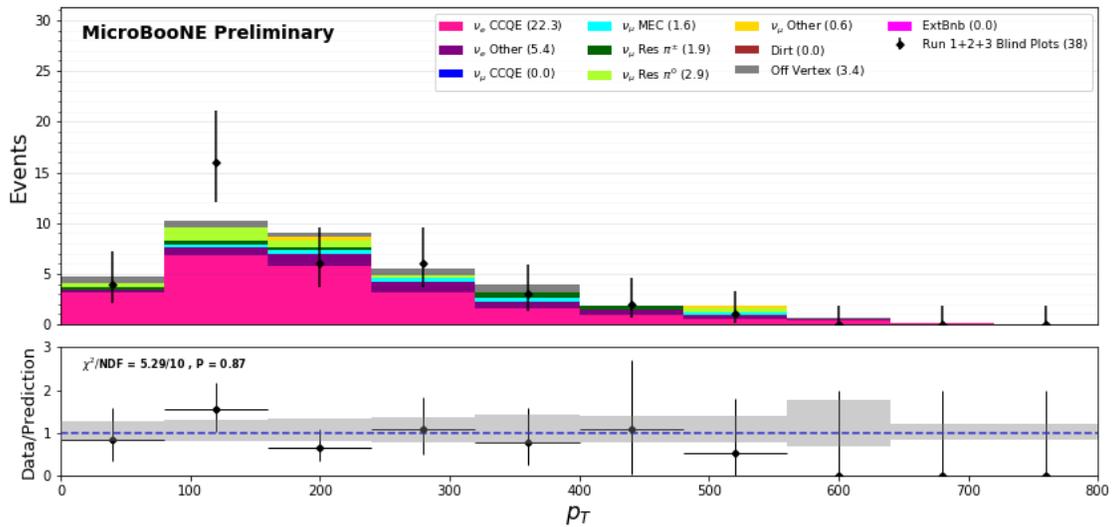


Figure 65: Blindness-safe plot ($1e1p$) from Run 1 + Run 3 : Final state transverse momentum of the event

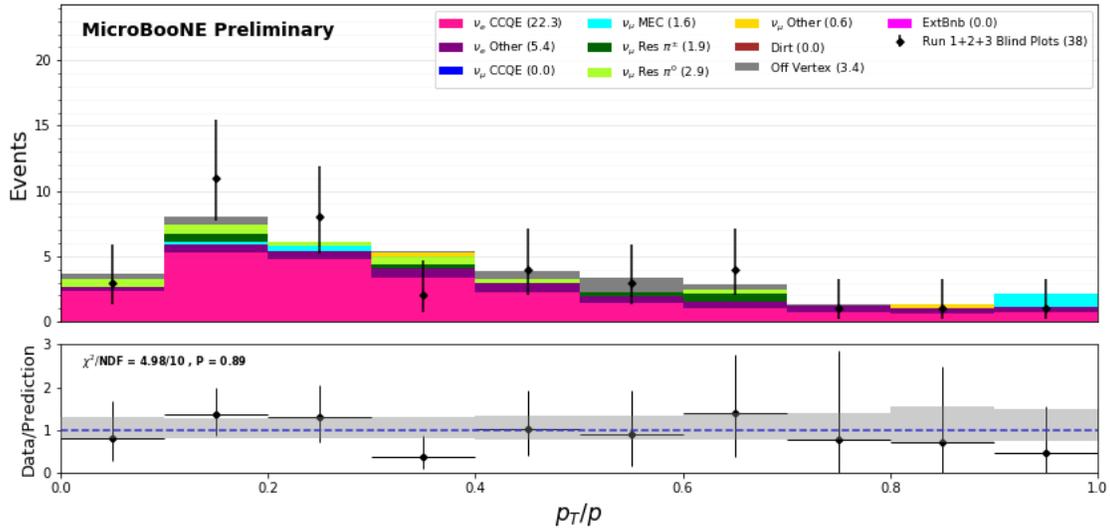


Figure 66: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Final state transverse momentum of the event divided by total momentum

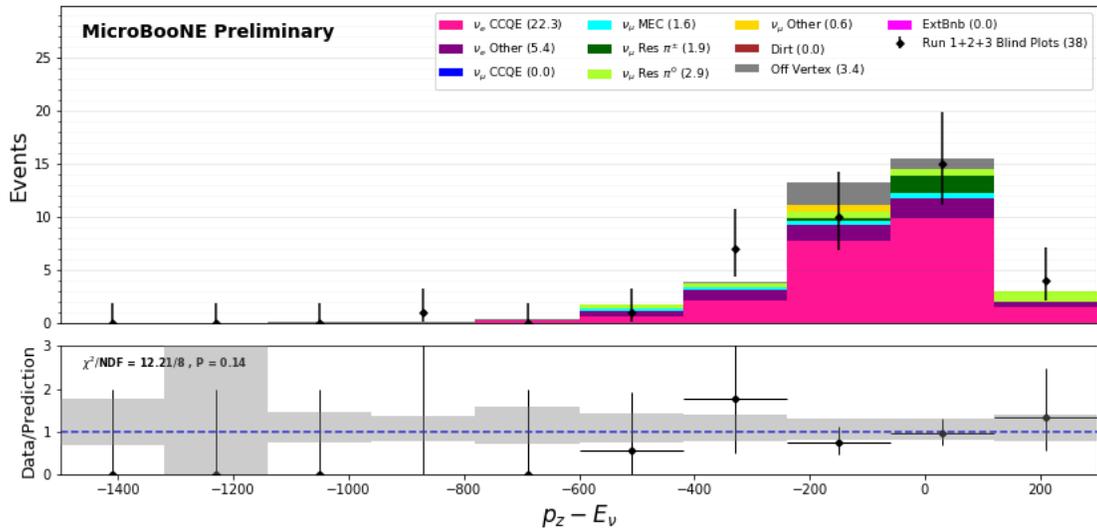


Figure 67: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : $p_z - E_\nu$

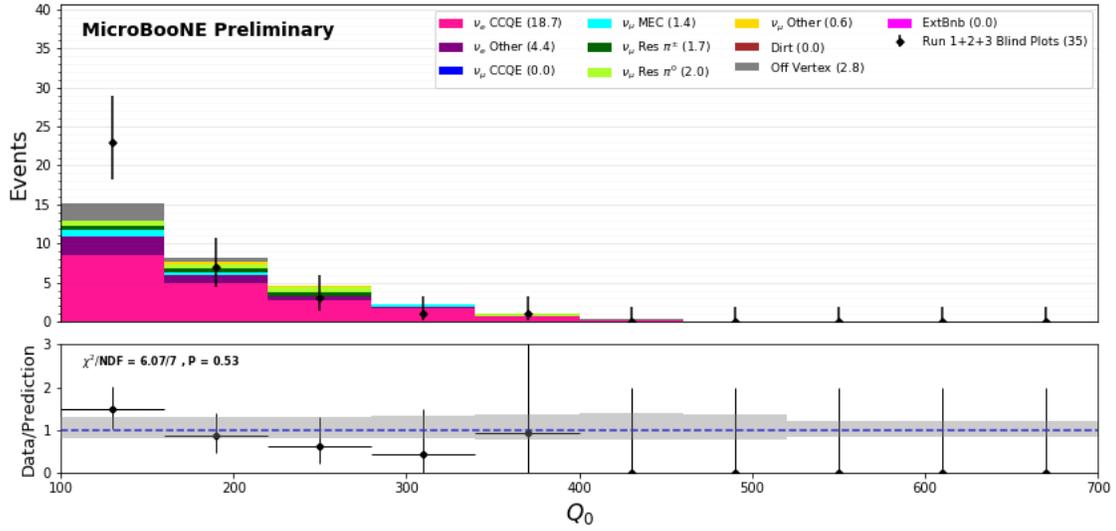


Figure 68: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Q_0

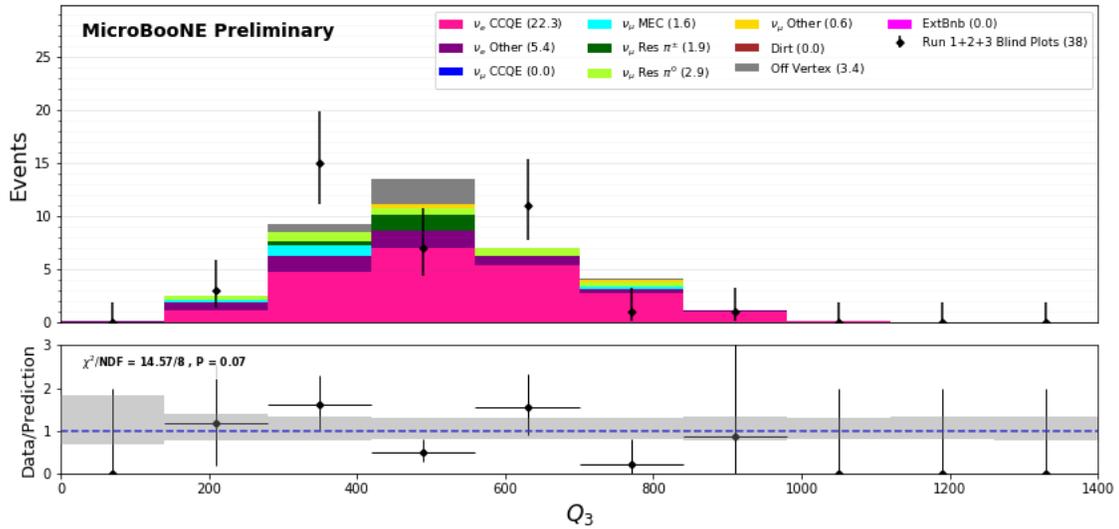


Figure 69: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Q_3

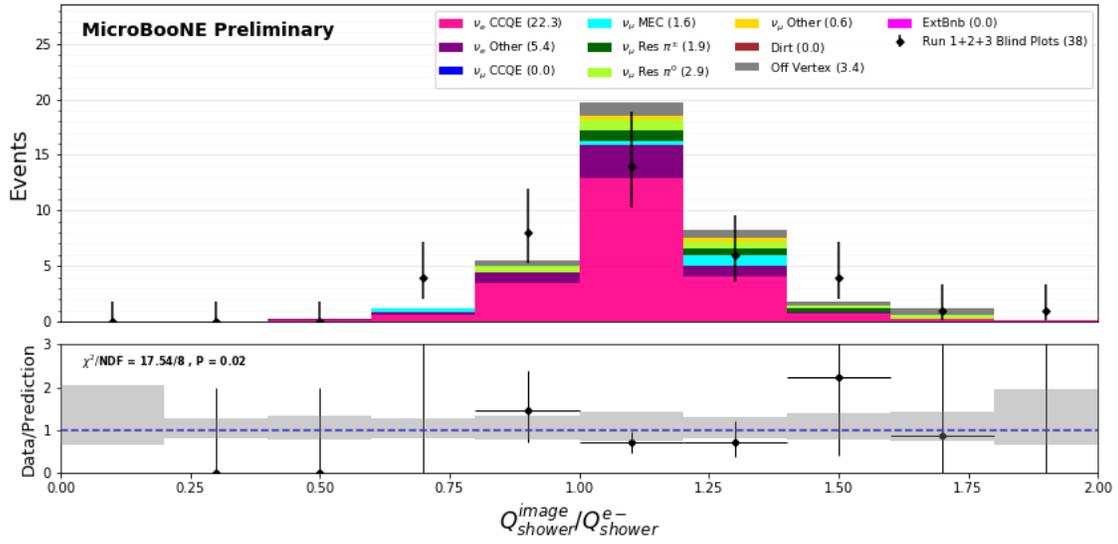


Figure 70: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Shower charge in image divided by shower charge in electron cluster

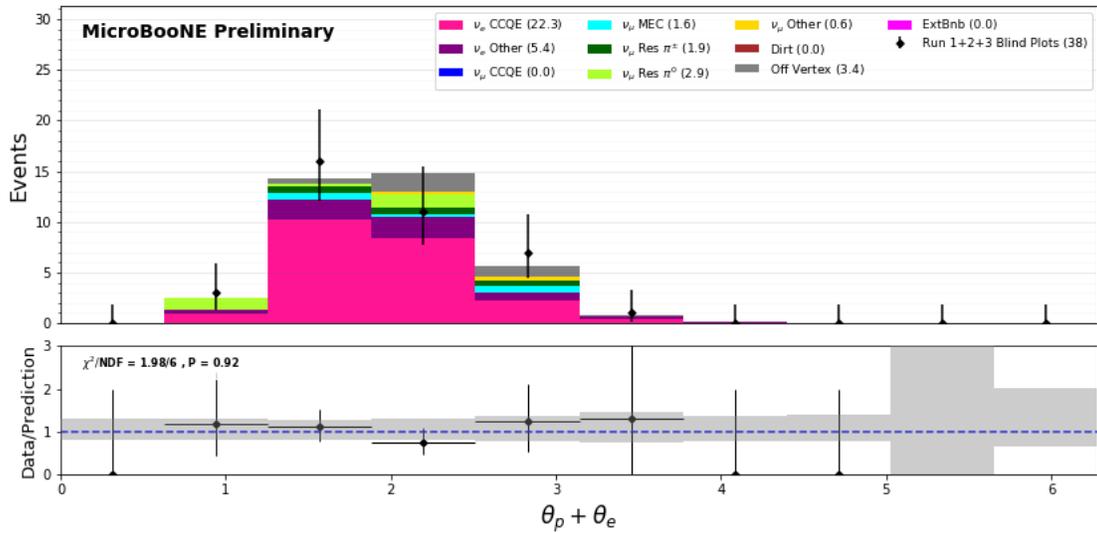


Figure 71: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : $\theta_p + \theta_e$

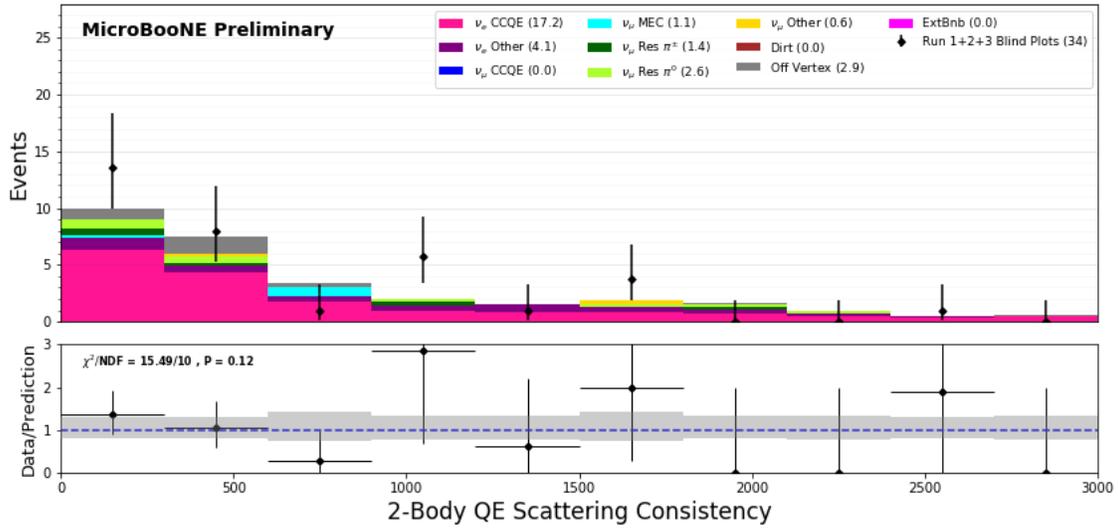


Figure 72: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Consistency between 3 QE reconstruction formulae

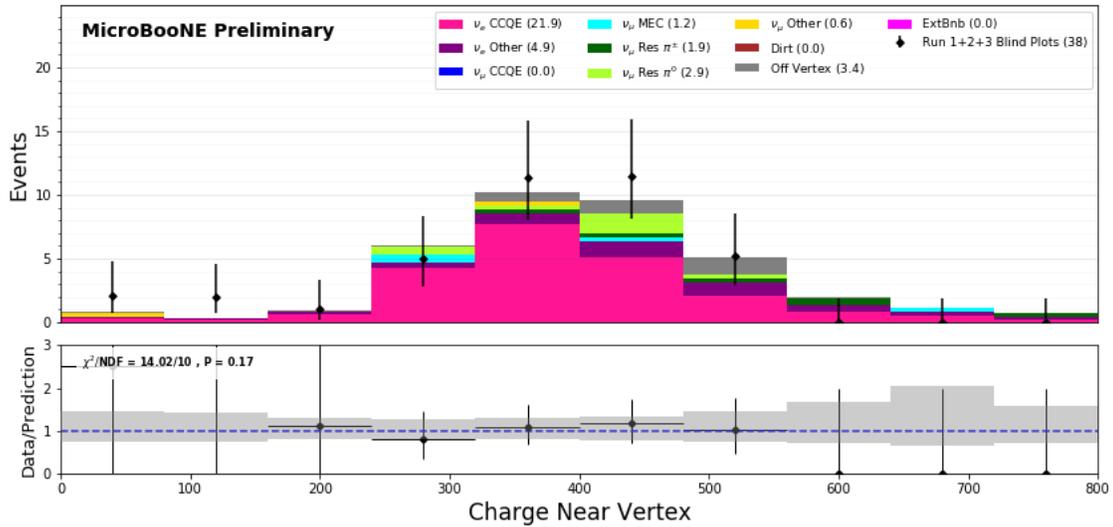


Figure 73: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Charge near vertex

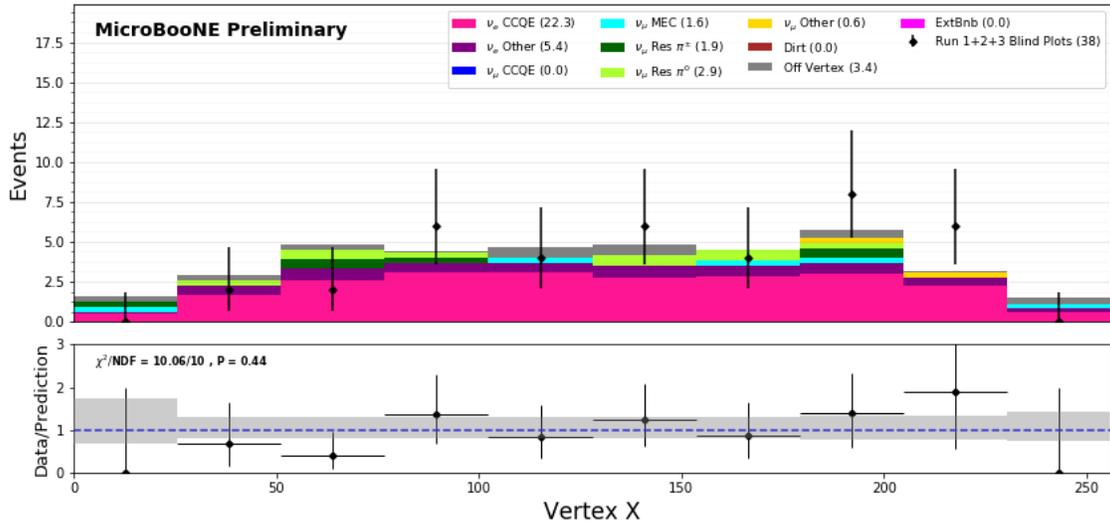


Figure 74: Blindness-safe plot from Run 1 + 2 + 3 : Vertex X

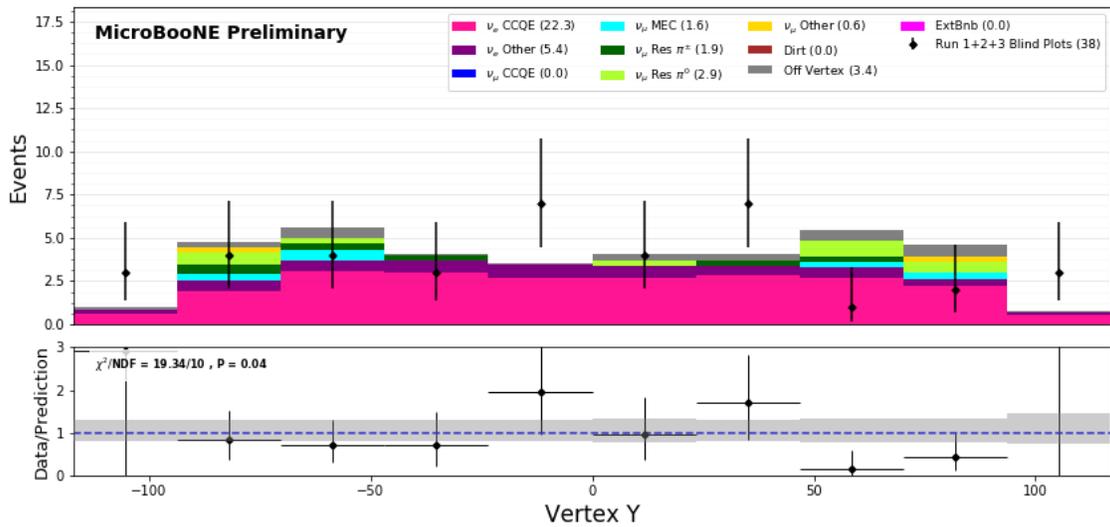


Figure 75: Blindness-safe plot ($1e1p$) from Run 1 + 2 + 3 : Vertex Y

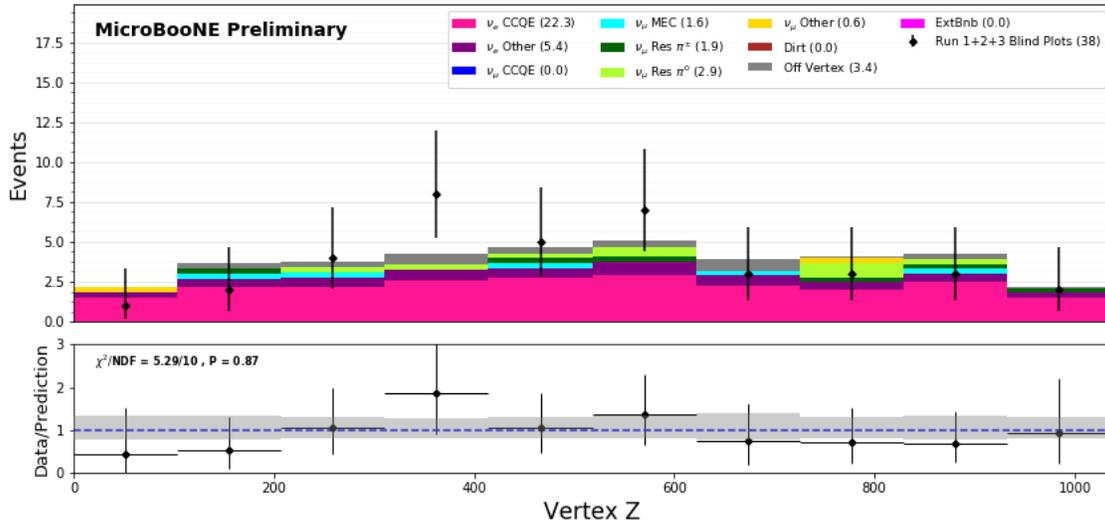


Figure 76: Blindness-safe plot (1e1p) from Run 1 + 2 + 3 : Vertex Z

References

- [1] M. Martini and M. Ericson, *Quasielastic and multinucleon excitations in antineutrino-nucleus interactions* Phys. Rev. C 87, 065501
- [2] MicroBooNE Collaboration, *Design and construction of the MicroBooNE detector*, JINST, 12,02, P02017–P02017, 2017
- [3] MicroBooNE Collaboration, “Neutrino Event Selection in the MicroBooNE Liquid Argon Time Projection Chamber using Wire-Cell 3-D Imaging, Clustering and Charge-Light Matching,” <https://microboone-docdb.fnal.gov/cgi-bin/private/ShowDocument?docid=28632>
- [4] C. Adams et al., *A Deep Neural Network for Pixel-Level Electromagnetic Particle Identification in the MicroBooNE Liquid Argon Time Projection Chamber*, arXiv:1808.07269
- [5] P. Abratenko et al., MicroBooNE Collaboration, *Vertex-Finding and Reconstruction of Contained Two-track Neutrino Events in the MicroBooNE Detector*, arXiv:2002.09375, <https://arxiv.org/abs/2002.09375>
- [6] MicroBooNE Collaboration, *Semantic Segmentation with Sparse Convolutional Neural Network for Event Reconstruction in MicroBooNE*, MICROBOONE-NOTE-1091-PUB
- [7] MicroBooNE Collaboration, *MicroBooNE low-energy excess signal prediction from unfolding MiniBooNE Monte-Carlo and data*, MICROBOONE-NOTE-1043-PUB, <https://microboone.fnal.gov/wp-content/uploads/MICROBOONE-NOTE-1043-PUB.pdf>

- [8] A. A. Aguilar-Arevalo et al., MiniBooNE Collaboration, *Unexplained Excess of Electronlike Events from a 1-GeV Neutrino Beam*, Phys. Rev. Lett. 102, 101802
- [9] MicroBooNE Collaboration, *Neutrino Interaction Model and Uncertainties for MicroBooNE Analyses*, MICROBOONE-NOTE-1074-PUB.
- [10] MicroBooNE Collaboration, *Novel Approach for Evaluating Detector Systematics in the MicroBooNE LArTPC*, MICROBOONE-NOTE-1075-PUB.