THE UNIVERSITY OF CHICAGO

MEASUREMENTS OF THE EXCLUSIVE DIFFERENTIAL CHARGED CURRENT CROSS SECTION FOR ELECTRON NEUTRINOS ON ARGON WITH THE NUMI DATASET IN MICROBOONE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
KATRINA RENEE MILLER

CHICAGO, ILLINOIS
JUNE 2023

For my grandmother, Wanda

There are years that ask questions and years that answer.

—Zora Neale Hurston, *Their Eyes Were Watching God*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to extend immense gratitude to my research mentor, David Schmitz, who gave me another chance to earn my Ph.D. by welcoming me with open arms into the University of Chicago neutrino group. I surely would not be here today without it. I am also extremely thankful to my MicroBooNE collaborators, especially Kirsty Duffy, Elena Gramellini, and Afroditi Papadopoulou, whose substantial guidance kept me focused and motivated to see this project through.

Thank you to Young-Kee Kim, Sam Zeller, Brian Nord, and Marcela Carena for the time and effort you spent being on my dissertation committee. Outside of our yearly meetings, each of you impacted my graduate experience in unique and essential ways that ensured it was the best it could possibly be. A special shout-out to David Reid—who had everything to do with me being admitted into the UChicago physics Ph.D. program—as well as Ramona Echols, Cassandra Heinz, Putri Kusumo, and the Duke physics department, notably Phillip Barbeau. Thank you all for looking out for me during my time here, and steadfastly believing in my potential.

I am so fortunate to have made and maintained so many close relationships throughout this journey. Andrea Bryant, I'm grateful we entered the bridge program together—how lucky was I to not have to ever experience being the "only one" in physics graduate school, because you were here, too. Thank you to Pratiti Deb, who has become an invaluable friend, confidant, and my favorite breath of fresh air from the clutches of academia. To Grey Batie, Bryan Ramson, and Kihana Wilson: I am deeply appreciative of your friendship, encouragement, and constant reminder that people like me *do* belong in this field. I express sincere gratitude as well to the rest of my #BlackInPhysics and #BlackInAstro fam for the never-ending support, especially Ashley Walker, who was here for me when no one else was. And to my lifelong friends outside of the physics sphere, notably Drew Underwood and Marleen Jones-Pool: Thank you for the perpetual love, patience, and grace.

As the only physicist and academic in my family, graduate school felt, at times, very isolating. My family couldn't always understand or value the work I was doing or relate to what I was experiencing. But they wholeheartedly supported me nonetheless, and for that I am forever grateful. There are far too many of you to name individually, but I will say that this experience, more than anything, has hit home for me the lesson that it really does take a village, and I am overwhelmingly blessed to have this one. I want to emphatically thank my grandmother, Chiyo Beasley, and my mother, Rachel Matthews, whose sacrifices paved the way for me to succeed; my brother, Alex Miller, who was always ready with the love and encouragement I needed in the way I needed to receive it; and my partner, David Dowd, who absolutely carried me through the latter half of graduate school. I could not have done this without you.

Last but not least, I humbly acknowledge the hidden figures whose shoulders I stand on: Willetta Greene-Johnson, Tonia Venters, and Cacey Stevens Bester. Thank you for blazing this trail, and making it easier for someone like me to walk along it.

ABSTRACT

Neutrino research is at the forefront of new and exciting experimental searches for physics beyond the Standard Model. MicroBooNE is the first of several detectors in Fermilab's leading-edge liquid argon time projection chamber (LArTPC) program working toward stringent measurements of neutrino oscillations and searches for new physics. At energies relevant to accelerator-based experiments, electron neutrino charged current interactions with at least one proton and no pions in the final topology are a dominant contribution to event rates. To date, however, limited experimental validation of this cross section exists, though such constraints are crucial for next-generation LArTPCs to reach discovery precision in the electron neutrino appearance oscillation channel. While MicroBooNE's primary physics analyses make use of the on-axis Booster Neutrino Beam, a significant off-axis neutrino flux is also received from the higher energy Neutrinos at the Main Injector (NuMI) beam. The greater electron neutrino content of the NuMI beam provides a unique opportunity for MicroBooNE to perform world-leading cross section measurements. This work presents the extraction of exclusive differential electron neutrino cross sections using a combination of NuMI data collected in neutrino and antineutrino mode by MicroBooNE. A state-of-the-art NuMI flux prediction, high purity event selection algorithm, and thorough evaluation of uncertainties are developed. Measurements are derived as a function of outgoing electron energy, total visible energy of the interaction, and opening angle between the electron and most energetic proton. The interaction rate as a function of proton multiplicity is also reported. Data-driven cross sections are compared to model predictions from neutrino event generators commonly employed in the field.

CHAPTER 1

NEUTRINOS: A PROBE INTO THE UNKNOWN

The neutrino, or *little neutral one*, is a chargeless, fundamental particle with mass so small it zips through the universe at nearly the speed of light. These ghostly particles are characterized by their low interaction rates and abundance in nature: Roughly 100 trillion neutrinos, for example, stream through our bodies every second, and nearly all will do so without a trace.

Neutrinos come from a variety of sources. Most that make it to Earth result from solar nuclear fusion, the reaction that generates sunlight. Ninety-nine percent of the energy from supernovae is carried away by neutrinos in the first ten seconds after the blast. Neutrinos are also produced in ultrahigh energy cosmic accelerators, like black holes, pulsars, gamma ray bursts, and more. The oldest source of these particles is the beginning of time itself: Neutrinos escaped from the hot sea of matter that emerged from the Big Bang, and it is believed that these low energy relics are still floating around the universe today.

On Earth, neutrinos are produced from the cascades of charged particles created when cosmic rays crash into atmospheric nuclei. Bananas, potatoes, and other foods can produce neutrinos via the radioactive decay of potassium. These particles can also be made synthetically: Nuclear reactors release intense fluxes of neutrinos via beta decay, and particle accelerators generate cascades that produce neutrinos in a manner not unlike how they occur naturally in the atmosphere.

Physicists are enamored with neutrinos because they are, quite frankly, everywhere—created almost any time atomic nuclei come together or break apart. They are cosmic messengers, carrying information across the stars for us to learn about the most fascinating astrophysical phenomena. Studying neutrinos could unravel the mystery of why we exist, and probing their properties may one day grant us entry into an otherwise inaccessible, so-called dark sector of our universe.

1.1 A brief history of the neutrino

The neutrino was first postulated in an open letter written in 1930 by theoretical physicist Wolfgang Pauli. It was an attempt to make sense of the continuous energy spectrum observed in beta decay, the process by which a nucleus-bound neutron converts into a proton via the emission of an electron. *Dear Radioactive Ladies and Gentlemen*, Pauli wrote in German, then proceeded to propose a "desperate remedy" that could explain why the ejected electrons were measured with a range of energies rather than a single value—an apparent violation of energy conservation. Pauli speculated that there was another particle, the neutrino, emitted alongside the electron, and that it must be electrically neutral, therefore escaping detection. The energetic sum of both emitted particles must be constant, and in this way conservation laws could be saved.

In 1934, Enrico Fermi expanded this idea into a quantitative description of beta decay [1], which subsequently became an early theory for the weak interaction. That same year, Hans Bethe and Rudolf Peierls penned a short paper computing the first neutrino cross section: $\sigma \approx 10^{-44}$ cm^2 for a ~ 2 MeV neutrino, corresponding to a mean free path of $\mathcal{O}(1000)$ light years in water [2]. *One can conclude*, they wrote, *that there is no practically possible way of observing the neutrino.*

But physicists are nothing if not persistent. In 1956, the neutrino—or more precisely, the electron antineutrino—was discovered by Clyde Cowan and Fred Reines using a specialized detector to capture the products of inverse beta decay ($\bar{\nu}_e + p^+ \rightarrow e^+ + n^0$) from the flux of a nuclear reactor [3]. Thus, the field of neutrino physics was born. That these particles come in more than one flavor was revealed with the discovery of the muon neutrino in 1962, in the first experiment to employ a dedicated neutrino beam [4]. (The last known flavor, the tau neutrino, was directly detected in 2001 [5].)

In the latter third of the 20$^{\text{th}}$ century, physicists were plagued by a discrepancy that came to be known as the solar neutrino problem: Upon measuring nuclear fusion in the sun with

the Homestake Solar Neutrino Detector, two-thirds of the electron neutrinos predicted by theoretical models were consistently absent [6]. The Kamiokande experiment confirmed the solar neutrino problem in 1996 [7], and the plot thickened when a shortage of atmospheric muon neutrinos was reported two years later by Super-Kamiokande [8]. This discrepancy was resolved in 2001, when an independent measurement by the Sudbury Neutrino Observatory confirmed the flavor-shifting quality of neutrinos [9]. A breakthrough recognized by the 2015 Nobel Prize in Physics, this discovery indicated that neutrinos have mass, and ushered in a new, exciting era of flavor oscillation research.

1.2 Physics beyond the Standard Model

Neutrinos are a part of the Standard Model, scientists' best theory for the behavior of the universe at subatomic scales. Within this framework, neutrinos form a doublet structure with the charged leptons (electron, muon, and tau) and weakly interact via the exchange of a $W^{\pm}$ or Z^0 boson. These interactions are known as charged current (CC) and neutral current (NC), respectively, and are represented by the Feynman diagrams shown in Figure 1.1. CC interactions with a nucleon produce a charged lepton of the same flavor as the incoming neutrino and flip the isospin of the nucleon, e.g. $\nu_l + n^0 \rightarrow l^- + p^+$. In NC interactions with a nucleon, the incoming neutrino elastically scatters and causes the nucleon to recoil ($\nu + n \rightarrow \nu + n'$). The CC channel is generally easier to detect because of the presence of a charged lepton in the final state.

The Standard Model also predicts neutrinos to be massless. That they oscillate, then, was an astonishing crack in a theory that has anticipated the outcome of countless experiments with exceptional accuracy. This discovery validated early models of flavor oscillation [10, 11], which interpret neutrinos as interacting via flavor eigenstates (ν_e, ν_μ, ν_τ) and propagating via non-zero mass eigenstates (ν_1, ν_2, ν_3). Using basic principles of quantum mechanics, the probability for a neutrino of flavor x oscillating into flavor y can be derived: $P(\nu_x \rightarrow \nu_y) =$

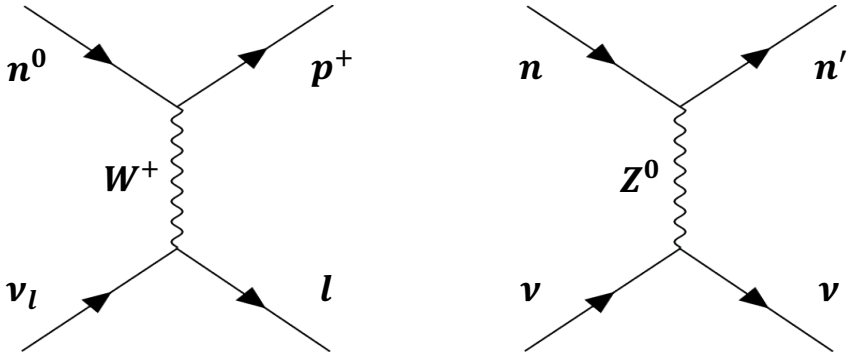


Figure 1.1: Feynman diagrams for charged current (left) and neutral current (right) neutrino interactions with a nucleon. Charged current interactions are mediated by a $W^{\pm}$ boson and produce a charged lepton of the same flavor as the incoming neutrino in the final state. Neutral current interactions are mediated by a Z^0 boson and produce a recoiling nucleon.

$|\langle\nu_x|\nu_y\rangle| \sim \sin^2(\frac{\Delta m^2_{ij} L}{E})$, where $\Delta m^2_{ij} = m^2_i - m^2_j$ represents the difference between the squares of two neutrino mass states , L is the distance propagated, and E is the neutrino energy. Accordingly, the probability goes to zero in the limit $\Delta m^2_{ij} \to 0$.

The so-called solar and atmospheric mass splittings are well constrained by oscillation experiments: $\Delta m^2_{21} \sim 10^{-5}\ eV^2$ and $\Delta m^2_{31} \sim 10^{-3}\ eV^2$, respectively [12]. The interplay between flavor and mass states is encapsulated in the 3×3 Pontecorvo–Maki–Nakagawa–Sakata mixing matrix U [10], i.e.:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \tag{1.1}$$

U can be parameterized as:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \times \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}\, e^{i\delta_{CP}} \\ 0 & 1 & 0 \\ -\sin\theta_{13}\, e^{i\delta_{CP}} & 0 & \cos\theta_{13} \end{pmatrix} \times \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.2}$$

This representation includes four free parameters that can also be experimentally determined. θ_{12}, θ_{23}, and θ_{13} have been measured through various solar, atmospheric, reactor, and accelerator neutrino experiments [12]. δ_{CP} is indicative of potential charge-parity (CP) violation in the lepton sector, which would manifest as an asymmetry between the oscillation behavior of neutrinos and antineutrinos.

The probability of electron neutrino appearance ($\nu_{\mu,\tau} \rightarrow \nu_e$) is dependent on δ_{CP}. Because of the sinusoidal nature of oscillations, a detector can be strategically positioned at a probability maximum relative to the neutrino source in order to measure this parameter. At the time of this writing, δ_{CP} has yet to be determined, due to a lack of sensitivity in current neutrino experiments. However, δ_{CP} is widely regarded as a tantalizing prospect for understanding the basis of the matter-antimatter asymmetry in our universe. Efforts are underway to establish δ_{CP} with the Deep Underground Neutrino Experiment (DUNE), which aims to measure ν_e appearance over long baselines of $\mathcal{O}(1000\,km)$ between Fermilab (Chicago, Illinois) and Sanford Lab (Lead, South Dakota) using liquid argon as a detection target [13]. DUNE is expected to come online within the next decade.

Beyond probing leptonic CP violation, ν_e appearance may also reveal a potential dark sector of unknown particles in our universe. Previous experiments at baselines of $\mathcal{O}(1\,km)$ have observed an excess of ν_e interactions inconsistent with oscillation rates expected from the standard 3-flavor paradigm [14, 15]. This ongoing mystery is referred to as the short baseline neutrino anomaly.

A possible explanation for this anomaly is that it is evidence of a new, "sterile" type of neutrino. First proposed in 1967 [11], these hypothetical particles do not interact via any known fundamental force (except gravity) but still participate in flavor mixing, thus driving detectable oscillations effects among active neutrinos of the Standard Model. Interpreting anomalous results as evidence of a sterile neutrino leads to an additional mass state ν_4 and mass-squared splitting $\Delta m^2_{41} \sim \mathcal{O}(1\ eV^2)$, which requires physics beyond the Standard

Model.

Discovering sterile neutrinos could confirm the existence of "dark" types of matter, and provide a portal for indirectly studying particles that are otherwise invisible to our detectors. The Short-Baseline Neutrino (SBN) Program, a system of three liquid argon detectors at Fermilab, will have the sensitivity to probe new types of oscillations and flavors that may shed light on this mystery. Like the solar neutrino problem of 50 years ago, particle physics may be on the brink of yet another groundbreaking discovery.

1.3 A lay of the cross section land

To enable the upcoming era of precision measurements, the neutrino physics community has identified a critical need for a more robust understanding of neutrino-nuclei interactions across all energies [16]. This will ensure the accuracy of theoretical models applied to meaningfully interpret experimental data and disentangle new physics from background processes. Knowledge of differential cross sections—i.e., those reported as a function of the kinematics of final state particles—is especially vital.

At energies relevant to accelerator-based experiments ($E_\nu \sim 0.1 - 20$ GeV), neutrino CC interactions are driven by three mechanisms. Quasielastic (QE) scattering dominates below $\sim$ 2 GeV, a process in which a neutrino scatters off and liberates a bound nucleon. (This type of interaction is illustrated in the left diagram of Figure 1.1.) With enough energy, a neutrino can excite the bound nucleon into a resonant (RES) state that quickly decays, most often into the ground-state nucleon and a single pion. Deep inelastic scattering (DIS) dominates at the highest energies, in which a neutrino scatters off a quark and generates a hadronic shower.

Figure 1.2 illustrates the suite of measurements collected over the past 50 years that have constrained theoretical models of ν_μ CC cross sections for QE (dashed), RES (dot-dashed), and DIS (dotted) processes, as well as the total, or inclusive, cross section (solid)

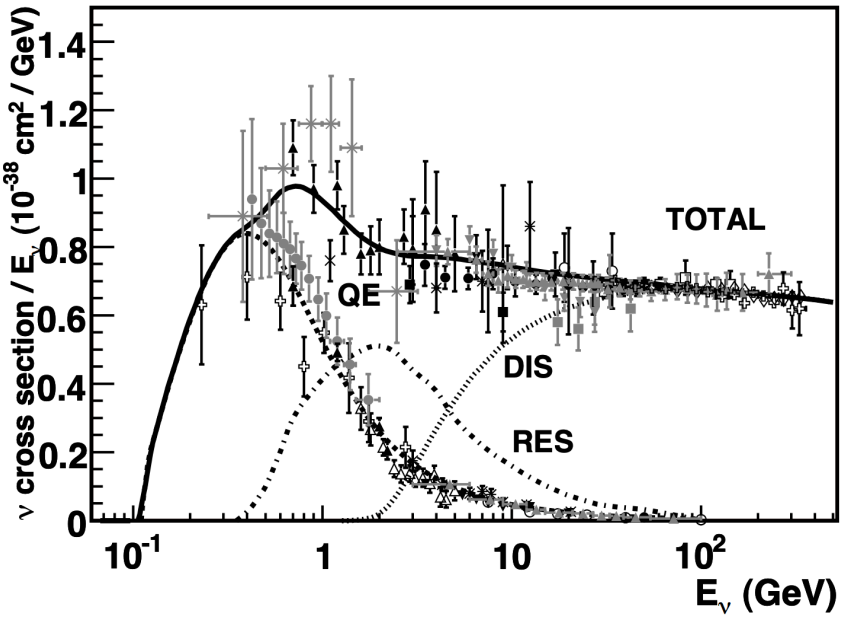


Figure 1.2: Muon neutrino charged current cross sections divided by neutrino energy and plotted as a function of neutrino energy. Theoretical contributions are shown from quasielastic (dashed), resonance (dot-dashed), and deep inelastic scattering (dotted) processes, as well as the total inclusive cross section (solid). Experimental measurements collected over the past 50 years that have constrained these predictions are overlaid. Figure from [17].

[17]. Predictions of ν_e cross sections, on the other hand, are generally constrained by ν_μ models and lack direct experimental verification. Extrapolating the kinematics and nuclear effects of ν_μ interactions to lighter lepton mass inflates experimental uncertainties in flavor oscillation measurements [18].

Only in the past decade have data-driven ν_e CC measurements begun to appear in the literature. The T2K and NOνA collaborations, for example, have reported ν_e [19, 20] and ν_e + $\bar{\nu}_e$ [21] inclusive cross sections as a function of neutrino energy, four-momentum transfer (Q^2), and outgoing lepton energy, momentum, and scattering angle. MINERνA has also published an exclusive (CCQE-like) ν_e + $\bar{\nu}_e$ cross section as a function of lepton energy, momentum, and Q^2 [22]. All of these results are derived for neutrino interactions on a carbon target.

More recently, experimental ν_e CC cross sections on argon are emerging, which will

be of particular importance for SBN and DUNE. ArgoNeuT published the first differential $\nu_e + \bar{\nu}_e$ inclusive cross section as a function of lepton angle in 2020 [23]. A year later, MicroBooNE reported a single-bin $\nu_e + \bar{\nu}_e$ inclusive result [24], quickly followed by differential measurements of the same channel as a function of lepton energy and scattering angle [25]. Exclusive ν_e CC cross sections for interactions with at least one proton and no pions in the final state ($\nu_e + {}^{40}Ar \rightarrow 1e + Np$) were extracted by MicroBooNE in 2022 [26]. The latter results used Fermilab's Booster Neutrino Beam and were reported as a function of energy and scattering angle for both the electron and leading (most energetic) proton.

In short, the neutrino community is still very much carving out the phase space for electron neutrino cross sections, which are crucial for reaching discovery potential in the ν_e appearance channel. This work adds to that effort, presenting measurements of ν_e CC cross sections with at least one proton and no pions in the final topology. Data is collected from Fermilab's Neutrinos at the Main Injector (NuMI) Beam, detailed in Chapter 2, by the MicroBooNE detector. An experimental design overview of MicroBooNE is found in Chapter 3, and the reconstruction of neutrino interactions and the simulation framework used to estimate event rates in the detector is outlined in Chapter 4. An algorithm, described in Chapter 5, is developed to select $\nu_e + {}^{40}Ar \rightarrow 1e + Np$ interactions from the NuMI dataset, and the statistical and systematic uncertainties associated with the selected event rates are quantified in Chapter 6. Flux-integrated differential cross sections are extracted as a function of outgoing electron energy, total visible energy, and opening angle between the electron and leading proton in Chapter 7, and compared to predictions from neutrino event generators commonly employed in the field. The interaction rate as a function of proton multiplicity is also reported. Chapter 8 concludes the work with a prospective outlook.

CHAPTER 2

THE NUMI FLUX AT MICROBOONE

MicroBooNE is a surface detector located in the Liquid Argon Test Facility (LArTF) building along the Booster Neutrino Beamline (BNB) at Fermilab. It also receives a significant off-axis neutrino flux from the NuMI beam. The higher energy of the NuMI beam lends itself to a greater production of electron neutrinos, making it an optimal source for independent measurements of ν_e interactions in the detector. This chapter provides a snapshot of the NuMI beamline as it pertains to MicroBooNE. A full report on the hardware, design considerations, and operations of NuMI can be found in [27].

2.1 Neutrinos at the Main Injector

The NuMI beam is generated by Fermilab's accelerator complex, a schematic of which is shown in Figure 2.1. The roughly 500-foot linear accelerator (Linac) accelerates H^- ions to 400 MeV. The ions are converted into protons via the stripping of two electrons and intermittently fed into the Booster Synchrotron in batches that are 1.6 μs long. The Booster accelerates the protons around its 1500-foot circumference ring to 8 GeV.

A portion of these protons are spilled onto a track to collide with a beryllium target and create the BNB. The rest of the protons are fed into the Main Injector, a synchrotron that is 7 times larger than the Booster thus can hold up to 7 batches. However, the Main Injector nominally accommodates 6 batches with one slot left empty to allow time for the kicker magnets, which extract the beam from the synchrotron, to ramp up. Six batches form a 9.6 μs beam spill, defining the pulsed timing structure of NuMI.

The Main Injector accelerates protons to 120 GeV. To create the NuMI beam, protons are extracted from the Main Injector and spilled onto a track with a graphite target. One batch contains 5.0×10^{12} protons, so a standard NuMI beam spill has an intensity of 3.0×10^{13}

protons on target, or POT [28].

A method called slip-stacking is implemented to enhance the intensity of the NuMI beam. With slip-stacking, a first set of X batches is fed into the Main Injector, then decelerated by one batch length. A second set of 6 batches is injected, and the first set eventually overlaps (or "slips") onto the second, producing X batches with double the intensity. Slip-stacking configurations are referred to as X+6, where X is the number of batches that have been doubled [29]. All X+6 batches are then spilled onto the NuMI beam track.

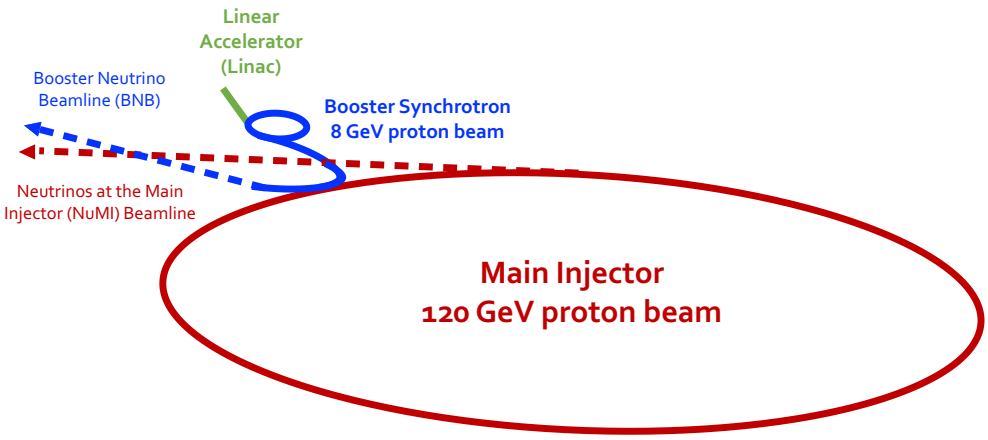


Figure 2.1: The Fermilab accelerator complex. Protons from the Linac are fed into the Booster Synchrotron, which accelerates them to 8 GeV. A portion of these protons are extracted from the Booster to create the BNB. The rest of the protons are fed into the Main Injector, where they are accelerated to 120 GeV and extracted into the NuMI beamline. In this graphic, MicroBooNE sits approximately at the intersection of the BNB and NuMI beamlines.

In MicroBooNE, NuMI data preceding the summer of 2016 was collected in 4+6 slip-stacking mode, corresponding to an intensity of 10 total batches or 5.0×10^{13} POT per beam spill. A single injection cycle takes 1.333 seconds to complete [29], so for 4+6 slip-stacking—which requires two injections—MicroBooNE receives 6.0×10^{20} POT per year from NuMI. After the summer of 2016, MicroBooNE's NuMI data was collected in the 6+6 slip-stacking configuration, corresponding to 6.0×10^{13} POT per spill or 7.0×10^{20} POT per year.

A schematic of the NuMI beamline from [27] is shown in Figure 2.2. The collision of protons with the graphite target generates cascades of hadrons (primarily pions, kaons, and nucleons) that are focused in or deflected from the forward direction based on the polarity of the current in a pair of magnetic horns immediately beyond the target. The focused hadrons travel down a 675-m tunnel where they decay, producing neutrinos. Charged particles that survive the decay pipe are terminated at a beam absorber comprised of aluminum, steel, and concrete.

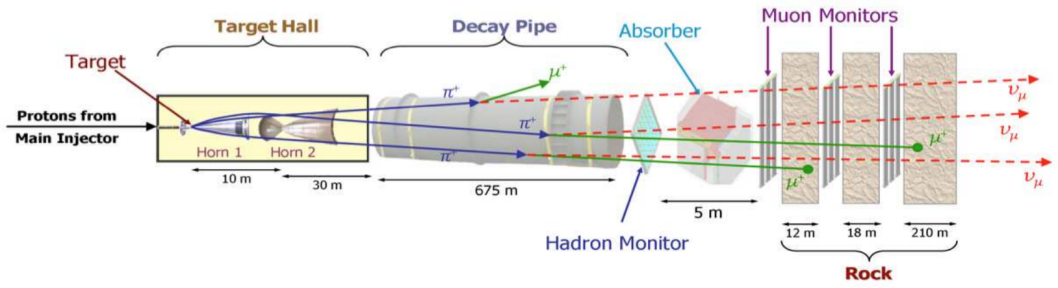


Figure 2.2: Individual components of the NuMI beamline with relevant dimensions. The collision of protons with the graphite target generates cascades of hadrons that travel down a decay pipe, where they will decay into neutrinos. Figure from [27].

NuMI was designed such that the relative positions between the target and focusing horns are flexible, allowing for a tunable energy beam. For data used in this work, NuMI was set in the medium energy beam configuration, corresponding to a 23-meter separation between the focusing horns [27] and a target that is 143.3 cm upstream from the start of the first focusing horn.

NuMI can also run in two beam modes, referred to as forward horn current (FHC) and reverse horn current (RHC). In FHC (RHC) mode, the polarity of the current running through the magnetic horns is +200 kA (-200 kA), and parent particles with positive (negative) charge are focused into the decay pipe to produce a primarily neutrino (antineutrino) beam.

2.2 Beamline geometry

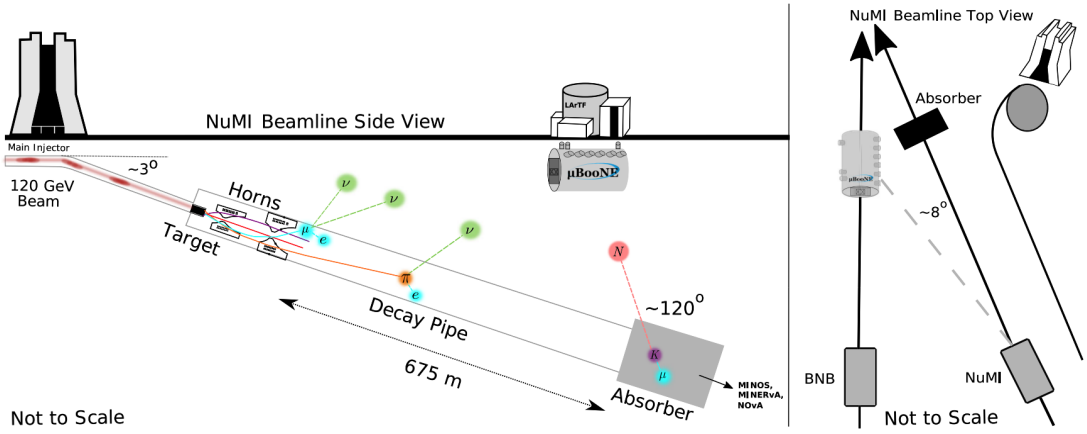


Figure 2.3: Illustration of the NuMI beamline relative to MicroBooNE from elevation (left) and bird's eye (right) views. Figure from [30].

MicroBooNE is located roughly at the intersection of the BNB and NuMI beamlines shown in Figure 2.1. A schematic of NuMI's geometry as it pertains to MicroBooNE, from [30], is shown in Figure 2.3. The beam points down into the Earth at an angle of approximately 3° from the surface. MicroBooNE lies ∼680 meters from the NuMI target, at an angle of ∼8° off the beamline from bird's eye view. Two distinct coordinate systems, of the beam and of the detector, are relevant for this work. The beam coordinate system has an origin located at the "upstream" end of the first horn (see Figure 2.2), where upstream refers to the part of the system that the beam encounters first. The origin of the detector coordinate system in beam coordinates is:

$$\vec{x}_0 = \begin{pmatrix} 5502 \\ 7259 \\ 67270 \end{pmatrix} \tag{2.1}$$

in units of `cm`. The rotation matrix that converts beam coordinates to detector coordinates is given by:

$$R = \begin{pmatrix} 0.9210385380402568 & 0.0227135048039241207 & 0.38880857519374290 \\ 0.0000462540012621546684 & 0.99829162468141475 & -0.0584279894529063024 \\ -0.38947144863934974 & 0.0538324139386641073 & 0.91946400794392302 \end{pmatrix}. \quad (2.2)$$

The position vector of a neutrino can be transformed between beam and detector coordinates using a rotation and a translation, i.e.:

$$\vec{x}_{beam} = R \times \vec{x}_{uB} + \vec{x}_0 \quad (2.3)$$

where $\vec{x}_{beam}$ is the position of the point in beam coordinates and $\vec{x}_{uB}$ is the position of the point in detector coordinates. Relatedly, the velocity vector transformation between beam and detector coordinates is the time derivative of Equation 2.3:

$$\vec{v}_{beam} = R \times \vec{v}_{uB}. \quad (2.4)$$

It is informative to think about the NuMI flux at MicroBooNE using spherical coordinates. Thus, we define the polar, or decay, angle θ off the beamline as:

$$\tan\theta = \frac{\sqrt{x^2 + y^2}}{z} \quad (2.5)$$

where x, y, and z are the elements of $\vec{x}_{beam}$. This is the angle between the neutrino ray and beamline horizontal of the elevation view in Figure 2.3. Similarly, we define an azimuthal angle ϕ in the xy plane:

$$\tan\phi = \frac{y}{x}. \quad (2.6)$$

These angles can be constructed in both detector and beam coordinate systems, though it is often more instructive to study distributions in the latter.

2.3 Flux prediction at MicroBooNE

The NuMI flux prediction for any experiment is generated using a combination of `Geant4` [31], a toolkit for simulating the passage of particles through matter, and a detailed geometric model of the beamline [32]. The simulation accounts for particle interactions and their propagation, beginning with protons colliding with the graphite target and ending with the hadrons that decay into neutrinos.

Primary interactions along the NuMI beamline involve the production of pions and kaons, which then decay via secondary reactions into neutrinos, i.e.:

$$
\begin{aligned}
p^+(120\ \mathrm{GeV}) + C &\rightarrow \pi^\pm + X \\
&\downarrow \\
&\pi^\pm \rightarrow \mu^\pm + \overset{(-)}{\nu}_\mu \\
&\pi^\pm \rightarrow e^\pm + \overset{(-)}{\nu}_e
\end{aligned}
$$

$$
\begin{aligned}
p^+(120\ \mathrm{GeV}) + C &\rightarrow K^\pm + X \\
&\downarrow \\
&K^\pm \rightarrow \mu^\pm + \overset{(-)}{\nu}_\mu \\
&K^\pm \rightarrow e^\pm + \pi^0 + \overset{(-)}{\nu}_e
\end{aligned}
$$

$$
\begin{aligned}
p^+(120\ \mathrm{GeV}) + C &\rightarrow \overset{(-)}{K}{}^0_L + X \\
&\downarrow \\
&\overset{(-)}{K}{}^0_L \rightarrow e^\pm + \pi^\mp + \overset{(-)}{\nu}_e
\end{aligned}
$$

where X represents other hadronic byproducts of the interaction, including nucleons.

In theory, precise knowledge of the beamline geometry and of π/K production cross sections should yield an accurate neutrino flux. In reality, however, the flux is more complicated to predict due to a variety of other processes that create neutrinos. Incident protons, for example, may bypass the carbon altogether and instead interact with air in the target hall, aluminum in the magnetic horns, or the helium gas or iron walls of the decay pipe:

$$p^{+}(120\ \text{GeV}) + Al/He/Fe \rightarrow \pi/K + X$$
$$\downarrow$$
$$\text{neutrino production.}$$

Nucleons from any of the above channels can also collide with materials along the beamline, creating their own hadronic cascades, i.e. $X(< 120\ \text{GeV}) + C/Al/He/Fe \rightarrow \pi/K$, which decay into neutrinos via tertiary processes. If they do not decay first, pions and kaons at any point along this production chain may do the same.

Traditionally, detailed simulations are built to account for the impact of these non-standard interactions on the flux prediction, though these models are largely unconstrained. To mitigate this, data-driven corrections to the raw NuMI flux prediction are applied using `PPFX`, an experiment-agnostic software package originally developed by the MINERνA collaboration. Using information about the neutrino ancestry and decay kinematics, `PPFX` modifies the `Geant4` simulation to reproduce independent experimental measurements of hadron production and absorption cross sections that inherently account for the above effects. The `PPFX` user can choose between using data collected by experiments employing thick (two interaction lengths) or thin (few percent interaction length) carbon targets.

Part of the work undertaken for this dissertation included upgrading `PPFX` tools for compatible integration into MicroBooNE's existing simulation software (see Chapter 4), and validating this implementation by reproducing published NuMI fluxes at MINERνA and the NOνA near detector. To simulate the NuMI flux prediction at MicroBooNE, the thin

target constraints are used, a decision motivated by results showing this option is in better agreement with *in situ* flux measurements taken by MINERνA [33]. External thin target datasets used by PPFX to constrain the flux prediction are thoroughly described in [32] and summarized in Table 2.1.

Hadron Production	
$p^+ + C \to \pi^\pm$ production	NA49 [34], Barton et al. [35]
$p^+ + C \to p^+$ or n^0 production	NA49 [34]
$p^+ + C \to K^\pm$ production	NA49 [36]
$p^+ + C \to K/\pi$ production ratios	MIPP [37]
Absorption/Inelastic Processes	
$K^\pm$ absorption by carbon or aluminum	Abrams et al. [38], Denisov et al. [39], Carroll et al. [40], Allaby et al. [41]
$\pi^\pm$ absorption by carbon or aluminum	Cronin et al. [42], Denisov et al. [39], Allaby et al. [41], Allardyce et al. [43], Longo et al. [44], Fedorov et al. [45], Bobchenko et al. [46], Carroll et al. [40]
Nucleon-carbon inelastic interactions	Bellettini et al. [47], NA61 [48], Denisov et al. [39], Carroll et al. [40], NA49 [49], Roberts et al. [50]

Table 2.1: Datasets used for PPFX thin target constraints of the NuMI flux prediction at MicroBooNE, summarized from [32]. A variety of processes at different energies are taken into account, including hadron production from proton-carbon collisions, kaon and pion absorption, and nucleon-carbon inelastic interactions.

A variety of processes at different energies are taken into account, including hadron production from proton-carbon collisions, kaon and pion absorption, and nucleon-carbon inelastic interactions. For cases where no data is available, PPFX attempts a theoretically guided extension of existing data or applies a 40% flat uncertainty on the interaction. PPFX-derived constraints come in the form of correction weights that can be applied to each event in the raw central value (CV) flux prediction. The package also generates a set of variation weights that can be used in place of the CV weight to evaluate the uncertainty on this correction (see Section 6.2).

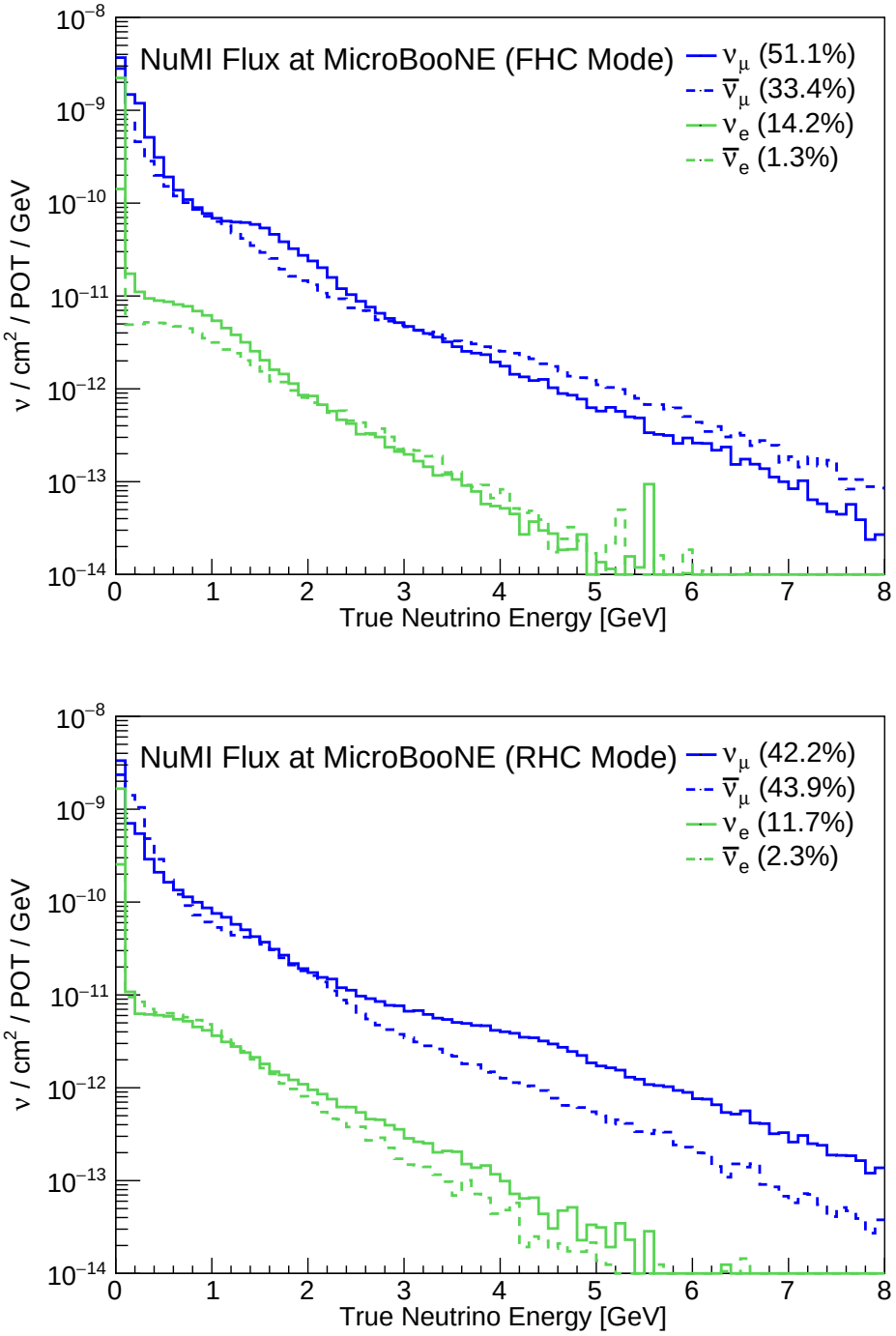


Figure 2.4: The PPFX-corrected NuMI flux prediction at MicroBooNE for FHC (top) and RHC (bottom) as a function of true neutrino energy. Electron neutrinos comprise about 14% (12%) of the total flux in FHC (RHC) mode, though this work only considers the flux greater than 60 MeV (excluding muon decay-at-rest interactions).

The PPFX-corrected NuMI flux prediction at MicroBooNE is shown as a function of true neutrino energy for FHC (top) and RHC (bottom) in Figure 2.4. Muons created at any point along the hadronic production chain can decay at rest, contributing to a low-energy (<60 MeV) peak in the ν_e and $\bar{\nu}_e$ flux predictions:

$$\mu^{\pm} \rightarrow e^{\pm} + \overset{(-)}{\nu}_e + \overset{(-)}{\nu}_{\mu}. \tag{2.7}$$

Higher energy ν_e and $\bar{\nu}_e$ result primarily from charged and neutral (long-lived) kaon decays:

$$K^{\pm} \rightarrow \overset{(-)}{\nu}_e + e^{\pm} + \pi^0 \tag{2.8}$$

$$K^0_L \rightarrow \nu_e + e^{\pm} + \pi^{\mp}. \tag{2.9}$$

Electron neutrinos comprise about 14% (12%) of the total flux in FHC (RHC) mode. Because the selection algorithm described in Chapter 5 is not sensitive to neutrinos below $\sim$ 200 MeV, the measurements described in this work only take into account the flux greater than 60 MeV, excluding the low energy ν_e contribution from decay-at-rest muons.

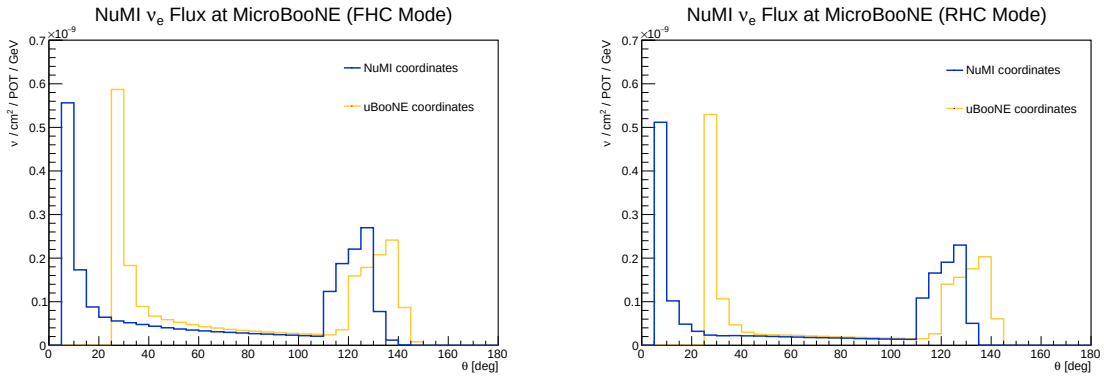


Figure 2.5: The ν_e flux at MicroBooNE for FHC (left) and RHC (right) beam modes as a function of decay angle. In beam coordinates, most ν_e enter MicroBooNE from the target ($\sim$8° off the beamline), while a secondary ν_e excess comes from the absorber ($\sim 110-140°$). Between these extremes, electron neutrinos arrive from along the decay pipe. (Note: these distributions are not constrained with PPFX.)

Figures 2.5 and 2.6 illustrate the ν_e flux dependence on decay angle (θ) and azimuthal

angle (ϕ) respectively, in both beam (blue) and detector (yellow) coordinate systems. Distributions of other neutrino flavors follow a similar pattern. The θ distribution is tripartite: most ν_e enter MicroBooNE at an angle of ~8° off the beamline, and a secondary excess comes from the beam absorber at angles ~110-140° in beam coordinates. Neutrinos between these two extremes arrive at MicroBooNE from the decay pipe.

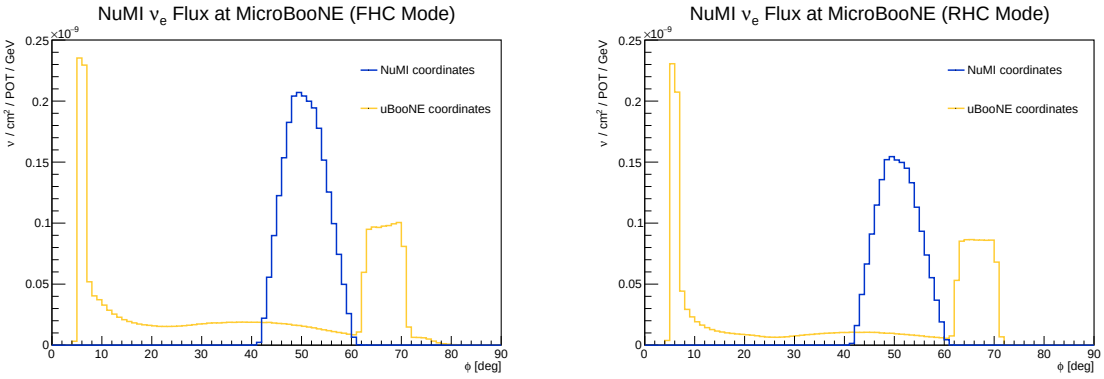


Figure 2.6: The ν_e flux at MicroBooNE for FHC (left) and RHC (right) beam modes as a function of azimuthal angle. In beam coordinates, only neutrinos with a direction of $\sim 40-60°$ make it into the detector. The width of this peak is roughly representative of the angular spread of MicroBooNE. (Note: these distributions are not constrained with `PPFX`.)

The azimuthal distribution of electron neutrinos entering MicroBooNE provides some intuition about where the detector sits in relation to NuMI. In beam coordinates, only ν_e with a direction of ~40-60° make it into the detector. The width of this peak is representative of the angular spread of MicroBooNE.

The NuMI flux at MicroBooNE presents a strong correlation between neutrino energy and decay angle θ. Figure 2.7 shows the `PPFX`-corrected ν_e flux in FHC and RHC as a function of both energy and decay angle in beam coordinates. There are four angular bins: $0-10°$ and $10-20°$ correspond to the NuMI target, $20-110°$ corresponds to the decay pipe, and $110-160°$ is the beam absorber. For energies above 0.25 GeV, the entire ν_e flux originates from the NuMI target in both FHC and RHC modes. The lowest energy neutrinos ($\leq$ 250 MeV) peak at both the target and the beam absorber.

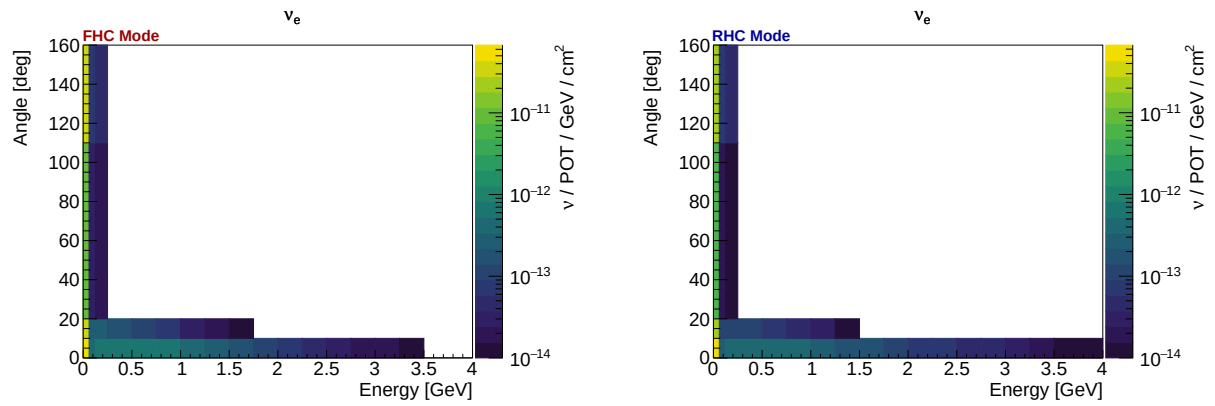


Figure 2.7: The NuMI ν_e flux prediction at MicroBooNE for FHC (left) and RHC (right) beam modes as a function of true neutrino energy and decay angle in beam coordinates. For energies above 0.25 GeV, the entire ν_e flux originates from the NuMI target.

2.4 MicroBooNE's NuMI dataset

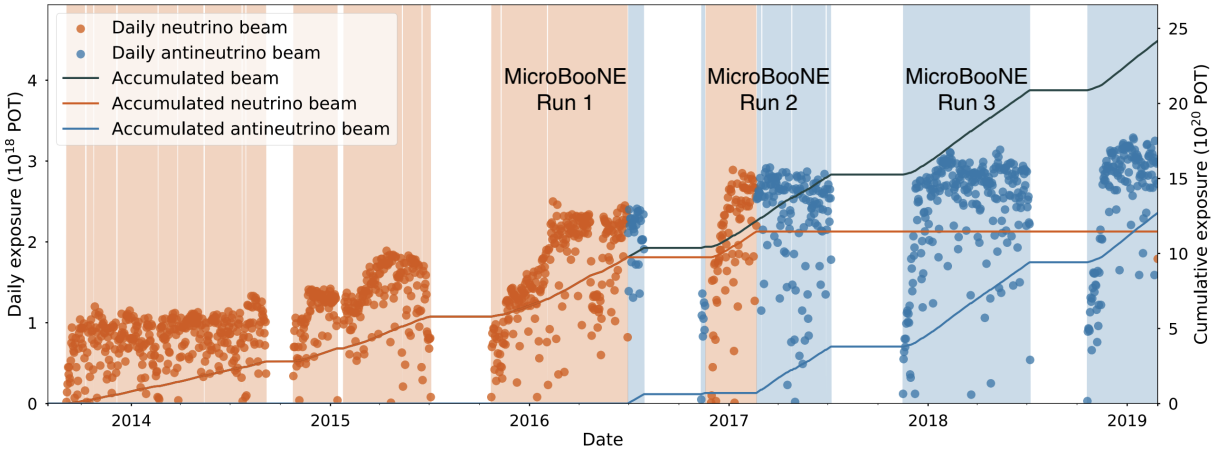


Figure 2.8: The total cumulative POT delivered by the NuMI beamline. Orange regions were operated in FHC mode, and blue regions were operated in RHC mode. The regions where no POT was delivered are due to accelerator complex shutdowns. Figure adapted from [51].

Figure 2.8 shows the total cumulative POT delivered by NuMI as a function of time. Orange regions are periods operated in FHC mode, and blue regions are periods operated in RHC mode. Periods where no POT were delivered are due to shutdowns of the accelerator complex for maintenance.

MicroBooNE collected data from October 2015 through March 2020. Table 2.2 lists the dates and operation modes of the NuMI beam with respect to MicroBooNE Run periods. The measurements described in this work make use of FHC Run 1 data (2.0×10^{20} POT) and RHC Run 3 data (5.0×10^{20} POT) for a total of 7.0×10^{20} POT.

Run Period	**FHC (+200 kA)**	**RHC (-200 kA)**
Run 1	October 23, 2015 - May 2, 2016	June 29, 2016 - July 29, 2016
Run 2	Nov 14, 2016 - Feb 20, 2017	Nov 11, 2016 - Nov 14, 2016
		Feb 20, 2017 - July 7, 2017
Run 3	–	Nov 7, 2017 - July 6, 2018
Run 4	Feb 26, 2019 - July 6, 2019	Oct 20, 2018 - Feb 26, 2019
Run 5	Oct 29, 2019 - March 20, 2020	–

Table 2.2: The dates and operation modes of the NuMI beam during MicroBooNE Run periods. This work uses a combination of FHC Run 1 and RHC Run 3 data.

2.5 Potential oscillation effects in the NuMI flux

MicroBooNE's distance away from the NuMI target creates the possibility of short baseline oscillations that, if significant enough, could alter the flux and any cross section measurements as a result. To ensure that this does not become an unaccounted source of uncertainty, an evaluation of potential oscillation effects is sketched in this section.

A binned analysis is performed considering the electron neutrino appearance channel, $\nu_\mu \rightarrow \nu_e$, for the NuMI beam in FHC mode. (The channel $\nu_\tau \rightarrow \nu_e$ is treated as negligible, since intrinsic ν_τ production in NuMI is minimal due to the relatively low energy of the beam.) Electron neutrino disappearance, due to intrinsic ν_e produced by the beam oscillating into other flavors, is also taken into account.

Figure 2.9 shows the NuMI FHC ν_μ (left) and ν_e (right) intrinsic flux predictions (sans oscillations), as a function of 50-MeV neutrino energy bins and 50-meter propagated distance

bins. These distributions are used to verify that active flavor oscillations between NuMI and MicroBooNE are trivial, as well as to quantify the detector's sensitivity to potential sterile mixing effects. An analogous study using the RHC flux yields similar results.

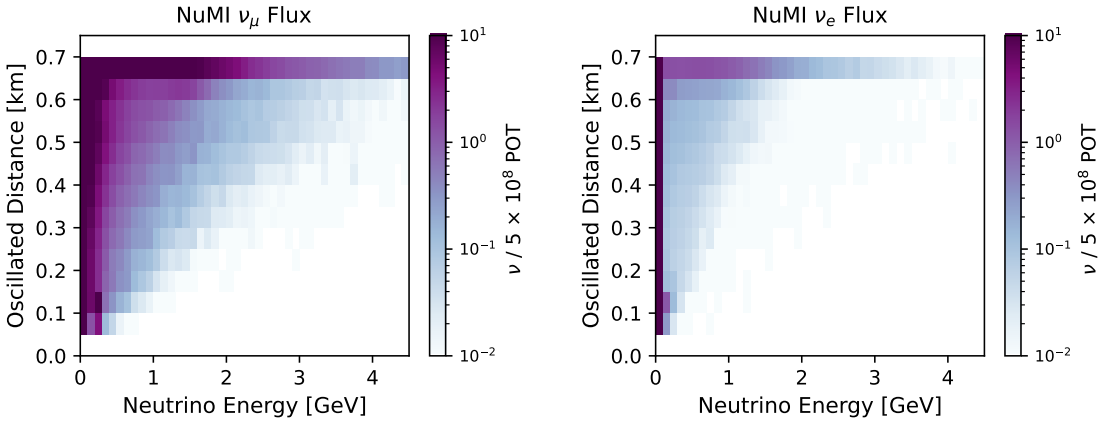


Figure 2.9: The NuMI ν_μ (left) and ν_e (right) flux prediction at MicroBooNE without any oscillations taken into account, shown as a function of energy and propagated distance of the neutrino.

2.5.1 Active flavor oscillations

Active flavor oscillations for neutrinos of $\mathcal{O}(GeV)$ that travel distances shorter than ~1 km should be negligible. We verify this limit using the leading oscillation terms from ν_e appearance and ν_e disappearance (or rather, survival) probabilities as derived from the standard 3-flavor framework:

$$P(\nu_\mu \to \nu_e) = \sin^2(\theta_{23})\sin^2(2\theta_{13})\sin^2\left(\frac{1.27\,\Delta m^2_{31} L}{E}\right) + \cos^2(\theta_{23})\cos^2(\theta_{13})\sin^2(2\theta_{12})\sin^2\left(\frac{1.27\,\Delta m^2_{21} L}{E}\right) \tag{2.10}$$

$$P(\nu_e \to \nu_e) = 1 - \sin^2(2\theta_{13})\sin^2\left(\frac{1.27\,\Delta m^2_{31} L}{E}\right) + \cos^2(\theta_{13})\sin^2(2\theta_{12})\sin^2\left(\frac{1.27\,\Delta m^2_{21} L}{E}\right). \tag{2.11}$$

Computing the oscillated NuMI ν_e flux at MicroBooNE requires combining the contributions from 1) intrinsic ν_μ oscillating into ν_e and 2) intrinsic ν_e that survive the trip to the detector. The first is acquired by multiplying the number of generated ν_μ in each bin of the flux shown in Figure 2.9 (left) by the appearance probability in Equation 2.10, which

is evaluated using the central value neutrino energy E and propagated distance L of each bin. The second contribution is similarly obtained, by multiplying the number of generated ν_e (right plot of Figure 2.9) in each bin by the survival probability (Equation 2.11) of that bin. The parameter values used for the calculations are taken from [12]: $\sin^2\theta_{12} = 0.310$, $\sin^2\theta_{13} = 0.02241$, and $\sin^2 2\theta_{23} = 0.058$ for the mixing angles; $\Delta m^2_{12} = 7.39 \times 10^{-5}$ eV^2 and $\Delta m^2_{13} = 2.523 \times 10^{-3}$ eV^2 for the mass splittings.

Figure 2.10 shows the ratio between the oscillated and unoscillated ν_e flux predictions as a function of neutrino energy. For active flavor mixing, the oscillated flux is maximally $\mathcal{O}(10^{-3})$ larger than the unoscillated flux, thus the effect is indeed negligible.

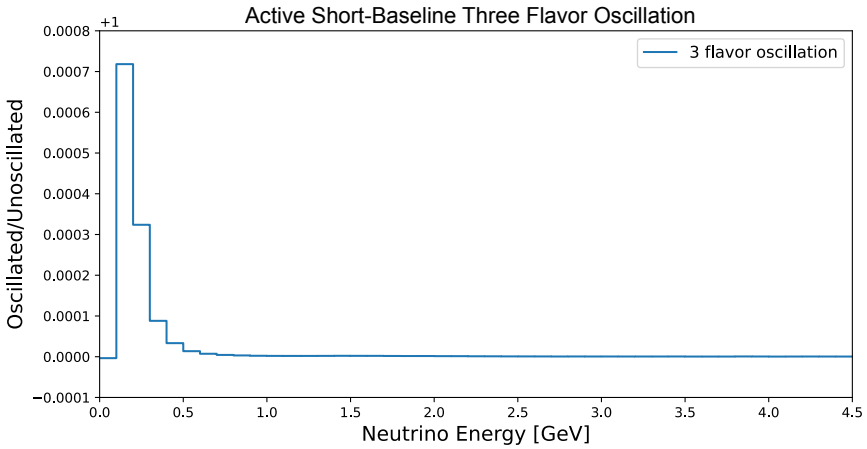


Figure 2.10: Ratio between oscillated and unoscillated ν_e flux predictions as a function of neutrino energy. The oscillated flux is maximally $\mathcal{O}(10^{-3})$ larger than the unoscillated flux, thus the effect is indeed negligible.

2.5.2 Sterile flavor oscillations

Oscillations involving sterile neutrino flavors could result in a more pronounced effect on the NuMI flux. To quantify the strength of this effect, a 3+1 model is assumed, consisting of 3 active flavors and 1 sterile flavor. The 3+1 mixing matrix is extended from Equation 2.12 to:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \\ \nu_s \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} & U_{e4} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} & U_{\mu 4} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} & U_{\tau 4} \\ U_{s1} & U_{s2} & U_{s3} & U_{s4} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix} \tag{2.12}$$

where ν_s is the sterile flavor and a fourth mass state ν_4 has been introduced.

In this framework, as long as $\Delta m^2_{41} \gg |\Delta m^2_{31}|, |\Delta m^2_{21}|$, the electron neutrino appearance channel at short baselines is well-approximated by a two-flavor vacuum oscillation formula [52]:

$$P(\nu_\mu \to \nu_e) = \sin^2(2\theta_{\mu e}) \sin^2\left(\frac{1.27\,\Delta m^2_{41} L}{E}\right) \tag{2.13}$$

where $\theta_{\mu e}$ is an effective mixing angle driving the oscillation channel $\nu_\mu \to \nu_e$. Analogously, the ν_e survival probability is:

$$P(\nu_e \to \nu_e) = 1 - \sin^2(2\theta_{ee}) \sin^2\left(\frac{1.27\,\Delta m^2_{41} L}{E}\right) \tag{2.14}$$

where θ_{ee} drives the channel $\nu_e \to \nu_e$. Unlike the active mixing case, 3+1 oscillation parameters $\sin^2 2\theta_{\mu e}$, $\sin^2 2\theta_{ee}$, and Δm^2_{41} are not confirmed, but exist as best-fit values in favored regions of phase space from different experiments.

The flux contribution from ν_e appearance is analyzed first, using the parameter set ($\sin^2 2\theta_{\mu e}$, Δm^2_{41}). Best-fit values from the LSND experiment $(0.003, 1.2\ eV^2)$ [53], the global ν_e appearance analysis $(0.00631, 0.559\ eV^2)$ [54], and the 2016 global inclusive analysis of both ν_e appearance and disappearance channels $(0.0015, 1.6\ eV^2)$ [55] are considered. In addition, the maximum potential oscillation effect is outlined by evaluating oscillation effects for the entire ($\sin^2 2\theta_{\mu e}$, Δm^2_{41}) phase space in the 3σ-favored region of the global ν_e appearance analysis, which is comprised of 90 parameter sets.

The oscillated NuMI ν_e flux with sterile mixing ν_e appearance effects is computed by adding the contribution from intrinsic ν_μ oscillating into ν_e with the unoscillated, intrinsic

ν_e flux. The procedure for evaluating the $\nu_\mu \rightarrow \nu_e$ contribution is the same as in the preceding section, except using Equation 2.13 to calculate the appearance probability. Results using best-fit parameters from LSND (blue), the global ν_e appearance analysis (red), and the global inclusive analysis (green) are shown in Figure 2.11, as well as for parameter sets spanning the global 3σ ν_e appearance phase space (yellow). Ratios to the unoscillated ν_e flux prediction are also displayed.

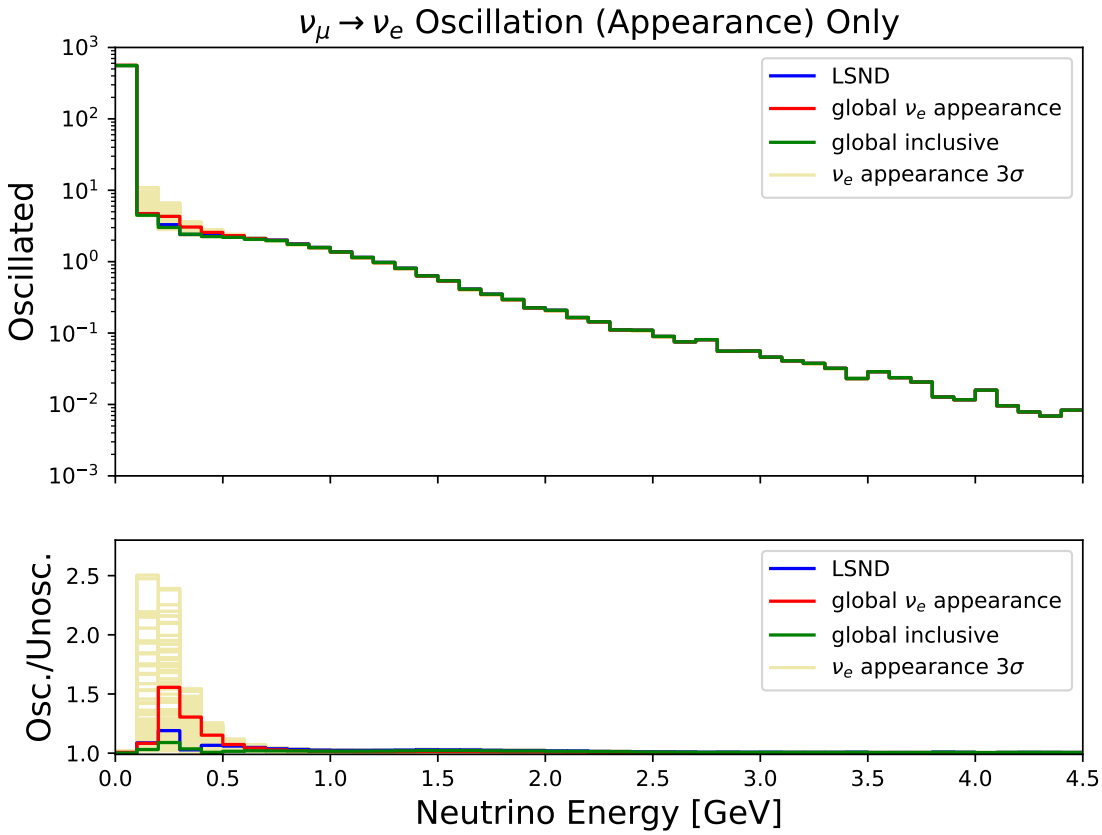


Figure 2.11: Oscillated NuMI fluxes with sterile mixing ν_e appearance effects using best-fit parameters from LSND (blue), the global ν_e appearance analysis (red), and the global inclusive analysis (green), as well as for parameter sets spanning the 3σ-favored phase space of the global ν_e appearance analysis (yellow). Ratios to the unoscillated ν_e flux prediction are also displayed.

Pronounced oscillation effects due to high L/E values for neutrino energies < 0.5 GeV are present, particularly in the 3σ phase space envelope, which can produce a distribution as high as 2.5 times the unoscillated flux. However, poor reconstruction and inefficiency of

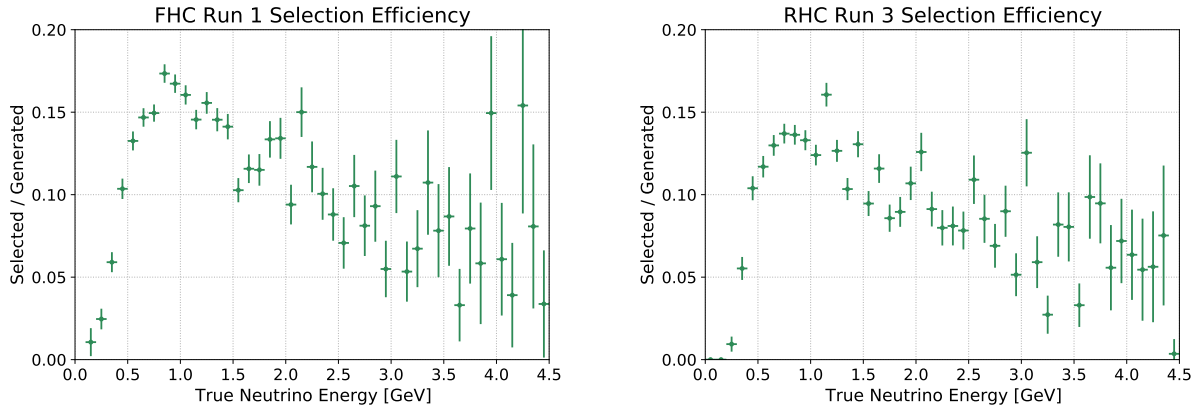


Figure 2.12: Efficiency of the algorithm designed to select $\nu_e + {}^{40}Ar \rightarrow 1e + Np$ interactions. Shown for FHC Run 1 (left) and RHC Run 3 (right) as a function of true neutrino energy. Low efficiency in the region < 0.5 GeV limits the sensitivity of this work to potential 3+1 oscillation effects.

the event selection algorithm described in Chapter 5 limits the sensitivity of this work to ν_e with energies < 0.5 GeV. This is illustrated in Figure 2.12. Systematic uncertainties in the flux also dominate in this region (see Section 6.2), reaching values $\sim$20% or higher.

Next, the change in the flux due to intrinsic ν_e disappearance ($\nu_e \rightarrow \nu_{\mu,\tau}$) is briefly examined. The oscillated NuMI flux in this scenario is computed by adding the $\nu_\mu \rightarrow \nu_e$ contribution to the intrinsic ν_e flux that does not disappear by the time it reaches MicroBooNE. The former is acquired in the same manner as above, using Equation 2.13 and the unoscillated ν_μ flux of Figure 2.9. The latter is obtained using the survival probability of Equation 2.14 with the unoscillated ν_e flux. For a rough quantitative evaluation, best-fit oscillation parameters from the global inclusive analysis are used: ($\sin^2 2\theta_{\mu e} = 0.00631$, $\Delta m^2_{41} = 0.559\, eV^2$) for ν_e appearance and ($\sin^2 2\theta_{ee} = 0.04038, \Delta m^2_{41} = 1.3\, eV^2$) for ν_e disappearance [55]. Correlations between the appearance and disappearance channels are not considered.

Figure 2.13 compares the oscillated fluxes taking into account ν_e appearance only (green) and ν_e appearance and disappearance effects (yellow). The flux excess due to ν_e appearance is notably suppressed by including the disappearance channel in the < 0.5 GeV region, and even results in a deficit below the unoscillated prediction for neutrinos with energies $< 0.5 - 1.5$

GeV, though the effect is slight enough to be covered by existing flux uncertainties. Similar reductions are anticipated for the entire allowed phase space.

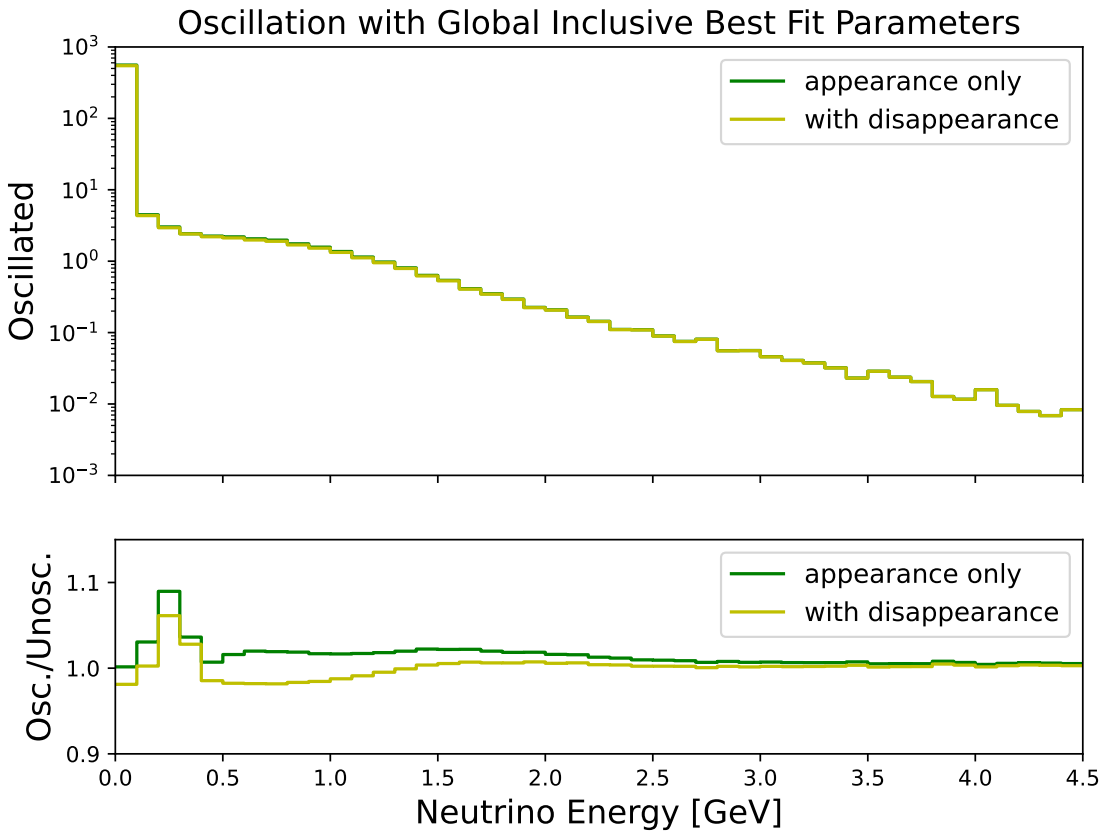


Figure 2.13: The oscillated ν_e flux prediction for only the appearance channel (green) and the inclusive (appearance + disappearance) channel (yellow), as well as the ratio to the unoscillated ν_e flux prediction, using uncorrelated best fit parameters from [55]. Inclusion of the disappearance channel slightly reduces sterile mixing effects that could be observed in the ν_e flux at MicroBooNE.

2.5.3 Sensitivity analysis

Taking it a step further, a rough estimate of MicroBooNE's sensitivity to observing sterile neutrino mixing in the FHC portion of the NuMI dataset can be derived. A prediction of the intrinsic ν_e event rate is first determined using the selection algorithm described in Chapter 5. This rate is multiplied by the 3σ phase space ratios shown in Figure 2.11 to obtain a set of 90 oscillated rate predictions. Next, a covariance matrix is constructed using the statistical

uncertainty of MicroBooNE's projected NuMI FHC sample (9.23×10^{20} POT from Runs 1 through 5) as well as major sources of systematic uncertainty associated with the ν_e event rate: hadron production in the flux simulation, cross section and re-interaction modeling, and detector response. A full description of these uncertainties can be found in Chapter 6.

A sensitivity surface can be mapped out by χ^2 goodness-of-fit values for each oscillated event rate:

$$\chi^2(\Delta m^2_{41}, \sin^2 2\theta) = \sum_{i,j} [N_i^{null} - N_i^{osc}(\Delta m^2_{41}, \sin^2 2\theta)] \, (E_{ij})^{-1} \, [N_j^{null} - N_j^{osc}(\Delta m^2_{41}, \sin^2 2\theta)] \quad (2.15)$$

where N_i^{null} is the intrinsic ν_e event rate in bin i, N_i^{osc} is the oscillated event rate in bin i for the parameter set $(\Delta m^2_{41}, \sin^2 2\theta)$, and E_{ij} is the covariance term in bin ij.

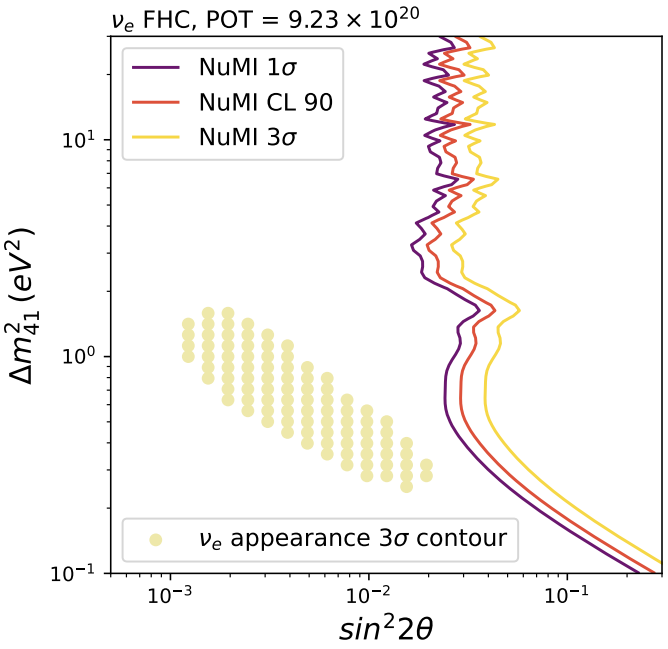


Figure 2.14: Sensitivity surface for the 3σ-favored phase space in the global ν_e appearance analysis [54] for the 3+1 model (yellow dots). For comparison, 1σ, 90%, and 3σ contours are derived in this work, using a prediction of the intrinsic ν_e event rate as well as statistical and major sources of systematic covariance.

Figure 2.14 shows the resulting sensitivity surface (yellow dots) as a function of $\sin^2 2\theta_{\mu e}$ and Δm^2_{41}, where each dot corresponds to one of the 90 parameter sets in the 3σ phase

space. Sensitivity contours for 1σ, 90%, and 3σ contours are also shown, computed using χ^2 distribution functions in the `SciPy` statistics library [56]. Even sensitivity at the level of 1σ lies far from the region mapped out by the 3σ-favored phase space from the global ν_e appearance analysis. This implies that the measurements described in this work are not sensitive to sterile flavor mixing effects of the assumed 3+1 model.

How might this sensitivity improve in the limits of no statistical uncertainty (infinite POT), or no systematic uncertainty (a perfect simulation)? Figure 2.15 (left) displays 90% confidence level contours for increasing orders of magnitude in the projected NuMI FHC sample. With one order of magnitude higher statistics, this sensitivity just skims the level required to observe sterile oscillation effects. The contours begin to converge to the case of infinite statistics (shown in black) at 9.23×10^{22} POT. At this point, MicroBooNE would be able to explore a significant region of the 3σ-favored phase space. The curve representing infinite statistics can be interpreted as the limit on sensitivity due to current systematic uncertainties.

On the other hand, the right plot of Figure 2.15 displays 90% confidence level curves for the projected 9.23×10^{20} POT with (red) and without (black) systematic uncertainties. Even with a perfect simulation of the experiment, MicroBooNE's sensitivity would not reach the 3σ phase space contour. This can be interpreted as the limit on sensitivity due to collected POT. Said another way, the largest limiting factor in this analysis is statistical.

The study sketched in this section is specific to the reconstruction, event selection, and statistics used for the measurements of this work, and does not represent what could generally be achieved in probing sterile flavor mixing with the NuMI beam at MicroBooNE. Notably, the projected sensitivity curves do not include the RHC portion of the NuMI dataset, which more than doubles the amount of statistics; this could have a positive impact on MicroBooNE's sensitivity. On the other hand, ν_e disappearance effects nor sub-dominant sources of systematic uncertainty are accounted for, both of which would suppress the sensitivity.

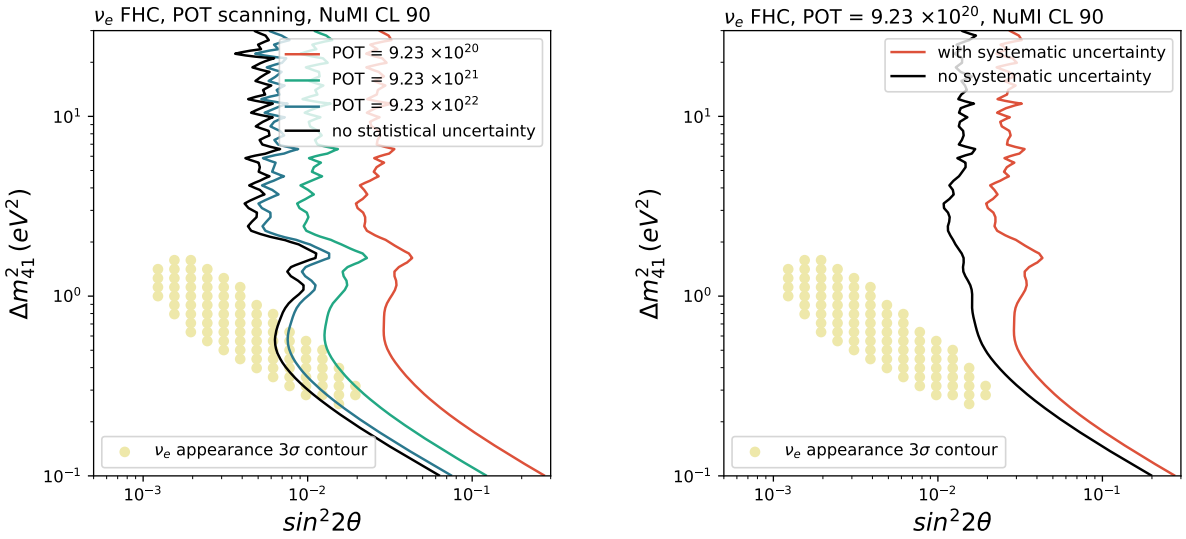


Figure 2.15: 90% confidence level sensitivity curves for the limits of no statistical uncertainty (left) and no systematic uncertainty (right), derived in this work. With enough statistics, MicroBooNE's sensitivity to sterile neutrino effects from the NuMI beam could be reached.

Also worth mentioning is that though the 3+1 model is arguably the most popular assumed explanation for the short baseline neutrino anomaly, other theories for sterile flavor mixing do exist that are not considered here.

The measurements described in this work use a combination of FHC and RHC data for a combined total of 7.0×10^{20} POT, which is $\sim$ 75% of the projected FHC POT used in this sensitivity analysis. The depletion in POT, increased systematic covariance, and the expectation of reduced efficiency in selecting ν_e events from the RHC sample are all anticipated to lessen the sensitivity of this work to sterile oscillation effects compared to the study shown here. All of that to say, the intrinsic NuMI ν_e flux prediction at MicroBooNE without oscillations—as shown in Figure 2.9—is a safe approximation for the following measurements.

CHAPTER 3

THE MICROBOONE DETECTOR

The liquid argon time projection chamber (LArTPC) is a type of detector recognized as particularly advantageous for studying neutrinos since the late 1970s [57]. In general, TPCs are equipped with the capability to collect both charge (ionization) and light (scintillation) resulting from incoming particles, and use this information to reconstruct three-dimensional images of the particle interactions. As a target, liquid argon is attractive because it is both dense and inert—maximizing the probability of a neutrino interaction while simultaneously minimizing the recombination of ionized particles that can suppress neutrino signals. It is also relatively cheap and abundant, allowing us to feasibly construct multi-ton scale neutrino detectors that further increase the probability of interaction.

This chapter provides a brief description of the experimental design and readout of the MicroBooNE detector. A detailed report can be found in [58].

3.1 Experimental design

MicroBooNE's LArTPC is housed inside of a cylindrical, stainless steel cryostat filled with 170 metric tons of liquid argon. Inside the cryostat is a rectangular field cage with dimensions $232.5 \times 256.0 \times 1036.8$ cm. The active volume inside of the field cage holds 85 metric tons of liquid argon. The field cage is also equipped with a cathode and a set of three anode planes on opposite facing sides. Behind the anode is an array of 32 8-inch Hamamatsu photomultiplier tubes (PMTs), which acts as MicroBooNE's light collection system. Fully immersed in liquid argon and facing into the active volume, the array is designed to detect scintillation from particles with kinetic energies as low as 40 MeV.

A cross section of this system is shown in Figure 3.1. In this perspective, the primary beamline (BNB) is directed out of the page in the z direction, x is the horizontal, and y is

the vertical. The cathode and anode are situated parallel to the y axis. The origin of this coordinate system lies at the vertical center of the upstream face of the LArTPC, at the x position of the innermost anode plane.

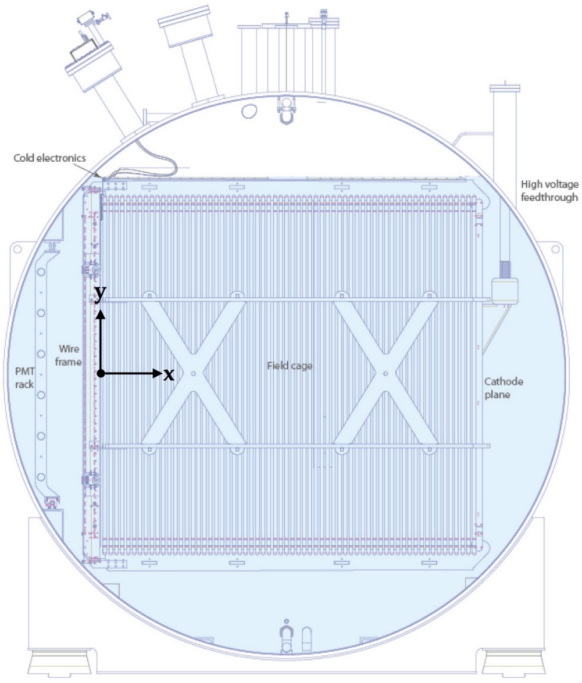


Figure 3.1: A cross section of the MicroBooNE LArTPC. In this perspective, the primary beam (BNB) is directed out of the page in the z direction. The origin of the detector coordinate system lies at the vertical center of the upstream face of the LArTPC, at the x position of the innermost anode plane. Figure from [58].

The cathode is made of steel and operated at -70 kV. The anode is comprised of three wire planes: two induction (U and V) planes, which are oriented at $\pm 60^\circ$ from vertical (y), and one collection (Y) plane with vertically oriented wires. A schematic of this setup is shown in Figure 3.2. There are a total of 8256 wires across all three planes. The anode is supplied with bias voltages of -110 volts (U plane), 0 volts (V plane), and $+230$ volts (Y plane). A uniform electric field of 273 V/cm is generated across the active volume in the x direction using the rectangular loops comprising the field cage. These loops are connected by a voltage divider chain linking the cathode and anode planes.

The detection scheme of MicroBooNE is also shown in Figure 3.2. Incoming neutrinos that interact with an argon atom will produce charged particles, which excite and ionize the surrounding argon. These particles traverse the active volume and leave behind trails of ionization electrons, like breadcrumbs revealing their paths. The electric field attracts liberated electrons to the anode, and the wires on all three planes act as sensing elements to collect information about the charge generated by the event. The anode is designed such that charge induces currents on the two innermost wire planes (U and V) and is collected on the back plane (Y).

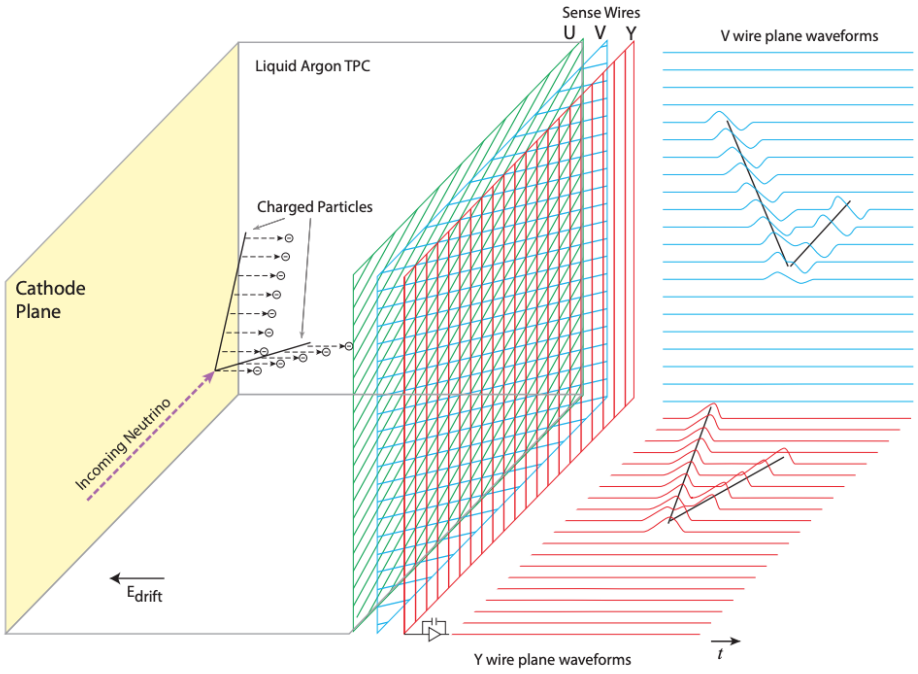


Figure 3.2: The operational principle of MicroBooNE. Incoming neutrinos that interact with the argon will liberate ionization electrons that can be detected using the set of three wire planes making up the anode. Figure from [58].

Two dimensions of the interaction (y and z) can be reconstructed based on where the ionization is deposited on the wire planes. Characteristics of the waveforms from these deposits allows for the extraction of information about particle energy, momentum, and

type.

Excited argon atoms resulting from charged particles will bond with surrounding ground state atoms to form excited argon dimers, or excimers. Excimers have $\mathcal{O}(ns)$ lifetimes thus will quickly decay via isotropic radiative emission, producing a vacuum ultraviolet (VUV) scintillation signal. Liquid argon is a prolific scintillator, producing some thousands of 128-nm photons per MeV, and it is also transparent to its own scintillation. Acrylic plates coated with tetraphenyl-butadiene are positioned in front of MicroBooNE's PMT array to shift the wavelengths of VUV photons into the visible spectrum, so that the scintillation can be converted into an amplified signal of photoelectrons for readout.

The x dimension of the interaction can be determined from the recorded drift time of the liberated electrons, i.e. the difference between the start of the event ($t = 0$) and the arrival of the ionization at the wire planes. MicroBooNE uses a combination of the beam spill timing and the scintillation signal to mark $t = 0$ of the event. The average drift velocity of the electrons is 1.076 mm/μs, corresponding to a maximum drift time of 2.3 ms in the x direction [59].

To avoid recombination with electronegative impurities (such as O_2 and H_2O) present in the active volume that can attenuate the charge signal, MicroBooNE employs a sophisticated cryogenics system to purify the liquid argon. This also aids in minimizing the amount of N_2 in the detector that can contribute to scintillation quenching effects. The pressure and temperature across the active volume are also strictly controlled as these quantities will affect the drift velocity of the electrons.

3.2 Event readout & triggering

The raw output from MicroBooNE must be processed in order to extract meaningful information from each event. Physical light signals are determined by first establishing the baseline of PMT waveforms, which express the observed number of photoelectrons in arbi-

trary ADC counts as a function of time (see Figure 3.3). A pulse-finding algorithm is then applied to identify and associate peaks above baseline with specific PMTs. Time-coincident pulses from neighboring PMTs are combined to reconstruct flashes in the detector.

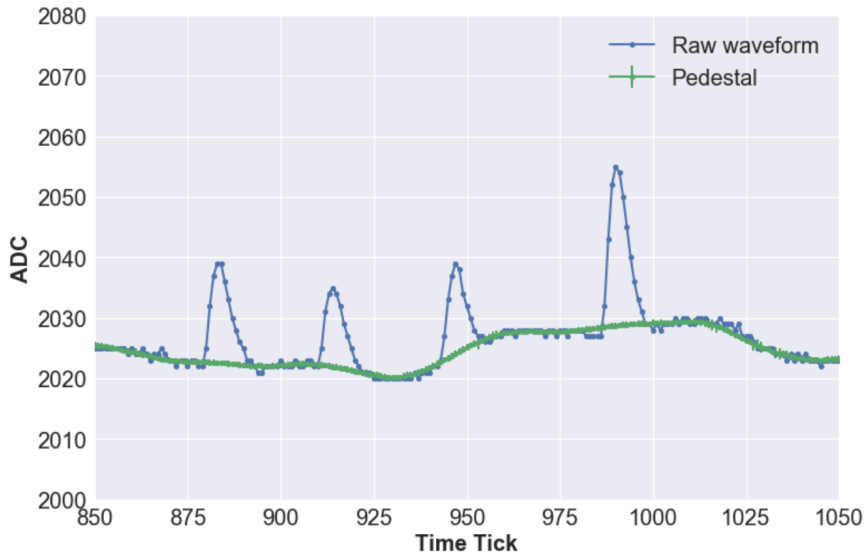


Figure 3.3: An example PMT waveform, expressed as the number of photoelectrons in arbitrary ADC counts observed as a function of time. Physical light signals are identified by establishing the baseline of the waveform (green), then using a pulse-finding algorithm to associate peaks above baseline with specific PMTs. Figure from [60].

Recorded TPC waveforms, on the other hand, are a convolution of three components: 1) the initial distribution of charge in the bulk of the detector and how it is shaped by diffusion and absorption effects during drift, 2) the response of currents induced on nearby U and V plane wires as the charge drifts toward the Y plane, and 3) the response of the electronics that receive, amplify, and shape the induced current signals. The number of ionization electrons arriving at each wire (in arbitrary ADC counts) as a function of time, as shown in Figure 3.4(a), is deconvolved from other effects shaping the waveform using the procedure detailed in [61]. Each pulse in the deconvolved waveform is fit with a Gaussian distribution to form a "hit," defined as the number of ionization electrons detected on a single wire at a definite drift time.

Figure 3.4(b) shows a single wire waveform expressed as a heat map, and Figure 3.4(c)

illustrates how positioning these expressions from adjacent wires on one of the anode planes can be used to visualize particle trajectories. The horizontal axis of (c) corresponds to wire number, and the vertical axis corresponds to electron drift time. Color is proportional to the height of the waveform pulse, or the amount of charge deposited on the wire, at a given time. These two-dimensional images are referred to as event displays in MicroBooNE.

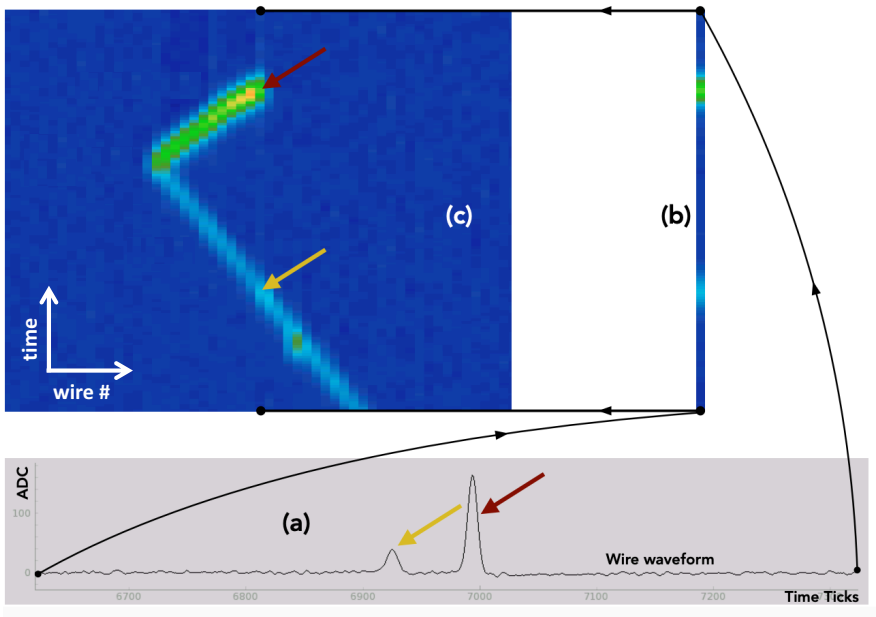


Figure 3.4: An deconvolved TPC waveform (a) showing the number of ionization electrons (in arbitrary ADC counts) recorded by a single wire as a function of electron drift time. This waveform can be expressed as a heat map (b). Positioning these expressions from adjacent wires on a single anode plane can be used to construct two-dimensional event displays in MicroBooNE (c). Figure adapted from [60].

The TPC and PMT readout systems in MicroBooNE are synchronized to a common 16 MHz clock with a frame size of 1.6 ms. Interactions in MicroBooNE are recorded using analog signals, which are then amplified and digitized with low-noise electronics. Digital signals are passed to a data acquisition system (DAQ) and written to disk. However, as a surface detector, MicroBooNE is subject to a significant amount of cosmic ray activity that is unrelated to neutrino interactions arising from NuMI or the BNB. To avoid the unnecessary storage of excessive amounts of data, events are filtered through the triggers described below.

The TPC readout consists of two data streams. The first is a continuous data stream that is used for analyses of data unrelated to the beam. The second, used in this work, is a triggered data stream based on the pulsed timing structure of the NuMI (as well as the BNB). When the beam spills, it alerts the DAQ to record 4.8 ms, or three frames worth, of data: 1.6 ms before and 3.2 ms after the spill. This flagging is called the hardware trigger, and ensures any ionization activity related to the event that occurs beyond the maximum drift time window (2.3 ms) is collected.

Roughly 1% of NuMI beam spills will cause a neutrino interaction. Thus, a software trigger is implemented to discard events with light levels below what is expected when a neutrino interacts in MicroBooNE. The PMT readout identifies scintillation that occurs in coincidence with a spill. If the event is above the configured threshold of the software trigger, the DAQ stores 6.4 ms of data: the frame containing the spill, the frame before the spill, and the two frames after the spill. For the NuMI dataset, the software trigger threshold was 9.5 photoelectrons during initial MicroBooNE Runs but was adjusted to 5.75 photoelectrons during Run 3 (see Table 2.2).

Additionally, MicroBooNE has an external (EXT) trigger that collects data when there is no beam, e.g. in between spills or when the Fermilab accelerator complex is shut down for maintenance. The EXT trigger alerts the DAQ to record three frames worth of beam-off data at a configurable frequency in a similar manner as the hardware trigger, but anti-coincident to a beam spill. To obtain a high statistics sample of pure cosmic activity, randomly selected EXT events are also written to disk upon passing the software trigger.

3.3 Interaction signatures

Arguably the greatest advantage of LArTPCs for neutrino studies is the ability to discriminate particles based on calorimetric information (energy deposit) in the detector. More specifically, electrons and heavier charged particles resulting from a neutrino interaction in

MicroBooNE produce distinct ionization signatures that allow us to filter, or select, events with a desired topology for cross section studies. This section summarizes the behavior of various particles in LArTPCs, which forms the foundation for algorithms used in the event reconstruction described in Section 4.2. A full explanation of radiation detection and measurement can be found in [62].

Charged particles in MicroBooNE collide and interact electrostatically—via the Coulomb force—with orbital (bound) electrons of the surrounding argon atoms.[1] These collisions transfer energy to bound electrons, causing them to excite or ionize and produce the telltale signatures of light and charge described in Section 3.1.

Due to their mass, heavy charged particles like protons and muons experience only a small fraction of energy loss and a slight decrease in velocity during each encounter with a bound electron. This, and the fact that heavy charged particles are interacting isotropically with many bound electrons at any given time, makes for little to no observable deflection along the particle's trajectory in the detector. As a result, heavy charged particles generally traverse MicroBooNE in straight tracks until they are depleted of energy and stopped.

The energy loss per unit length ($-\frac{dE}{dx}$) for a heavy charged particle is given by the Bethe formula:

$$\frac{dE}{dx} = -\frac{4\pi e^4 z^2}{m_0 v^2} NB \tag{3.1}$$

where $B \equiv Z\left[ln\frac{2m_0v^2}{I} - ln\left(1-\frac{v^2}{c^2}\right) - \frac{v^2}{c^2}\right]$. In these equations, v and z are the velocity and charge number of the heavy charged particle, e is the elementary charge constant, and m_0 is the electron rest mass. N, Z, and I represent the number density, atomic number, and average ionization energy of the target medium. Liquid argon has values of $N = 2.1 \times 10^{16}$ atoms/m^3, $Z = 18$, and $I = 23.6$ eV/pair [58].

Note that the rate of energy loss is inversely proportional to v^2, meaning that heavier

1. In theory, interactions also occur with argon nuclei, though these are rare and do not contribute significantly to detector response.

(therefore slower) particles deposit more energy per unit length in the detector. This makes sense given that a slower particle inherently spends more time in the vicinity of a given bound electron, thus exerts a larger force and transfers a larger amount of energy. In MicroBooNE, this feature allows us to distinguish muons (106 MeV/c^2 rest mass) from protons (938 MeV/c^2 rest mass): Muons tend to travel further and deposit less energy along their tracks than protons.

Figure 3.4(c) is an example of a ν_μ CC interaction candidate: $\nu_\mu + {}^{40}Ar \rightarrow \mu^- + p^+$. While the neutrino is invisible due to its neutral charge, a short proton-like track and longer muon-like track emanate from the interaction vertex. The color of the proton, indicating its higher energy deposit per unit length relative to the muon, is a result of its heavier mass.

In contrast, free electrons resulting from neutrino interactions are lighter (0.511 MeV/c^2) and quicker. Free electrons overall deposit less energy per unit length than heavier charged particles, but a larger percentage of their energy can be transferred during collisions with bound electrons. The rate of collisional energy loss for free electrons is given by:

$$\left(\frac{dE}{dx}\right)_c = -\frac{2\pi e^4}{m_0 v^2} NB \tag{3.2}$$

where $B \equiv Z\left[ln\frac{m_0 v^2 E}{2I^2(1-\beta^2)} - (ln2)(2\sqrt{1-\beta^2} - 1 + \beta^2) + (1-\beta^2) + \frac{1}{8}\left(1 - \sqrt{(}1-\beta^2)^2\right)\right]$ and $\beta \equiv \frac{v}{c}$. Because the mass of the free electron is equivalent to that of its collision partner, it is prone to significant, observable deflections from its original path.

Like all charged particles, free electrons emit *bremsstrahlung*, or electromagnetic "braking" radiation, when their direction changes.[2] The rate of radiative energy loss for free electrons is given by:

$$\left(\frac{dE}{dx}\right)_r = -\frac{NEZ(Z+1)e^4}{137 m_0^2 c^4}\left(4ln\frac{2E}{m_0 c^2} - \frac{4}{3}\right). \tag{3.3}$$

2. Heavy charged particles can produce *bremsstrahlung* as well, though the energy loss is greatly suppressed because it is inversely proportional to the rest mass of the particle, as indicated in Equation 3.3. In MicroBooNE, this effect is negligible.

This rate is directly proportional to E and the square of Z, indicating that more *bremsstrahlung* is emitted with higher free electron energy and denser target material. The total energy loss of free electrons traversing MicroBooNE is $\left(\frac{dE}{dx}\right)_c + \left(\frac{dE}{dx}\right)_r$.

Bremsstrahlung photons will travel a few centimeters before interacting. Lower energy photons will either be absorbed by an argon atom ($E_\gamma <\sim 0.1$ MeV) or Compton scatter off a bound electron ($E_\gamma \sim 0.1 - 10$ MeV), but most photons generated in MicroBooNE are energetic enough ($E_\gamma >\sim 10$ MeV) to pair produce, i.e. $\gamma \rightarrow e^+ + e^-$ where the full energy of γ is split between the positron and electron. The positron will immediately annihilate ($e^+ \rightarrow 2\gamma$) and the electron will subsequently lose energy via the collisional and radiative processes described above, effectively creating an electromagnetic shower of successive photons and electrons until energy is depleted.

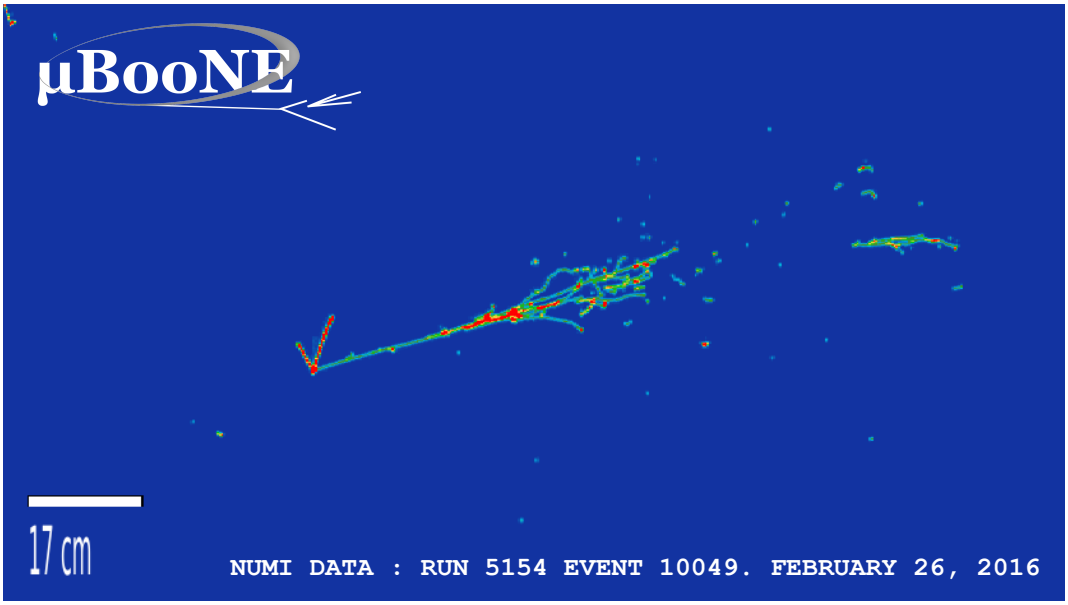


Figure 3.5: A ν_e event candidate reconstructed by MicroBooNE, comprised of one electron-like electromagnetic shower and $N = 2$ proton-like tracks, from the NuMI dataset.

Figure 3.5 shows an electron neutrino interaction candidate: $\nu_e + {}^{40}Ar \rightarrow e^- + p^+$. An unambiguous electromagnetic shower emanating from the interaction vertex indicates the presence of an electron. Also identifiable is a short, bright red track characteristic of a proton-like trajectory. Longer tracks with lower energy deposit per unit length, observed on

either side of the neutrino interaction, are likely cosmic rays (muons).

Muon neutrino interactions that produce a π^0 can mimic ν_e signatures in the detector due to neutral pion decay ($\pi^0 \rightarrow 2\gamma$), generating photons that can set off their own electromagnetic showers, as shown in Figure 3.6. This makes interactions involving the creation of a π^0 one of the largest sources of background in MicroBooNE's electron neutrino studies. However, photons are distinct from electrons because they travel some distance away from the interaction vertex before pair producing, and the showers generally come in pairs with a higher $\frac{dE}{dx}$ at the starting point. One or more of these handles can be used to discriminate ν_e interactions from π^0 events, demonstrating the advantage LArTPCs have over other types of neutrino detection mechanisms.

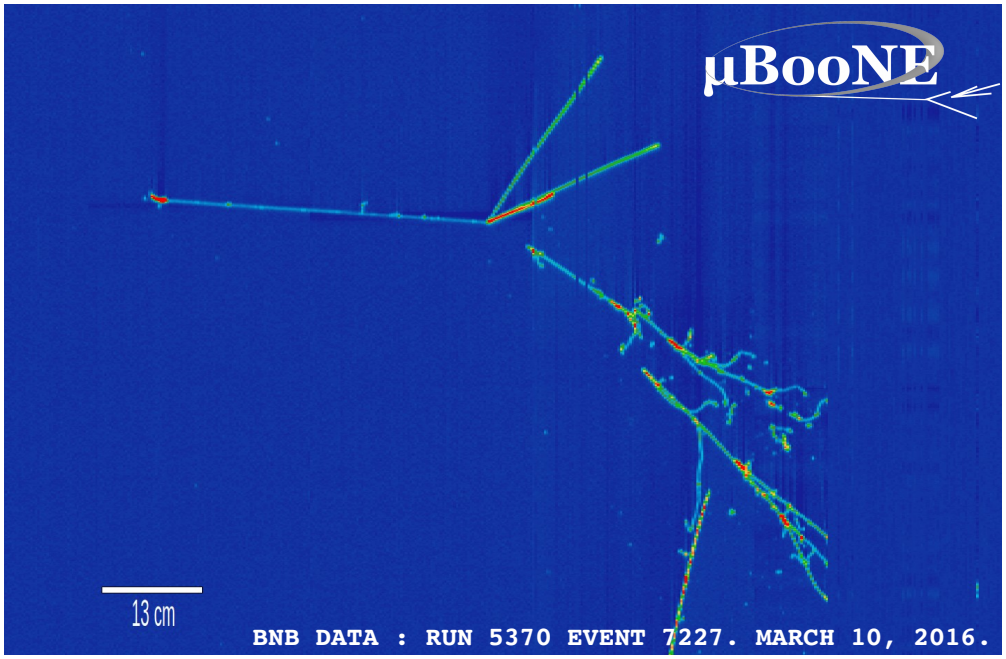


Figure 3.6: A $\pi^0 \rightarrow 2\gamma$ event candidate event display from the BNB dataset. The photons set off electromagnetic showers that become one of the largest sources of background in MicroBooNE's ν_e studies. As shown, photons generally travel some centimeters away from the neutrino interaction vertex before pair producing, and the resulting showers come in pairs with high energy deposit at the shower starting point. One or more of these handles can be used to discriminate ν_e and π^0 events in the LArTPC.

CHAPTER 4

EVENT SIMULATION AND RECONSTRUCTION

In general, the analysis of data collected by MicroBooNE relies on theoretical simulations of neutrino behavior, as well as the reconstruction of events in the detector, to accurately interpret experimental observations. Simulated event samples help characterize the detector response, i.e. how effects in the LArTPC "smear" reconstructed observables—position, charge deposit, etc.—away from their true values. They are also used to understand the expected signal and background distributions of collected data, information from which a robust selection algorithm can be developed. The simulated flux, on the other hand, is an essential ingredient in extracting the cross section.

MicroBooNE's simulation and reconstruction of neutrino events is performed within the LArSoft framework [63], a set of software tools built for the analysis of data from liquid argon neutrino experiments. This chapter summarizes the standard simulation and reconstruction workflows, and describes the NuMI event samples used for this work.

4.1 Simulation workflow

The simulation of neutrino events is performed as follows. Proton interactions with the NuMI target, hadronic re-interactions, and the decay of these particles into neutrinos are modeled using `Geant4 v4.9.2.p03` and a geometric model of the beamline [33]. This flux serves as input to `GENIE` [64], a Monte Carlo (MC) neutrino event generator that simulates initial $\nu + {}^{40}Ar$ weak interactions as well as the propagation of resulting hadrons and leptons through the target nuclear medium (so-called final-state interactions), returning a list of final-state particles entering the detector and their kinematics. Specifically, MicroBooNE uses `GENIE v3.0.6 G18_10a_02_11a` [65], a tune developed to fix an underestimation of simulated ν_μ CC events as compared to data. (This discrepancy carries over to the ν_e CC

simulation because the same underlying cross section models are used in `GENIE` for both neutrino flavors.) The propagation of final-state particles and any secondaries they produce in the detector is simulated with `Geant4 v10.3.03c`. This includes the amount of energy deposited at each point along the particle paths.

Custom algorithms in LArSoft are employed to simulate the readout of scintillation and ionization signals, referred to as the detector response. For each simulated event, the position and amount of deposited energy is used to estimate the number of photons at each point along the particle's path. A photon visibility library of the detector, which takes into account the attenuation and scattering effects influencing light traversing the TPC, is then used to estimate the number of photoelectrons measured by the PMT array. This results in a simulated PMT waveform. The energy deposit is also used to estimate the amount of ionization electrons liberated along the particle trajectories. Electron drift to the wire planes is simulated, accounting for ionic recombination and electric field distortion due to the accumulation of argon ions at the cathode (known as space charge).

Simulated neutrino signals are then overlaid onto data from MicroBooNE's EXT sample in order to replicate cosmic ray activity occurring in-time with neutrino interactions in a data-driven way. This bypasses the need to use the air shower simulation toolkit `CORSIKA` [66], which has demonstrated limitations regarding angular coverage, as well as modeling of the LArTF building, space charge effects in the detector, and interaction rates for shorter cosmic muons [67].

At this point, simulated waveforms are processed in the same manner as data, following the procedure outlined in Section 3.2. EXT events with pure cosmic activity, absent of a simulated neutrino interaction, also undergo this processing for a complete simulation (MC+EXT) of MicroBooNE data. Notably, reconstructed hits are derived from TPC waveforms which serve as the input for event reconstruction.

4.2 Reconstruction workflow

The purpose of reconstruction is to build physics-relevant quantities from the detector's PMT and TPC readout systems, which become the foundation for data analysis. To reconstruct interactions in the LArTPC, MicroBooNE employs an automated, multi-algorithm approach to pattern recognition called `Pandora` [68]. This toolkit uses topological and calorimetric information to assemble three-dimensional particles traversing through the detector and identify candidate neutrino interactions. The two reconstruction paths shown in Figure 4.1, `PandoraCosmic` and `PandoraNu`, have been developed for use within MicroBooNE.

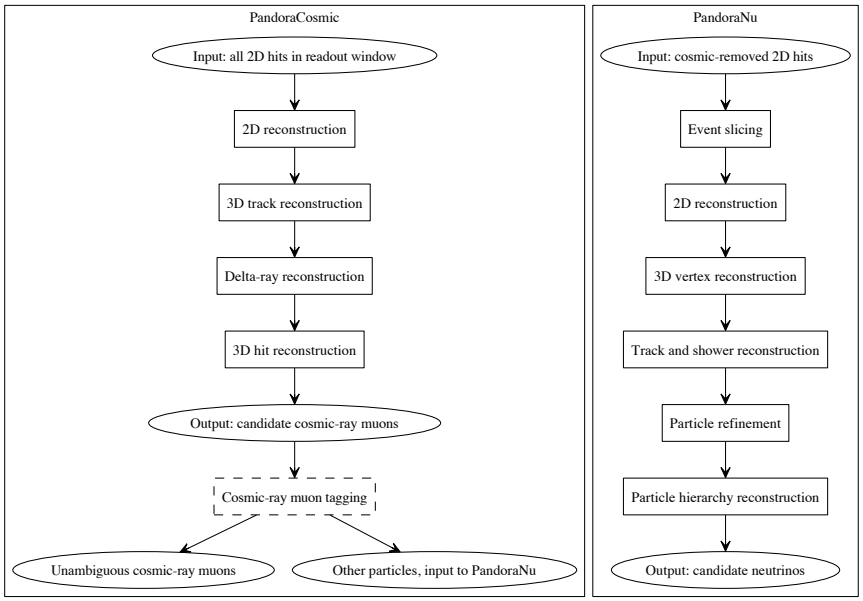


Figure 4.1: A summary of the two reconstruction paths developed by MicroBooNE using `Pandora`. Hit information from TPC event readout becomes the input for `PandoraCosmic`, which removes obvious cosmic activity. The cosmic-removed hit collection is the input to `PandoraNu`, which outputs a neutrino candidate, including a reconstructed vertex position and hierarchy of charged particles associated with the interaction. Figure from [68].

`PandoraCosmic` takes all hits from a TPC event readout as input. Associated hits from each anode plane are grouped together to form 2D clusters. Representations of 3D charged particles and their positions (known as spacepoints) are built by matching associated clusters across planes. `PandoraCosmic` classifies these representations under a cosmic hypothesis,

identifying start and end points for reconstructed tracks and assuming that any nearby showers are delta rays, low energy electrons emitted from the passage of muons through the LArTPC. The output of `PandoraCosmic` is a list of candidate cosmic interactions. This list is examined by an external tagger that flags unambiguous background and generates a new, cosmic-removed hit collection that is used as input for `PandoraNu`.

For each event, `PandoraNu` organizes information into slices—lists of hits belonging to the same interaction—with the intent of isolating neutrino-induced signals from cosmic ray activity. Each slice is processed using the same 2D track-oriented clustering algorithms of `PandoraCosmic`. Slices are analyzed to build a list of possible interaction vertex candidates, which are fed into a Support Vector Machine (SVM) to determine the most likely vertex for the event. The SVM also returns a score (0 or 1) based on how closely the candidate slice resembles a neutrino. `PandoraNu` then reconstructs tracks and primary showers emerging from this vertex, and matches them with any secondaries produced. A refinement algorithm is run over the reconstructed particles, with a focus on completing the hit collections of particularly sparse showers. The output of `PandoraNu` is a candidate neutrino, which includes a reconstructed vertex position and hierarchy of charged particles associated with the interaction, an example of which is shown in Figure 4.2. Charged particles are also assigned a score indicating their likeness to a shower (0 to 0.5) or track (0.5 to 1) topology.

Calorimetric energy reconstruction is performed differently depending on whether `Pandora` classifies an object as a shower or a track. For showers, the charge for all hits in the object is summed, then converted to E_{shower} using a fixed calibration factor that scales ADC counts to MeV with the assumption that the electrons and positrons creating the shower are minimally ionizing, thus lose energy at a fixed rate of 2.3 MeV/cm. This assumption leads to a $\sim 2\%$ contribution to the overall shower energy resolution [69]. E_{shower} is subsequently corrected to account for an observed energy bias from charge under-clustering in `Pandora`, as well as detector non-uniformities, such as space charge effects and electron attachment to

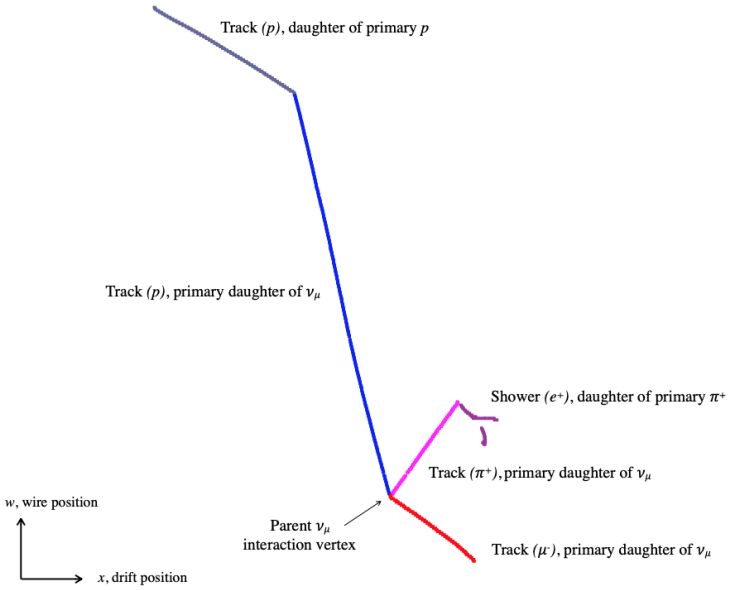


Figure 4.2: An example of the reconstructed `PandoraNu` output for a simulated ν_μ CC event with a muon (red), proton (blue), charged pion (pink) in the final topology. The parent neutrino has a reconstructed interaction vertex and three track-like particles emanating from this point. The charged pion decays into a μ^+, which subsequently decays into an e^+ that manifests as a shower-like secondary. The proton scatters off a nucleus and generates a secondary particle reconstructed as a track. Figure from [68].

impurities [70].

For tracks, the charge of each hit dQ_{track} is converted to dE_{track} using the inverse modified Box model:

$$\frac{dE}{dx} = \frac{e^{\beta \times C W_{ion} \frac{dQ}{dx}} - \alpha}{\beta} \tag{4.1}$$

where $W_{ion} = 23.6$ eV [71], denoting the work function for ionizing an argon atom, and C is a gain factor that converts ADC counts into the number of collected electrons. The constants α and β are free parameters experimentally determined by the ArgoNeuT collaboration [72]:

$$\alpha = 0.93 \pm 0.02 \tag{4.2}$$

$$\beta = \frac{k_b}{\rho \times E_{field}}, \ where \ k_b = 0.212 \pm 0.002 \tag{4.3}$$

with ρ representing the density of liquid argon, 1.383 g/cm^3, and E_{field} obtained from MicroBooNE's position-dependent electric field maps [73]. E_{track} is calculated by integrating dE_{track} over all hits associated with the track. Like shower objects, this energy is also corrected to account for detector non-uniformities.

4.3 Event samples

NuMI data and simulation samples processed through the reconstruction framework are output into `ROOT` files containing information organized in the `TTree` data structure [74]. This section details the types of samples used, the reweighting factors applied, and the normalization procedure enacted to scale the prediction to data.

4.3.1 Simulated neutrino samples

Simulated neutrino samples, commonly referred to as "overlay" in reference to the merging with EXT data, come in three types: standard, intrinsic, and dirt. Standard overlay samples consist of `GENIE`-generated ν_μ, $\bar{\nu}_\mu$, ν_e, and $\bar{\nu}_e$ interactions inside the MicroBooNE cryostat, and are simulated in relative fractions and energies based on the input flux. Table 4.1 lists the size of the FHC Run 1 and RHC Run 3 standard overlay samples used in this work.

Type	**POT**
FHC Run 1 Standard Overlay	2.33652×10^{21}
RHC Run 3 Standard Overlay	6.06172×10^{21}

Table 4.1: The size of the standard overlay samples used in this work.

Intrinsic overlay samples consist only of `GENIE`-generated ν_e and $\bar{\nu}_e$ CC interactions inside MicroBooNE's active volume, simulated with energies based on the input flux. Once created, $\nu_e/\bar{\nu}_e$ CC events existing in standard overlay are removed and replaced with those from the intrinsic overlay sample, and weighted appropriately to scale to the correct POT. This is

done to boost the signal-to-background ratio of the event samples and overcome statistical limitations inherently associated with studying electron neutrinos from a primarily ν_μ or $\bar{\nu}_\mu$ beam. Table 4.2 lists the size of the intrinsic overlay samples used in this work.

Type	**POT**
FHC Run 1 Intrinsic Overlay	2.37838×10^{22}
RHC Run 3 Intrinsic Overlay	2.53451×10^{22}

Table 4.2: The size of the intrinsic overlay samples used in this work. These are generated to replace $\nu_e/\bar{\nu}_e$ CC interactions in standard overlay in order to boost the signal-to-background ratio in the event samples.

Dirt overlay samples consist of `GENIE`-generated ν_μ, $\bar{\nu}_\mu$, ν_e, and $\bar{\nu}_e$ interactions occurring outside of the MicroBooNE cryostat that produce enough light to pass the NuMI software trigger. Table 4.3 shows the size of the dirt overlay samples used in this work.

Type	**POT**
FHC Run 1 Dirt Overlay	1.42143×10^{21}
RHC Run 3 Dirt Overlay	10.32262×10^{20}

Table 4.3: The size of the dirt overlay samples used in this work.

Events in the overlay samples are based on underlying flux and cross section models that are insufficiently constrained. This can lead to poor agreement between simulation and data, an inaccuracy that carries over into the final measurements. To alleviate this, the following corrections are applied to each interaction in the overlay samples.

Section 2.3 described data-driven constraints on the `Geant4` hadron production and absorption models used to generate MicroBooNE's NuMI flux prediction. `PPFX` can also be used to constrain the underlying flux models used to simulate the event rates. Using information about the neutrino ancestry and parent decay kinematics, a CV reweighting factor is derived for each event in the overlay samples. Relevant to the measurements in this work,

applying the PPFX reweight decreases the number of simulated ν_e CC interactions in the active volume by 9% (10%) in FHC Run 1 (RHC Run 3).

Regarding cross section models, MicroBooNE analyses employ the GENIE tune referenced in Section 4.1, which also manifests as a CV weight applied to each event in the overlay samples. The GENIE tune weight increases the simulated number of ν_e CC interactions in the active volume by 13% (11%) in FHC Run 1 (RHC Run 3).

The final correction stems from a flat overestimation of simulated dirt events observed when comparing with NuMI data. To mitigate this, the dirt overlay sample is scaled down by 35% (55%) in FHC Run 1 (RHC Run 3), and a 100% uncertainty on the dirt contribution to the event rate is assumed. As nearly all dirt interactions are removed due to the selection algorithm described in Chapter 5, this scaling does not have a significant impact on the final results.

4.3.2 Data

As described in Section 3.2, two types of data samples are collected in MicroBooNE. The beam-on sample consists of any detector activity that sets off MicroBooNE's hardware trigger: neutrino events and cosmic ray interactions that sometimes produce signals occurring in-time with the beam. The collected NuMI beam-on samples are shown in Table 4.4, along with the collected POT and the number of times the hardware trigger was prompted.

Type	**POT**	**Hardware Triggers**
FHC Run 1 Beam ON	2.000×10^{20}	5268051.0
RHC Run 3 Beam ON	5.014×10^{20}	10372943.0

Table 4.4: The NuMI beam-on samples, including the collected POT and number of hardware triggers.

The EXT sample, on the other hand, is taken intentionally when the beam is off with an external trigger that fires anti-coincident to a spill. A portion of the EXT sample is used

to create the overlay samples listed in Section 4.3.1. The rest is saved for estimating the portion of the beam-on sample consisting of in-time cosmic background, in order to build a full prediction (MC+EXT) of collected data. The collected EXT samples and number of external triggers that flagged an interaction are shown in Table 4.5. Only a specified fraction of EXT triggers are written to disk, so these values have been multiplied by a prescaling factor to accurately reflect the number of events recorded.

Type	**EXT Triggers**
FHC Run 1 EXT	9199232.74
RHC Run 3 EXT	32905506.725

Table 4.5: The NuMI EXT samples including the collected number of hardware triggers, which have been multiplied by a prescaling factor.

4.3.3 Normalization procedure

Normalizing the overlay and EXT samples is necessary to build an accurate representation of beam-on data. The scaling applied to overlay is simply the ratio between the POT values:

$$s_{overlay} = \frac{\mathrm{POT}_{\text{beam-on}}}{\mathrm{POT}_{\text{overlay}}}$$

where $s_{overlay}$ is distinct for the standard, intrinsic, and dirt samples.

EXT does not have associated POT values, so these samples are scaled using the ratio between the number of triggers:

$$s_{EXT} = \frac{trig_{\text{beam-on}}}{trig_{\text{EXT}}}$$

where the EXT triggers have been multiplied by a prescaling factor. The (seemingly absurd) expectation inherent in this scaling is that none of the beam-on triggers contain neutrinos. This is a fine assumption for the on-axis BNB, but breaks down for NuMI because of its greater beam intensity, which results in a larger proportion of beam spills creating a neutrino

interaction. For this reason, the EXT samples are scaled down by an additional 2% to account for neutrino occupancy.

The reweighting factors and normalization procedure can be verified by comparing the MC+EXT prediction to beam-on data for events that have passed the software trigger. Figure 4.3 shows the distributions of FHC Run 1 (left) and RHC Run 3 (right) beam-on data as a function of the timing of the largest flash in the event on top of a stacked histogram of the normalized EXT, overlay (In Cryo MC), and dirt predictions. Central value `PPFX` and `GENIE` reweights have been applied to the overlay events. The dirt and EXT events have also been scaled down to account for unconstrained models and neutrino occupancy, respectively.

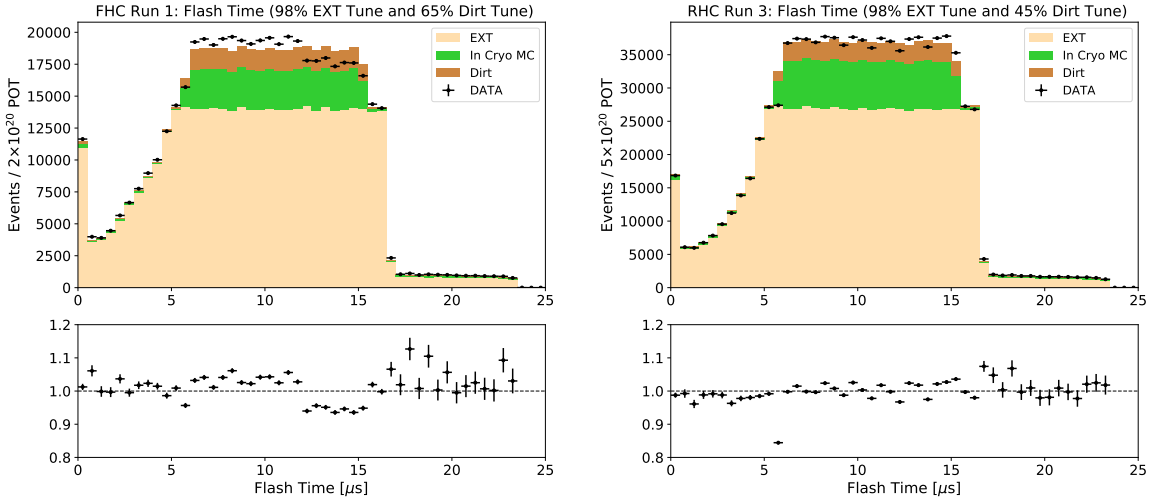


Figure 4.3: Beam-on event distributions as a function of the timing of the largest flash in the event for FHC Run 1 (left) and RHC Run 3 (right) on top of a stacked histogram of the normalized EXT, overlay (In Cryo MC), and dirt predictions.

The NuMI beam spill window for both runs occurs between 5.64 and 15.44 μs and is the primary range where neutrino-induced events populate. In FHC Run 1, the step-like shape of the spill is due to the composition of the data: roughly 70% exists in the 4+6 slip-stacking configuration, while the other 30% was collected in the 6+6 configuration. The NuMI prediction simulates neutrinos evenly over the time window, thus no step-like shape is present. This feature is not present in RHC Run 3 because all of the data was collected in the same slip-stacking configuration.

CHAPTER 5

ELECTRON NEUTRINO EVENT SELECTION

The event selection is an algorithm designed to isolate electron neutrino interactions with the $1e + Np$ topology from various sources of background. It is built using reconstructed variables available in the simulated event samples, then applied to NuMI beam-on data.

Two metrics are used to evaluate the strength of the selection. The efficiency quantifies how well the algorithm retains signal and is defined as:

$$\epsilon = \frac{\text{true signal events selected}}{\text{true signal events generated}}. \tag{5.1}$$

The purity quantifies how effective the algorithm is at removing background, defined as the fraction of selected events that are signal:

$$p = \frac{\text{true signal events selected}}{\text{total events selected}}. \tag{5.2}$$

Both metrics are calculated purely from the simulated event rate. The aim of the selection is to maximize purity while achieving as high of an efficiency as possible.

An overview of the selection workflow is illustrated in Figure 5.1. Initial requirements ensure that the events are of quality and match the signal topology at the reconstructed level. Together, these two stages are referred to as the preselection. Next, a set of constraints is loosely applied to reject obvious ν_μ CC and π^0 events. This increases the signal-to-background ratio of the event sample for input into a Boosted Decision Tree (BDT) trained to discriminate signal from background using a multivariate assessment. The BDT assigns a probability score to each event based on how likely it is to be signal. The final stage of the selection constrains this score distribution. Separate BDT models are developed for different beam operating modes; thus, while analogous and similar, the selections run over FHC Run 1 and RHC Run 3 are distinct.

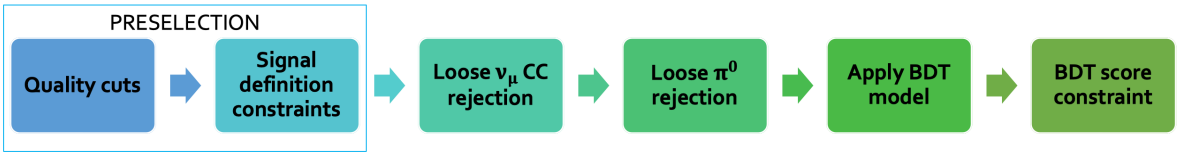


Figure 5.1: A summary of the NuMI $1e + Np$ event selection. Separate BDT models are developed for different beam operating modes; thus, while analogous, the selections run over FHC Run 1 and RHC Run 3 are distinct.

Agreement between data and prediction for reconstructed variables used in the selection is checked at each stage, so as to avoid using parameters that are poorly modeled. The goodness-of-fit is computed as:

$$\chi^2 = \sum_{ij}(x_i - y_i) * Cov_{ij}^{-1} * (x_j - y_j) \tag{5.3}$$

where x_i and y_i are the number of beam-on and MC+EXT counts in a bin i, respectively. Cov_{ij}^{-1} is the ij^{th} element of the inverse of a covariance matrix encoding the statistical and systematic uncertainty of the prediction, as well as the statistical uncertainty of the beam-on dataset.

5.1 Signal and background definitions

In this work, signal events are ν_e CC interactions with at least 1 proton and no charged or neutral pions in the final topology. Protons and charged pions must have kinetic energies greater than 40 MeV to be counted, a requirement arising from the minimum detection threshold of the LArTPC. To avoid border effects leading to poor quality reconstruction at the edges of the active volume, interaction vertices must also take place inside of a fiducial volume (FV) that exists 10 cm inward from any side of the LArTPC. An example of a selected signal candidate is shown in Figure 3.5.

Primary sources of background in the event samples are other types of neutrino interactions: ν_e events that do not fully match the signal definition and ν_μ interactions, especially

those that produce a neutral pion, whose decay products manifest as electromagnetic showers. Cosmics can also produce shower-like topologies in the detector, though these events are suppressed via the requirement of a proton-like track in the signal definition.

Event distributions in this chapter are differentiated into the categories listed in Table 5.1. EXT and DATA refer to beam-off and beam-on events, respectively, while the other categories characterize simulated interactions in the overlay samples by their truth-level information. The label ν_e CC0πNp specifically refers to the signal events of this work.

Event type	**Description**
ν_e CC0πNp	ν_e CC interactions in the FV with at least 1 proton $>$ 40 MeV kinetic energy, no charged pions $>$ 40 MeV kinetic energy, and no neutral pions. These are the signal events of the selection.
$\bar{\nu}_e$ CC0πNp	$\bar{\nu}_e$ CC interactions in the FV with at least 1 proton $>$ 40 MeV kinetic energy, no charged pions $>$ 40 MeV kinetic energy, and no neutral pions. These are topologically identical to our signal events and considered an irreducible background.
ν_e CC other	ν_e and $\bar{\nu}_e$ CC interactions in the FV that contain at least 1 neutral pion, at least 1 charged pion $>$ 40 MeV kinetic energy, and/or no protons $>$ 40 MeV kinetic energy.
ν_e NC	ν_e and $\bar{\nu}_e$ neutral current interactions in the FV.
ν_μ CC	ν_μ and $\bar{\nu}_\mu$ CC interactions in the FV with no π^0 present.
ν_μ NC	ν_μ and $\bar{\nu}_\mu$ NC interactions in the FV with no π^0 present.
ν_μ CC Nπ^0	ν_μ and $\bar{\nu}_\mu$ CC interactions in the FV that contain a π^0.
ν_μ NC Nπ^0	ν_μ and $\bar{\nu}_\mu$ NC interactions in the FV that contain a π^0.
Out FV	Neutrinos of any flavor with an interaction vertex outside of the FV, including but not limited to all dirt interactions.
EXT	Pure cosmic background from the beam-off sample.
DATA	All interactions from the beam-on sample.

Table 5.1: Plots shown in this chapter are differentiated into the following categories. EXT and DATA refer to beam-off and beam-on events, respectively, and the rest of the categories characterize simulated interactions in the overlay samples by their truth-level information.

All events in the beam-on data sample must pass the software trigger described in Section 3.2 in order to be saved. However, the MC+EXT prediction contains interactions below the light levels required by this threshold. As a result, only the subset of the MC+EXT sample passing the software trigger is used to compare to beam-on data in the plots shown in the following sections. The efficiency and purity are also estimated using the same subset of the prediction. (This leads to a negligible effect on the final results, since according to simulation,

99.98% of signal passes the software trigger threshold.)

Unless otherwise specified, the grey shaded region on the event distributions represents the statistical and systematic uncertainty of the MC+EXT event rate. (This includes the systematic event rate change due to `GENIE` uncertainties on the signal channel, a contribution that is turned off when extracting the cross section.) Error bars on the data represent the statistical uncertainty associated with the beam-on sample.

For most of the plots of this chapter, the grey error band and χ^2 value do not include a contribution from systematic uncertainty due to detector response. This is a result of statistical limitations inhibiting the ability to accurately assess bin-to-bin variance and correlations for this source of uncertainty. In lieu of this, late stage selection distributions shown in Section 5.7 incorporate a flat detector uncertainty of 12.1% for the FHC Run 1 sample and 12.9% for the RHC Run 3 sample. A full description of how these values are derived, and the evaluation of other sources of uncertainty, can be found in Chapter 6.

5.2 Quality cuts

The initial stage of the selection aims to select well-reconstructed, contained neutrino events away from the edges of the LArTPC.

The first constraint involves the `Pandora` slice identification score assigned by the Support Vector Machine as described in Section 4.2. Slices that favor the neutrino hypothesis are assigned a value of 1; those that favor the cosmic hypothesis are given a score of 0. Figure 5.2 shows the event rate as a function of `Pandora` slice ID for events that have passed the software trigger. The selection removes events that are assigned a score of 0 by `Pandora`.

Next, the reconstructed position of the neutrino interaction vertex is constrained to exist within the FV. Figures 5.3 and 5.4 display the event rates as a function of the reconstructed `x`, `y`, and `z` vertex coordinates after the `Pandora` slice identification cut. Events pass this requirement if the reconstructed vertex is located within bounds of the FV: `10 cm < x <`

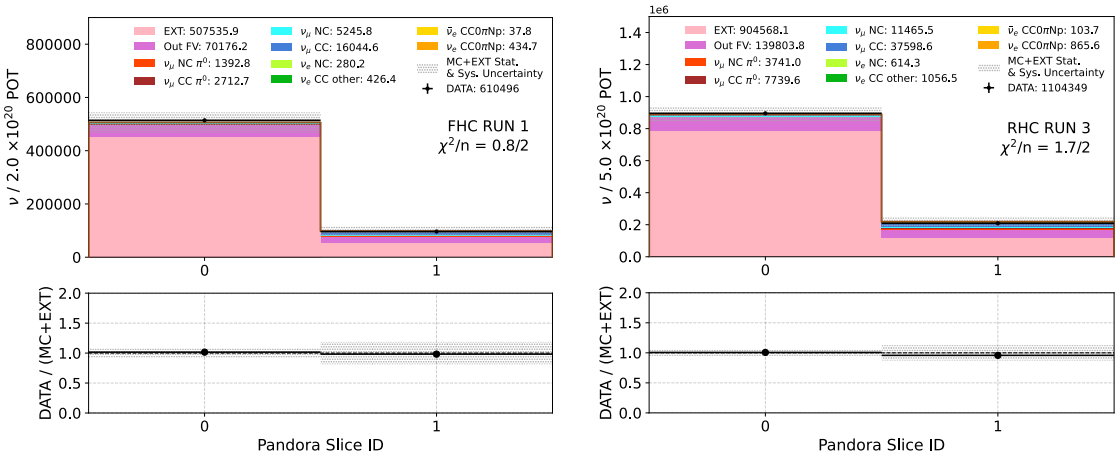


Figure 5.2: The FHC Run 1 (left) and RHC Run 3 (right) event rates as a function of the `Pandora` slice identification score for interactions that pass the software trigger.

`246 cm`, `-106 cm < y < 106 cm`, and `10 cm < z < 1026 cm`.

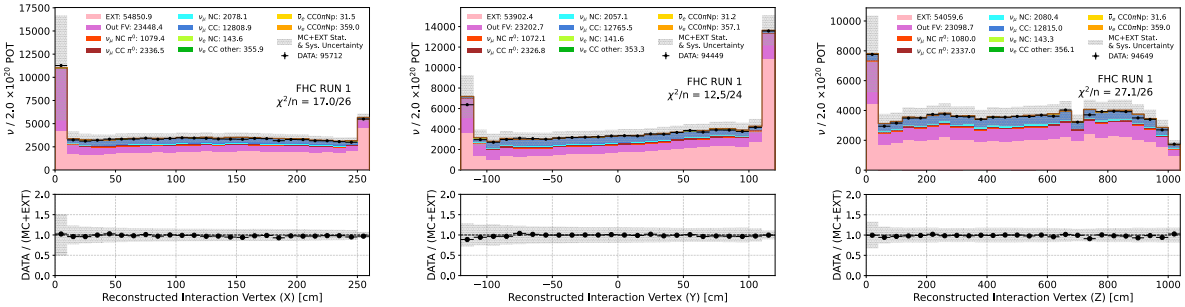


Figure 5.3: Event rates as a function of reconstructed `x`, `y`, and `z` vertex coordinates after the `Pandora` slice identification constraint for FHC Run 1.

The final cut in this stage concerns the fraction of hits in a slice that are contained within the FV. Figure 5.5 displays the event rates as a function of the contained fraction, after the slice identification and reconstructed vertex constraints have been applied. Events less than 90% contained are removed from the samples. This requirement ensures selected interactions are well-reconstructed, and also removes a significant amount of background, namely cosmic activity and higher energy ν_μ events with long tracks extending beyond the walls of the LArTPC.

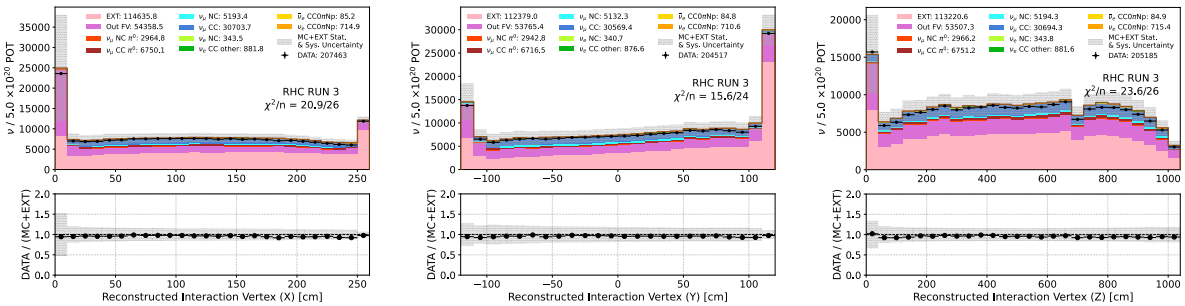


Figure 5.4: Event rates as a function reconstructed `x`, `y`, and `z` vertex coordinates after the `Pandora` slice identification constraint for RHC Run 3.

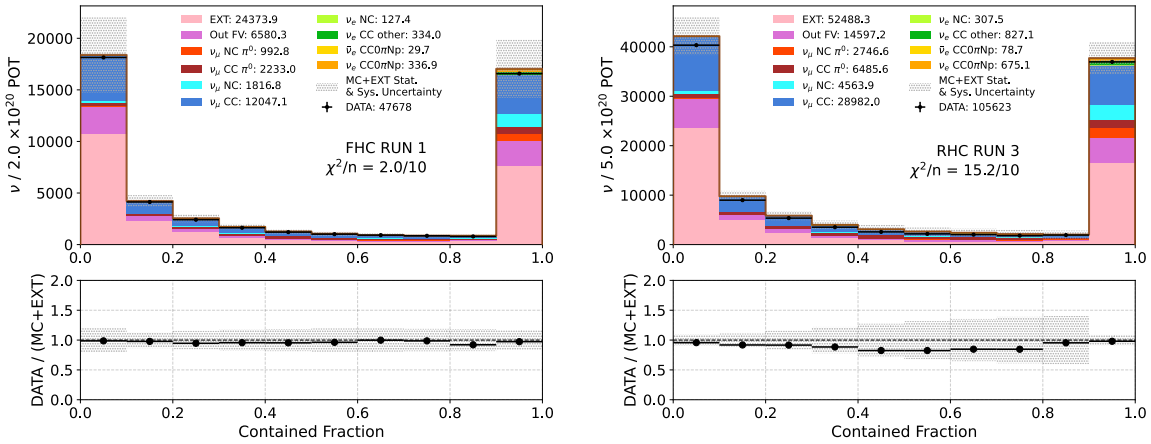


Figure 5.5: FHC Run 1 (left) and RHC Run 3 (right) event rates as a function of contained fraction for interactions that pass the `Pandora` slice identification and reconstructed vertex constraints.

After this stage of the selection, the purity for FHC Run 1 (RHC Run 3) is 1.6% (1.4%) and the efficiency is 62.0% (61.5%).

5.3 Signal definition constraints

The second stage of the selection attempts to restrict shower and track topologies based on the signal definition.

The characteristic topology of ν_e CC interactions is the presence of a single electromagnetic shower. Figure 5.6 shows the event rates as a function of the number of showers with

reconstructed vertex inside the FV after the quality cuts of the previous section have been applied. Events are required to have exactly one shower to pass this stage of the selection.

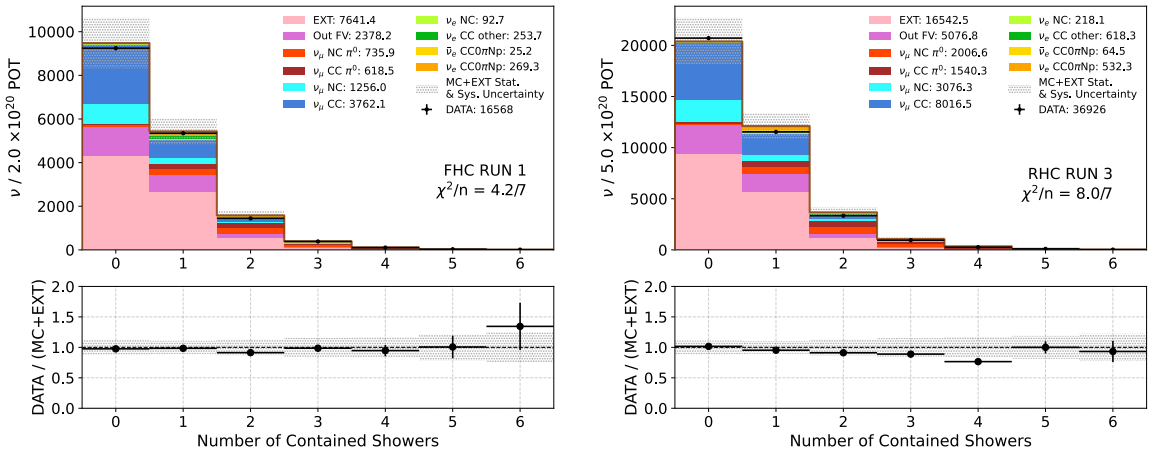


Figure 5.6: Event rates as a function of the number of showers with reconstructed vertex within the FV. Shown for FHC Run 1 (left) and RHC Run 3 (right) after quality cuts have been applied.

On the hadronic side, the signal definition requires the presence of at least one proton above detection threshold. Figure 5.7 displays the distributions of events passing quality cuts as a function of the number of tracks with all associated hits contained in the FV.

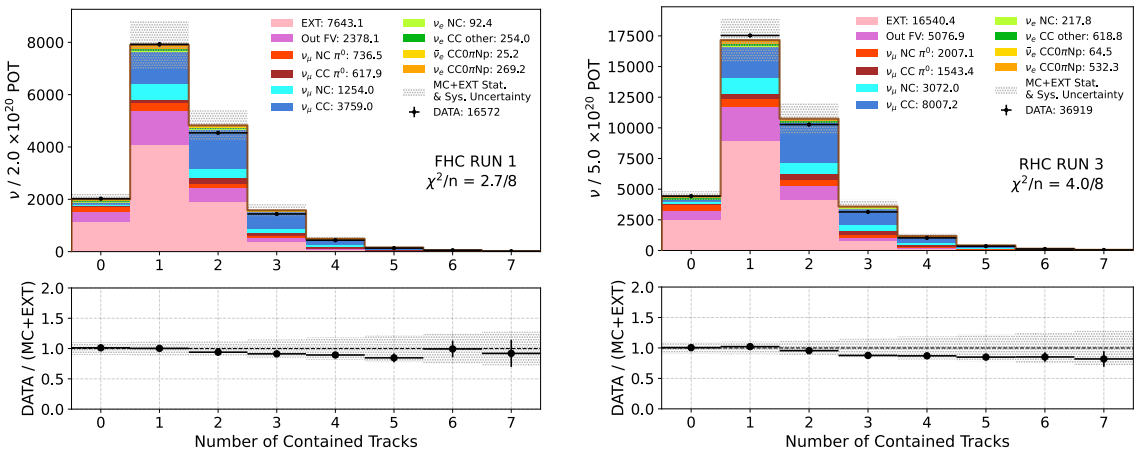


Figure 5.7: Event rates as a function of the number of tracks with all associated hits contained in the FV after quality cuts have been applied. Shown for FHC Run 1 (left) and RHC Run 3 (right).

Events with no reconstructed tracks are removed by the selection. The energy of the leading (longest) track is analyzed for events that survive this cut, as shown in Figure 5.8. Interactions with a leading track energy less than 40 MeV are removed from the event samples.

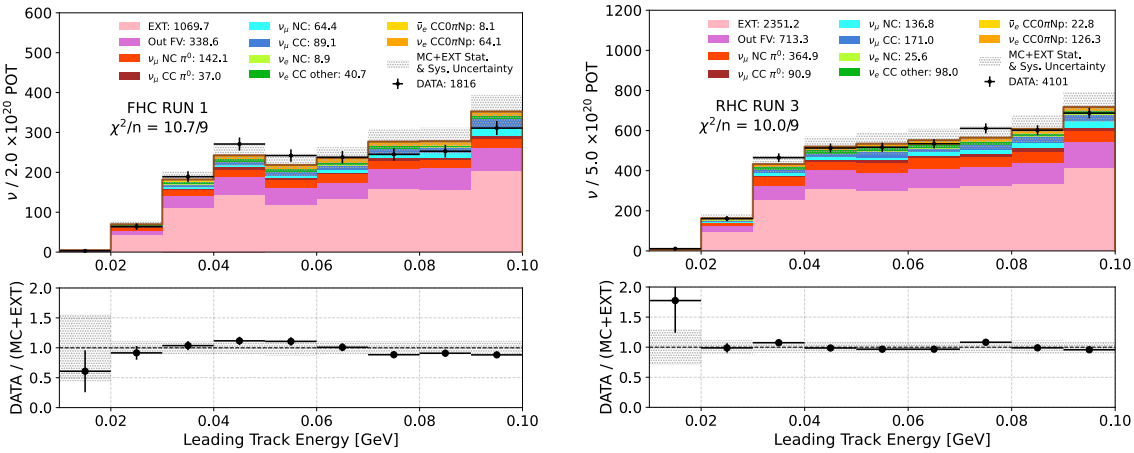


Figure 5.8: Event rates for FHC Run 1 (left) and RHC Run 3 (right) as a function of the leading (longest) track energy. Shown for interactions that have passed quality cuts and have at least one track present in the final topology.

After this stage of the selection, the purity for FHC Run 1 (RHC Run 3) is 4.5% (3.8%) and the efficiency is 39.1% (37.6%).

5.4 Loose ν_μ CC rejection

After quality cuts and the signal definition constraints—together referred to as the preselection—ν_μ CC backgrounds are loosely constrained in order to boost the signal-to-background ratio of the event samples for input into a Boosted Decision Tree. First, the selection aims to suppress ν_μ CC events without a π^0 in the final topology. This background channel is displayed in dark blue in the plots of this chapter.

As explained in Section 4.2, objects classified as showers by `Pandora` are given a score ranging from 0 to 0.5, based on how shower-like their topologies are. Event rates as a function

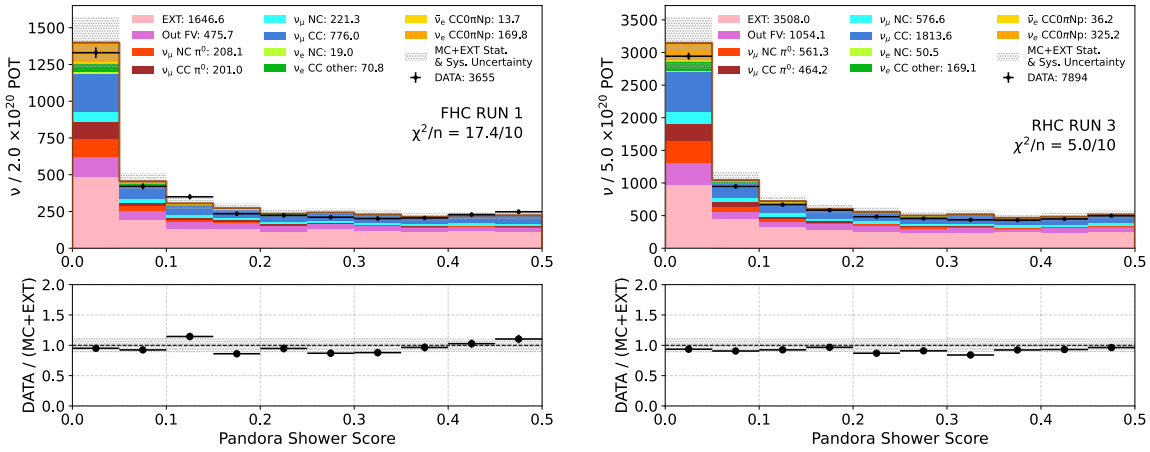


Figure 5.9: Event rates after the preselection as a function of `Pandora` shower score after preselection for FHC Run 1 (left) and RHC Run 3 (right).

of this shower score are shown in Figure 5.9 for interactions that pass the preselection. This stage of the selection tightens this classification by removing events with a shower score greater than 0.3.

For track objects, a particle identification (PID) tool developed by the MicroBooNE collaboration [75] is applied. In a nutshell, this tool uses calorimetry measurements (i.e., $\frac{dE}{dx}$) to analyze the leading reconstructed track of each event, constructing a likelihood that the track is a proton and a separate likelihood that the track is a muon. The test statistic to determine the classification of the track is defined as the ratio between these likelihoods. The output of the PID tool is the assignment of a score to the leading track in each event, where -1 represents tracks that are completely proton-like, and $+1$ represents tracks that are completely muon-like.

The track PID score distributions for events passing the preselection are shown in Figure 5.10. A peak at $\sim$ 0 is present, attributed to short tracks for which there is too little information to adequately discriminate between the muon and proton hypotheses. (This peak slightly favors the muon score because it is mostly comprised of cosmic rays, displayed in pink.) Events with a track PID score greater than 0.35 are removed from the samples.

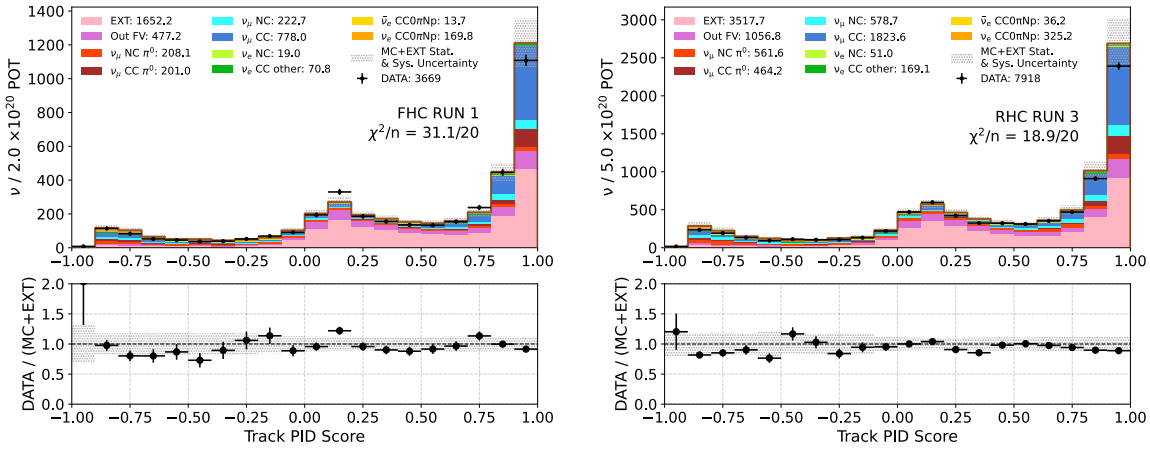


Figure 5.10: Event rates after the preselection as a function of the track PID score for FHC Run 1 (left) and RHC Run 3 (right).

After this stage of the selection, the purity for FHC Run 1 (RHC Run 3) is 12.3% (10.3%) and the efficiency is 30.5% (29.2%).

5.5 Loose π^0 rejection

Next, ν_μ events that create a π^0 are targeted, backgrounds that are shown in red in the plots of this chapter. Neutral pions can decay into photons that can mimic the shower signature of signal interactions via e^+/e^- pair production. Though π^0 interactions manifest as double showers in the LArTPC, `Pandora` frequently reconstructs them as a single object. The constraints in this section aim to suppress that reconstruction failure.

The separation between the neutrino interaction point and the start of the shower is a distinguishing factor between π^0 and ν_e CC events, as photons tend to travel a measurable distance away from the vertex before pair producing. Figure 5.11 shows the distribution of events passing the preselection and loose ν_μ rejection as a function of the 3D separation between the starting points of the shower and the leading track in the interaction. Assuming the track originates at the interaction vertex, events with separation greater than 12 centimeters are removed by the selection.

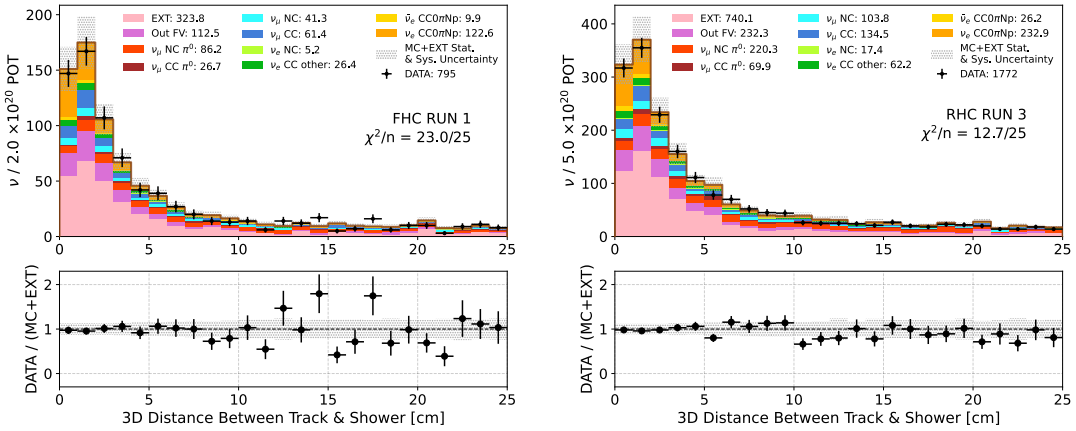


Figure 5.11: FHC Run 1 (left) and RHC Run 3 (right) event rates after the preselection and loose ν_μ CC rejection as a function of the separation between the leading track and shower objects in the neutrino interaction.

Shower objects are further classified by analyzing their associated spacepoints, the 3D clusters of hits that have been matched across 2D planes. The Molière angle is defined as the angle between the shower's momentum vector and a vector pointing from the shower's start point to one of its spacepoints. This variable is visualized in Figure 5.21b, where the angles between $\vec{p}_{shower}$ and $\vec{x}_1$, $\vec{x}_2$ are two examples of the Molière angle. The distributions of the average of all such angles in the shower object are shown in Figure 5.12 for events passing the preselection and loose ν_μ CC rejection.

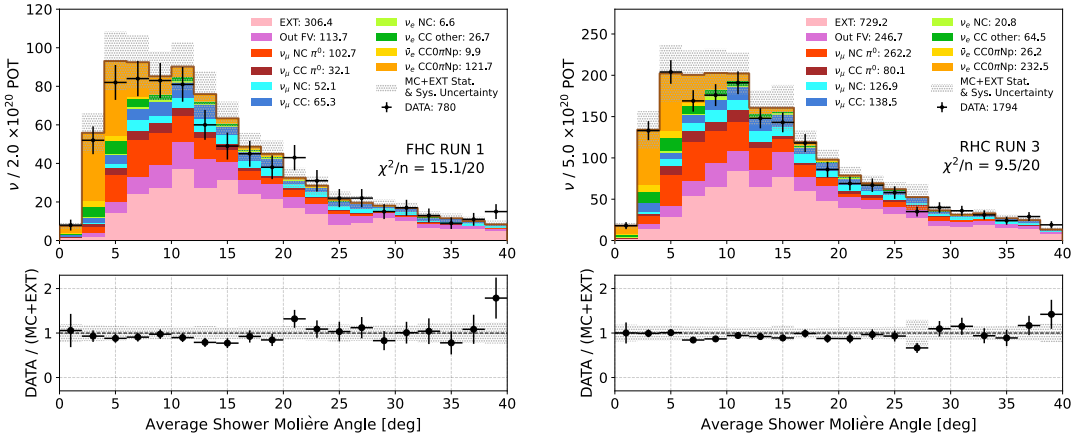


Figure 5.12: FHC Run 1 (left) and RHC Run 3 (right) event rates after the preselection and loose ν_μ CC rejection as a function of the average Molière angle for showers in the neutrino interaction.

In general, when two showers have been boosted and merged together by `Pandora`, the average Molière angle is larger than expected from a single shower (hence why signal interactions, in orange, pile up at low angular values). The selection loosely constrains these distributions by removing events with an average Molière angle of greater than 15°. Fortuitously, this cut also removes a substantial portion of poorly reconstructed cosmic activity.

Arguably the most powerful advantage of LArTPCs for neutrino studies is the ability to discriminate particles based on calorimetric information deposited in the detector. This feature provides the opportunity to distinguish between showers arising from electrons and those resulting from photons. Electrons are minimally ionizing with a mean energy loss of 2.1 MeV/cm, which is a measurable observable at the trunk of a shower. Photon showers, on the other hand, originate from an e^+/e^- pair, so $\frac{dE}{dx}$ at the shower trunk is consistent with two minimally ionizing particles, i.e. 4.2 MeV/cm. To compute $\frac{dE}{dx}$ at the trunk of a shower, the first four centimeters of the object is fit with a track, and the inverse modified Box model formula as given in Equation 4.1 is used.

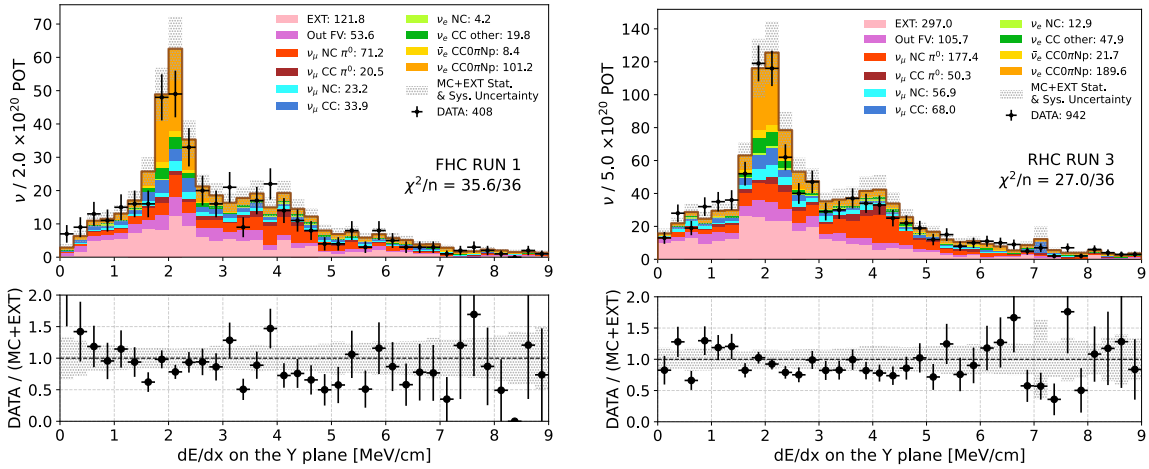


Figure 5.13: FHC Run 1 (left) and RHC Run 3 (right) event rates after the preselection, loose ν_μ CC rejection, and the average Molière angle constraint. Shown as a function of $\frac{dE}{dx}$ at the trunk of the shower belonging to the neutrino interaction.

Figure 5.13 shows the distributions of $\frac{dE}{dx}$ at the shower trunk, computed using hit clusters on the collection plane, for events passing the preselection, loose ν_μ CC rejection, and also

the average Molière angle constraint, applied to suppress the number of EXT events in the distribution. As anticipated, signal events pile up around 2.1 MeV/cm, and a notable peak of π^0 interactions is present at 4.2 MeV/cm. Events are conservatively required to have shower objects with $\frac{dE}{dx}$ less than 7 MeV/cm to pass this stage of the selection.

After this stage of the selection, the purity for FHC Run 1 (RHC Run 3) is 23.3% (20.5%) and the efficiency is 23.0% (21.6%).

5.6 Development of the BDT model

Here, a detour is taken from the nominal selection workflow of Figure 5.1 to describe the development of the BDT models used to apply the final constraint in the selection algorithm.

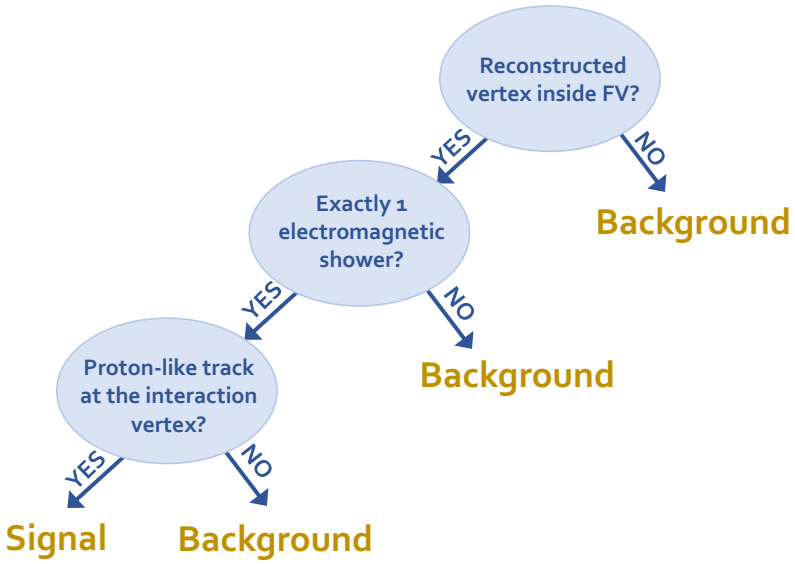


Figure 5.14: Event selection is a binary classification problem: The interaction is either signal or background. Linear selections sort events into these two categories by passing them through a decision tree structure, an example of which is shown above.

At its core, event selection is a binary classification problem: The interaction is either signal, or it is background. Traditional selections aim to sort events into one of these categories by linearly passing them through a given set of criteria in a structure that can be described as a decision tree. Figure 5.14 is an example of the decision tree structure for a

few of the selection requirements detailed in this chapter. At each tree node, input data is split based on the value of the variable being assessed. If an event is consistent with the node requirement, it moves on to the next stage of the algorithm. If not, it is assigned a negative class label (specific to this work, it is flagged as background). Events that pass every node are assigned a positive class label (flagged as signal).

Boosting is a machine learning technique that identifies multivariate trends across a given set of criteria. It combines a multitude of weak learners to create a strong learner that can more accurately classify data by way of additive training, in which each new iteration of the model seeks to fix the misclassifications of the previous one. In this work, the weak learners involved come in the form of decision trees whose splits represent different selection criteria. Successive trees are developed using interactions that were wrongly classified by the preceding tree. Each weak learner improves the final model, as illustrated in Figure 5.15.

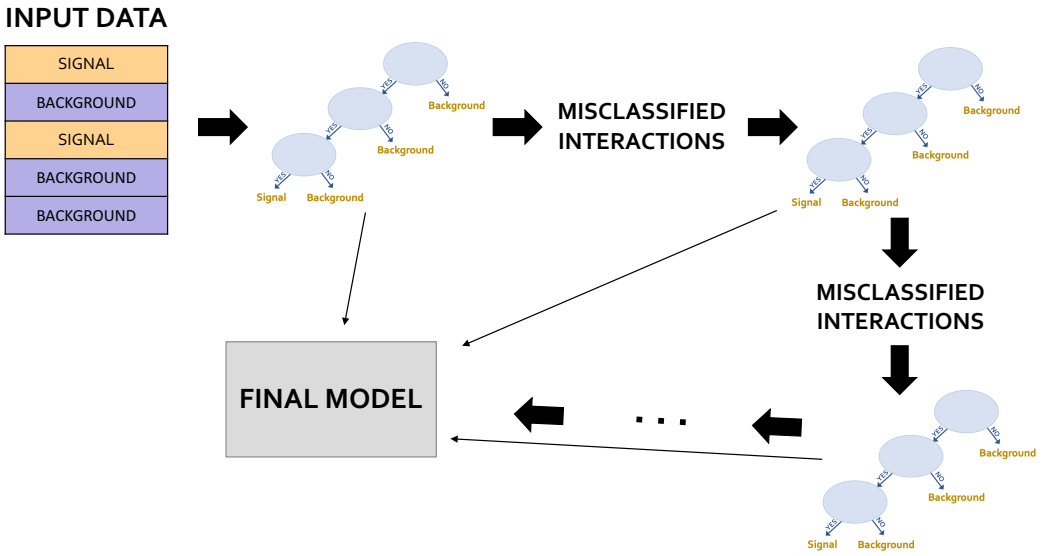


Figure 5.15: A schematic of the process to develop a Boosted Decision Tree model. Successive decision trees are developed using interactions that were misclassified by the preceding tree. The final BDT model is run over input event distributions and computes a score for each interaction, representing how likely it is to be signal.

The gradient boosting framework `XGBoost` [76] is used to develop the BDT models for this work. The final BDT model runs over input event distributions and computes a score for

each interaction that represents how likely it is to be signal.

Figure 5.16 illustrates the workflow for the development of the BDT models in this work. Note that the first four steps mirror the nominal selection workflow shown in Figure 5.1. The rest of this section details various decisions taken into account to train and test the BDT model. Results as they pertain to the nominal workflow are reported in the following section.

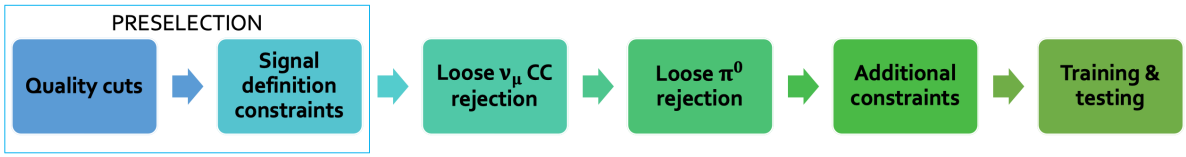


Figure 5.16: The workflow for the development of our BDT models. Note that the first four steps mirror the nominal workflow shown in Figure 5.1.

5.6.1 Training and testing samples

Ideally, the BDT development process makes use of a dedicated training sample for the model to learn from, and a separate testing sample for evaluation of the model's performance. Independent samples help to avoid over-training, which can lead to an overly optimistic estimate of performance. However, this work did not have access to additional samples for dedicated training and testing, thus it was necessary to train and test using the simulated FHC Run 1 and RHC Run 3 event samples used in the nominal selection workflow.

As illustrated in Figure 5.16, the BDT training and testing occurs on samples that have first been passed through the preselection and loose ν_μ CC and π^0 rejection constraints. This is done so as to overcome a low signal-to-background ratio which can negatively impact BDT performance.

The phase space of the BDT training sample is further restricted with criteria not present in the nominal selection workflow. Specifically, the event distributions as a function of reconstructed shower energy and opening angle (between the leading track and shower) are

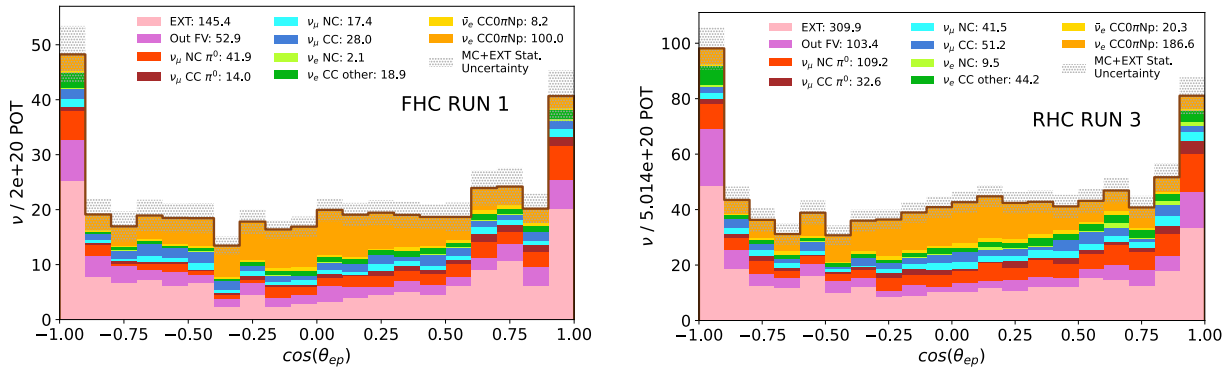


Figure 5.17: Event rate distributions as a function of reconstructed opening angle for FHC Run 1 (left) and RHC Run 3 (right) after preselection and loose ν_μ CC/π^0 background rejection constraints have been applied. Events with poor reconstruction pile up on either side of the distribution, which can skew how the BDT learns.

constrained. The purpose of this is to optimize the quality of the events that the BDT learns from. The model, trained on a subset of events in the full shower energy and opening angle distributions, is tested and applied to the full phase space of these variables. This method avoids model dependence associated with constraining the distributions of parameters that the cross section will be extracted on.

The distributions of events as a function of reconstructed opening angle are shown in Figure 5.17. The training phase space is restricted to events within $\cos\theta_{ep} = \pm 0.9$. This removes interactions with poorly reconstructed topologies occurring at the edges of this distribution. Events in this region are not representative of generalizable topologies so will likely be misclassified, and as a result, be given more importance in the training procedure as the BDT attempts to fix mistakes from previous iterations of the model.

The distributions of events as a function of reconstructed shower energy are shown in Figure 5.18. The training phase space is restricted to events with a shower energy greater than 70 MeV. This is done to suppress the population of delta rays and low energy Michel electrons produced by decay-at-rest cosmic muons, further boosting the signal-to-background ratio for optimal training.

Additional training constraints remove 41.2% of events (5.8% of signal) in the FHC Run 1

sample, and 39.6% of events (6.1% of signal) in the RHC Run 3 sample after the preselection and loose ν_μ CC and π^0 rejection has been applied.

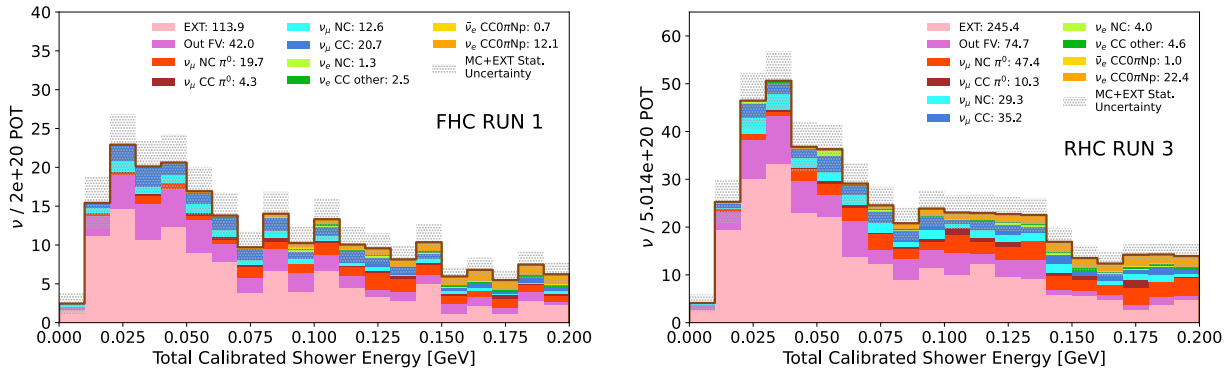


Figure 5.18: Events as a function of reconstructed shower energy for FHC Run 1 (left) and RHC Run 3 (right) after the preselection and loose ν_μ CC/π^0 background rejection have been applied. Interactions containing delta rays and Michel electron showers pile up at low energies.

A concern regarding BDT training and testing is the statistical limitations of the samples used. On the training side, it is necessary for the BDT to have enough events to firmly learn the topologies of signal and background. On the testing side, it is equally important to have enough statistics to properly evaluate the BDT's performance. Both samples should reflect the data it will later classify as best as possible. To ensure the latter is true, a method called stratification is employed, which preserves the signal-to-background ratio when splitting the simulated event rates into training and testing samples. (In general, because the split is random, it is expected that the signal-to-background ratio be inherently preserved. However, due to the limited sizes of the event samples, this may not be the case. Stratification guarantees distribution uniformity in both.)

To determine how much of the simulated event rates should be devoted to training vs. testing, the performance of the pre-BDT selection criteria (quality cuts, signal topology constraints, and loose ν_μ CC and π^0 rejection) on test samples of different sizes is analyzed. Figures 5.19 and 5.20 show the efficiencies and purities of the pre-BDT selection as a function of the fractional test sample size for FHC Run 1 and RHC Run 3, respectively.

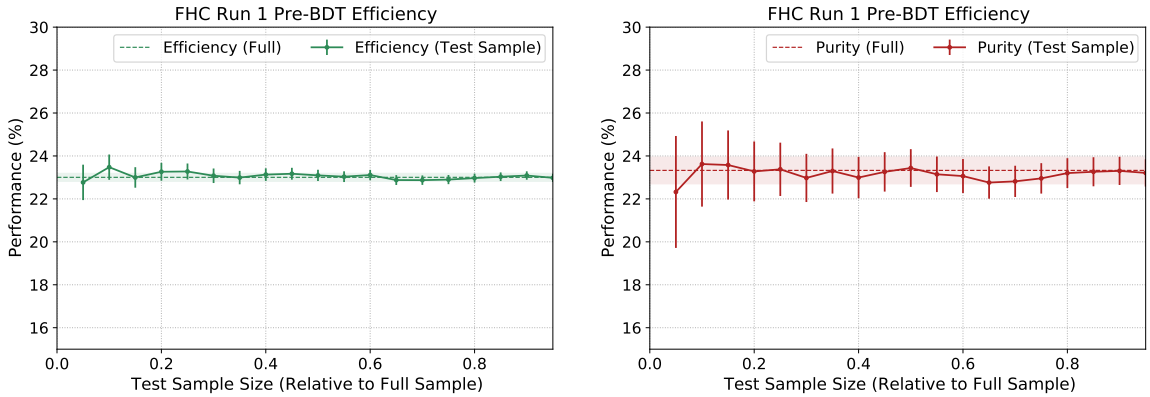


Figure 5.19: The FHC Run 1 efficiency (left) and purity (right) of the pre-BDT selection for test samples of different fractions relative to the full FHC Run 1 sample. The performance converges when the test sample is about 30% of the full sample size.

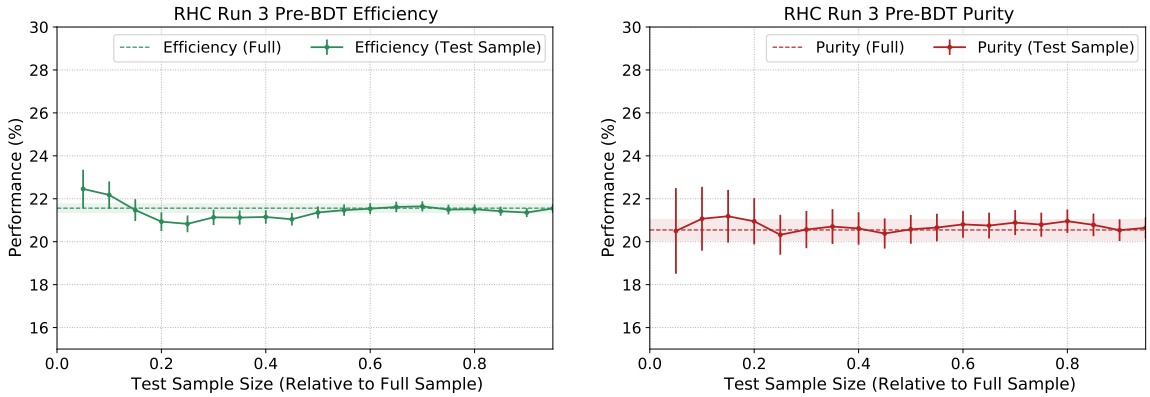
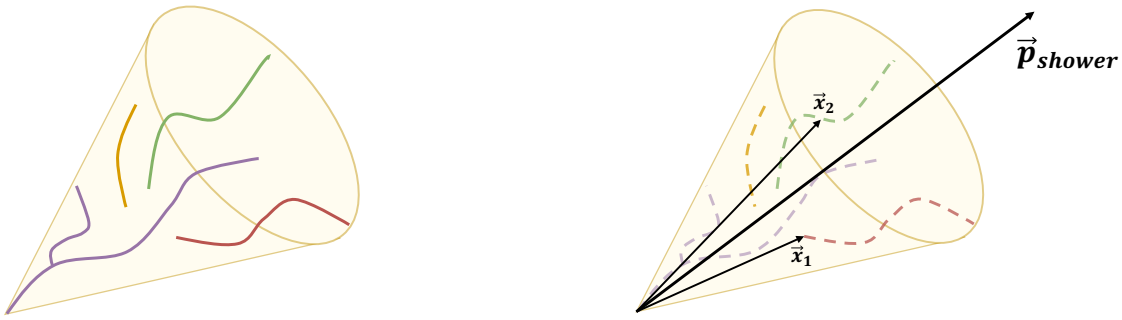


Figure 5.20: The RHC Run 3 efficiency (left) and purity (right) of the pre-BDT selection for test samples of different fractions relative to the full RHC Run 3 sample. The performance converges when the test sample is about 50% of the full sample size.

The assumption is that the test samples are no longer statistically limited when their efficiencies and purities converge to that of the full sample. This occurs when the FHC Run 1 (RHC Run 3) test sample is $\sim 30\%$ ($\sim 50\%$) the size of the full sample. For consistency, a 50/50 test/train split is chosen for both FHC Run 1 and RHC Run 3.

5.6.2 Training parameters



(a) Subclusters (b) Molière angle

Figure 5.21: Subclusters (a) are isolated 2D charge segments that exist because electromagnetic showers are made up of branches separated by gaps caused by photons traversing through the LArTPC. The Molière angle (b) is the angle between the shower's momentum ($\vec{p}_{shower}$) and a vector from the shower start point to a 3D spacepoint ($\vec{x}_1$, $\vec{x}_2$).

Reconstructed parameters fed to the BDT must have strong selective power, so that the model can learn to discriminate signal topologies from various sources of background. Most of these parameters are present as loosely applied criteria to remove ν_μ backgrounds earlier in the selection.

To target ν_μ CC events without a π^0, for example, the BDT trains on the `Pandora` shower score, the track PID score, and the number of shower subclusters present across all three anode planes. A subcluster is an isolated 2D charge segment within the reconstructed shower object. These exist because electromagnetic showers are made up of branches separated by gaps caused by photons traversing through the LArTPC. This feature is visualized in Figure 5.21a, which illustrates a shower reconstructed with 4 subclusters. This variable addresses the situation where ν_μ CC track objects are misidentified by `Pandora` as a shower. Signal typically has a higher number of subclusters than events with misidentified tracks.

To target π^0 background, the BDT trains on the average Molière angle, $\frac{dE}{dx}$ at the shower trunk (computed using hit clusters on the collection plane), and both the 3D and 2D separation between the track and shower. The latter parameter provides an additional handle

on evaluating the distance between the reconstructed interaction vertex and shower, and is derived from charge deposit on the collection plane.

Figures 5.22, 5.23, 5.24, 5.25, 5.26, 5.27, and 5.28 show the training parameter distributions for events passing all pre-BDT selection criteria and the additional opening angle and shower energy training constraints. The distributions are plotted using the full samples to illustrate goodness-of-fit with data; however, only 50% is used as input for training.

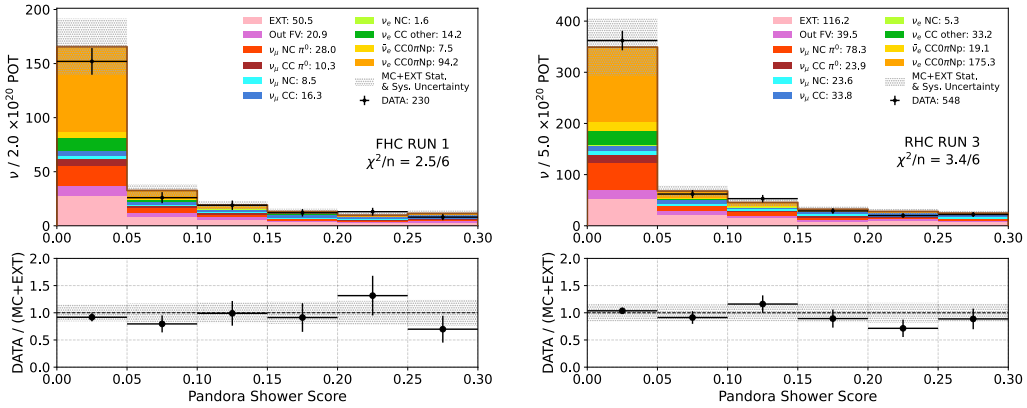


Figure 5.22: `Pandora` shower score distributions for FHC Run 1 (left) and RHC Run 3 (right).

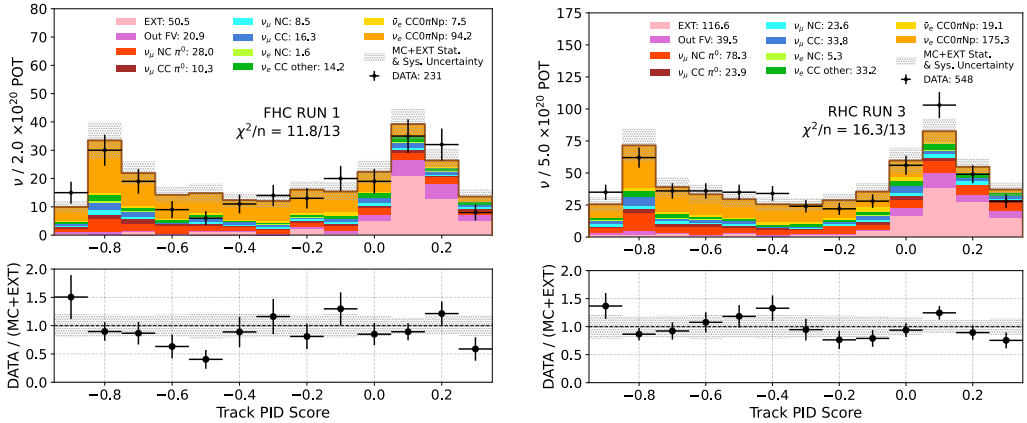


Figure 5.23: Track PID score distributions for FHC Run 1 (left) and RHC Run 3 (right).

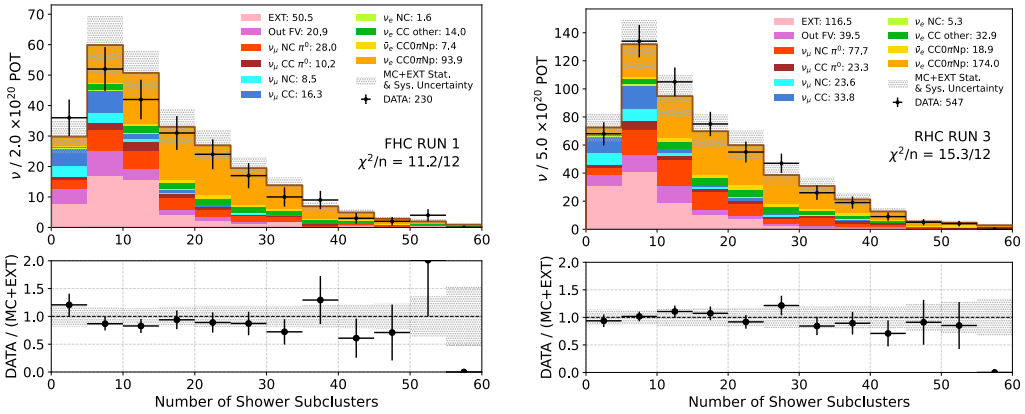


Figure 5.24: Distributions of the number of shower subclusters for FHC Run 1 (left) and RHC Run 3 (right).

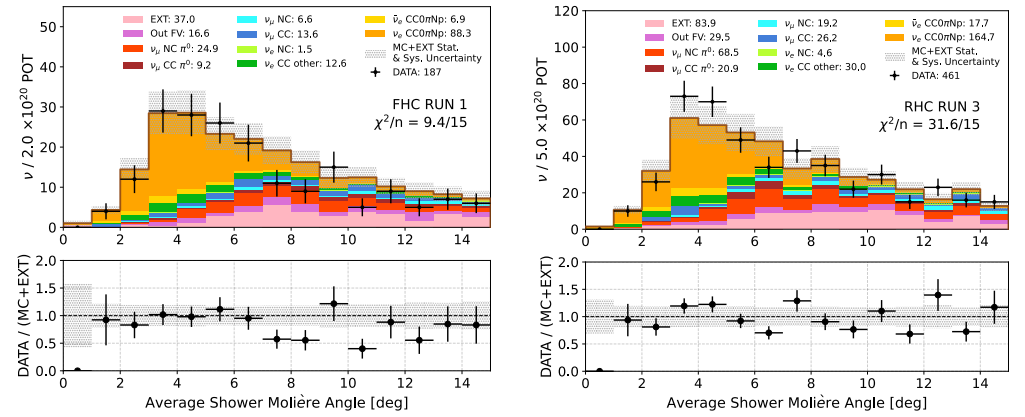


Figure 5.25: Average shower Moliere angle distributions for FHC Run 1 (left) and RHC Run 3 (right).

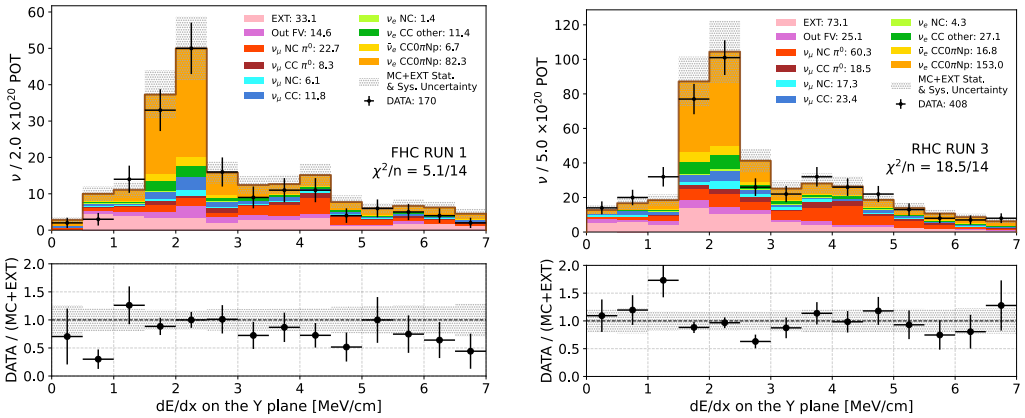


Figure 5.26: $\frac{dE}{dx}$ distributions for FHC Run 1 (left) and RHC Run 3 (right).

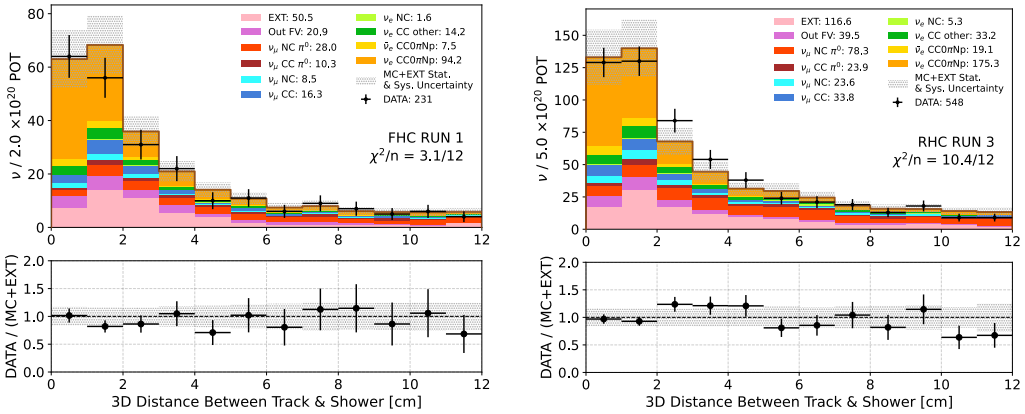


Figure 5.27: Distributions of the 3D distance between track and shower for FHC Run 1 (left) and RHC Run 3 (right).

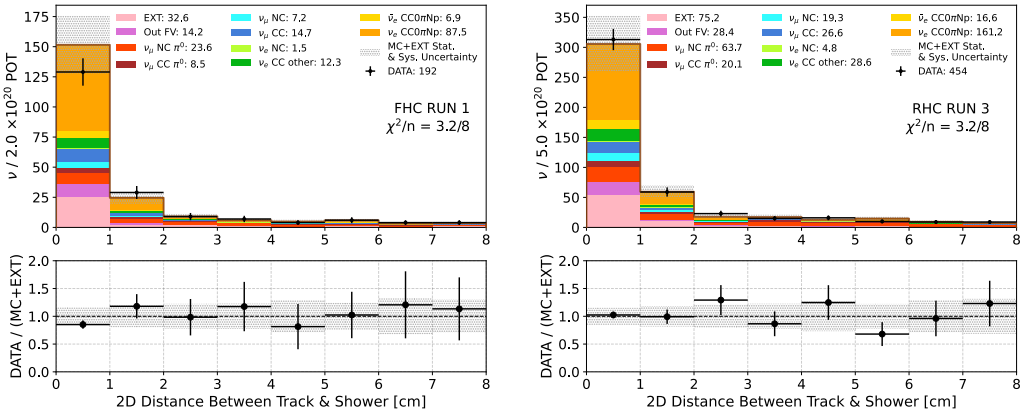


Figure 5.28: Distributions of the 2D distance between the track and shower as measured from the collection plane for FHC Run 1 (left) and RHC Run 3 (right).

5.6.3 Model hyperparameters

A hyperparameter is a variable external to the BDT model that is set before the learning process begins, such as the maximum number of nodes the decision trees can have, or the maximum number of reconstructed variables that can be used in a given tree. In summary, hyperparameters dictate exactly how the algorithm trains the dataset, and can be tuned for optimal performance either with an independent validation sample or a portion of the training sample.

Statistical limitations in the training sample do not allow for robust tuning of the hyperparameters. Thus, to avoid introducing bias into the BDT performance evaluation, all but two of the hyperparameters are set using the same values as a parallel cross section result by the MicroBooNE collaboration, which measured $\nu_e + {}^{40}Ar \rightarrow 1e + Np$ interactions using data collected from the BNB [26]. These results employed a BDT with hyperparameters tuned from an independent validation dataset.

In this work, two hyperparameters are optimized: `scale_pos_weight`, defined as the ratio between the negative class and positive class, and `nrounds`, the number of boosting rounds. The values of these hyperparameters are listed in Table 5.2.

Hyperparameter	**FHC Training**	**RHC Training**
`scale_pos_weight`	0.384	0.506
`nrounds`	200	150

Table 5.2: The optimized values for the two BDT hyperparameters tuned for the selection algorithm: `scale_pos_weight`, defined as the ratio between the negative class and positive class, and `nrounds`, the number of boosting rounds.

The hyperparameter `scale_pos_weight` represents the ratio between the number of background events and the number of signal events. This is computed using the raw number of interactions from the simulated event rates of the training samples (no `GENIE`, `PPFX`, or other correction weights as described in Section 4.3.1 are applied).

The hyperparameter `nrounds` is the maximum number of decision trees the BDT model can use to improve its performance. As each iteration focuses on improving misclassifications of the previous tree, it is expected that the BDT performance on the training sample will increase as `nrounds` increases. However, because of the nature of this type of boosting, the model will reach a certain point where it begins to memorize the data rather than recognize generalizable topologies in the testing sample. This is commonly referred to as overfitting.

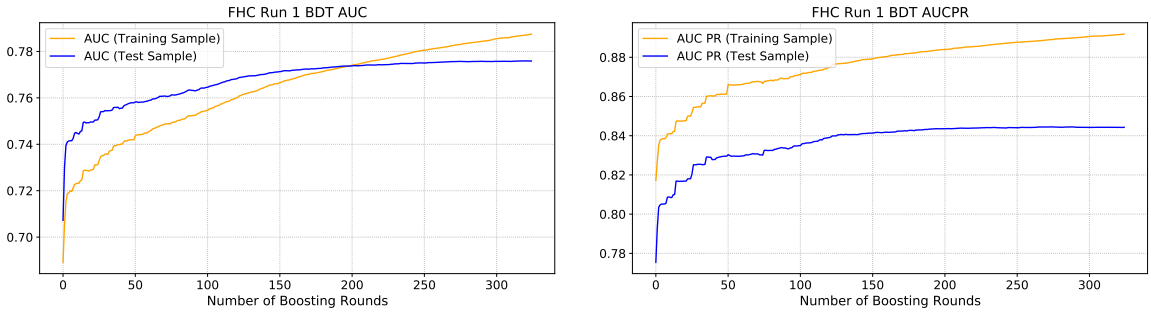


Figure 5.29: `AUC` (left) and `AUCPR` (right) metrics for the FHC Run 1 training and testing samples. As expected, the performance of the training sample increases as the number of boosting rounds increases. However, the performance of the testing sample stagnates around 200 rounds, becoming susceptible to overfitting.

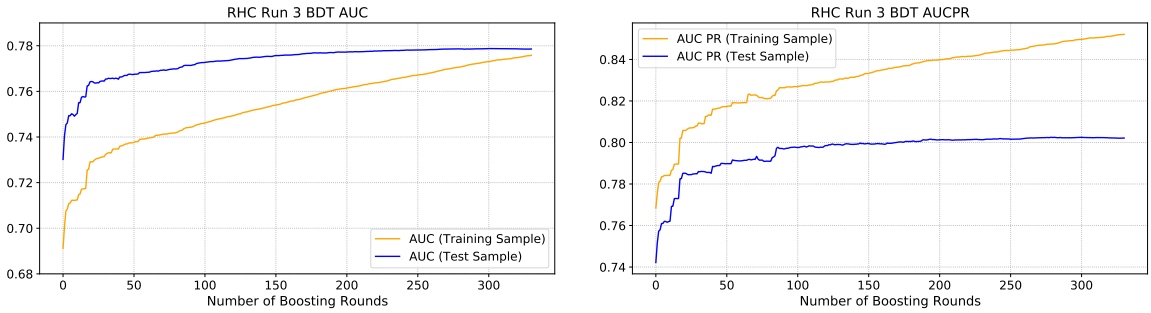


Figure 5.30: `AUC` (left) and `AUCPR` (right) metrics for the RHC Run 3 training and testing samples. As expected, the performance of the training sample increases as the number of boosting rounds increases. However, the performance of the testing sample stagnates around 150 rounds, becoming susceptible to overfitting.

Two diagnostic curves come in handy for determining the appropriate number of rounds, because they evaluate how well the BDT classifies the training sample (its performance) irrespective to the classification threshold (the BDT score constraint). The first is a receiver operating characteristic (ROC) curve, which plots the true positive rate of the BDT as a function of the false positive rate. The second is called a precision-recall (PR) curve. "Precision" refers to the purity of a model, defined as the ratio between the number of true positives and the total number of positives (true and false). "Recall" refers to the efficiency of the model, defined as the ratio between the number of true positives and the total number of events belonging to the positive class in the input dataset. The PR curve plots the precision

as a function of the recall. For both diagnostics, the area under the curve is proportional to the skill of the model. (A perfect BDT would have an area under the curve of 1.) ROC curves are best for evaluating skill when the positive and negative classes are balanced, equivalently when the signal-to-background ratio is close to 1. PR curves, on the other hand, are best when the classes are not balanced.

To determine the optimal value for `nrounds`, the area under the curve for both the ROC (`AUC`) and PR (`AUCPR`) curves is studied. Figures 5.29 and 5.30 show the `AUC` (left) and `AUCPR` (right) metrics as a function of the number of boosting rounds for the FHC Run 1 and RHC Run 3 training and testing samples, respectively. As expected, the performance of the training sample, displayed in orange, increases with the number of boosting rounds. However, the performance of the testing sample, displayed in blue, stagnates around 200 (150) rounds for FHC Run 1 (RHC Run 3). At this point, the samples become susceptible to overfitting. To avoid this, `nrounds` is set to 200 (150) for the FHC Run 1 (RHC Run 3) training sample.

5.6.4 Testing the BDT model

Once the training parameters are determined and the hyperparameters optimized, separate BDT models are trained for FHC Run 1 and RHC Run 3, and their performance can be tested. To further reduce the effects of statistical fluctuations, the BDT is evaluated using a process called k-fold cross validation. Cross validation is a re-sampling procedure used to estimate the performance of a machine learning algorithm with less bias than a single train/test split. This is particularly advantageous when working with statistically limited samples.

In general, k-fold cross validation splits the dataset into k sets and iterates over each one, taking the chosen group as a test sample and the rest of the sets together as the training sample. A model is fit on the training sample and performance is evaluated from the testing

sample. The model is discarded after each iteration. The end result is k evaluation scores from which the skill of the model can be quantified more reliably, because the algorithm being trained and tested multiple times in a random manner.

Due to the limited statistics available in this work, cross validation is performed with $k = 2$, such that the event samples are subject to a (stratified) 50/50 train/test split. This procedure is repeated a total of 20 times. Specifically, the method is as follows: The BDT is trained on the first set, and the efficiency and purity as a function of BDT score is evaluated on the second set. The model is discarded, and the intended purpose of the sets are flipped: the BDT is freshly trained on the set previously devoted to testing, and tested on the set previously devoted to training. The efficiency and purity are again evaluated, and the model is again discarded. The event sample is then shuffled and split, and the procedure begins again. In the end, there are 40 independent evaluations of the BDT's performance (2 per cross validation scheme). The average of the 40 performance values are plotted as a function of the BDT score, shown in Figure 5.31 for FHC Run 1 (left) and RHC Run 3 (right).

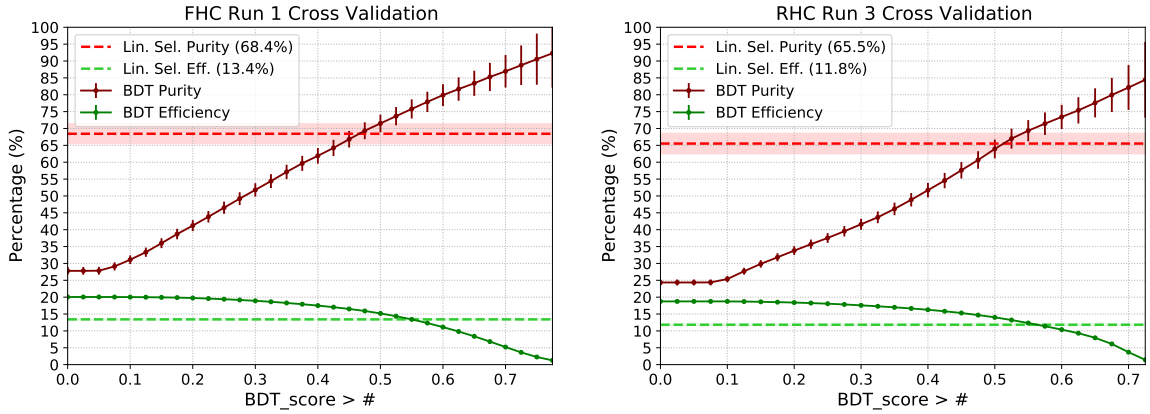


Figure 5.31: Cross validated purity (red) and efficiency (green) results evaluated from the FHC Run 1 (left) and RHC Run 3 (right) simulated event samples. For comparison, the performance of a corresponding linear selection is also displayed.

Results from the cross validation can be used to determine the optimal BDT score constraint without bias arising from statistical fluctuations. The BDT is compared to the performance from a linear selection, listed in Table 5.3, developed for this work before the

implementation of a BDT. Purity and efficiency values derived from the linear selection are shown in Figure 5.31 as flat dashed lines. Uncertainty on the purity and efficiency is quantified assuming Poisson and binomial errors, respectively.

The BDT score constraint is chosen to improve the purity of the final selected event rates as compared to what the linear selection can accomplish, while maintaining the same (or better) efficiency. Event rates are restricted to a BDT score greater than 0.55 for FHC Run 1 and 0.575 for RHC Run 3. Final performance values are reported in the following section, derived after the trained BDT models have been applied to the full event samples as illustrated in the nominal selection workflow (Figure 5.1).

Linear Selection Criteria
Pandora slice ID == 1
Interaction vertex within FV
Event at least 90% contained within FV
Exactly 1 shower with vertex contained in FV
At least 1 track with hits fully contained in FV
Leading track energy greater than 40 MeV
`Pandora` shower score < 0.125
Average Molière angle $< 8°$
$\frac{dE}{dx}$ of the shower trunk < 4 MeV/cm
FHC: 3D separation between track and shower < 5 cm
RHC: 3D separation between track and shower < 4 cm

Table 5.3: Linear selection criteria used to compare with the BDT selection performance.

5.7 Selection performance and final event rates

Once the BDT is trained and a score threshold is chosen, the models are applied to the entire available NuMI dataset. Figure 5.32 shows the distributions of events passing the preselection and loose ν_μ CC and π^0 rejection as a function of the BDT score. The grey

error band on these plots, and those thereafter, now includes an estimated flat contribution to account for the systematic uncertainty on the detector response: 12.1% for FHC Run 1 and 12.9% for RHC Run 3. Figure 5.33 displays the BDT score distribution for the prediction only with finer binning.

After the BDT score cut, the selection achieves a final efficiency of of 13.7% (11.4%) and a final purity of 77.0% (73.5%) in the FHC Run 1 (RHC Run 3) sample. Summaries of the performance at different selection stages are given in Tables 5.4 and 5.5.

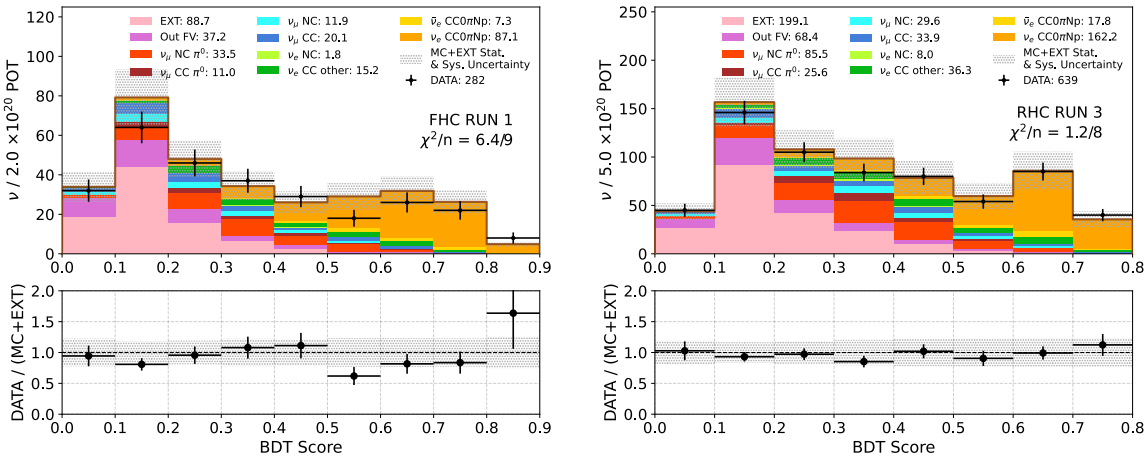


Figure 5.32: Event rates as a function of the BDT score for FHC Run 1 and RHC Run 3 after the preselection and loose ν_μ CC and π^0 rejection have been applied.

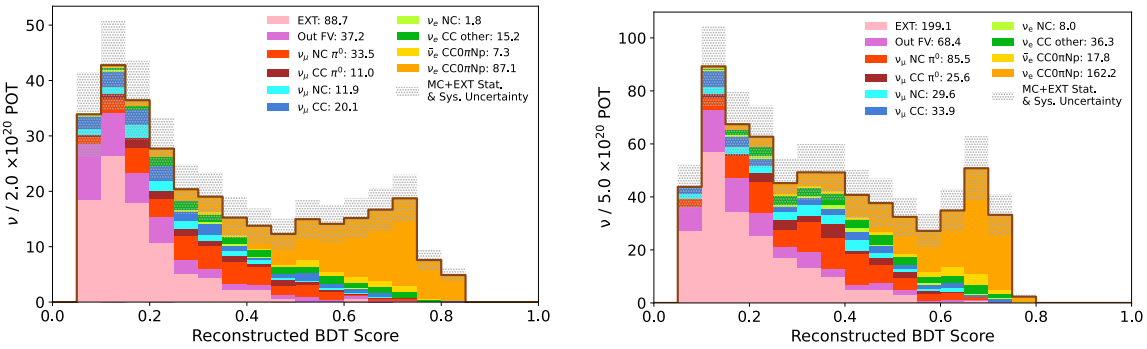


Figure 5.33: The simulated event rate as a function of BDT score for FHC Run 1 (left) and RHC Run 3 (right) after the preselection and loose ν_μ CC and π^0 rejection have been applied, shown with finer binning for the prediction only.

Selection stage	Efficiency	Purity
Quality cuts	62.0%	1.6%
Signal definition constraints	39.1%	4.5%
Loose ν_μ CC rejection	30.5%	12.3%
Loose π^0 rejection	23.0%	23.3%
BDT score constraint	13.7%	77.0%

Table 5.4: Event selection performance evolution for FHC Run 1.

Selection stage	Efficiency	Purity
Quality cuts	61.5%	1.4%
Signal definition constraints	37.6%	3.8%
Loose ν_μ CC rejection	29.2%	10.3%
Loose π^0 rejection	21.6%	20.5%
BDT score constraint	11.4%	73.5%

Table 5.5: Event selection performance evolution for RHC Run 3.

The differential cross section is extracted as a function of four variables: the electron energy, total visible energy (defined as the sum of all kinetic energy associated with the event detected by the LArTPC), opening angle between the electron and leading proton momenta, and proton multiplicity. Final selected event rates can be plotted as a function of the reconstructed versions of these variables. Binning for reconstructed electron energy, visible energy, and opening angle is chosen by first computing the resolution R between the true (x_{true}) and reconstructed (x_{reco}) values of the variable for each event in the sample:

$$R = \frac{x_{true} - x_{reco}}{x_{true}}. \tag{5.4}$$

(For opening angle, R is computed as the numerator of Equation 5.4, so as to avoid division by zero.) For each cross section variable, the standard deviation of the selected signal distribution as a function of R is computed across the entire phase space and taken as the minimum permissible bin width. This threshold mitigates sizable bin-to-bin migrations in the final measurements. This width is increased accordingly to ensure that there are at least

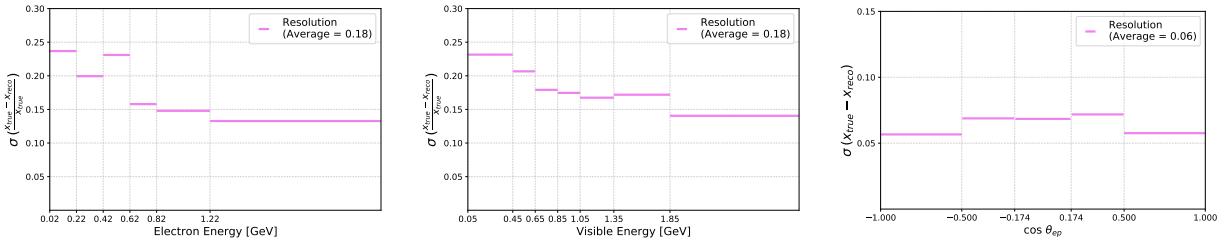


Figure 5.34: The resolution between true and reconstructed observables in MicroBooNE as a function of the chosen binning for electron energy, visible energy, and opening angle. The average resolution across the entire phase space its taken as the minimum permissible bin width.

10 total estimated signal events per bin in the combined (FHC+RHC) distributions. The resolution as a function of the chosen binning for electron energy, visible energy, and opening angle are displayed in Figure 5.34.

The reconstructed electron energy is defined as the energy of the shower object in the event, computed as described in Section 4.2. Figure 5.35 displays the final selected event rates as a function of reconstructed electron energy. The bin edges for these distributions are: `[0.02, 0.22, 0.42, 0.62, 0.82, 1.22]` in GeV. The final bin is overflow and includes events with reconstructed electron energy greater than 1.22 GeV.

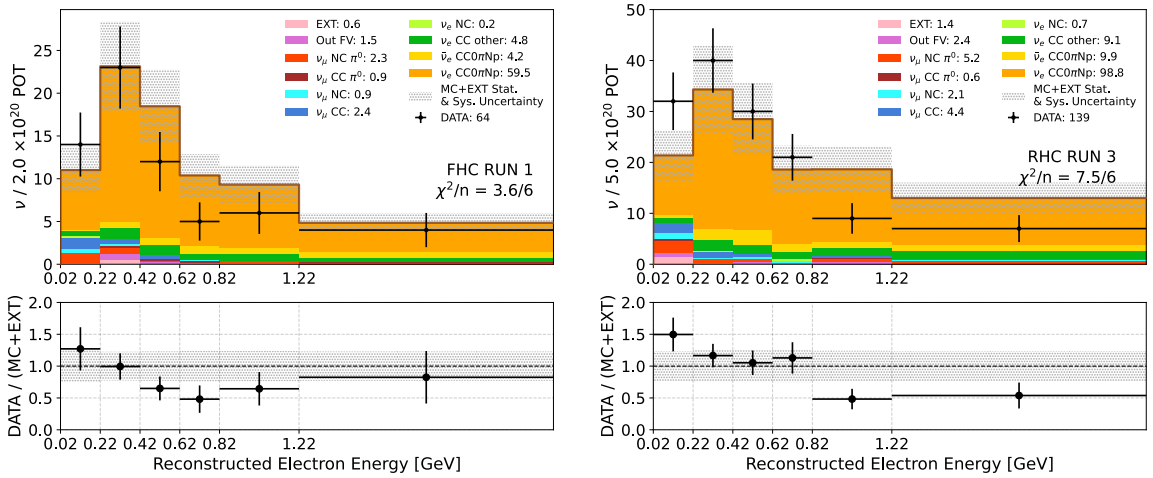


Figure 5.35: FHC Run 1 (left) and RHC Run 3 (right) selected event rates as a function of reconstructed electron energy.

The reconstructed visible energy is the sum of the shower energy and all the energies of all tracks associated with the interaction:

$$E_{visible} = E_{shower} + \sum E_{track}. \tag{5.5}$$

This variable can be considered a model-independent proxy for true neutrino energy. Figure 5.36 displays the final selected event rates as a function of reconstructed visible energy. The bin edges for these distributions are: `[0.05, 0.45, 0.65, 0.85, 1.05, 1.35, 1.85]` in GeV. The final overflow bin includes events with visible energy greater than 1.85 GeV.

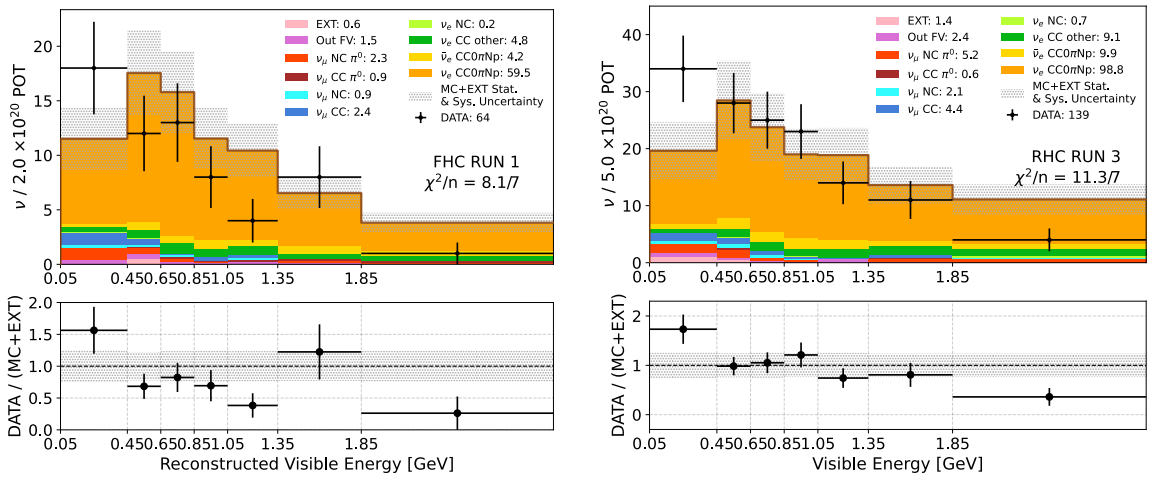


Figure 5.36: FHC Run 1 (left) and RHC Run 3 (right) selected event rates as a function of reconstructed visible energy.

The reconstructed opening angle is the cosine between the shower ($\vec{p}_{shower}$) and leading track ($\vec{p}_{track}$) momentum vectors:

$$\cos\theta_{ep} = \frac{\vec{p}_{shower} \cdot \vec{p}_{track}}{\|\vec{p}_{shower}\| \|\vec{p}_{track}\|}. \tag{5.6}$$

Figure 5.37 displays the final selected event rates as a function of reconstructed opening angle. The bin edges for these distributions are: `[`$\cos(\pi)$, $\cos(\frac{2\pi}{3})$, $\cos(\frac{5\pi}{9})$, $\cos(\frac{4\pi}{9})$, $\cos(\frac{\pi}{3})$, $\cos(0)$`]`.

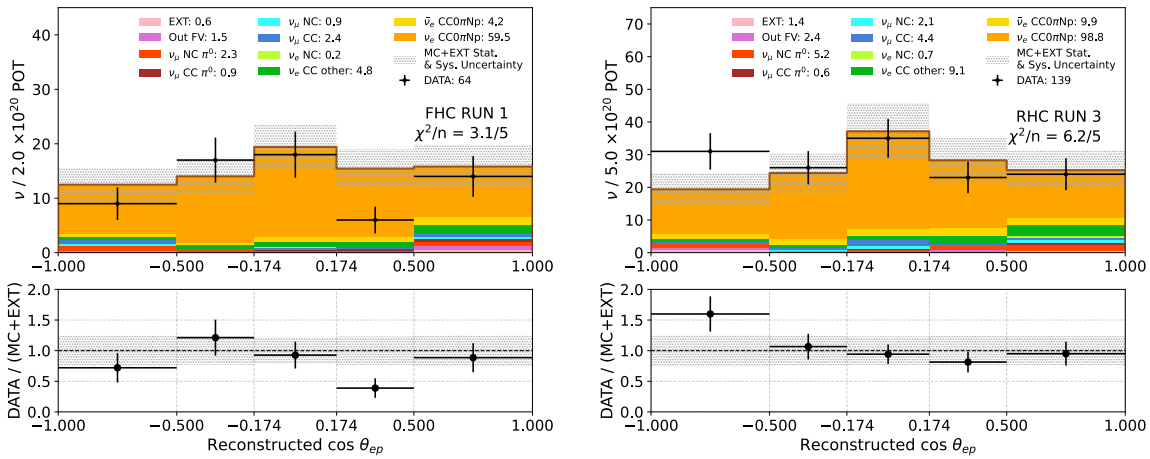


Figure 5.37: FHC Run 1 (left) and RHC Run 3 (right) selected event rates as a function of reconstructed opening angle.

Finally, the reconstructed proton multiplicity is defined as the number of tracks in the event. Figure 5.38 shows the final selected event rates as a function of reconstructed proton multiplicity. The bin edges for these distributions are: `[1, 2, 3]`. The final bin is overflow and includes all interactions with 3 or more tracks.

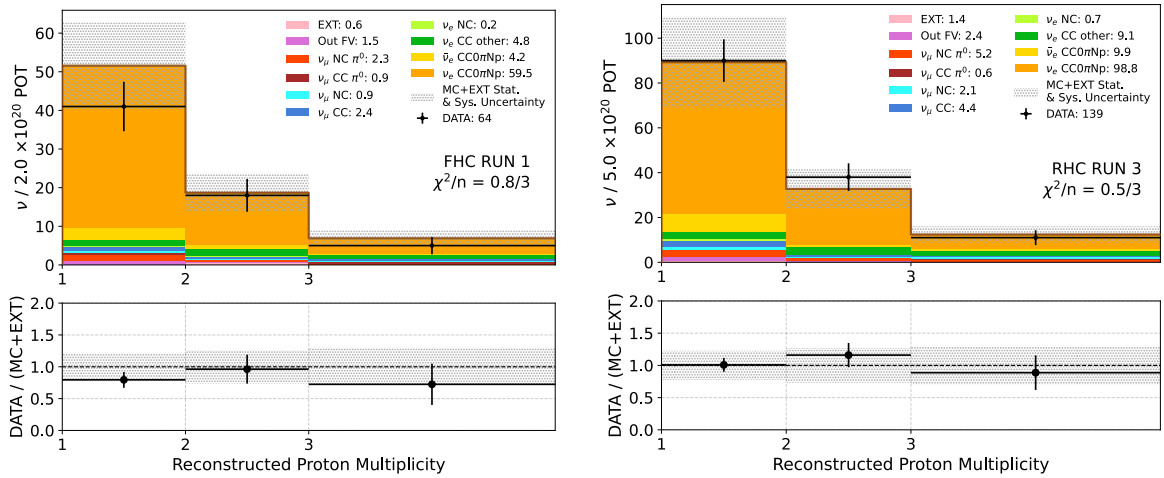


Figure 5.38: FHC Run 1 (left) and RHC Run 3 (right) selected event rates as a function of reconstructed proton multiplicity.

The combined (FHC+RHC) selected event rates are produced by adding the individual FHC Run 1 and RHC Run 3 contributions from each bin together. Figure 5.39 shows the combined selected event rates as a function of the cross section variables. In these plots, various sources of background have been collapsed into a single legend category.

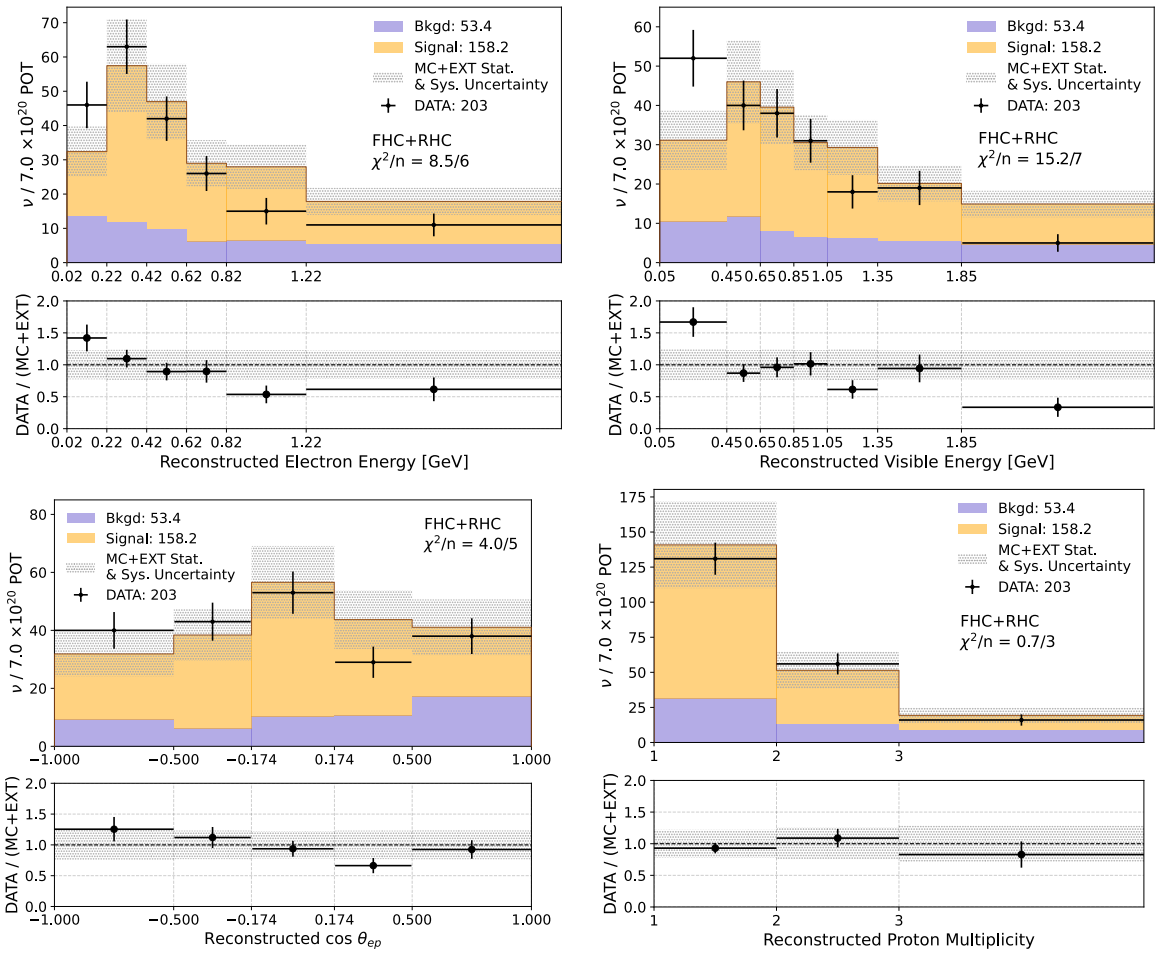


Figure 5.39: The combined (FHC+RHC) selected event rates as a function of the cross section variables. A total of 203 beam-on events survive the selection.

The efficiency as a function of the cross section variables is evaluated on truth-level information and displayed in Figure 5.40 for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) distributions. As anticipated, the FHC+RHC efficiency lies between those of the individual samples.

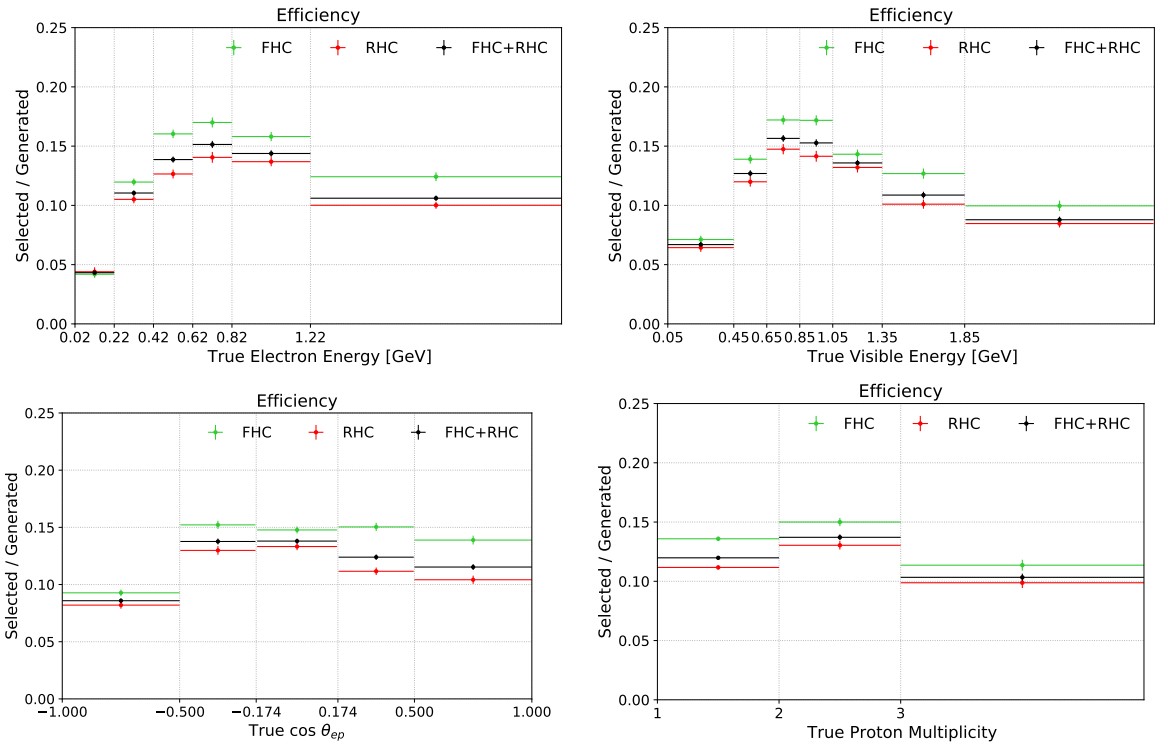


Figure 5.40: The efficiency as a function of the cross section variables for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black). As anticipated, the FHC+RHC efficiency lies between those of the individual samples.

CHAPTER 6

SOURCES OF UNCERTAINTY

Uncertainty, which comes from finite event samples and imperfections in the modeling of different parts of the experiment, can lead to discrepancies between our simulated event rates and real data. The evaluation of uncertainties shown in this chapter concerns event rates that have been subtracted by MC+EXT-estimated background, since this is a primary input for cross section extraction (see Equation 7.1). However, the procedure is analogous for handling uncertainties on the full event rates displayed in Chapter 5 at earlier stages of the selection.

Several sources of uncertainty are associated with MicroBooNE's NuMI dataset. There are systematic effects that carry over from the hadron production and geometric beamline models used in the flux prediction, as well as from neutrino cross section and secondary particle re-interaction models used to simulate event rates. Similar uncertainties arise from limitations in understanding the detector's response to interacting particles, unconstrained knowledge regarding out-of-cryostat (dirt) events, and uncertainty in the estimated POT delivered to the NuMI beamline. Finally, statistical fluctuations affect both the beam-on data and MC+EXT simulation.

Systematic uncertainties in the prediction can change the event count, how the detector responds to interactions within the active volume, and the estimated efficiency of the selection algorithm—all of which impact a cross section measurement. To quantify this impact, parameters associated with each uncertainty are identified and varied to generate alternate universe (UV) event rates distinct from the central value distribution. Variations can be created by re-simulating entire event samples with alternate parameters, or by employing a reweighting scheme, in which interactions in the CV distribution are weighted to produce the UV event rate. Reweighting factors are generated using a multisim or unisim approach. Multisim reweights take into account correlations between dependent parameters by ran-

domly sampling them all at once. Unisim reweights, on the other hand, treat the parameters as independent and are constructed by varying each one individually, usually by $\pm 1\sigma$.

The exclusive signal definition makes the measurements in this work particularly prone to statistical limitations, motivating the desire to add the FHC Run 1 and RHC Run 3 event rates (and associated uncertainties) before extracting cross section results. This requires careful combination of FHC and RHC systematic variations, which are generated independently for each beam mode using the same random seed, to construct a FHC+RHC covariance matrix. For systematic variations produced from new MC simulations, this means adding together FHC Run 1 and RHC Run 3 universes from the same generation. For systematic variations produced using the reweighting scheme, this means adding together FHC Run 1 and RHC Run 3 universes created by weights of the same index. Statistical uncertainties, on the other hand, are evaluated using Poisson errors on the FHC+RHC event rates.

6.1 Uncertainty calculation procedure

For each source of systematic uncertainty, a representative set of FHC+RHC variations is created. The effect on the simulated background-subtracted event rate x_i in bin i for each universe is evaluated:

$$x_i \equiv N^{UV}_{reco\,i} - B^{CV}_{reco\,i} = (S^{UV}_{reco\,i} + B^{UV}_{reco\,i}) - B^{CV}_{reco\,i} \tag{6.1}$$

where $N^{UV}_{reco\,i}$, $S^{UV}_{reco\,i}$, and $B^{UV}_{reco\,i}$ are the selected total, signal, and background predictions in universe bin i, and $B^{CV}_{reco\,i}$ is the selected background prediction of the CV event rate in bin i. This approach quantifies the uncertainty on the signal channel as well as the estimated background, by assessing the residual $(B^{UV}_{reco\,i} - B^{CV}_{reco\,i})$ left after the CV background subtraction. Note that $S^{UV}_{reco\,i}$ can be rewritten as a linear combination of the signal events in true space smeared across reconstructed bin i:

$$S^{UV}_{reco\,i} = \sum_j R^{UV}_{ij} \times \hat{S}^{CV}_{true\,j} \tag{6.2}$$

where $\hat{S}^{CV}_{true\,j}$ denotes the number of generated CV signal events in true bin j, before the selection has been applied. R^{UV}_{ij} is a response matrix term, defined as:

$$R^{UV}_{ij} = \frac{S^{UV}_{ij}}{\hat{S}^{CV}_{true\,j}} \tag{6.3}$$

where S^{UV}_{ij} is the number of signal events in reconstructed bin i and true bin j of the universe. This can effectively be thought of as the localized 2D efficiency of bin ij, as summing over j returns the efficiency of the selection in true bin j.

Equation 6.3 can be expanded into two terms:

$$\frac{S^{UV}_{ij}}{\hat{S}^{CV}_{true\,j}} = \frac{S^{UV}_{ij}}{\hat{S}^{UV}_{true\,j}} \times \frac{\hat{S}^{UV}_{true\,j}}{\hat{S}^{CV}_{true\,j}} \tag{6.4}$$

where $\hat{S}^{UV}_{true\,j}$ denotes the number of generated signal events in true bin j of the universe, before the selection has been applied.

This expansion is useful to understand what the background-subtracted event rate variation is accounting for. The first term on the right side of Equation 6.4 assesses the systematic changes in event rate reconstruction: the detector response (smearing) in bin ij and the efficiency of the selection in true bin j for each universe. The second term quantifies the systematic event rate change at the generator level: the difference in the number of generated signal interactions between the UV and CV event rates in true bin j, before any selection has been applied. For flux uncertainties, the first term is equal to 1 because the flux has no effect at the reconstruction level—rather, it only affects the rate of generated neutrino interactions. On the other hand, for detector and re-interaction (`Geant4`) systematics, the second term is equal to 1 because these sources have no effect at the generator level; rather, they only affect the detector's response and selection performance. Cross section, i.e. `GENIE`,

uncertainties nominally affect both; however, the second term is manually set to 1 for the measurements in this work, since this is the interaction channel attempted to be constrained.

Thus, the response matrix term for `GENIE` uncertainties is reduced to:

$$R_{ij}^{UV} = \frac{S_{ij}^{UV}}{\hat{S}_{true\,j}^{UV}} \tag{6.5}$$

such that the evaluation accounts for variations in the smearing and selection efficiency only. Cross section uncertainty on the background channel is still assessed through the residual $B_{reco\,i}^{UV} - B_{reco\,i}^{CV}$.

Once the systematic variations have been created, a covariance matrix can be constructed as follows:

$$cov(i,j) = \frac{1}{N_{UV}} \sum_{k=0}^{N_{UV}} (x_i^k - x_i^{CV})(x_j^k - x_j^{CV}) \tag{6.6}$$

where x_i^k is the event rate in universe k subtracted by the CV background (as defined in Equation 6.1), x_i^{CV} is the CV event rate subtracted by the CV background, and N_{UV} is the total number of systematic universes for each source of uncertainty. (For paired unisim variations—parameters that are symmetrically varied by $\pm 1\sigma$—N_{UV} is equal to 2. In the case of single unisim variations, $N_{UV} = 1$.) Note that the background term $B_{reco\,i}^{CV}$ cancels out in the construction of the covariance matrix due to the subtraction present in the numerator.

It is useful to construct a correlation matrix:

$$corr(i,j) = \frac{cov(i,j)}{\sqrt{cov(i,i)}\sqrt{cov(j,j)}} \tag{6.7}$$

as well as a fractional covariance matrix:

$$\frac{cov(i,j)}{x_i^{CV} x_j^{CV}}. \tag{6.8}$$

For statistical uncertainties in the beam-on data samples, the covariance matrix is constructed as a diagonal matrix of the Poisson variance, which is simply the number of events

in each bin. For the statistical uncertainty on the MC and EXT samples, the covariance is constructed in a similar manner but using the sum of the squared weights, given that these events are weighted.

The total covariance matrix encodes information from both the statistical and systematic uncertainties and is constructed by summing the fractional covariance matrices from all sources, then converting back into absolute covariance units. The standard deviation of each bin with respect to the CV is evaluated by taking the square root of the diagonals of the total covariance matrix.

6.2 Flux models

As described in Section 4.1, a `Geant4`-based simulation of the NuMI flux is generated using a detailed geometric model of the beamline. This flux serves as input into an event generator to simulate neutrino interaction rates in MicroBooNE. `PPFX` [32, 33] constrains the underlying hadron production, scattering, and absorption cross sections used to estimate the flux as best as possible; however, there is still some uncertainty associated with these interaction models that carries over into the event rates. In addition, there is also uncertainty in the modeling of the NuMI beamline geometry itself. This section describes how these limitations affect the simulated background-subtracted event rates.

Hadron production, scattering, and absorption

Interaction probabilities for the processes listed in Table 2.1 affect the prediction of hadronic behavior along the NuMI beamline. `PPFX` constructs a multivariate Gaussian distribution with a mean equal to the central value probabilities:

$$N(\vec{x}, \vec{u}, \mathbf{V}) = \frac{1}{(2\pi)^{M/2}\sqrt{det(\mathbf{V})}} \exp(-\frac{1}{2}(\vec{x} - \vec{u}) \cdot \mathbf{V}^{-1} \cdot (\vec{x} - \vec{u})) \tag{6.9}$$

where $\mathbf{V}$ is the covariance matrix of all M parameters, $\vec{u}$ is the vector of central values, and

$\vec{x}$ is a free parameter of deviate values [32]. This distribution is randomly sampled to choose deviate parameters in place of the central values used in the neutrino flux determination. The effect on the resulting neutrino event rates is encapsulated in the form of a UV correction weight that is applied, instead of the CV `PPFX` reweighting factor described in Section 4.3.1, to events in the overlay samples. This procedure is repeated 600 times for the FHC+RHC event rates to produce a representative set of multisim variations that can be plotted as a function of the cross section variables, as shown in Figure 6.1.

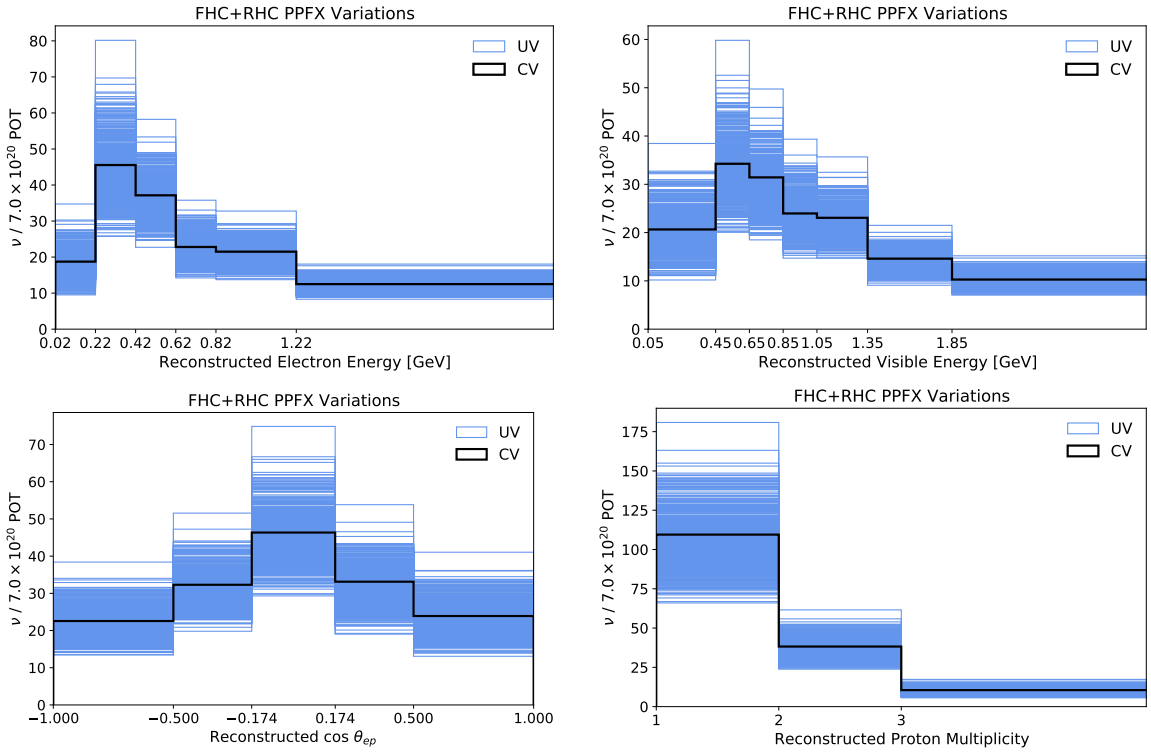


Figure 6.1: `PPFX` multisim variations (blue) for the background-subtracted FHC+RHC selected event rates as a function of each cross section variable. Shown in comparison with the CV event rate (black).

Figures 6.2, 6.3, 6.4, and 6.5 display the covariance and correlation matrices derived from the `PPFX` multisim variations for reconstructed electron energy, visible energy, opening angle, and proton multiplicity, respectively. Fractional uncertainties are shown in Figure 6.6, comparing curves derived from FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC

(black) multisim variations. Overall, uncertainty due to hadronic behavior along the NuMI beamline ranges from 15-20% across all bins for each cross section variable.

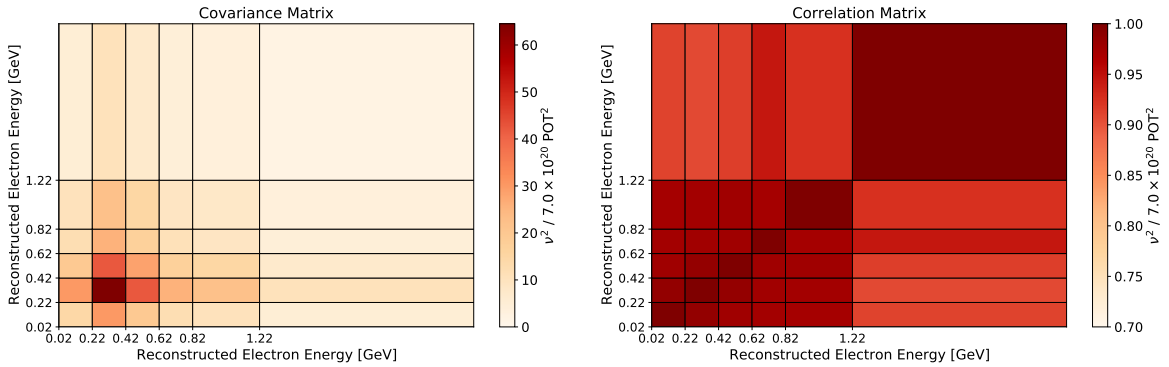


Figure 6.2: Covariance (left) and correlation (right) matrices as a function of reconstructed electron energy, derived from 600 PPFX multisim universes.

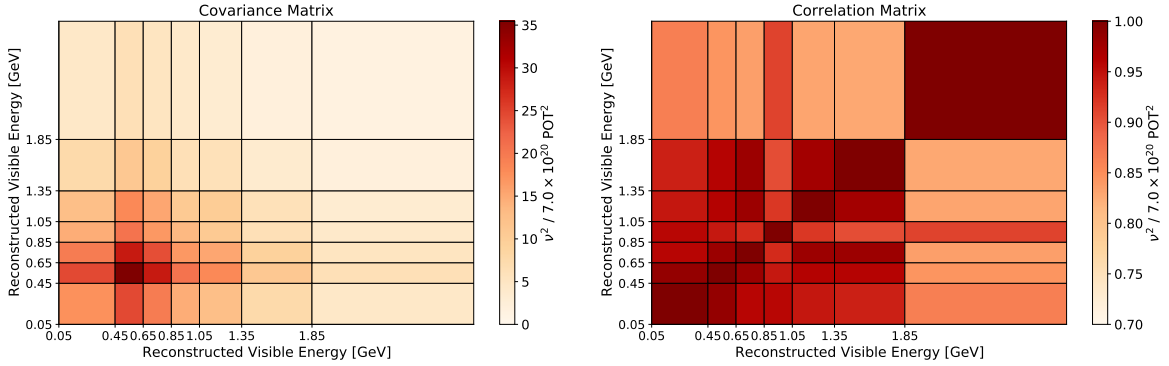


Figure 6.3: Covariance (left) and correlation (right) matrices as a function of reconstructed visible energy, derived from 600 PPFX multisim universes.

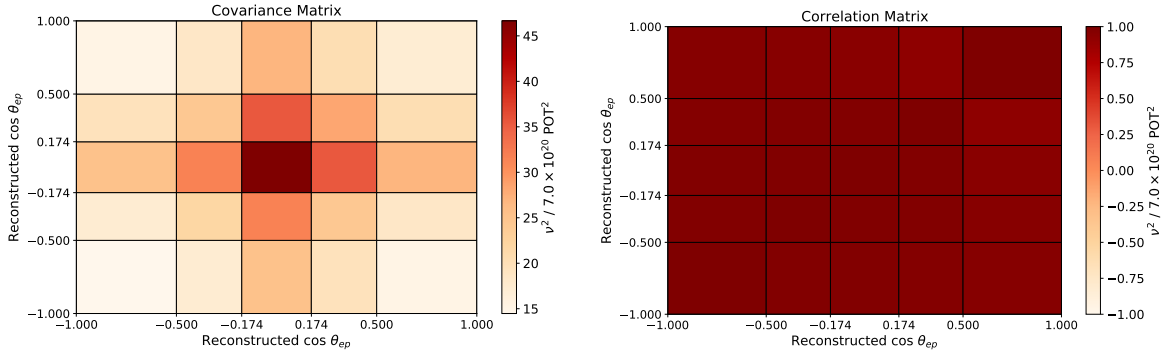


Figure 6.4: Covariance (left) and correlation (right) matrices as a function of reconstructed opening angle, derived from 600 PPFX multisim universes.

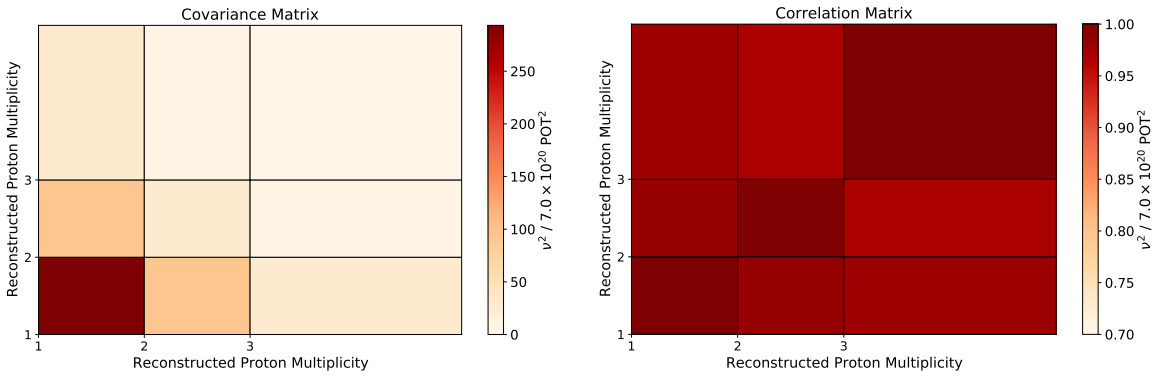


Figure 6.5: Covariance (left) and correlation (right) matrices as a function of reconstructed proton multiplicity, derived from 600 PPFX multisim universes.

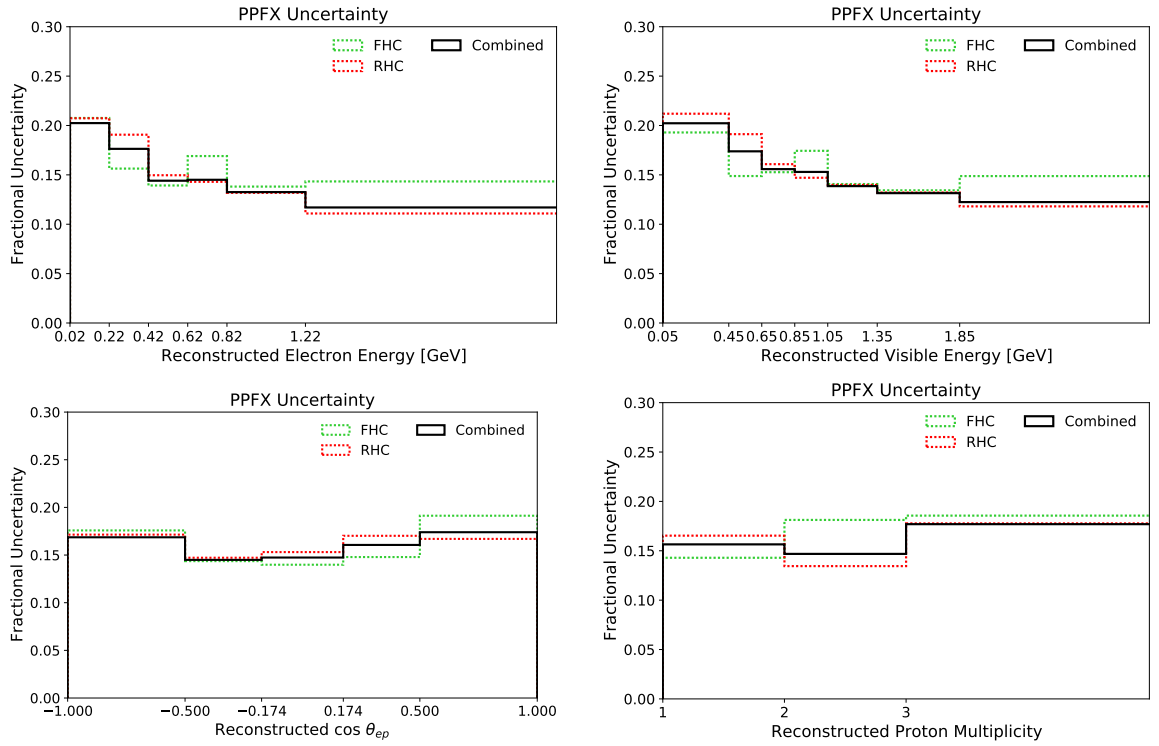


Figure 6.6: Fractional uncertainty on the background-subtracted selected event rates due to hadronic behavior along the beamline. Shown for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black).

Beamline geometry

Uncertain parameters in the modeling of the beamline geometry include the positions and currents of the focusing horns, the position of the graphite target, and the position and size of the beam spot incident upon it. Because of their current load, the horns are constantly sprayed for cooling, which creates a thin layer of water around their inner conductors—the thickness of this layer is an additional source of uncertainty. Central values for these parameters and their $\pm 1\sigma$ variations are listed in Table 6.1.

Parameter	**Central value**	**Variation**
Horn current	200 kA	± 2 kA
Horn 1 `x` position	0 mm	± 3 mm
Horn 1 `y` position	0 mm	± 3 mm
Beam spot size	1.3 mm	± 0.2 mm
Horn 2 `x` position	0 mm	± 3 mm
Horn 2 `y` position	0 mm	± 3 mm
Water on horns	1 mm	± 1 mm
Beam `x` position	0 mm	± 1 mm
Beam `y` position	0 mm	± 1 mm
Target `z` position	143.3 cm	± 7 mm

Table 6.1: Uncertain parameters in the geometric beamline model.

These parameters are considered to be uncorrelated, so uncertainty in the geometric beamline model is quantified using a unisim reweighting method. Two variations for each uncertain parameter are generated by swapping out the central value for its $+1\sigma$ or -1σ value in the neutrino flux determination. The effect on the resulting neutrino event rate is extrapolated into a UV reweight that is applied to each interaction in the overlay samples. Figures 6.7, 6.9, 6.11, and 6.13 show the FHC+RHC beamline geometry variations for each uncertain parameter as a function of reconstructed electron energy, visible energy, opening angle, and proton multiplicity, respectively. The ratios to the CV event rate are also displayed in lieu of the CV distribution, since the differences are only a few percent.

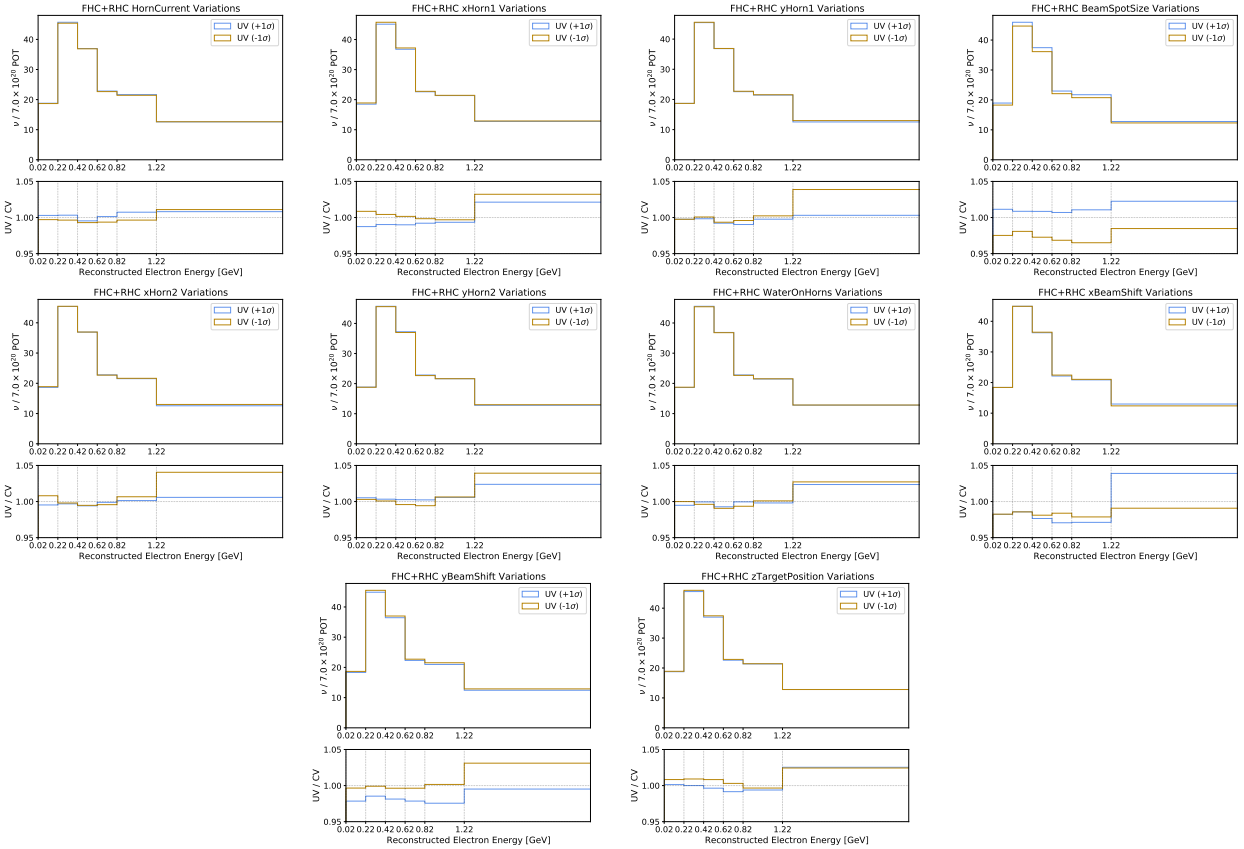


Figure 6.7: Beamline geometry variations of the background-subtracted selected event rate and the ratio to CV as a function of reconstructed electron energy.

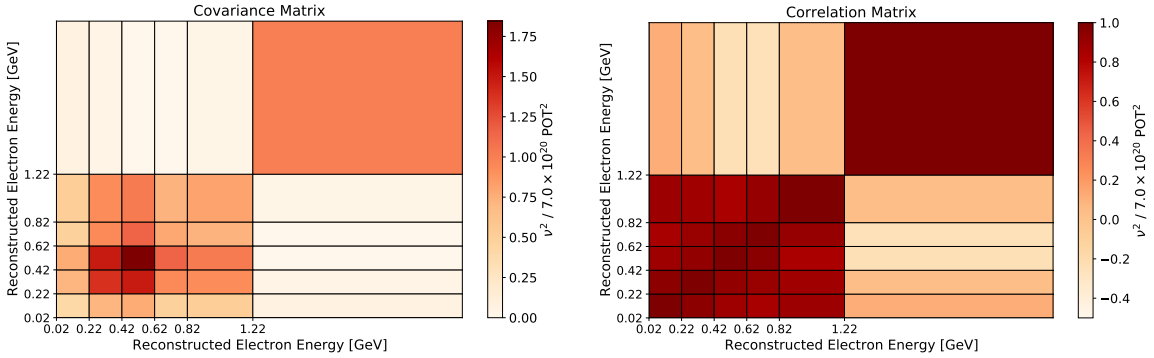


Figure 6.8: Covariance (left) and correlation (right) matrices as a function of reconstructed electron energy, derived from 10 $\pm 1\sigma$ beamline unisim universes.

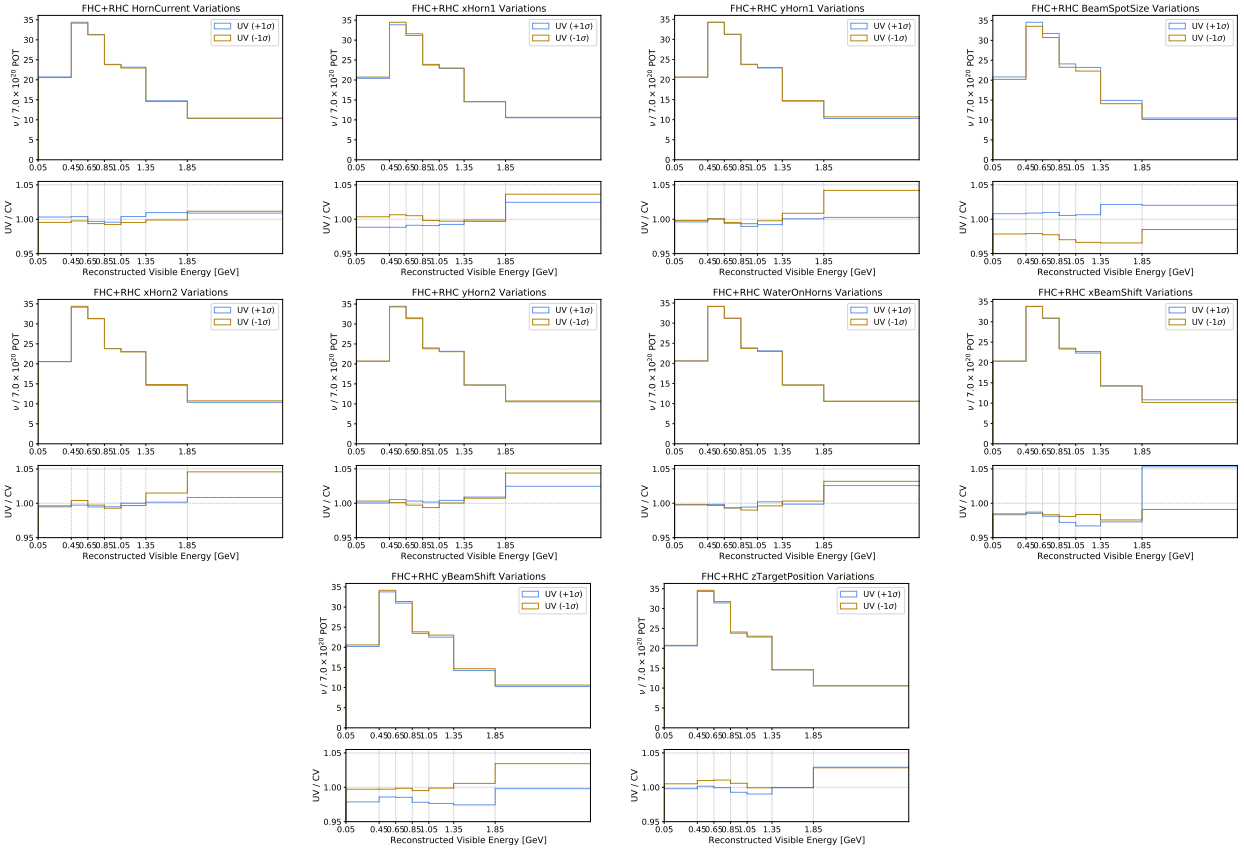


Figure 6.9: Beamline geometry variations of the background-subtracted selected event rate and the ratio to CV as a function of reconstructed total visible energy.

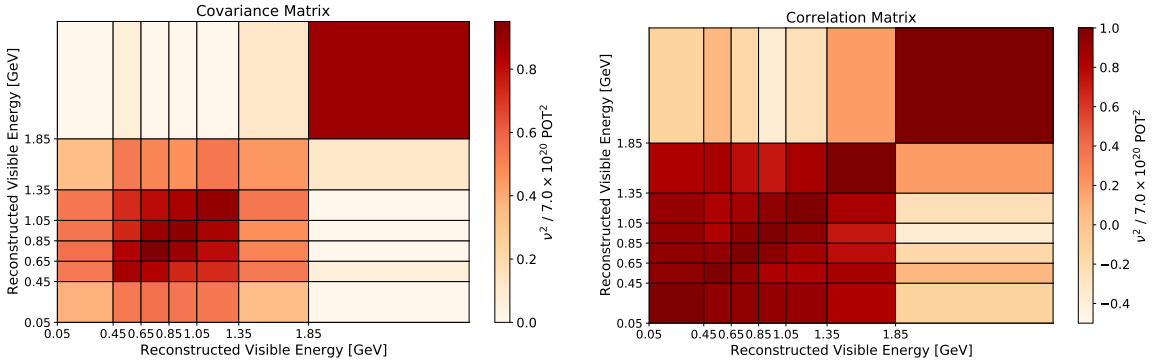


Figure 6.10: Covariance (left) and correlation (right) matrices as a function of reconstructed total visible energy, derived from 10 $\pm 1\sigma$ beamline unisim universes.

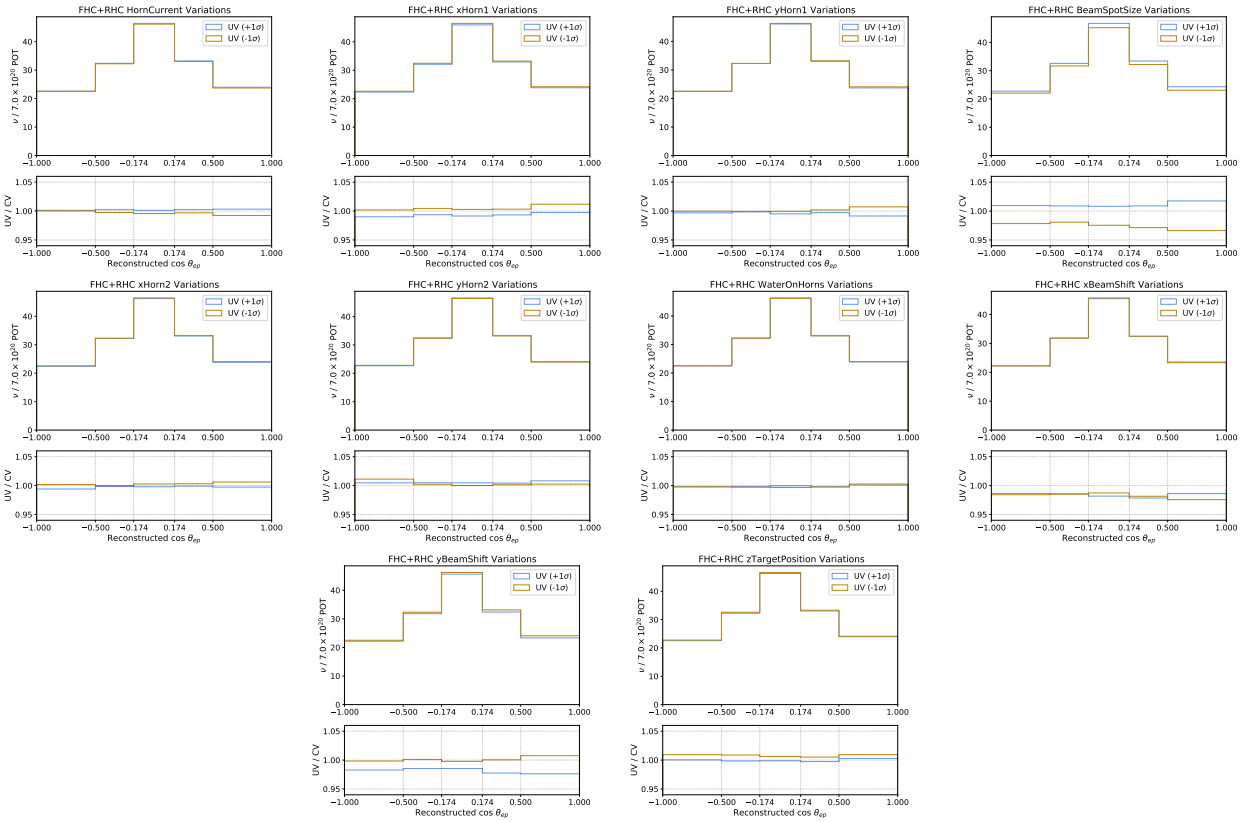


Figure 6.11: Beamline geometry variations of the background-subtracted selected event rate and the ratio to CV as a function of reconstructed opening angle.

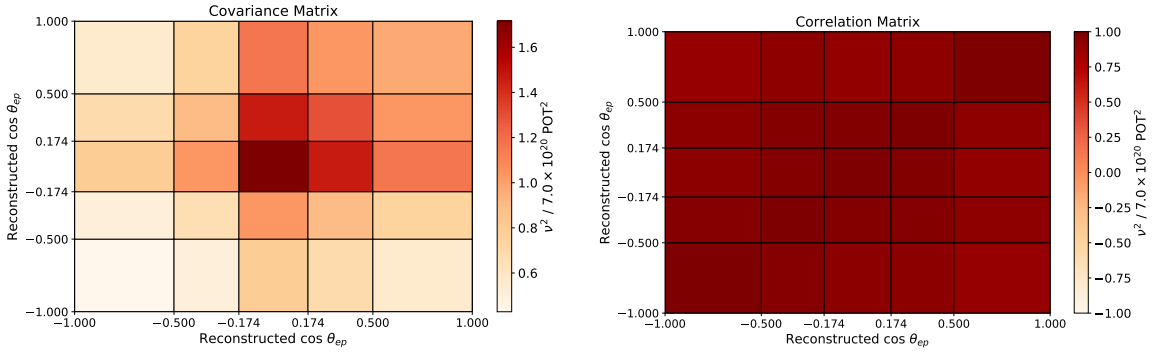


Figure 6.12: Covariance (left) and correlation (right) matrices as a function of reconstructed opening angle, derived from 10 $\pm 1\sigma$ beamline unisim universes.

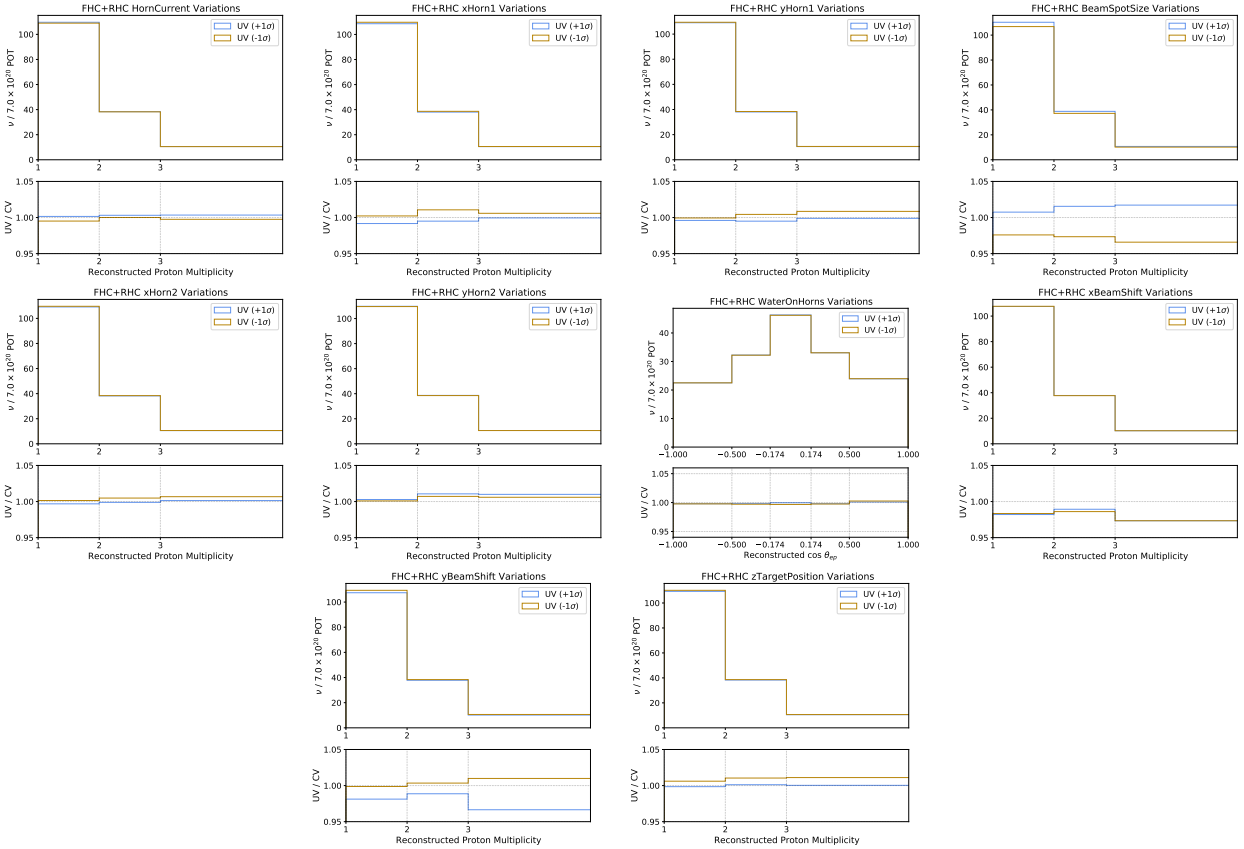


Figure 6.13: Beamline geometry variations of the background-subtracted selected event rate and the ratio to CV as a function of reconstructed proton multiplicity.

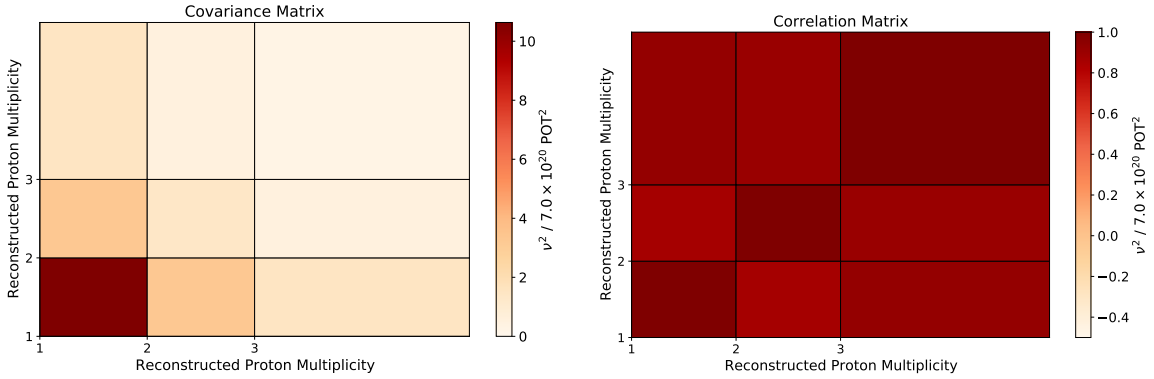


Figure 6.14: Covariance (left) and correlation (right) matrices as a function of reconstructed proton multiplicity, derived from 10 $\pm 1\sigma$ beamline unisim universes.

A covariance matrix is constructed for each unisim pair. These matrices are added together to create the total beamline geometry covariance and correlation matrices shown in Figures 6.8, 6.10, 6.12, and 6.14 for reconstructed electron energy, visible energy, opening angle, and proton multiplicity, respectively. Fractional uncertainties are displayed in Figure 6.15, comparing curves derived for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) unisim variations. Overall, uncertainty from modeling of the beamline geometry ranges from 2-9% across all bins for each cross section variable.

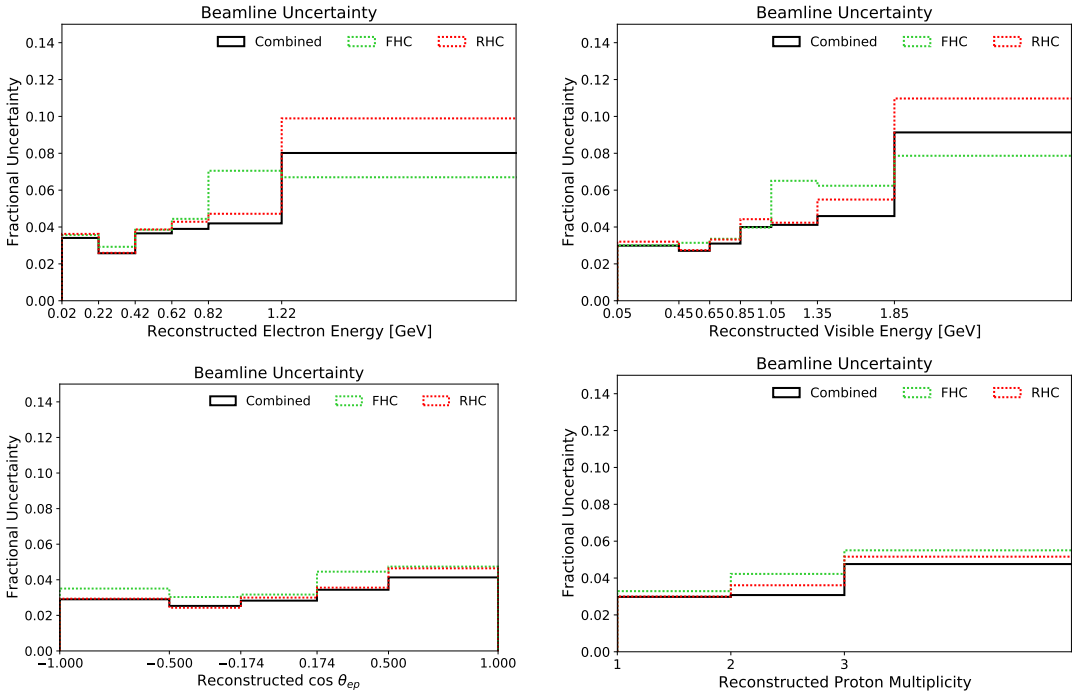


Figure 6.15: Fractional beamline geometry uncertainty on the background-subtracted selected event rate as a function of reconstructed electron energy, total visible energy, opening angle, and proton multiplicity for FHC (green), RHC (red), and FHC+RHC (black).

6.3 Interaction models

Cross section models are employed in the `GENIE v3.0.6` neutrino interaction generator as well as in the `Geant4`-based simulation of charged particles propagating and re-interacting in the MicroBooNE detector. This section reports the associated uncertainties in these models,

which impacts both the true and reconstructed event rates.

GENIE uncertainties

As described in Section 6.1, the systematic change in the background-subtracted event rates arising from GENIE uncertainties on the signal channel is not taken into account in the measurements of this work. Thus, the covariance matrices evaluated in this section encode the GENIE systematic uncertainty as it pertains to changes in the event rate, detector smearing, and selection efficiency on the background channel, but only the impact of the latter two effects on the signal channel.

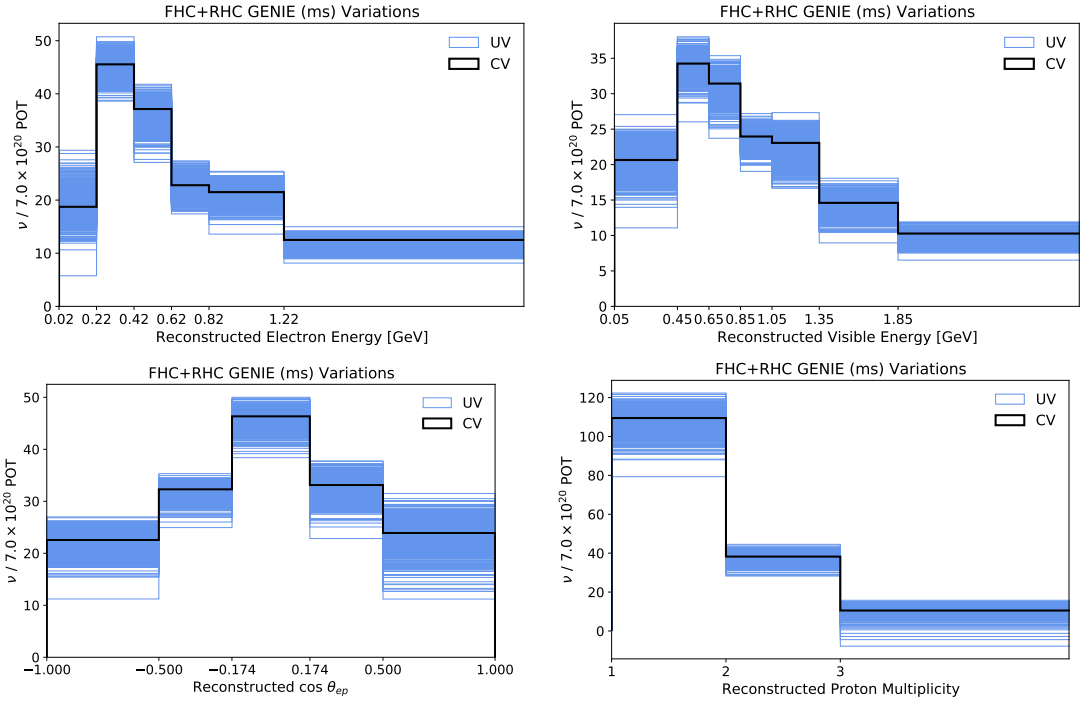


Figure 6.16: GENIE multisim variations (blue) for the background-subtracted FHC+RHC selected event rates as a function of each cross section variable. Shown in comparison with the CV event rate (black).

Most GENIE uncertainties are evaluated with 600 multisim universes, as shown in Figure 6.16. These are created by varying parameters that belong to the suite of interactions listed in Table 6.2. Taken into account are quasi-elastic and resonant processes, as well as meson

exchange current (MEC) interactions, in which a muon neutrino interacts with a bound pair of nucleons and produces a $1\mu + 2p^+$ signature. Deep inelastic scattering and final state interactions (when hadrons get released before leaving the nucleus) are also included.

Parameter	Description
Quasi-elastic interactions	
`MaCCQE`	CCQE axial mass parameter
`CoulombCCQE`	Value of Coulomb potential used in CCQE corrections
`MaNCEL`	NCEL axial mass
`EtaNCEL`	Empirical parameter used to account for sea quark contribution to NCEL form factor
MEC interactions	
`NormCCMEC`	Energy-independent normalization for CCMEC
`NormNCMEC`	Energy-independent normalization for NCMEC
`FracPNCCMEC`	Fraction of initial nucleon pairs that are pn
`FracDeltaCCMEC`	Relative contribution of Δ diagrams to total MEC cross section
Resonant interactions	
`MaCCRES`	CCRES axial mass
`MvCCRES`	Shape-only CCRES axial mass
`MaNCRES`	NCRES axial mass
`MvNCRES`	NCRES vector mass
Non-resonant interactions	
`NonRESBGvpNC1pi`	Non-resonant background normalization for νp NC1π
`NonRESBGvpNC2pi`	Non-resonant background normalization for νp NC2π
`NonRESBGvnNC1pi`	Non-resonant background normalization for νn NC1π
`NonRESBGvnNC2pi`	Non-resonant background normalization for νn NC2π
`NonRESBGvbarpNC1pi`	Non-resonant background normalization for $\bar{\nu}$p NC1π
`NonRESBGvbarpNC2pi`	Non-resonant background normalization for $\bar{\nu}$p NC2π
`NonRESBGvbarnNC1pi`	Non-resonant background normalization for $\bar{\nu}$n NC1π
`NonRESBGvbarnNC2pi`	Non-resonant background normalization for $\bar{\nu}$n NC2π
`NonRESBGvpCC1pi`	Non-resonant background normalization for νp CC1π
`NonRESBGvpCC2pi`	Non-resonant background normalization for νp CC2π
`NonRESBGvnCC1pi`	Non-resonant background normalization for νn CC1π
`NonRESBGvnCC2pi`	Non-resonant background normalization for νn CC2π
`NonRESBGvbarpCC1pi`	Non-resonant background normalization for $\bar{\nu}$p CC1π
`NonRESBGvbarpCC2pi`	Non-resonant background normalization for $\bar{\nu}$p CC2π
`NonRESBGvbarnCC1pi`	Non-resonant background normalization for $\bar{\nu}$n CC1π
`NonRESBGvbarnCC2pi`	Non-resonant background normalization for $\bar{\nu}$n CC2π
`AhtBY`	A_{HT} higher-twist parameter in the Bodek-Yang model scaling variable ξ_w
`BhtBY`	B_{HT} higher-twist parameter in the Bodek-Yang model scaling variable ξ_w
`CV1uBY`	CV1u valence GRV98 PDF correction parameter in the Bodek-Yang model
`CV2uBY`	CV2u valence GRV98 PDF correction parameter in the Bodek-Yang model
Hadronization	
`AGKYxF1pi`	Hadronization parameter, applicable to true DIS interactions only
`AGKYpT1pi`	Hadronization parameter, applicable to true DIS interactions only
Final state interactions	
`MFP`$_\pi$	π mean free path
`MFP`$_N$	Nucleon mean free path
`FrCEx`$_\pi$	Fractional cross section for π charge exchange
`FrInel`$_\pi$	Fractional cross section for π inelastic scattering
`FrAbs`$_\pi$	Fractional cross section for π absorption
`FrCEx`$_N$	Fractional cross section for nucleon charge exchange
`FrInel`$_N$	Fractional cross section for nucleon inelastic scattering
`FrAbs`$_N$	Fractional cross section for nucleon absorption
Delta resonant decays	
`RDecBR1gamma`	$\Delta \to \gamma$ decay branching ratio
`RDecBR1eta`	$\Delta \to \eta$ decay branching ratio

Table 6.2: Uncertain parameters in `GENIE` that are varied to produce 600 multisim universes.

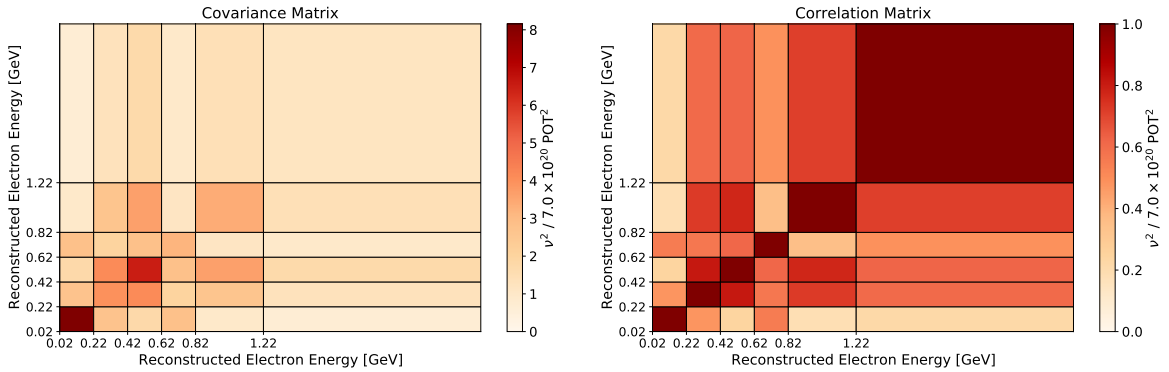


Figure 6.17: Covariance (left) and correlation (right) matrices as a function of reconstructed electron energy, derived from 600 `GENIE` multisim universes.

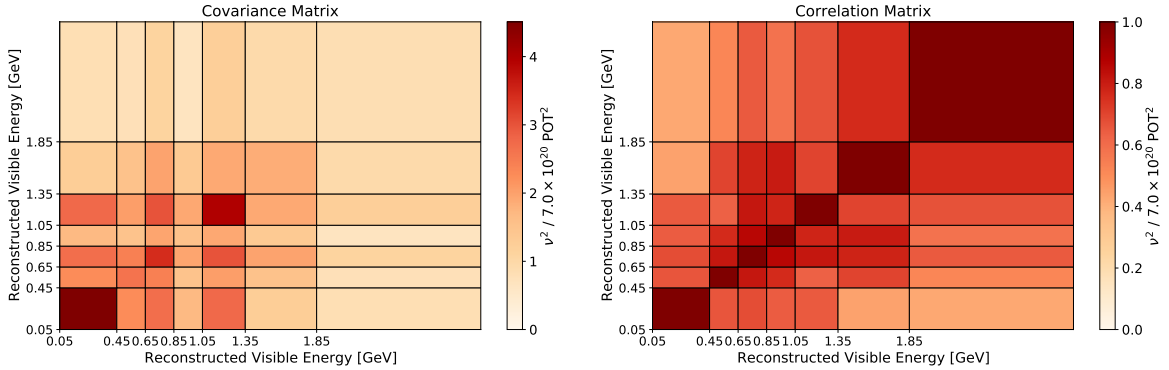


Figure 6.18: Covariance (left) and correlation (right) matrices as a function of reconstructed visible energy, derived from 600 `GENIE` multisim universes.

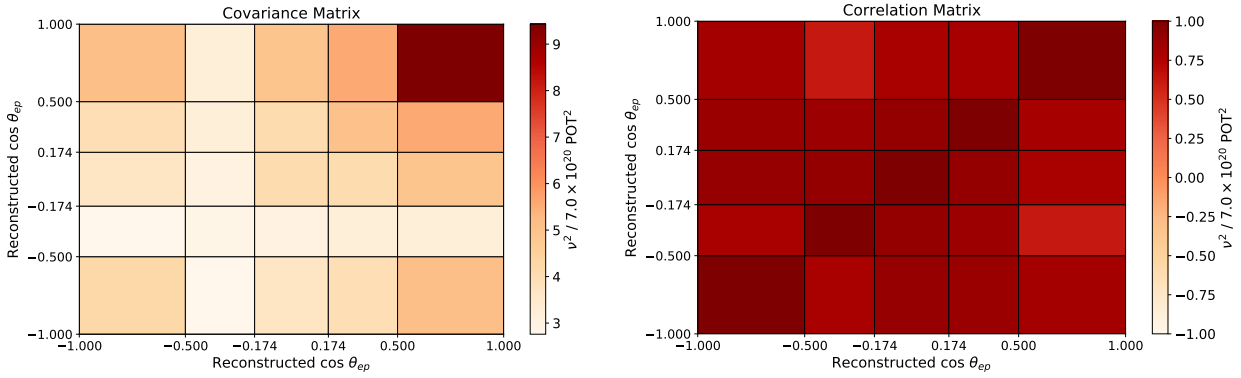


Figure 6.19: Covariance (left) and correlation (right) matrices as a function of reconstructed opening angle, derived from 600 `GENIE` multisim universes.

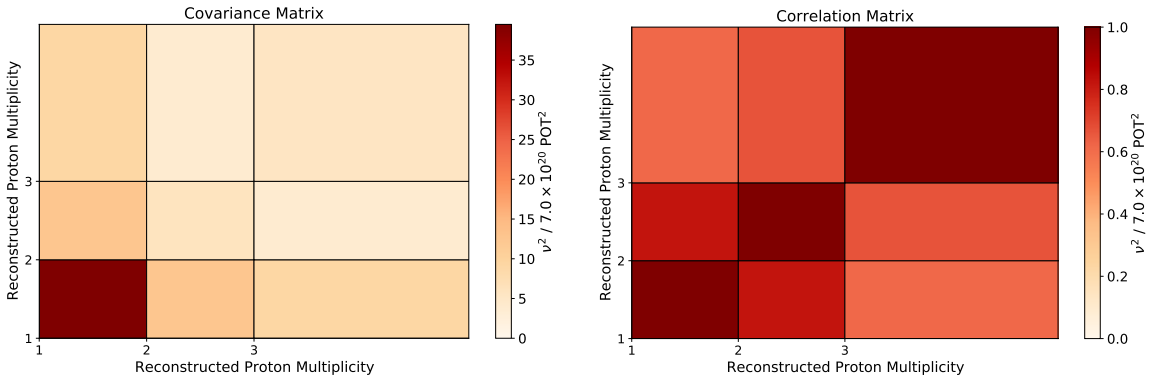


Figure 6.20: Covariance (left) and correlation (right) matrices as a function of reconstructed proton multiplicity, derived from 600 GENIE multisim universes.

Covariance and correlation matrices derived from these variations are shown in Figures 6.17, 6.18, 6.19, and 6.20 for the reconstructed electron energy, visible energy, opening angle, and proton multiplicity, respectively.

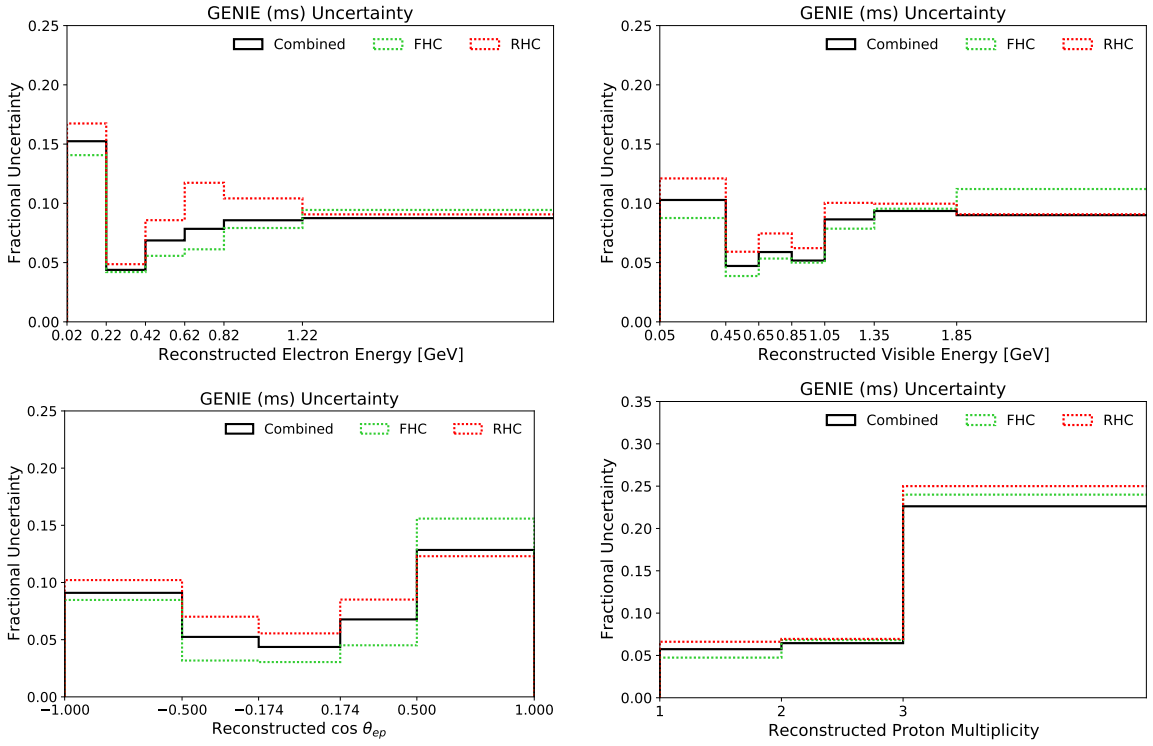


Figure 6.21: Fractional uncertainty on the background-subtracted event rates due to cross section models employed in the GENIE event generator. Shown for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black).

Fractional uncertainties are shown in Figure 6.21, comparing curves derived from FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) multisim variations. Overall, uncertainty from the parameters in Table 6.2 ranges from about 5 to 15% across all bins for each cross section variable.

Additional uncertain parameters are used in GENIE that are not included in the above multisim universes. These are shown in Table 6.3. Most of these are switches to alternative quasi-elastic, MEC, and resonant interaction models that manifest as single unisim variations. Parameters related to coherent pion production from neutrino-nucleus scattering are also included. Second class currents concern potential charge or time symmetry violation in the form factors that correct for the treatment of protons and neutrons as point-like [77]. These are not a part of the GENIE simulation, but are accounted for here because they contribute to uncertainty in the cross section models. Only the strength of the random phase approximation (RPA) correction, which suppresses the QE processes at low four-momenta, is varied with a $\pm 1\sigma$ unisim pair.

Knob	Description	CV	Variation
QE interactions			
RPA	Strength of the RPA correction for long-range nucleon correlations	0.4	± 1.0
AxFFCCQE	Switches parameterization of the nucleon axial form factor	Dipole	Z-expansion
VecFFCCQE	Switches the parameterization of the nucleon vector form factors	BBA07	Dipole
MEC interactions			
CCMEC	Switches to different CCMEC differential cross section model	Nieves	GENIE empirical
DecayAngMEC	Switches decay distribution of nucleon pairs in MEC interactions	Isotropic	$\cos^2\theta$
Resonant interactions			
ThetaDelta2Npi	Varies the pion angular distribution for $\Delta \to N + \pi$	Rein-Sehgal	Isotropic
ThetaDelta2NRad	Varies the angular distribution of the photon in radiative Δ decays	Isotropic	$\cos^2\theta$
Coherent π production			
NormCCCOH	Scaling factor for CC coherent π production	Nominal	100% increase
NormNCCCOH	Scaling factor for NC coherent π production	Nominal	100% increase
Second class currents			
xsr_scc_Fv3	Reweight CCQE to include second class vector currents	Nominal	Include SCC
xsr_scc_Fa3	Reweight CCQE to include second class axial currents	Nominal	Include SCC

Table 6.3: Uncertain parameters used to produce GENIE unisim variations.

Unisim variations for each cross section variable are shown in Figures 6.22, 6.24, 6.26, and 6.28. Ratios to the CV event rate are also displayed in lieu of the CV distribution, since the differences are only a few percent.

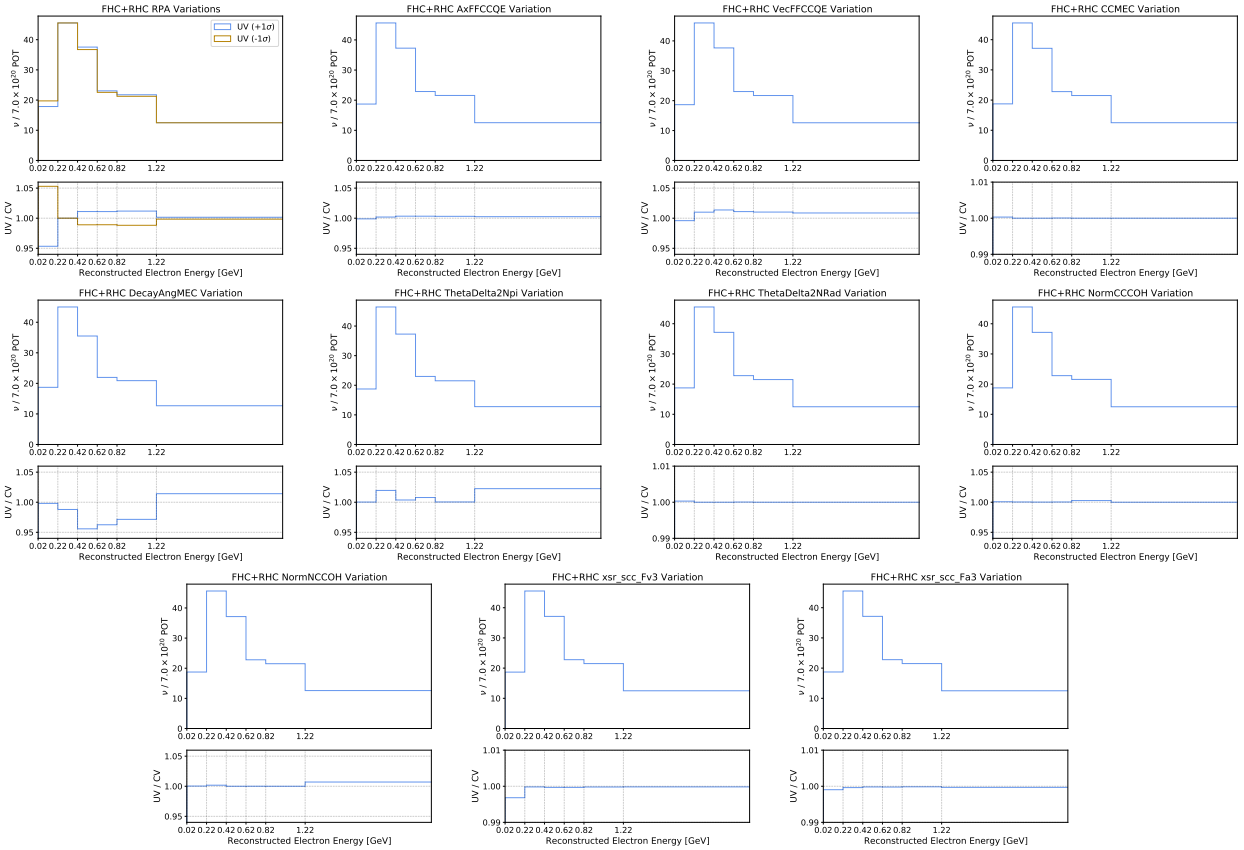


Figure 6.22: `GENIE` unisim variations of the background-subtracted selected event rate and the ratio to CV as a function of reconstructed electron energy.

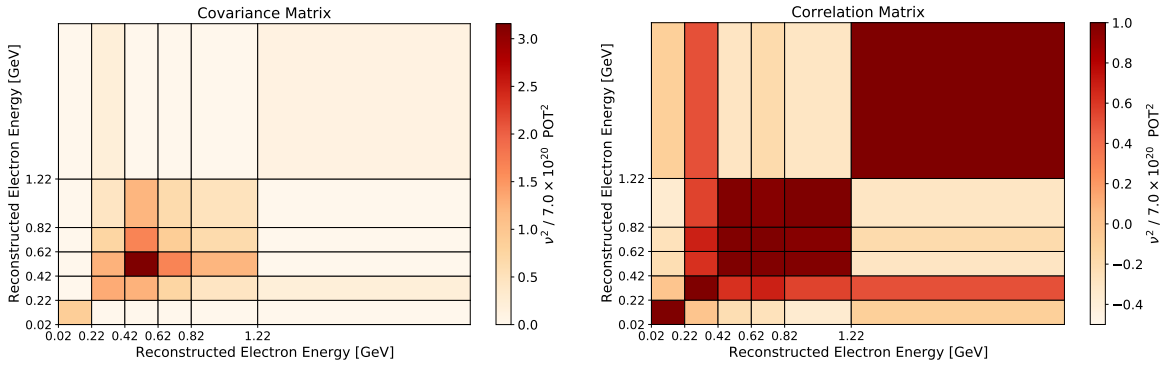


Figure 6.23: Covariance (left) and correlation (right) matrices as a function of reconstructed electron energy, derived from the `GENIE` unisim variations.

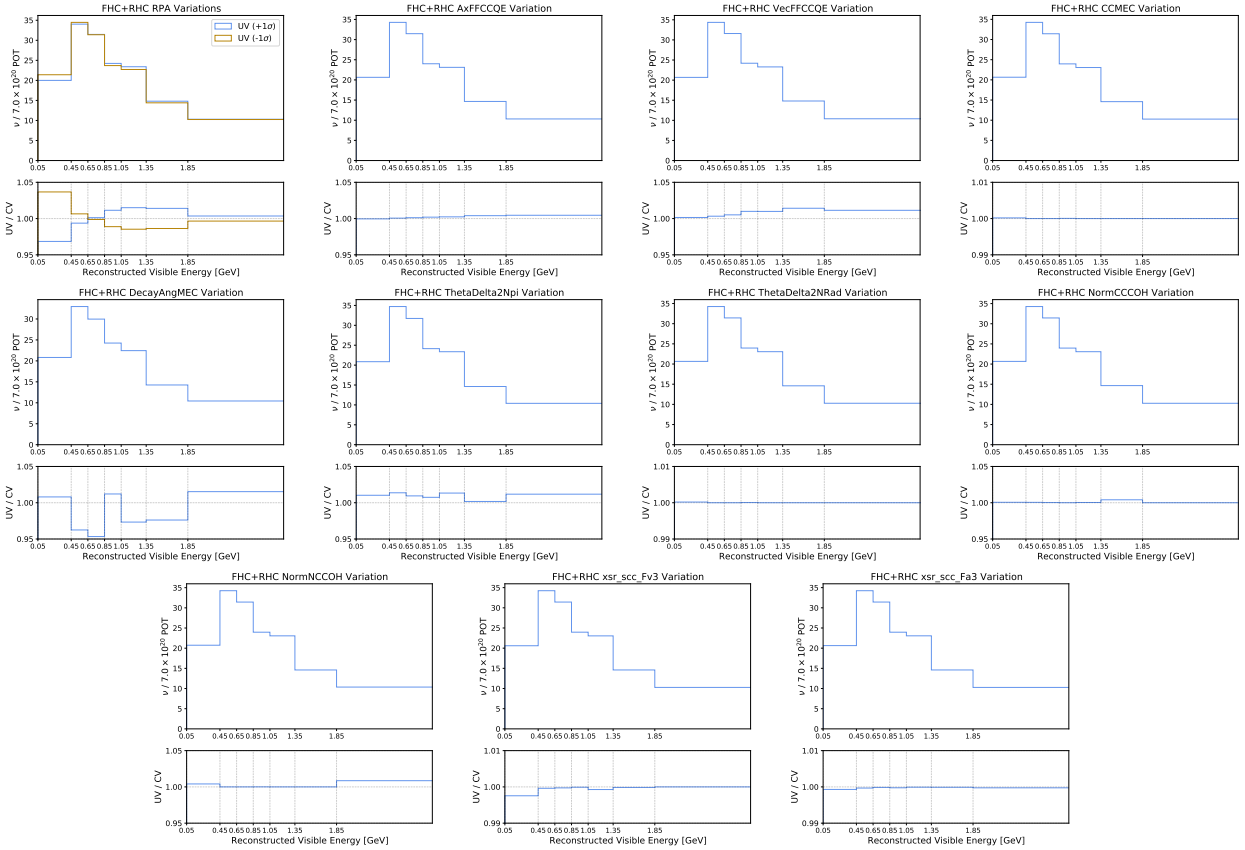


Figure 6.24: GENIE unisim variations of the background-subtracted selected event rate and the ratio to CV as a function of reconstructed visible energy.

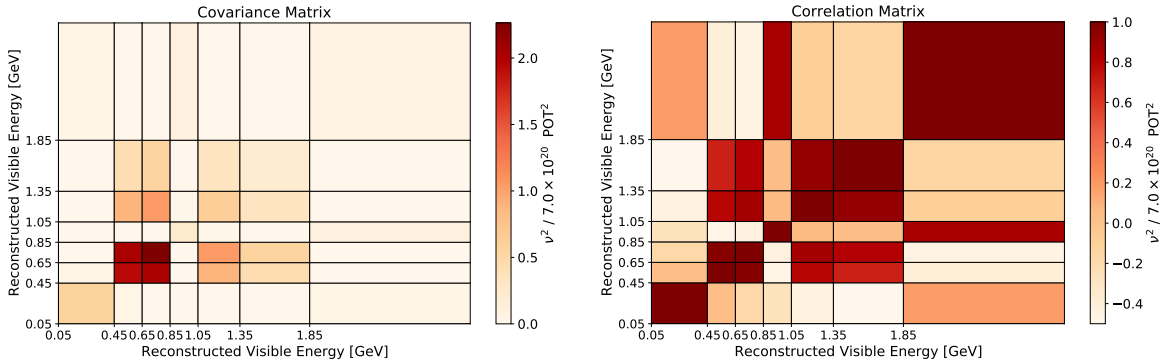


Figure 6.25: Covariance (left) and correlation matrices as a function of reconstructed visible energy, derived from the GENIE unisim variations.

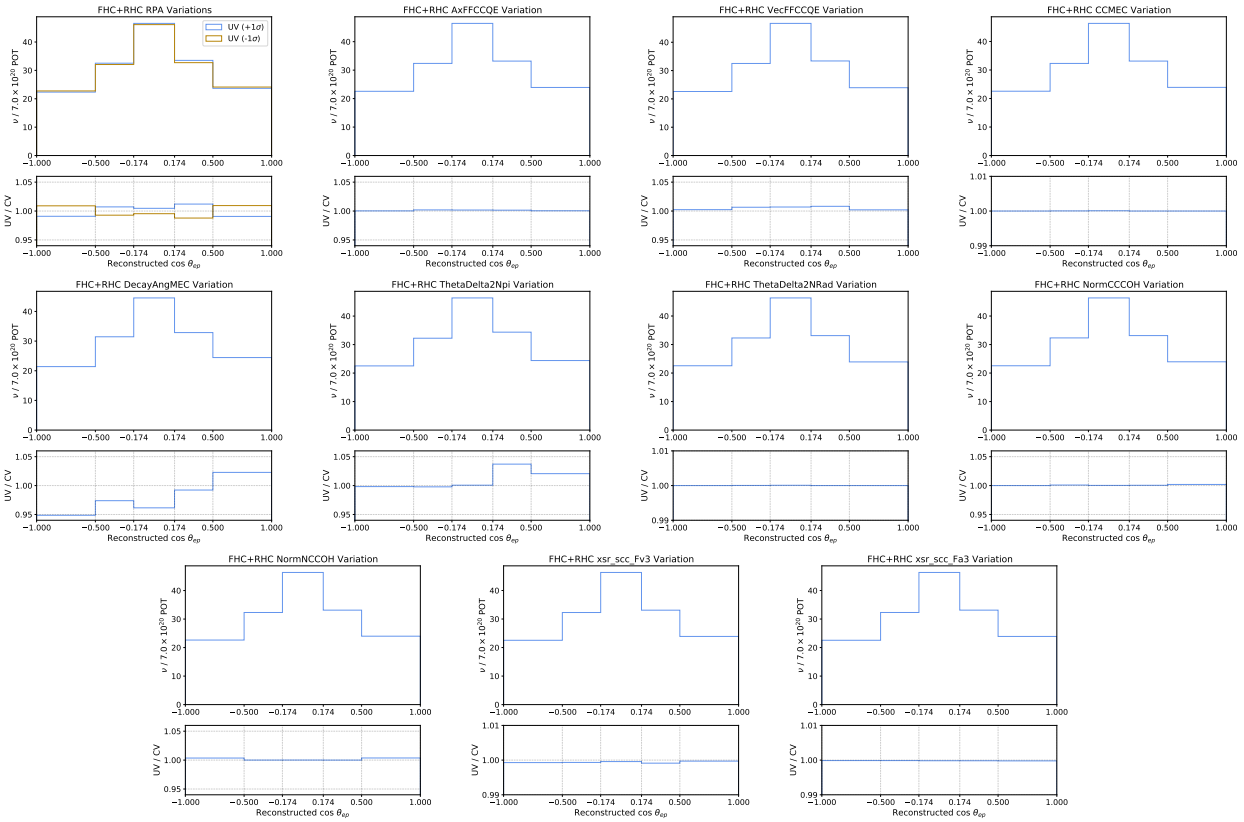


Figure 6.26: GENIE unisim variations of the background-subtracted selected event rate and the ratio to CV as a function of reconstructed opening angle.

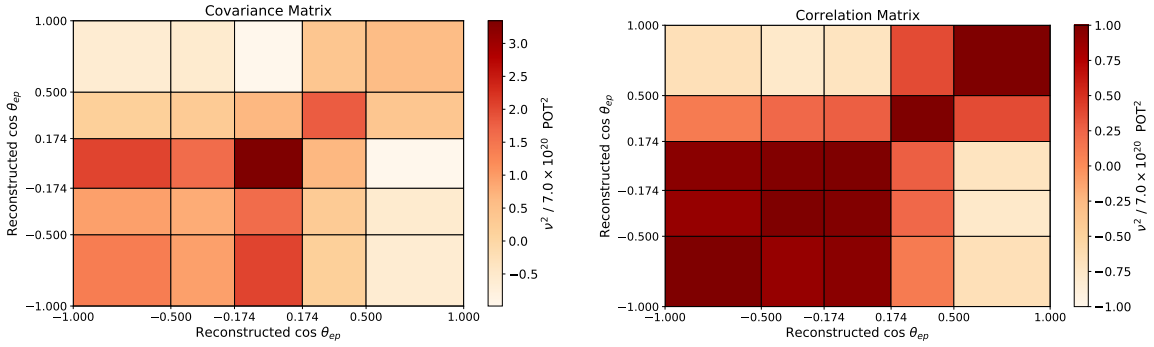


Figure 6.27: Covariance (left) and correlation (right) matrices as a function of reconstructed opening angle, derived from the GENIE unisim variations.

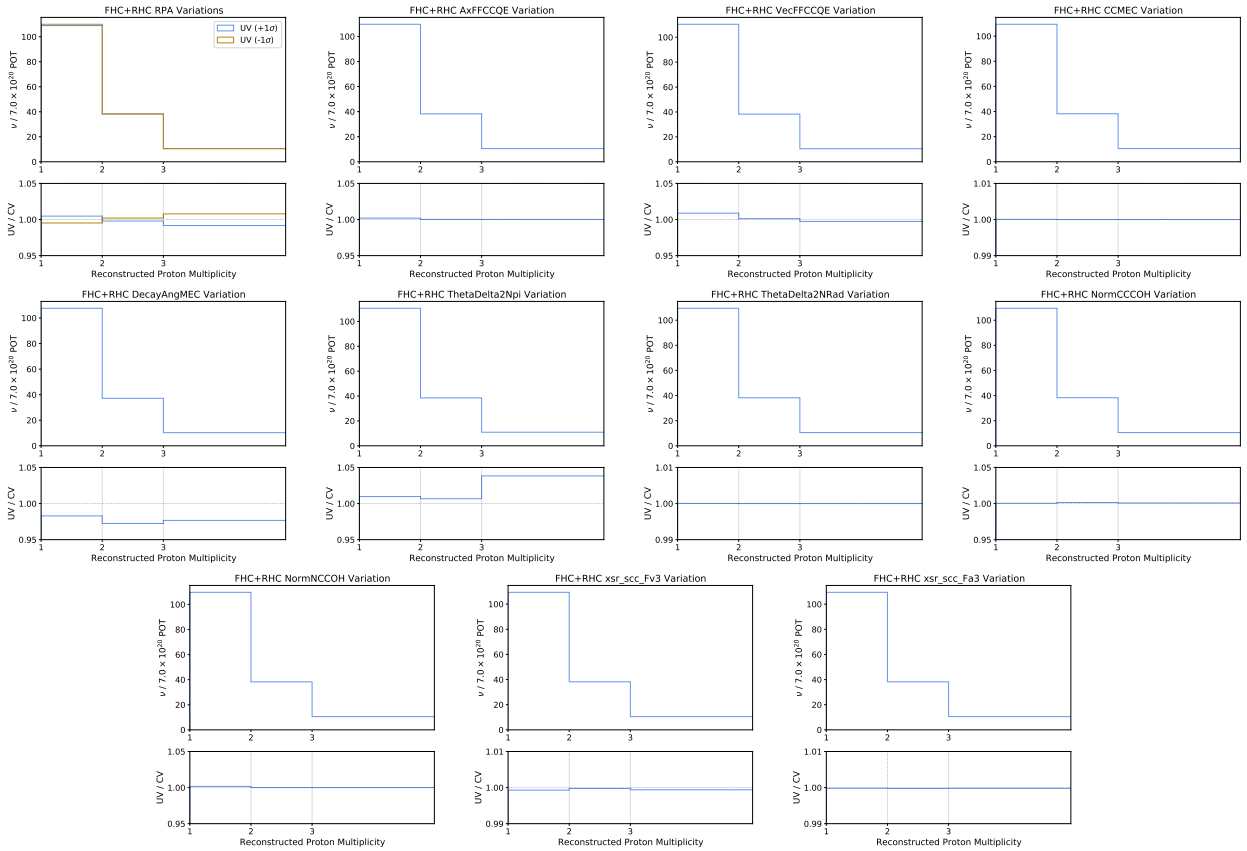


Figure 6.28: `GENIE` unisim variations of the background-subtracted selected event rate and the ratio to CV as a function of reconstructed proton multiplicity.

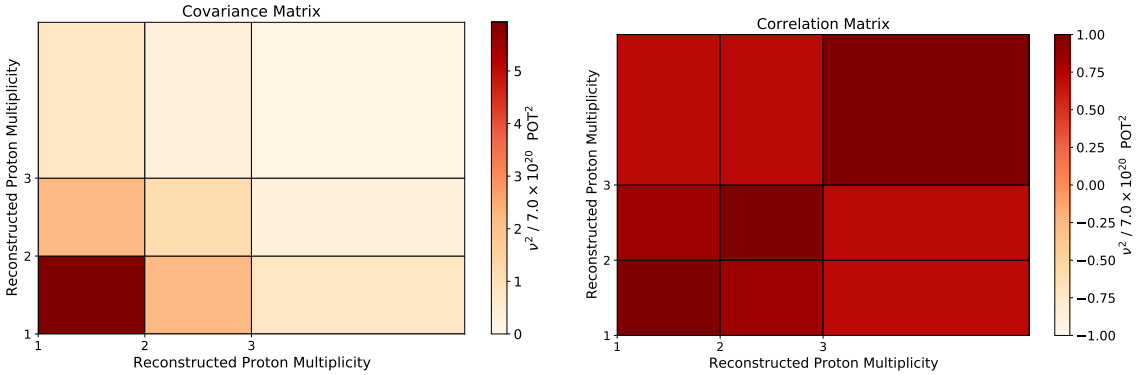


Figure 6.29: Covariance (left) and correlation (right) matrices as a function of reconstructed proton multiplicity, derived from the `GENIE` unisim variations.

A covariance matrix is constructed for each unisim. These matrices are then added together to produce the total `GENIE` unisim covariance and correlation matrices shown in

Figures 6.23, 6.25, 6.27, and 6.29 for reconstructed electron energy, visible energy, opening angle, and proton multiplicity, respectively. Fractional uncertainties are displayed in Figure 6.30, comparing curves derived for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) unisim variations. Overall, uncertainty from parameters in Table 6.3 ranges between 2-6% across all bins for each cross section variable.

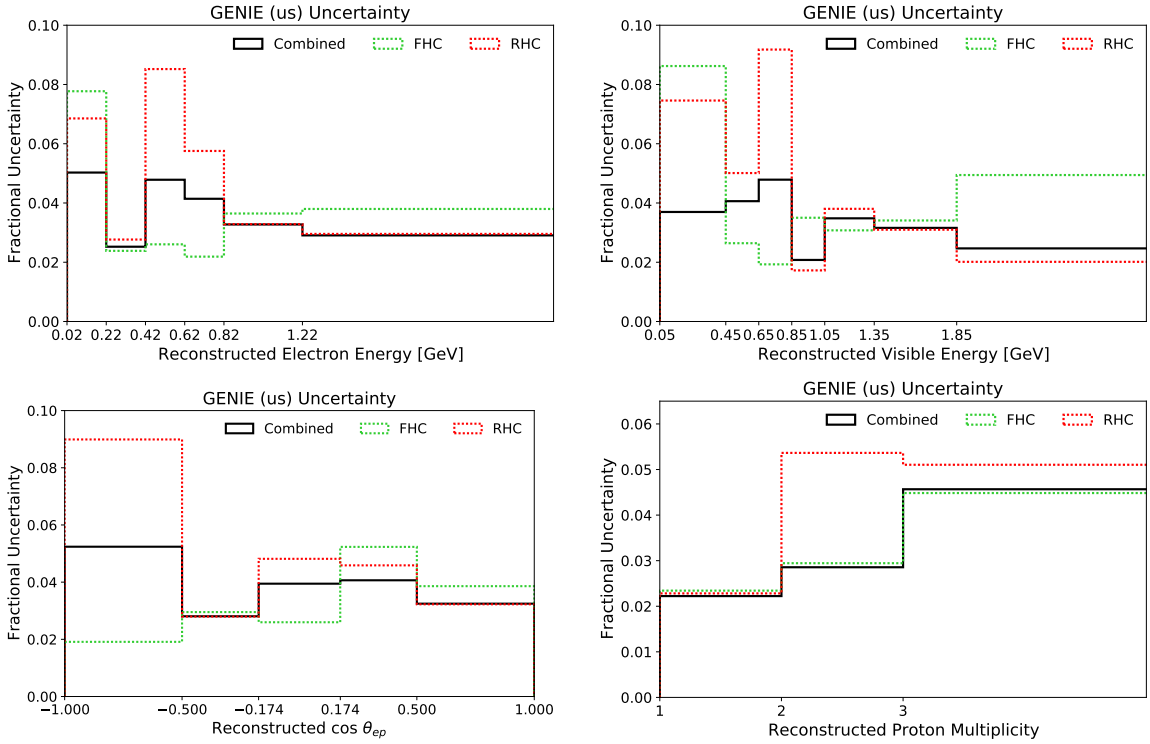


Figure 6.30: Fractional `GENIE` unisim uncertainties on the background-subtracted selected event rate. Shown as a function of electron energy, visible energy, opening angle, and proton multiplicity for FHC (green), RHC (red), and FHC+RHC (black).

Re-interaction uncertainties

The total cross section for neutrino-induced protons and charged pions in MicroBooNE can be written as the sum of each exclusive interaction channel:

$$\sigma_{total} = \sigma_{elast} + \sigma_{react} = \sigma_{elast} + (\sigma_{inel} + \sigma_{abs} + \sigma_{cex} + \sigma_{dcex} + \sigma_{prod}) \tag{6.10}$$

where σ_{elast} is the elastic cross section and σ_{react} is the reaction cross section, comprised

of inelastic, absorption, single and double charge exchange, and production cross sections, respectively. Table 6.4 provides a description for each of these channels as they pertain to `Geant4` charged pion re-interactions. Protons undergo similar processes in the LArTPC.

Variable	**Interaction**	**Description**
σ_{elast}	$\pi^{\pm} + N \rightarrow \pi^{\pm} + N$	Elastic scattering where the kinetic energy of the incident particle is preserved.
σ_{inel}	$\pi^{\pm} + N \rightarrow \pi'^{\pm} + N'$	Inelastic/quasi-elastic scattering where the kinetic energy of the incident particle is not preserved.
σ_{abs}	$\pi^{\pm} + N \rightarrow N'$	Absorption by a nucleus.
σ_{cex}	$\pi^{\pm} + N \rightarrow \pi^{0} + N'$	Single charge exchange that produces a neutral pion in the final topology.
σ_{dcex}	$\pi^{\pm} + N \rightarrow \pi^{\mp} + N'$	Double charge exchange that produces a charged pion of flipped charge in the final topology.
σ_{prod}	$\pi^{\pm} + N \rightarrow n\pi + N'$	Interaction with a nucleus that produces multiple pions in the final topology.

Table 6.4: `GEANT4` charged pion re-interaction channels. N' refers to the nucleus as well as any nuclear fragments on the outgoing side. Protons undergo similar processes.

The elastic and total reaction cross sections, σ_{elast} and σ_{react}, are varied for proton re-interactions. For pions, on the other hand, each exclusive channel listed in Table 6.4 is varied. (Electron scattering uncertainties are not implemented in this procedure.) The `Geant4Reweight` framework [78] is employed to generate the 1000 multisim universes shown in Figure 6.31 as a function of the cross section variables.

Covariance and correlation matrices derived from these variations are displayed in Figures 6.32, 6.33, 6.34, and 6.35 for reconstructed electron energy, visible energy, opening angle, and proton multiplicity, respectively. Fractional uncertainties are given in Figure 6.36, comparing curves from FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) multisim variations. Overall, uncertainty due to re-interaction models is no more than 3% across all bins for each cross section variable.

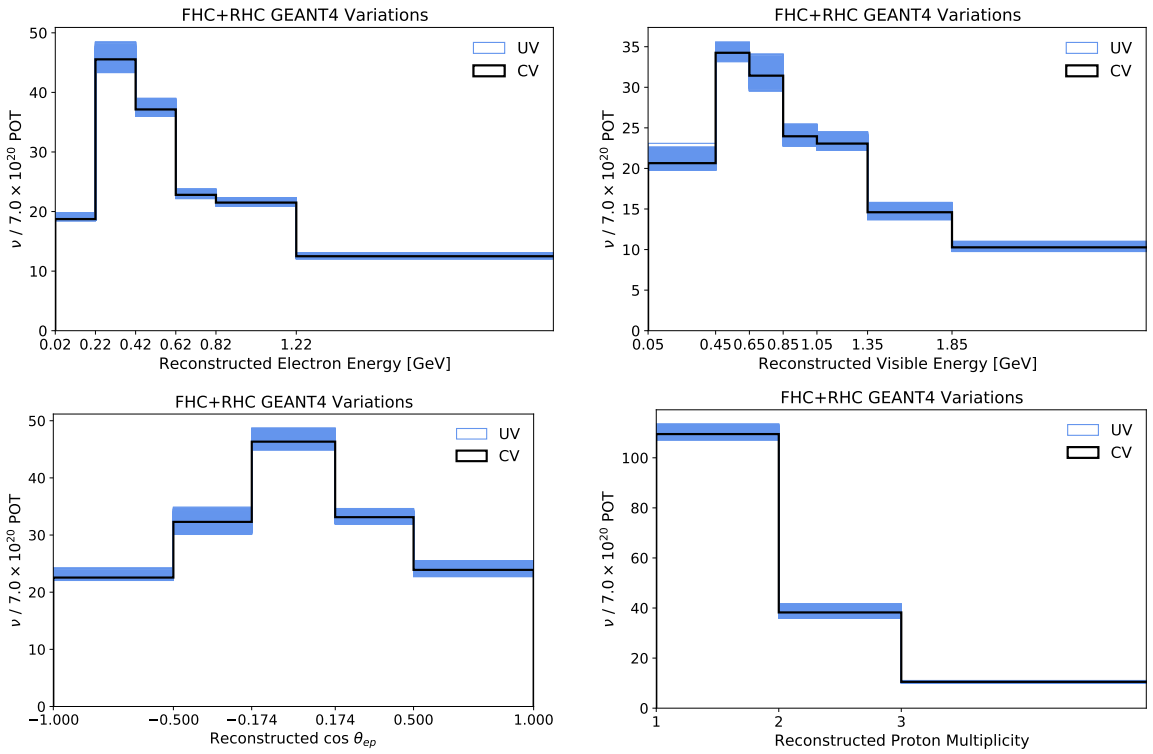


Figure 6.31: `Geant4` multisim uncertainties variations (blue) for the background-subtracted FHC+RHC selected event rates as a function of each cross section variable. Shown in comparison with the CV event rate (black).

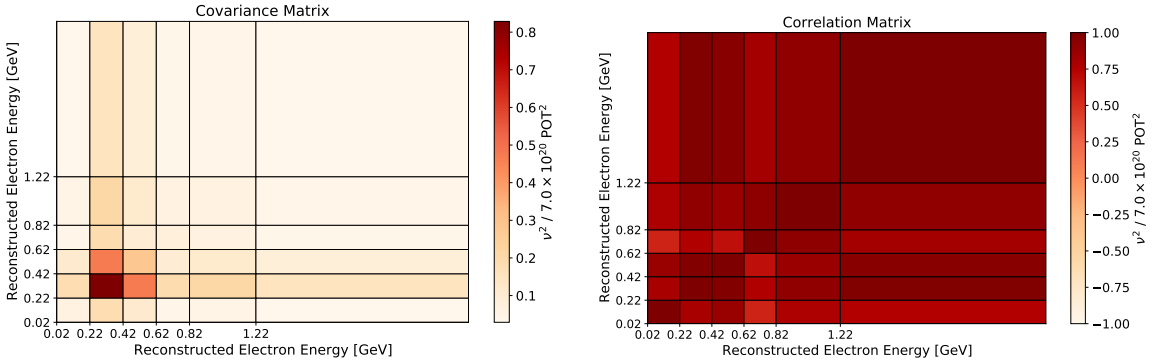


Figure 6.32: Covariance (left) and correlation (right) uncertainties as a function of reconstructed electron energy, derived from 1000 `Geant4` multisim universes.

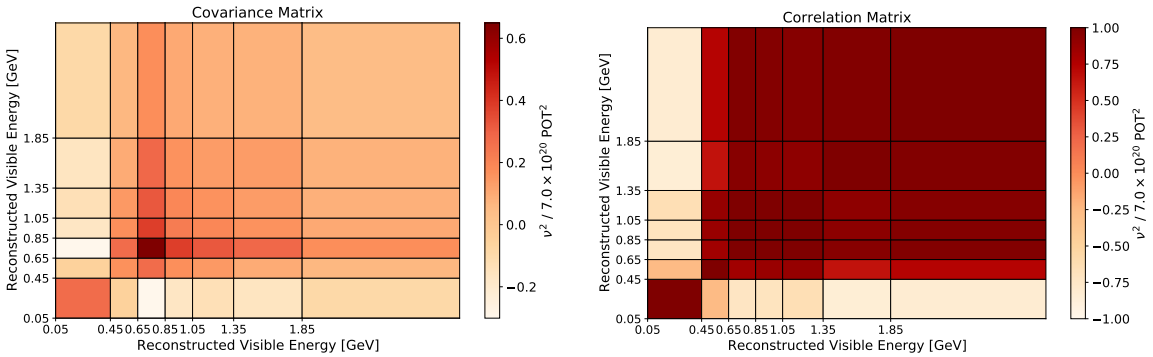


Figure 6.33: Covariance (left) and correlation (right) matrices as a function of reconstructed visible energy, derived from 1000 `Geant4` multisim universes.

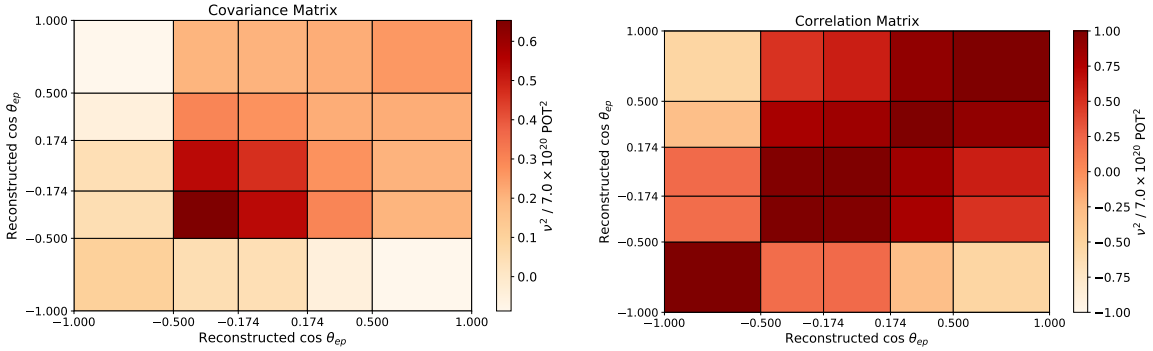


Figure 6.34: Covariance (left) and correlation (right) matrices as a function of reconstructed opening angle, derived from 1000 `Geant4` multisim universes.

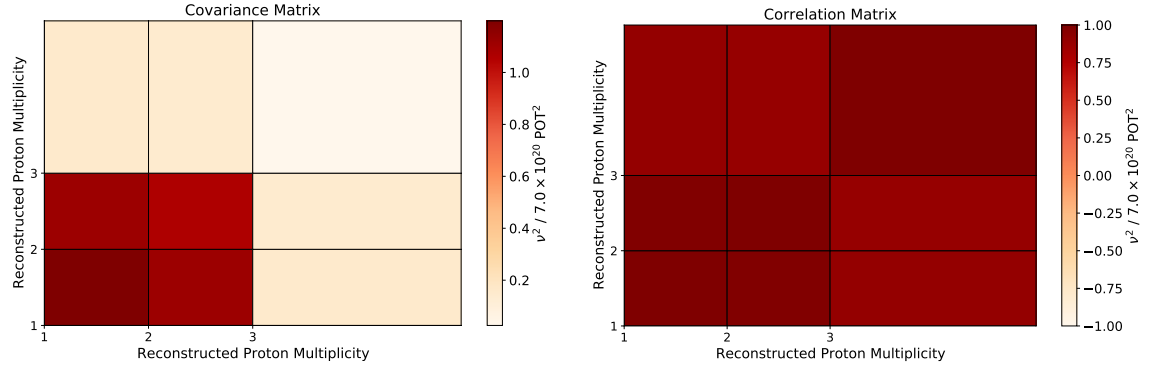


Figure 6.35: Covariance (left) and correlation (right) matrices as a function of reconstructed proton multiplicity, derived from 1000 `Geant4` multisim universes.

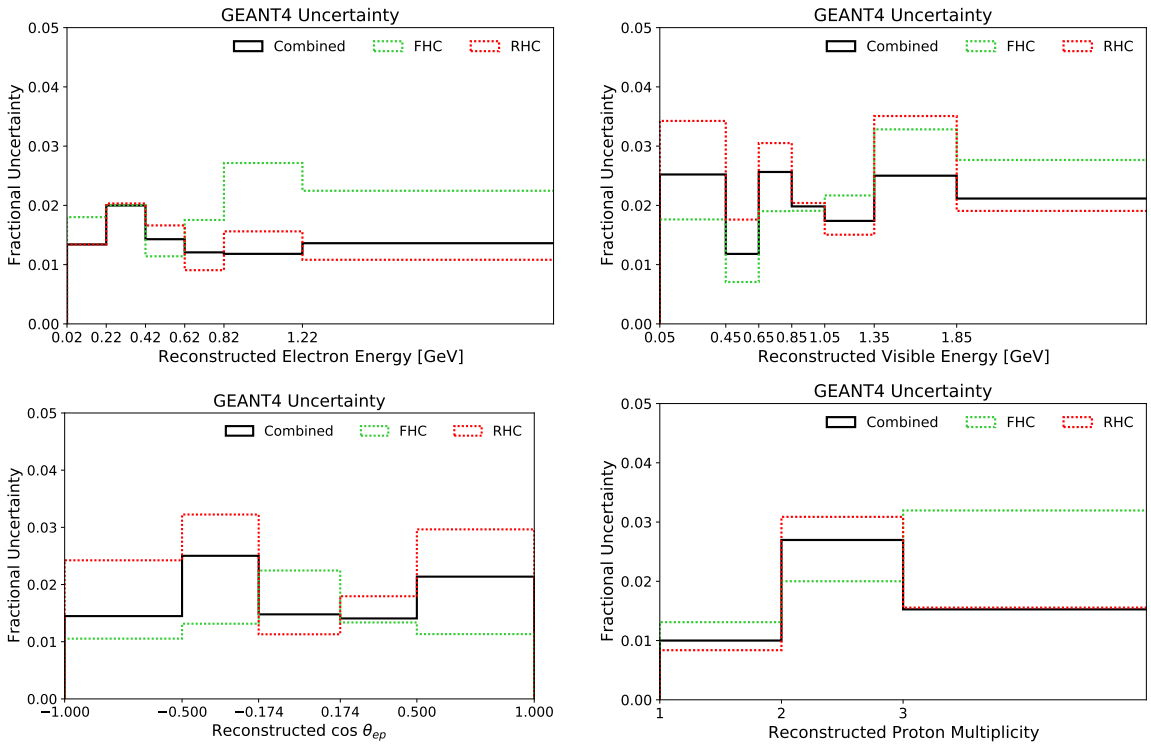


Figure 6.36: Fractional uncertainty on the background-subtracted event rates due to re-interaction models employed in the `Geant4Reweight` framework. Shown for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black).

6.4 Detector response

There is uncertainty associated with the modeling of how the detector responds to charge and light activity within the active volume. Rather than employ the reweighting technique of the previous sections, entire event samples with deviate parameters related to the detector response are simulated. Specifically, nine pairs of standard and intrinsic overlay samples are produced, where events in the latter replace ν_e CC interactions in the former, as described in Section 4.3.1. A single parameter is changed in each pair of samples, so the set corresponds to nine unisim variations that can be used to assess the detector uncertainty.(Instead of using the nominal CV event rates, a dedicated pair of samples with CV parameters is also created for direct comparison to the unisims.)

Table 6.5 lists the parameters adjusted to an alternate value or model to create the detector unisim samples. The first three knobs are related to light yield (LY) in the LArTPC: the attenuation as it travels (which affects RHC Run 3 only), Rayleigh scattering effects by particles much smaller than the photon wavelength, and the amount of photoelectrons produced by the PMTs. The next four knobs alter the shape and height of the wire waveforms resulting from incident charge on the anode planes. The final two knobs deal with space charge effects that distort the shape and strength of the electric field, and the modeling of how argon ions and free electrons recombine in the detector, which can reduce charge quantity and, as a result, signal intensity.

Knob	**Description**
Light	
LY Attenuation	Include attenuation effects in the light yield calculation (RHC Run 3 only)
LY Rayleigh	Increase the length of Rayleigh scattering from 60 cm to 90 cm
LY Down	25% decrease in the production of PMT photoelectrons
Wire modification	
WireMod X	Modification to the deconvolved waveforms in `x`
WireMod Y	Modification to the deconvolved waveforms in `yz` plane
WireMod ThetaXZ	Modification to the deconvolved waveforms in θ_{xz} plane
WireMod ThetaYZ	Modification to the deconvolved waveforms in θ_{yz} plane
Charge	
Space charge effects	Switch to alternative space charge map
Recombination	Use $\beta = 0.184$ for inverse modified Box model (Equation 4.1)

Table 6.5: Parameters are adjusted to create the detector response unisim samples.

Each detector variation sample is fed into the selection algorithm detailed in Chapter 5. However, due to the limited size of the standard overlay samples, selected event rates are dominated by statistical fluctuations. Figure 6.37 illustrates this concern for FHC Run 1, comparing the selected ν_e CC events from the intrinsic overlay samples (left) to selected non-ν_e CC interactions in the standard overlay samples (middle) as a function of reconstructed electron energy. The standard overlay generated for detector uncertainties has a POT $\sim 10^{20}$, one order of magnitude smaller than the nominal sample counterparts listed in Section 4.3.1. This, as well as the application of a robust selection targeting the ν_μ and ν_e NC interactions

present in these samples, prevents meaningful uncertainty assessment of these backgrounds. Instead, the result is a set of detector variations driven by statistical fluctuations rather than systematic changes in the event rate, which produces the inflated uncertainty estimation ($\sim$ 27%) shown on the right of Figure 6.37. This limitation is not mitigated by the combination of FHC Run 1 and RHC Run 3 samples.

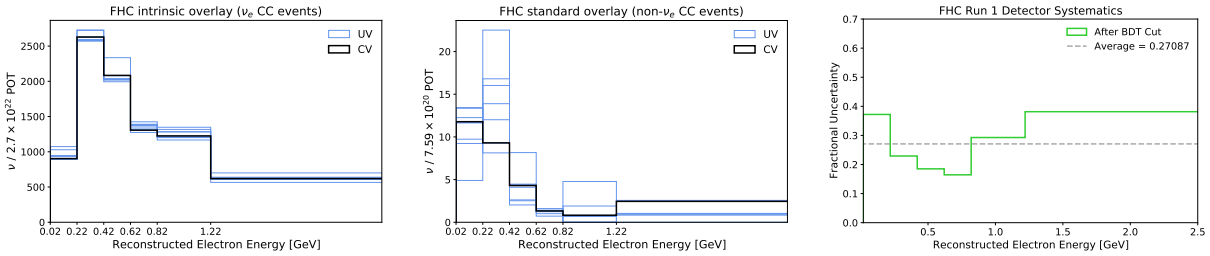


Figure 6.37: FHC Run 1 selected event rates of the detector variation intrinsic overlay samples (left), which is comprised of ν_e CC interactions, and the standard overlay samples (middle), which consist of ν_μ and ν_e NC background. Low statistics in the latter superficially inflates uncertainty on the background-subtracted selected event rate (right).

The strategy to resolve this begins with removing as much artificial shape influence as possible. The detector variations are investigated on the reconstructed `x` position of the neutrino interaction vertex, which is a more uniform distribution than any of the cross section variables. Figure 6.38 shows the evolution of the fractional uncertainty for the FHC Run 1 intrinsic overlay (left) and standard overlay (right) samples after the preselection in blue, after loose ν_μ CC and π^0 rejection in orange, and after the final BDT constraint in green. The legend reports the number of CV events passing the selection stage used to create each curve, as well as the average uncertainty across all bins. The error bars on each curve illustrate the statistical uncertainty in each bin, estimated by the square root of the event count.

It is observed that both bin-to-bin fluctuations and the average uncertainty across all bins increases as the selection progresses. After loose ν_μ CC and π^0 rejection, the size of the standard overlay samples is already reduced enough such that the uncertainty on ν_μ and ν_e NC backgrounds is not meaningful. Though the event counts are higher in the intrinsic

overlay samples, the statistics are not substantial enough to overcome bin-to-bin fluctuations that drive up the average uncertainty.

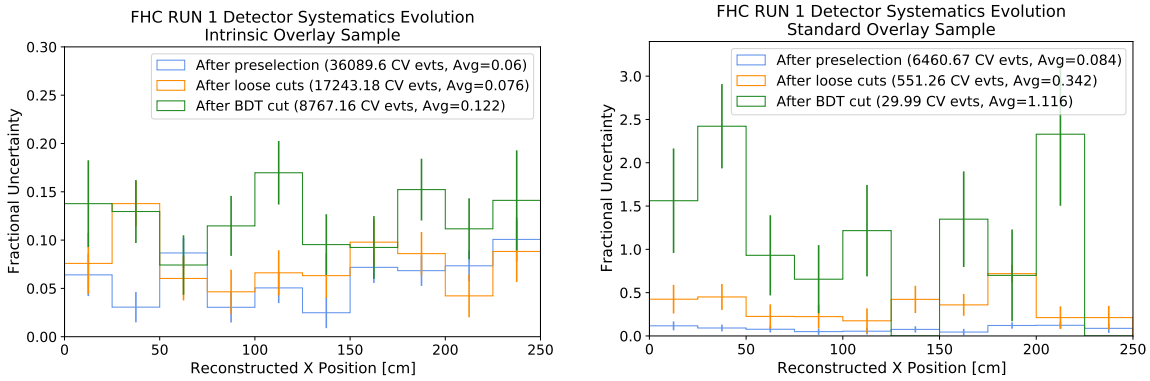


Figure 6.38: Evolution of FHC Run 1 fractional detector uncertainty for the intrinsic overlay (left) and standard overlay (right) samples, shown as a function of reconstructed `x` vertex position after the preselection (blue), loose cuts (orange), and BDT cut (green).

As a result of this study, the standard overlay samples are deemed unusable for this work. Instead, the FHC+RHC background-subtracted event rates, constructed from the intrinsic overlay samples, are analyzed as a function of the reconstructed `x` vertex position. Figure 6.39 displays the FHC+RHC background-subtracted selected event rate variations for the intrinsic overlay samples of each uncertain detector parameter as a function of the reconstructed `x` vertex position. A covariance is computed for each unisim, and the combined sum of these matrices is shown in Figure 6.40 along with the corresponding correlation matrix. The fractional uncertainty derived from the covariance is presented in Figure 6.41 (left). Fractional uncertainties for FHC Run 1 (middle) and RHC Run 3 (right) are calculated in an analogous manner.

As anticipated, statistical limitations are still present in the final uncertainty values, even when analyzing the combined (FHC+RHC) intrinsic overlay samples as a function of a uniform variable. This is because, while the POT of the intrinsic overlay is two orders of magnitude larger than that of standard overlay, these samples are rich in ν_e CC interactions, which contribute to only a few percent of the total NuMI event rate.

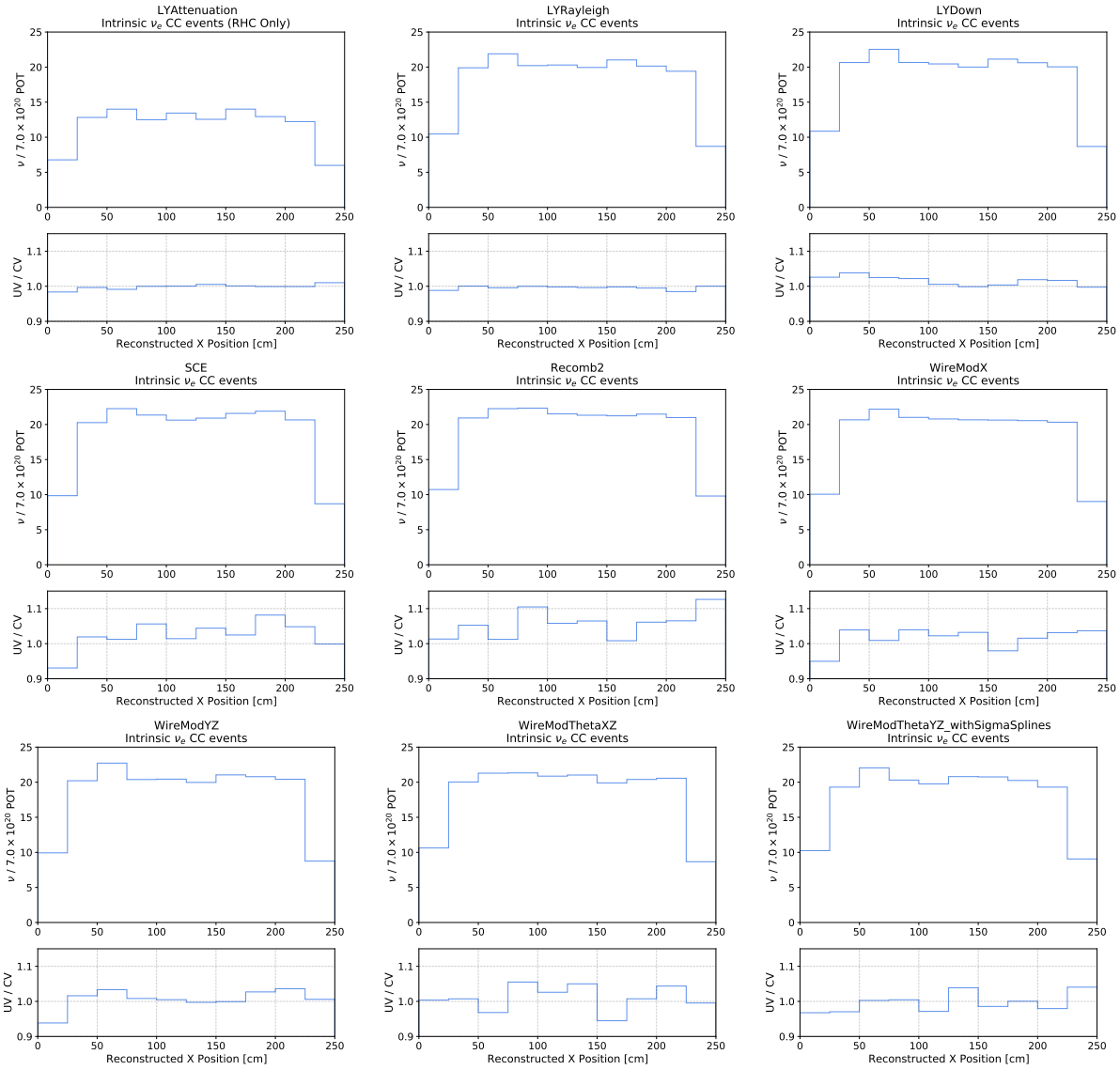


Figure 6.39: Detector unisim variations on the background-subtracted selected event rate as a function of reconstructed `x` position. Note that the LYAttenuation variation is only relevant for the RHC event rate.

In lieu of bin-to-bin covariance and correlations, then, a flat detector uncertainty is derived from the FHC+RHC background-subtracted event rates as a function of the reconstructed `x` vertex position, and incorporated into the uncertainty budget of the final cross section results. An analogous solution is implemented for the individual FHC Run 1 and RHC Run 3 background-subtracted event rates. These flat estimates are also included in

the error bands of late stage selection plots (all figures in Section 5.7).

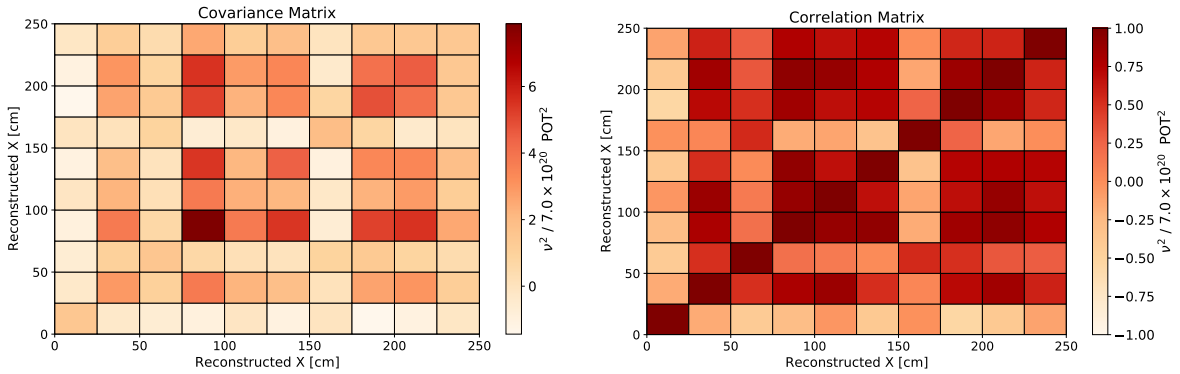


Figure 6.40: Covariance (left) and correlation (right) matrices as a function of the reconstructed `x` position of the neutrino interaction vertex.

From Figure 6.41, the detector uncertainty on the FHC+RHC background-subtracted selected event rate is estimated as 10.0%. The corresponding FHC Run 1 detector uncertainty is 12.2%, and 12.9% for RHC Run 3.

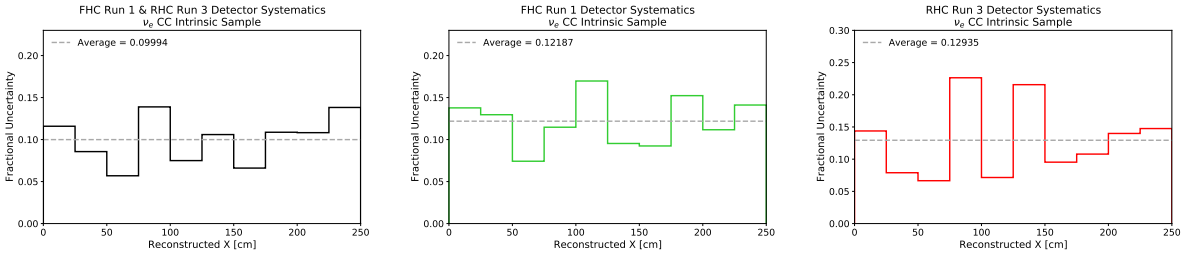


Figure 6.41: Fractional uncertainties for FHC+RHC (left), FHC Run 1 (middle), and RHC Run 3 (right) derived from the detector unisim variations for the intrinsic overlay sample. Shown as a function of the reconstructed `x` position of the neutrino interaction vertex.

6.5 POT counting

The number of protons delivered to the NuMI target is measured using two toroids encircling the beam at different locations [27]. At the guidance of [32], a conservative 2% flat uncertainty in POT counting for FHC Run 1, RHC Run 3, and FHC+RHC event rates is applied across all bins. This uncertainty is added to the total covariance matrix.

6.6 Dirt interaction models

As discussed in Section 4.3.1, the simulation of out-of-cryostat, or dirt, interactions is not well constrained. To account for this, a 100% uncertainty on dirt events is assumed. In the RHC Run 3 sample, no dirt interactions survive the selection, so contributions to the FHC+RHC event rate come from the FHC Run 1 sample only. Figure 6.42 shows the single FHC+RHC dirt variation as a function of the cross section variables in comparison with the CV dirt event rate that survives the selection.

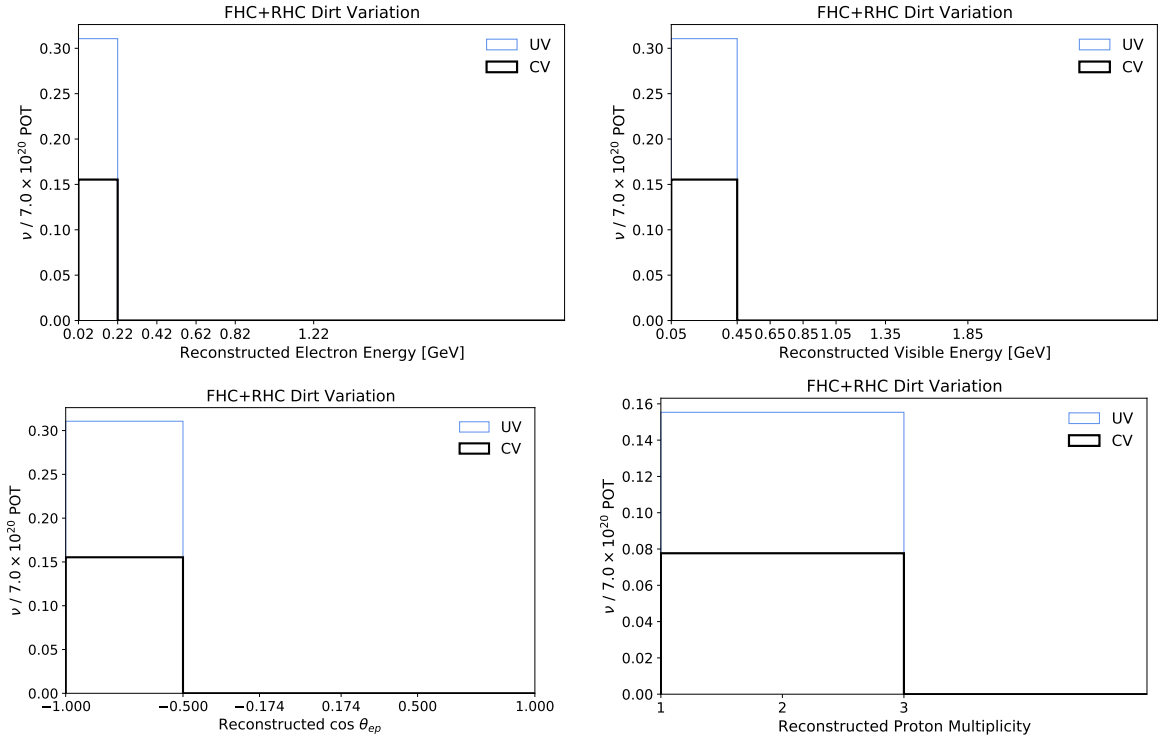


Figure 6.42: Assuming a 100% uncertainty on dirt interactions, a single variation is produced, shown in comparison with the CV dirt event rate that survives the selection.

Fractional uncertainties derived from the single variation are shown in Figure 6.43, comparing curves derived from FHC Run 1 (green) and FHC+RHC (black) selected dirt event rates. The total number of dirt interactions in these samples are the same, but because the FHC+RHC distribution has a higher event count in each bin, the fractional uncertainty is reduced compared to that of FHC Run 1 only.

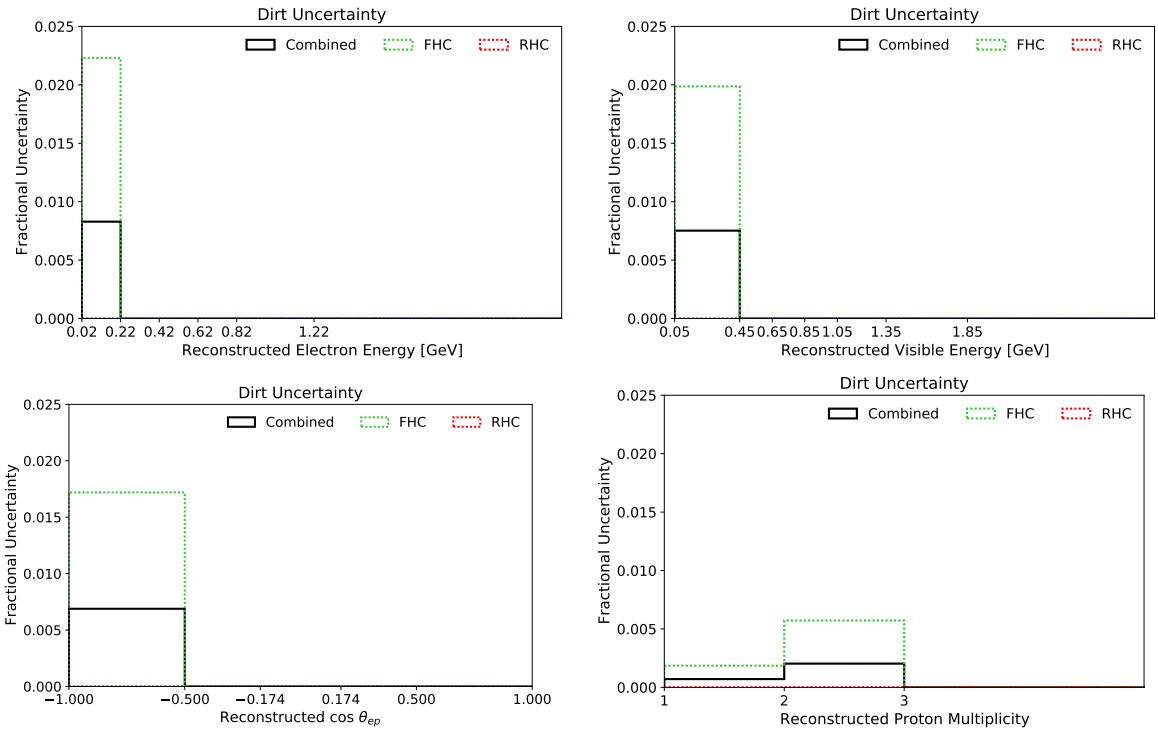


Figure 6.43: Fractional dirt uncertainties on the selected event rate. Shown as a function of reconstructed electron energy, visible energy, opening angle, and proton multiplicity for FHC Run 1 (green) and FHC+RHC (black).

6.7 Statistical uncertainty

The finite number of events in beam-on data and the MC+EXT prediction contributes to statistical uncertainty in both the background-subtracted event rates and in the construction of the response matrix, which encodes the smearing of reconstructed observables away from their true values as well as the efficiency of the selection algorithm. This section describes the quantification of statistical uncertainty for both of these sources.

Of the background-subtracted event count

The background-subtracted event rate for beam-on data is given by the following:

$$N_{beam\,on} - N_{EXT} - B_{MC} \tag{6.11}$$

where $N_{beam\ on}$ and N_{EXT} are the total event rates in beam-on data and the EXT sample, respectively, and B_{MC} is the neutrino background contribution estimated by `GENIE`-generated overlay samples.

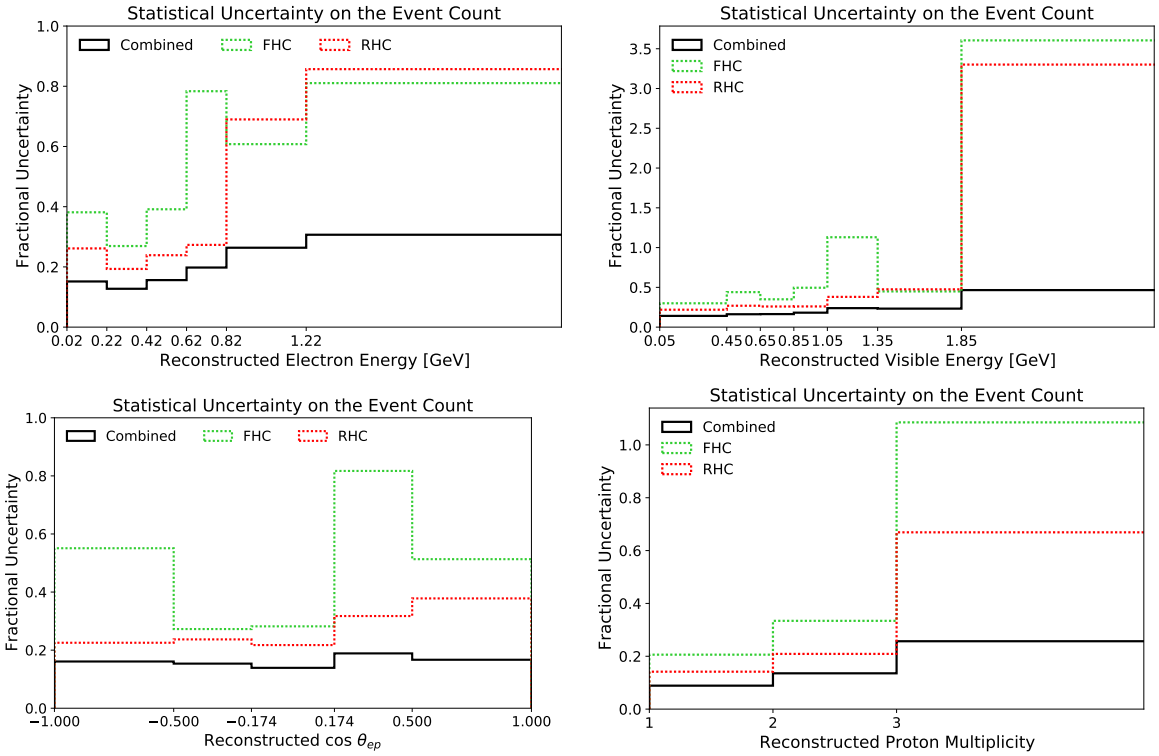


Figure 6.44: Fractional uncertainty on the beam-on background-subtracted event count resulting from finite statistics in data and simulation. Shown as a function of the cross section variables for events that survive the selection in FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) samples.

The uncertainty on Equation 6.11 is derived using the standard error propagation formula:

$$\sqrt{(\Delta N_{beam\ on})^2 + (\Delta N_{EXT})^2 + (\Delta B_{MC})^2} = \sqrt{N_{beam\ on} + \Sigma w_{EXT}^2 + \Sigma w_{MC}^2} \qquad (6.12)$$

where Poissonian statistics sets $\Delta N_{beam\ on} = \sqrt{N_{beam\ on}}$, and ΔN_{EXT} and ΔB_{MC} as the square root of the sum of the squared weights for the EXT and MC background events, respectively. The square of Equation 6.12 becomes the diagonals of a covariance matrix that accounts for statistical uncertainty associated with the beam-on dataset and the estimation of background.

Fractional uncertainties on the beam-on background-subtracted event rates are shown in Figure 6.44 as a function of the cross section variables. These curves quantify the effect of statistical limitations in the FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) samples of events that survive the selection detailed in Chapter 5. As expected, combining the FHC and RHC event samples significantly suppresses the statistical uncertainty in each bin.

Of the response matrix

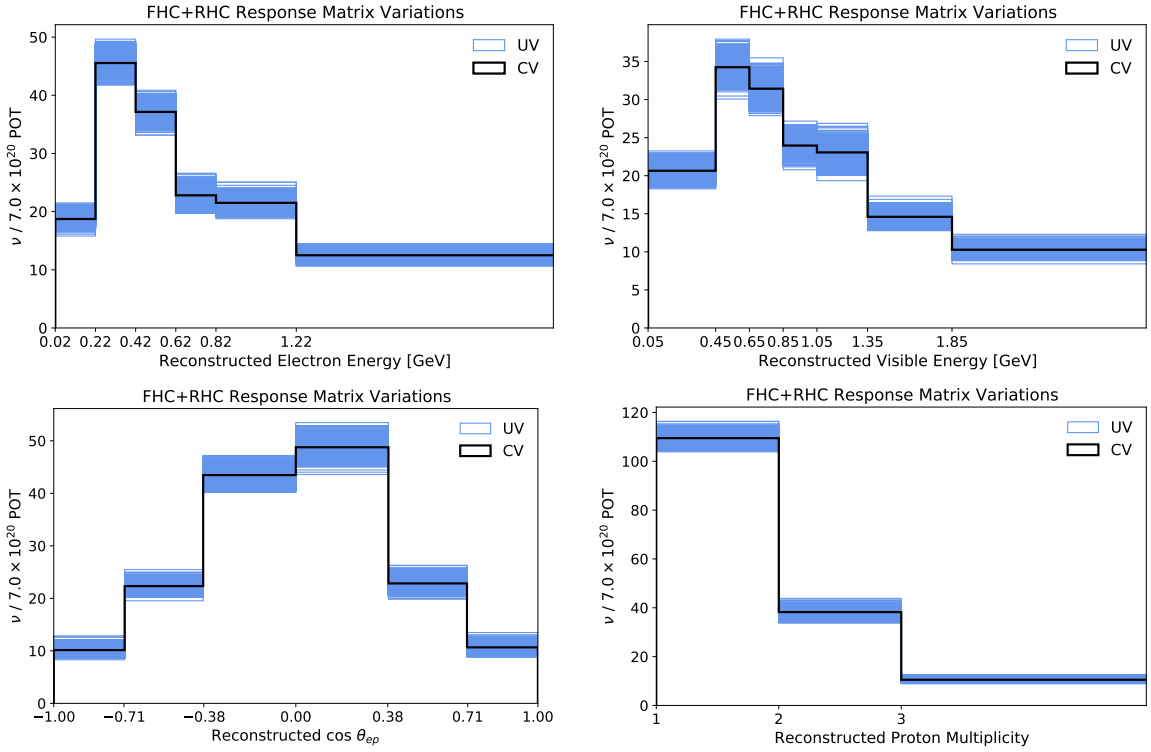


Figure 6.45: Variations on the selected signal prediction, derived from a bootstrapping technique, to assess statistical uncertainty associated with the estimation of the response matrix. Shown in comparison with the CV event rate (black).

The statistical uncertainty associated with the estimation of the response matrix is evaluated with bootstrapping. Bootstrapping is a random re-sampling procedure that is used to create variations of a single dataset. As it pertains to this work, 1000 values for each event in the

selected signal prediction are assigned by randomly sampling a Poisson distribution with a mean of 1.0. Each event uses its own random seed to ensure that the sampled set of values is unique to the interaction. These effectively serve as reweights that can be applied to each event in the distribution to produce the variations shown in Figure 6.45.

Fractional uncertainties derived from these variations are displayed in Figure 6.46, comparing curves from the FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) samples. As anticipated, combining the FHC and RHC event rates leads to a decreased statistical uncertainty, which ranges from 2-7% across all bins.

Note that the response matrix uncertainty does not contribute to the error band on the event rate prediction of plots shown in Chapter 5. This source of uncertainty is taken into account only for the final cross section results.

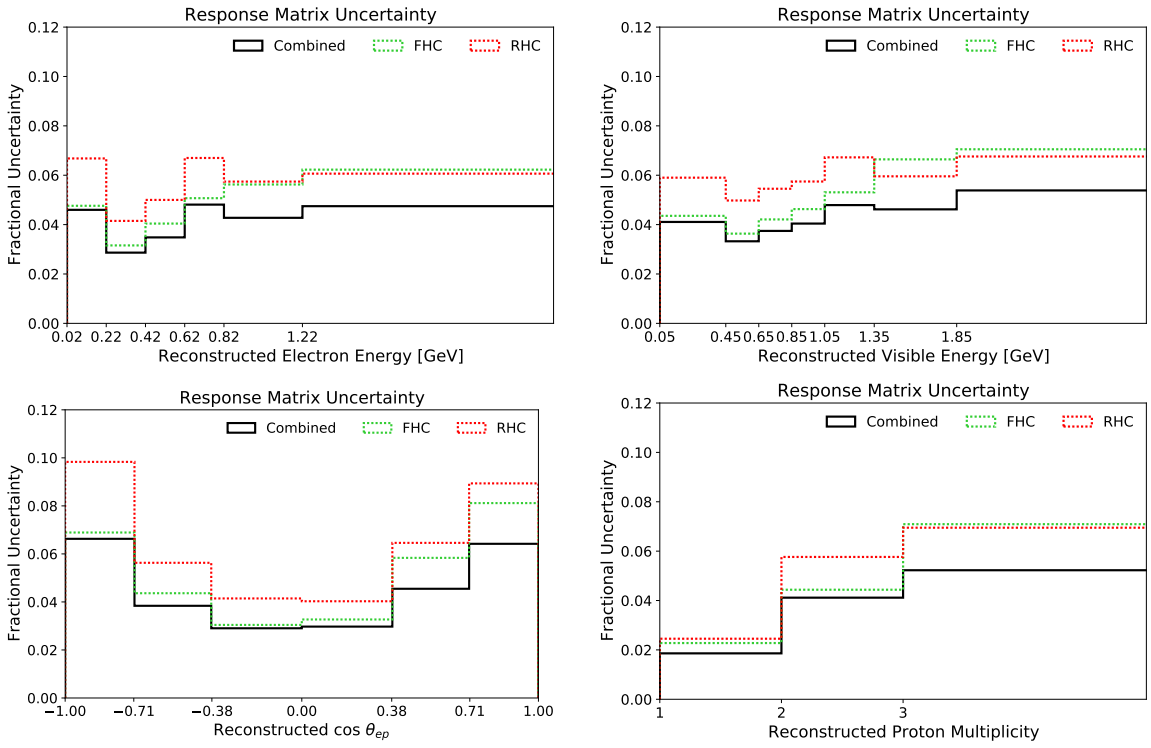


Figure 6.46: Fractional uncertainty on the selected signal prediction resulting from finite statistics used to construct the response matrix. Shown as a function of the cross section variables for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black).

6.8 Total uncertainty

From each source of uncertainty, the covariance matrices are added together to produce the total covariance for the FHC+RHC background-subtracted selected event rates. Figures 6.47, 6.48, 6.49, and 6.50 show the total covariance along with the fractional uncertainty breakdown for reconstructed electron energy, visible energy, opening angle, and proton multiplicity, respectively. For the uncertainty breakdown, the flux curve (blue) includes both `PPFX` and beamline geometry sources, the `GENIE` curve (orange) includes both multisim and unisim variations, and the statistical curve (pink) includes uncertainty from the beam-on and MC+EXT event counts as well as the estimation of the response matrix. Each source of fractional uncertainty is added in quadrature to obtain the total curve (black). At low energies, the largest source of uncertainty is from the flux, specifically the modeling of hadronic behavior along the beamline. At higher energies, however, statistical uncertainty dominates due to the limited number of events in the beam-on data samples. These contributions are comparable across the bins of non-energetic variables.

Figure 6.51 shows the total FHC+RHC uncertainty (black) in comparison with curves derived for FHC Run 1 (green) and RHC Run 3 (red) as a function of the cross section variables. Combining the samples does in fact decrease the total uncertainty in each bin, primarily due to increased statistics.

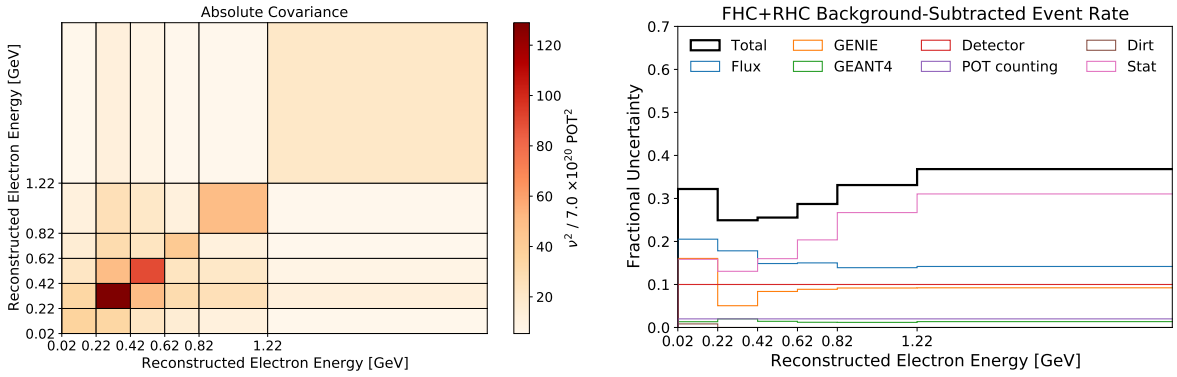


Figure 6.47: Total covariance and breakdown of fractional uncertainties for the FHC+RHC background-subtracted event rate as a function of reconstructed electron energy.

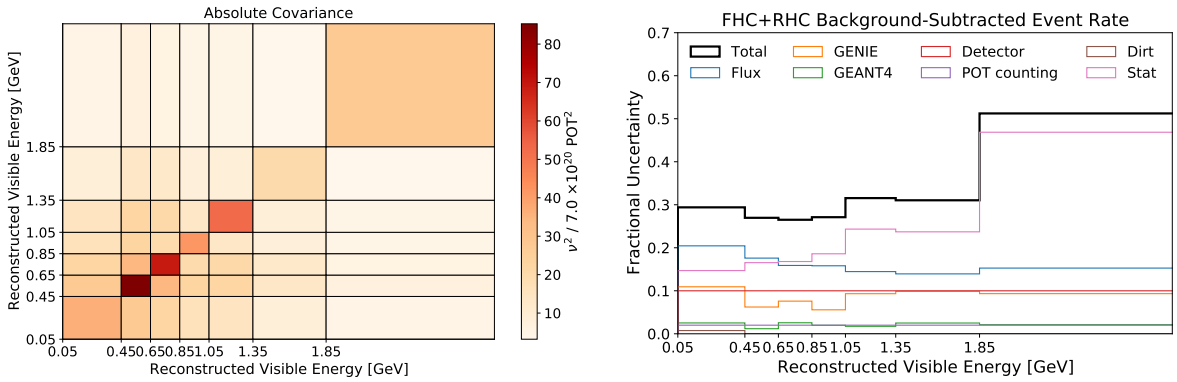


Figure 6.48: Total covariance and breakdown of fractional uncertainties for the FHC+RHC background-subtracted event rate as a function of reconstructed visible energy.

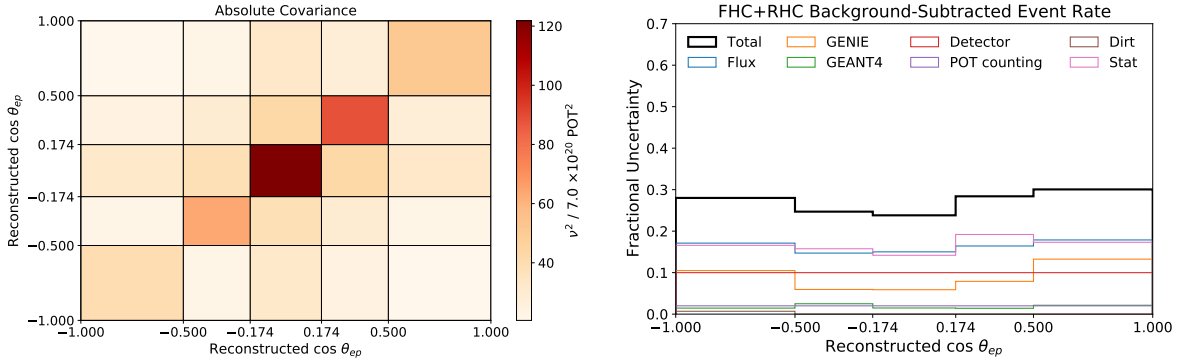


Figure 6.49: Total covariance and breakdown of fractional uncertainties for the FHC+RHC background-subtracted event rate as a function of reconstructed opening angle.

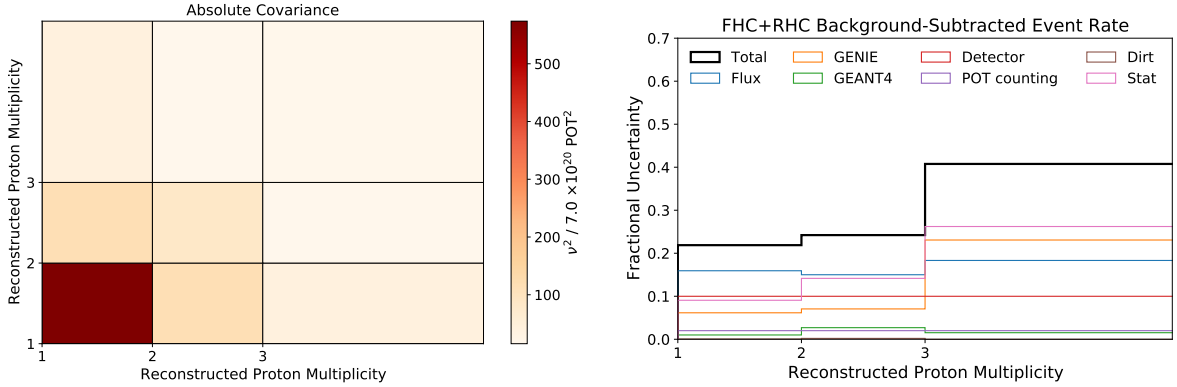


Figure 6.50: Total covariance and breakdown of fractional uncertainties for the FHC+RHC background-subtracted event rate as a function of reconstructed proton multiplicity.

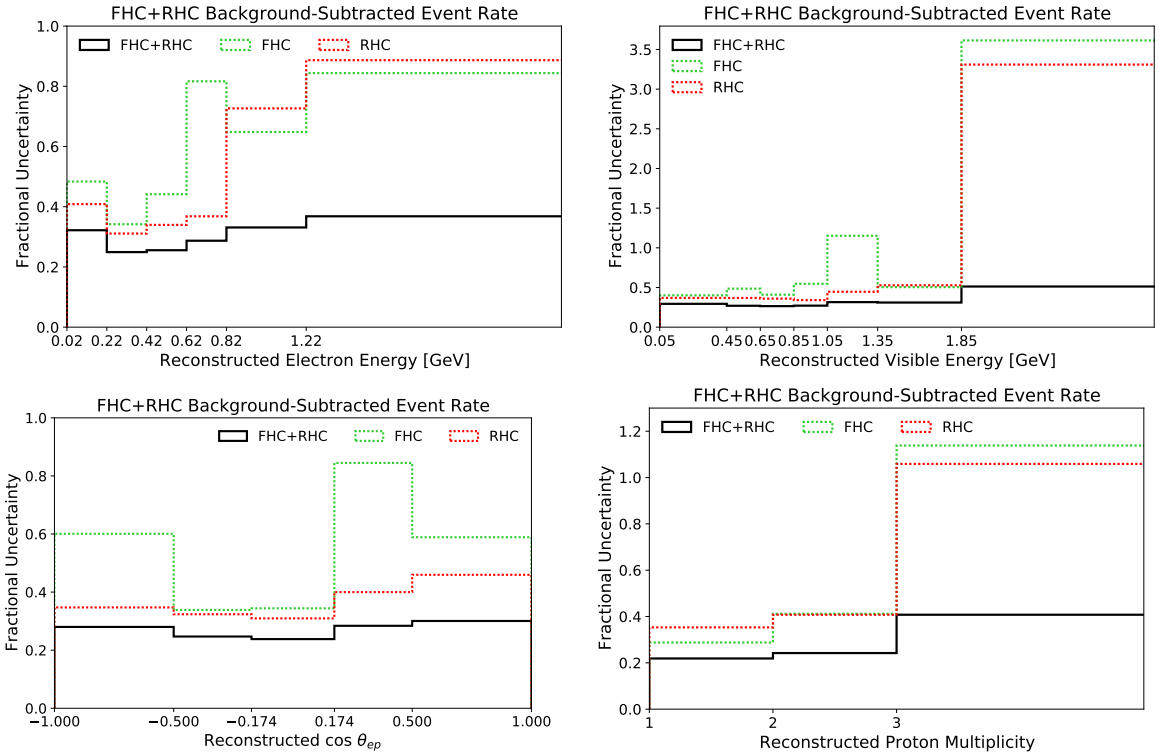


Figure 6.51: Total fractional uncertainty on the background-subtracted selected event rates as a function of the cross section variables. Shown for FHC Run 1 (green), RHC Run 3 (red), and FHC+RHC (black) distributions.

CHAPTER 7

DIFFERENTIAL CROSS SECTION EXTRACTION

The flux-integrated cross section in a bin i is given by:

$$\left(\frac{d\sigma}{dx}\right)_i = \frac{N_i - B_i}{\epsilon_i \times N_{target} \times \Phi_{\nu_e} \times \Delta x_i} \tag{7.1}$$

where x is the differentiating variable, N_i is the selected total event rate, B_i is the estimated selected background rate, ϵ_i is the efficiency of the selection, N_{target} is the number of collision targets in the active volume, Φ_{ν_e} is the integrated ν_e flux prediction, and Δx_i is the bin width. ($N_i - B_i$ is the background-subtracted event rate for which the uncertainties are computed for in Chapter 6.) With the exception of N_i, which comes from the beam-on dataset, all parameters in Equation 7.1 are estimated from simulations of the flux and event rate.

Horn Current	**Integrated ν_e flux [ν/cm^2/POT]**	**Beam-on POT**
FHC Run 1	$1.1864531 \times 10^{-11}$	2.000×10^{20}
RHC Run 3	$8.6283762 \times 10^{-12}$	5.014×10^{20}

Table 7.1: Integrated flux values and POT of FHC Run 1 and RHC Run 3.

Φ_{ν_e} is estimated using a POT-weighted sum of the integrated FHC and RHC flux predictions shown in Figure 2.4. The flux is computed over the full angular phase space; however, only true neutrino energies > 60 MeV are considered so as to exclude muon decay-at-rest interactions that the selection is not sensitive to. The lowest selected true neutrino energy is estimated to be 198 MeV, far above this threshold.

Integrated flux values and POT contributions from FHC Run 1 and RHC Run 3 are shown in Table 7.1. The FHC+RHC central value flux is computed as:

$$\Phi_{\nu_e} = (1.1864531 \times 10^{-11} * 2.000 \times 10^{20}) + (8.6283762 \times 10^{-12} * 5.014 \times 10^{20}) = 6699174027$$

with units of $[\nu/\mathrm{cm}^2]$. Uncertainty on Φ_{ν_e} is accounted for through the effect the flux has on the simulated event rate, i.e. in the second term on the right side of Equation 6.4.

N_{target} represents the number of nucleons within the FV that serve as potential targets for neutrino interactions. This is determined with the following formula:

$$N_{target} = \frac{\rho_{Ar} \times V \times N_A \times N_{nucleons}}{m_{mol}} = 4.241 \times 10^{31} \tag{7.2}$$

where ρ_{Ar} is the density of liquid argon, V is the fiducial volume, N_A is Avogadro's number, $N_{nucleons}$ is the number of nucleons per argon atom, and m_{mol} is the number of grams per mole for argon. Table 7.2 shows the values used for deriving N_{target}. The uncertainty on the number of targets is considered negligible.

Parameter	**Value**
ρ_{Ar}	1.3836 g/cm^3
V	5.0832512×10^7 cm^3
N_A	6.022×10^{23} atoms/mol
$N_{nucleons}$	40 nucleons/atom
m_{mol}	39.95 g/mol

Table 7.2: Values used to compute the number of targets.

7.1 Wiener-SVD unfolding procedure

The background-subtracted event rate M is measured in reconstructed quantities affected by the selection efficiency, which impacts the event count in each bin, and the detector response, which can smear observables away from their true values. To meaningfully compare and constrain theoretical models with data-driven results, such effects need to be unfolded from the event rate via a response matrix R encoding the strength of the inefficiency and smearing. In theory, if R is invertible, this can be done via direct inversion, i.e.:

$$M = R \times s$$

$$s = R^{-1} \times M$$

where s is the true underlying signal distribution. In practice, however, random fluctuations arising from statistical and systematic uncertainties are inflated by directly inverting R, which can lead to meaningless results. To mitigate this issue, it is common to regularize by introducing a small amount of bias in the final result to suppress the amplified variance.

In this work, differential cross sections are unfolded with the Wiener singular value decomposition (Wiener-SVD) method [79], which takes advantage of deconvolution techniques traditionally used in digital signal processing. The Wiener-SVD method has been employed for several results within the MicroBooNE collaboration [25, 80, 81]. The degree of regularization is determined by minimizing the mean square error between the variance and the introduced bias of the result. An advantage of the Wiener-SVD method is that χ^2 between the data and model prediction is independent of the regularization, thus remains consistent before and after unfolding. This is important to accurately assess the validity of model predictions used in neutrino studies.

As input, the Wiener-SVD method requires the following:

- The **measured event rate** as a function of the reconstructed differentiating variable.
- The **estimated signal event rate** before the selection has been applied as a function of the true differentiating variable.
- An **estimated response matrix** encoding the post-selection smearing between the true and reconstructed differentiating variables. Each truth bin in the matrix is normalized by the efficiency in that bin.
- An **estimated covariance matrix** encoding the total statistical and systematic uncertainty of the measured event rate.

The measured event rate comes from beam-on data that has passed the selection described in Chapter 5 and from which the estimated background has been subtracted, i.e. the $N_i - B_i$ term in the numerator of Equation 7.1. The estimated signal event rate is used as somewhat of a prior for the unfolding algorithm; it and the latter two inputs are estimated using the `GENIE v3.0.6 G18_10a_02_11a` event generator. The method returns an efficiency-corrected, unfolded event rate $\hat{s}$ and unfolded covariance matrix, both of which can be converted to cross section units via normalization by flux, bin width, and N_{target}. The goodness of fit between the unfolded result and the simulated cross section is explicitly computed as:

$$\chi^2 = \sum_{ij} (\hat{s}_i - s_i) * Cov_{ij}^{-1} * (\hat{s}_j - s_j) \tag{7.3}$$

where $\hat{s}_i$ and s_i are the unfolded result and the generated prediction in bin i, respectively, and Cov_{ij}^{-1} is the ij^{th} element of the inverse of the unfolded covariance matrix. An additional smearing matrix that encodes the introduced bias, A_c, is also returned by the Wiener-SVD method such that:

$$\hat{s} = (A_c\, R^{-1}) \times M = A_c \times s + A_c \times noise$$

where the second term represents the suppression of terms contributing to high variance. Independent theoretical predictions of the cross section can be smeared by A_c to enable a direct comparison with $\hat{s}$.

The Wiener-SVD method includes a free parameter C, which is the derivative between adjacent bins that the algorithm attempts to smooth. For tests performed to validate the Wiener-SVD implementation, the recommendation of [79] is followed, with $C = 2$ to smooth the curvature of the distribution. However, for unfolded results with the NuMI dataset, C is instead set to 1, i.e. the slope of the distribution is smoothed. Note that because χ^2 is independent of the regularization, this choice does not bias the final result. More information

about the impact of this choice can be found in Appendix A.

Wiener-SVD unfolding requires a number of criteria to hold, including that the event rates are smooth and continuous. This is obviously not the case for proton multiplicity—it is a discrete distribution for which the Wiener-SVD approach does not converge. In lieu of an unfolded result, then, interaction rates and total covariance as a function of proton multiplicity are also reported. The response matrix is also included, which may prove useful for future analyzers to employ alternate unfolding algorithms.

7.2 Closure tests

Instead of beam-on data, closure tests use a `GENIE`-simulated distribution for the measured event rate input into the Wiener-SVD algorithm. In this scenario, the background-subtracted event rate is equivalent to the selected signal prediction. Because this comes from the same simulation as the estimated signal event rate, the expectation is that the unfolded spectra should perfectly close onto this prediction. This provides a way of verifying that the unfolding is performing as anticipated.

This section reports results of FHC+RHC closure tests for electron energy, visible energy, and opening angle. Note that, because the measured event rate is replaced with simulation, the statistical contribution to the covariance is given as follows:

$$(\Delta N_{MC})^2 + (\Delta N_{EXT})^2 + = \Sigma w_{MC}^2 + \Sigma w_{EXT}^2 \tag{7.4}$$

where w_{EXT}^2 are the squared weights of EXT events, and, in contrast to Equation 6.12, w_{MC}^2 are the squared weights of both the MC signal and background channel.

Electron energy

Figure 7.1 shows the input distributions for the Wiener-SVD closure test on electron energy. The selected signal distribution, estimated with `GENIE`, is shown in the left plot in blue as a

function of the reconstructed electron energy. This serves as the measured event rate input into the Wiener-SVD algorithm. The estimated signal event rate is the generated GENIE signal prediction before the selection has been applied, shown in the left plot in orange as a function of the true electron energy.

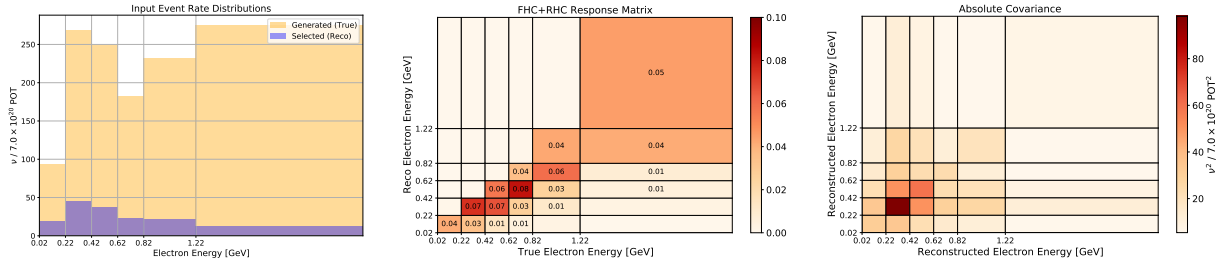


Figure 7.1: The measured event rate and estimated signal event rate (left) which, in the case of a GENIE closure test, are equivalent to the selected and generated signal prediction. The response matrix (middle) and covariance of the measured event rate (right) are also shown, estimated by GENIE as well.

Figure 7.2 displays the output of the closure test on electron energy: the unfolded result compared with the input event distributions, additional smearing matrix A_c (middle), and unfolded covariance (right). To properly assess the goodness of fit of the unfolded spectrum with the estimated signal event rate, the latter distribution is smeared by A_c.

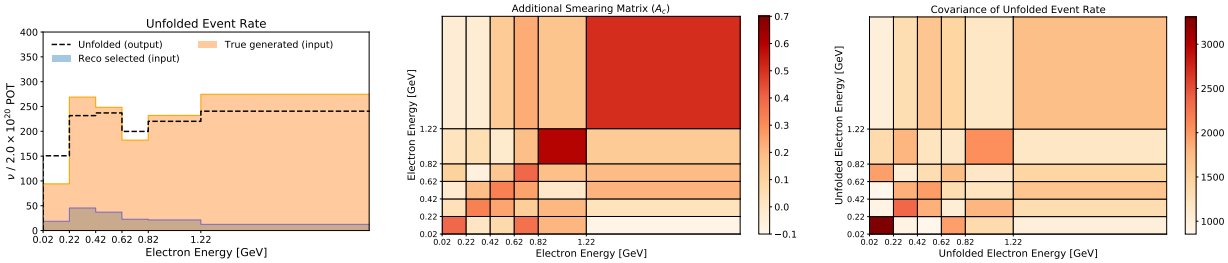


Figure 7.2: Results of the GENIE closure test on electron energy. The input event rate distributions (orange and blue) and the unfolded result are shown on the left. The additional smearing matrix A_c is in the middle, the unfolded covariance is given on the right.

Figure 7.3 compares the unfolded result (black) to the estimated signal event rate, now smeared by A_c (green). The distributions have been converted into cross section units by performing the flux, bin width, and N_{target} normalization given in Equation 7.1. As

anticipated, the unfolded distribution closes perfectly onto the generated signal prediction with $\chi^2 = 0$.

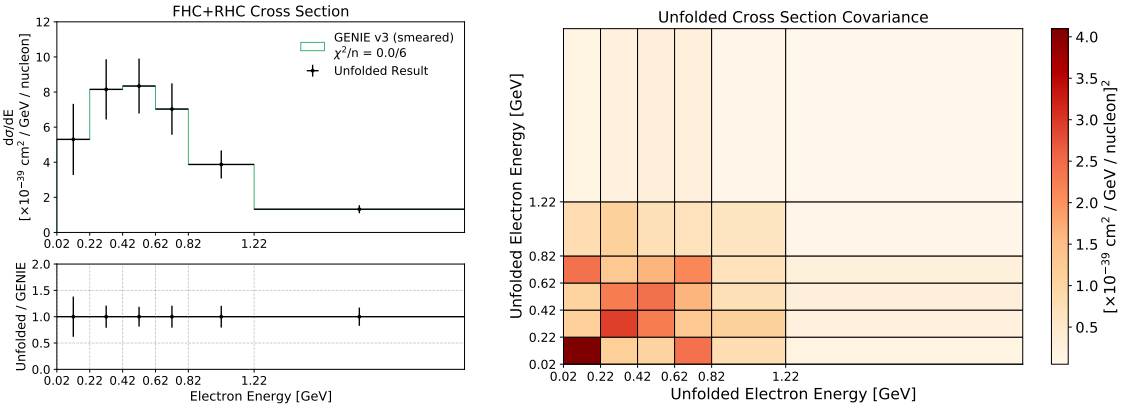


Figure 7.3: Results of the `GENIE` closure test on electron energy, now converted into cross section units. The estimated signal event rate has been smeared by A_c (green), compared to the unfolded distribution (black). The unfolded covariance, also converted into cross section units, is shown on the right.

Visible energy

Figure 7.4 shows the input distributions for a Wiener-SVD closure test as a function of visible energy.

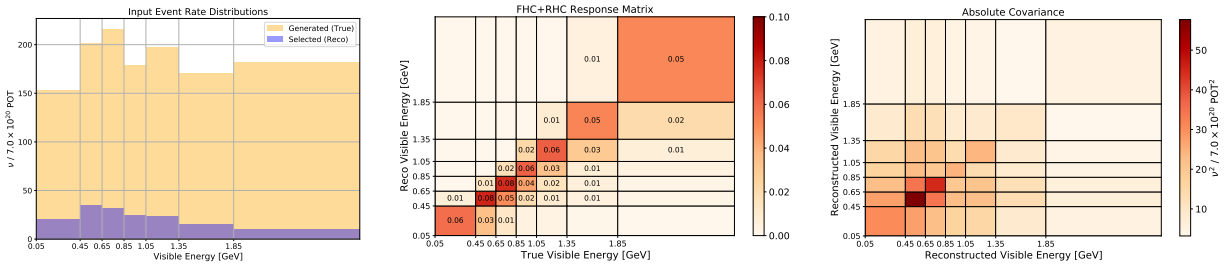


Figure 7.4: The measured event rate and estimated signal event rate (left) which, in the case of a `GENIE` closure test, are equivalent to the selected and generated signal prediction. The response matrix (middle) and covariance of the measured event rate (right) are also shown, estimated by `GENIE` as well.

Figure 7.5 displays the output of the closure test on visible energy: the unfolded result compared with the input event distributions (left), additional smearing matrix A_c (middle),

and the unfolded covariance (right).

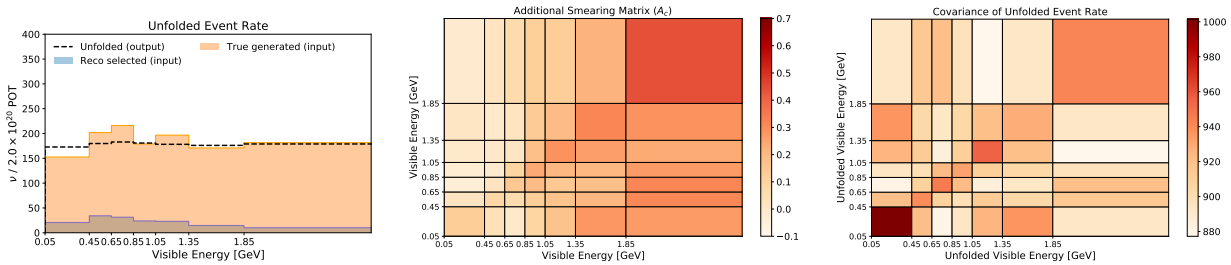


Figure 7.5: Results of the GENIE closure test on visible energy. The input event rate signal distributions (orange and blue) compared with the unfolded result (black) are on the left. The additional smearing matrix A_c is shown in the middle, and the unfolded covariance is given on the right.

Figure 7.6 compares the unfolded spectrum (black points) to the estimated signal event rate, now smeared by A_c (green). Both distributions have been converted into cross section units. As anticipated, the unfolded distribution closes perfectly onto the generated signal prediction with $\chi^2 = 0$.

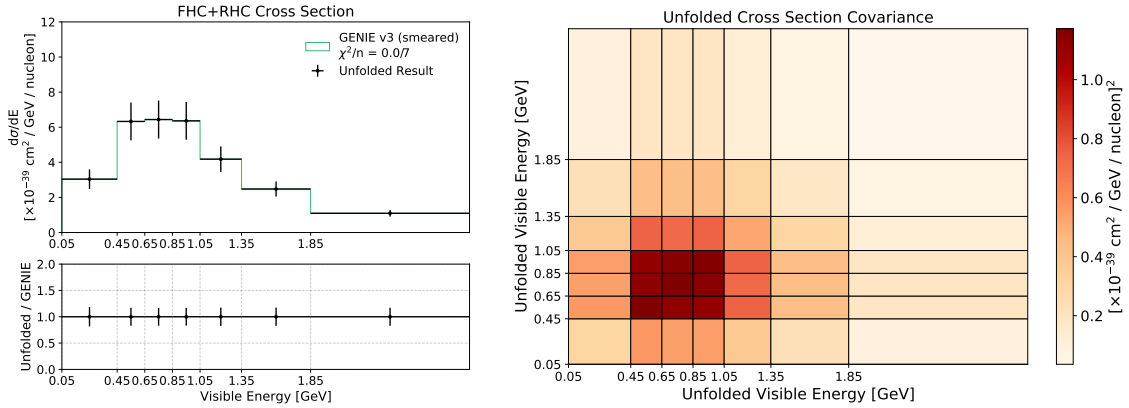


Figure 7.6: Results of the GENIE closure test as a function of visible energy, now converted into cross section units. The true prediction (smeared by A_c) is shown in green, with the unfolded distribution in black. The unfolded covariance, also converted into cross section units, is shown on the right.

Opening angle

Figure 7.7 shows the input distributions for Wiener-SVD closure test as a function of opening angle.

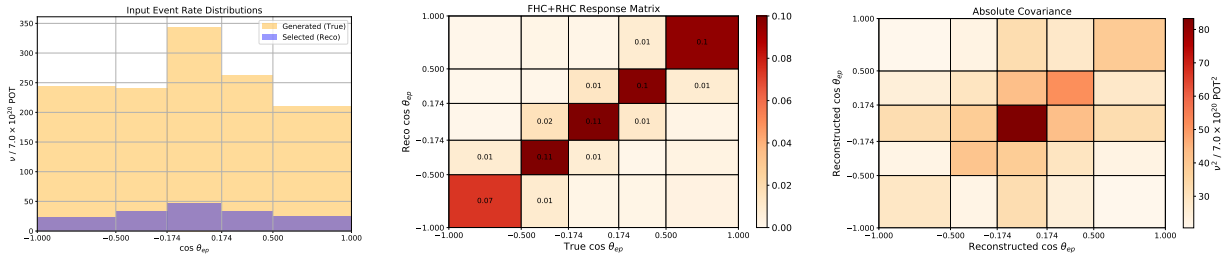


Figure 7.7: The measured event rate and estimated signal event rate (left) which, in the case of a GENIE closure test, are equivalent to the selected and generated signal prediction. The response matrix (middle) and covariance of the measured event rate (right) are also shown, estimated by GENIE as well.

Figure 7.8 displays the output of the closure test on opening angle: the unfolded result compared with the input event distributions (left), additional smearing matrix A_c (middle), and unfolded covariance (right).

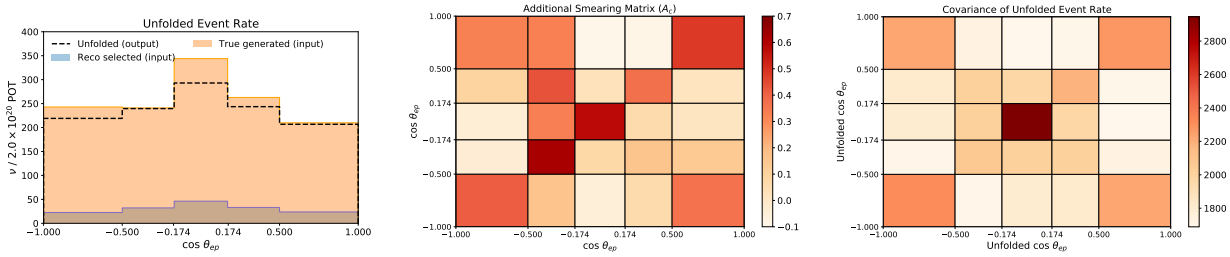


Figure 7.8: Results of the GENIE closure test as a function of opening angle, including the input event rate signal distributions compared with the unfolded, efficiency-corrected result (left), the additional smearing matrix (middle), and the unfolded covariance (right).

Figure 7.9 compares the unfolded result (black) to the estimated signal event rate, now smeared by A_c (green). Both distributions have been converted into cross section units. As anticipated, the unfolded distribution closes perfectly onto the generated signal prediction with $\chi^2 = 0$.

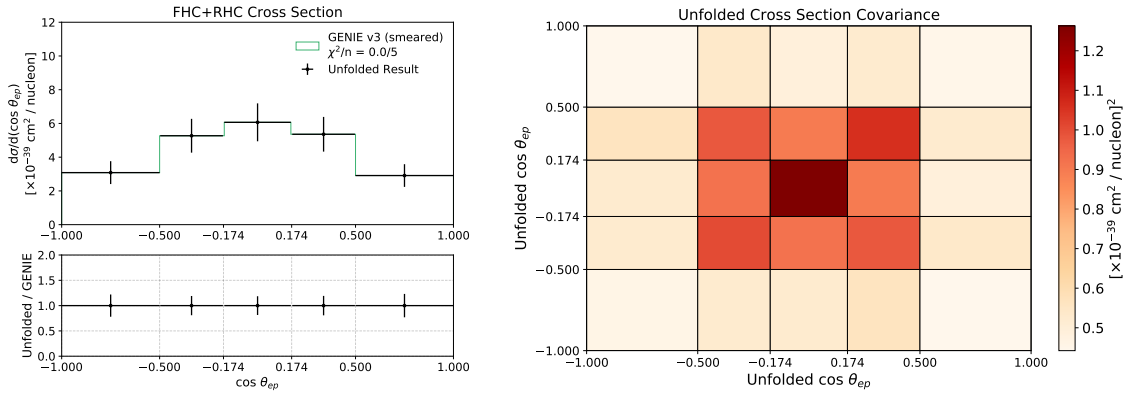


Figure 7.9: Results of the GENIE closure test on opening angle, now converted into cross section units. The estimated signal event rate has been smeared by A_c (green), compared to the unfolded distribution (black). The unfolded covariance, also converted into cross section units, is shown on the right.

7.3 Fake data studies

Another test that can be performed to validate the unfolding procedure is with the use of fake data. The measured event rate is estimated with an event distribution simulated by an alternate event generator, while the backgrounds, systematic covariance, response matrix, and signal event rate are still estimated using GENIE.

The expectation is that the unfolded result should, within uncertainty, close onto a prediction of the signal event rate by the alternate generator—despite feeding the unfolding algorithm a GENIE-estimated prediction with GENIE-estimated uncertainties. A successful fake data test lends confidence that the Wiener-SVD method does not produce unfolded results skewed toward the primary event generator. More specifically, this states that the final results using NuMI data will be a better representation of nature's true cross section, rather than what is predicted by GENIE.

A modest FHC Run 1 event sample (6.64×10^{20} POT of standard overlay and 1.68×10^{22} POT of intrinsic overlay) generated using the NuWro neutrino generator [82, 83] was available at the time of this work for use as fake data. This section reports results of FHC Run 1 fake

data tests for electron energy, visible energy, and opening angle. Because the measured event rate is replaced with simulation, the statistical contribution to the covariance is as follows:

$$(\Delta N_{NuWro})^2 + (\Delta N_{EXT})^2 + (\Delta B_{MC})^2 + = \Sigma w^2_{NuWro} + \Sigma w^2_{EXT} + \Sigma w^2_{MC} \tag{7.5}$$

where w^2_{NuWro}, w^2_{EXT}, and w^2_{MC} are the squared weights for events in the `NuWro`, EXT, and `GENIE` MC background samples, respectively.

`NuWro` overlay samples do not contain the required neutrino ancestry information to generate `PPFX` central value reweights on an event-by-event basis. In lieu of this, ratios between the `PPFX`-corrected and uncorrected flux predictions are created for each flavor, and binned as a function of true neutrino energy and decay angle, as given in Equation 2.5. Each neutrino in the raw `NuWro` event rate is corrected with the flux ratio value corresponding to the particle's flavor, energy, and decay angle.

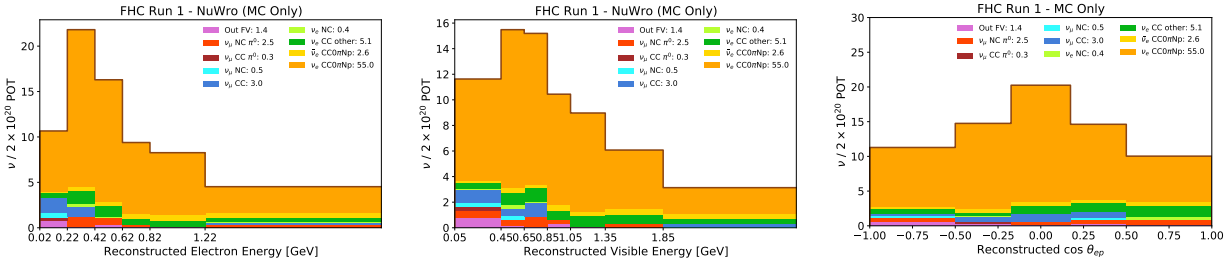


Figure 7.10: The `NuWro` neutrino event rate passing the selection as a function of reconstructed electron energy, visible energy, and opening angle.

Figure 7.10 shows the breakdown of the `NuWro`-estimated event rate that passes the selection as a function of reconstructed electron energy, visible energy, and opening angle. Figures 7.11 and 7.12 compare the estimated signal event rates generated by `GENIE` and `NuWro`, and the selected event rates of `GENIE` and `NuWro` that have been subtracted by `GENIE`-estimated backgrounds.

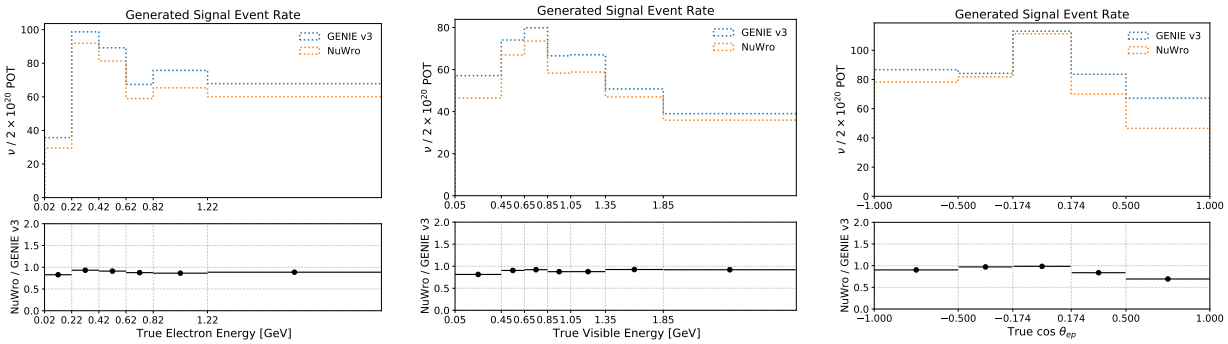


Figure 7.11: Comparisons between the estimated signal event rates generated by `GENIE` (blue) and `NuWro` (orange) as a function of reconstructed electron energy, visible energy, and opening angle.

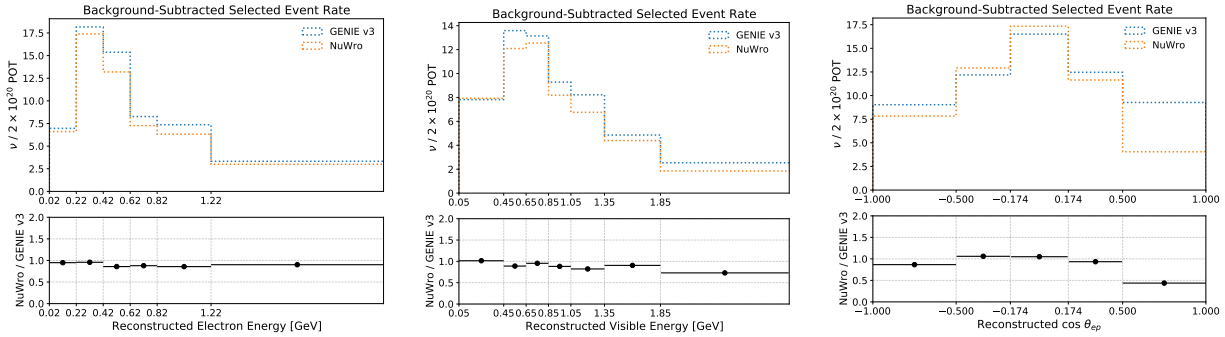


Figure 7.12: Comparisons between `GENIE` (blue) and `NuWro` (orange) selected event rates subtracted by `GENIE`-estimated backgrounds as a function of reconstructed electron energy, visible energy, and opening angle.

Electron energy

Figure 7.13 shows the input distributions for the Wiener-SVD fake data test on electron energy. The selected signal distribution, estimated with `NuWro`, is shown in blue on the left plot as a function of reconstructed electron energy. This serves as the measured event rate input into the unfolding algorithm. The estimated signal event rate is the generated `GENIE` signal prediction before the selection has been applied, shown in the left plot in orange as a function of true electron energy. The covariance matrix (right) is also estimated with `GENIE`. The same response matrix from Figure 7.1 is used as input.

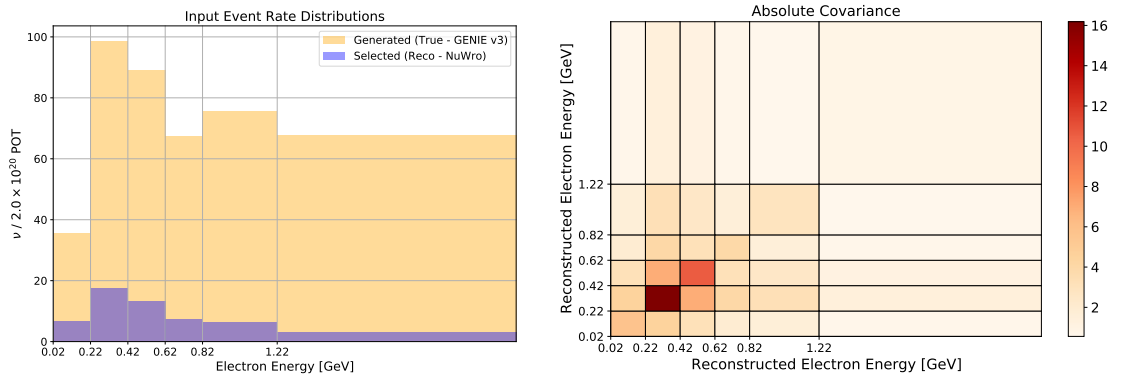


Figure 7.13: The event rates and covariance used as input for the Wiener-SVD fake data study on electron energy. The response matrix from Figure 7.1 is used.

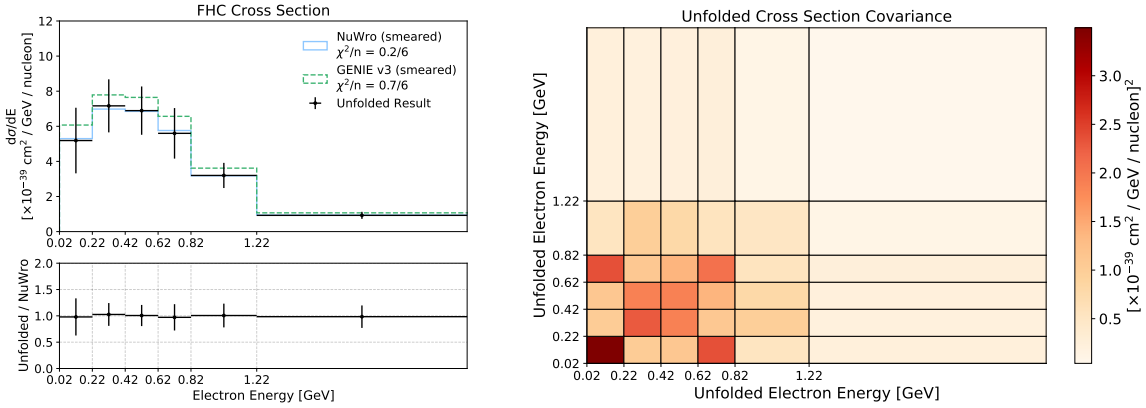


Figure 7.14: Results of the `NuWro` fake data test on electron energy (left), along with the unfolded covariance (right).

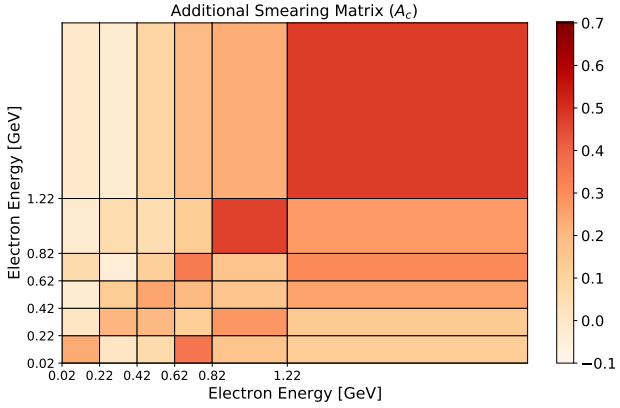


Figure 7.15: The additional smearing matrix A_c produced by the `NuWro` fake data test on electron energy.

Figure 7.14 (left) compares the unfolded fake data cross section to curves estimated by `NuWro` and `GENIE`, shown in blue and green, respectively, and the unfolded covariance (right). The `NuWro` and `GENIE` distributions have been smeared by the output A_c matrix in Figure 7.15. The test performs as expected: The unfolded result agrees with the underlying `NuWro` prediction ($\chi^2/n = 0.2/6$) better than the underlying `GENIE` prediction ($\chi^2/n = 0.7/6$), where n is the number of bins in the distribution.

Visible energy

Figure 7.16 shows the input distributions for the Wiener-SVD fake data test on visible energy. The same response matrix from Figure 7.4 is used as input.

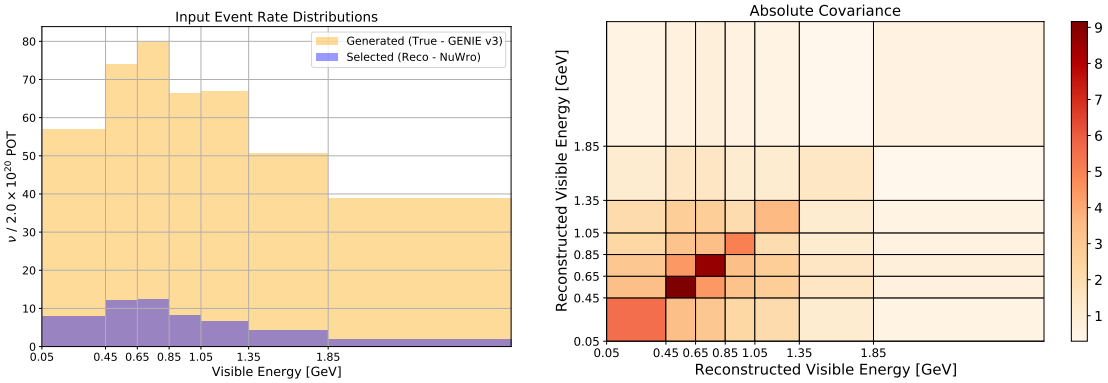


Figure 7.16: The event rates and covariance used as input for the Wiener-SVD fake data study on visible energy. The response matrix from Figure 7.4 is used.

Figure 7.17 (left) compares the unfolded fake data cross section to curves estimated by `NuWro` and `GENIE`, shown in blue and green, respectively, and the unfolded covariance. The `NuWro` and `GENIE` distributions have been smeared by the output A_c matrix in Figure 7.18. The test performs as expected: The unfolded result agrees with the underlying `NuWro` prediction ($\chi^2/n = 1.2/7$) better than the underlying `GENIE` prediction ($\chi^2/n = 1.4/7$).

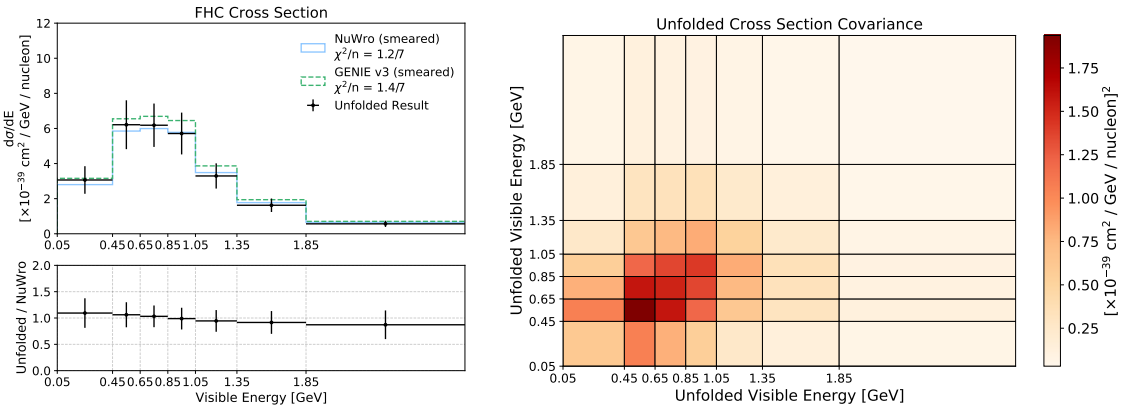


Figure 7.17: Results of the `NuWro` fake data test on visible energy (left), along with the unfolded covariance (right).

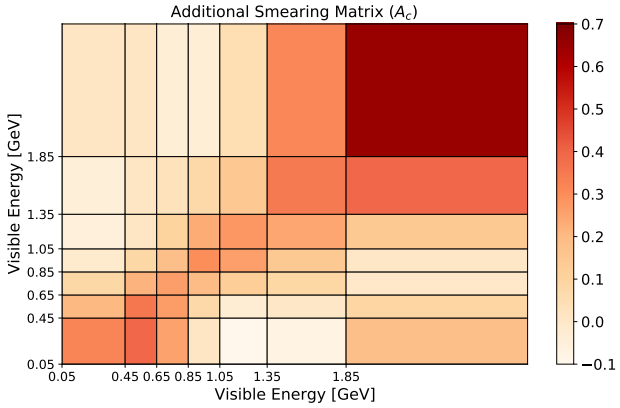


Figure 7.18: The additional smearing matrix A_c produced by the `NuWro` fake data test on visible energy.

Opening angle

Figure 7.19 shows the input distributions for the Wiener-SVD fake data test on opening angle. The same response matrix as in Figure 7.7 is used as input.

Figure 7.20 (left) compares the unfolded fake data cross section to curves estimated by `NuWro` and `GENIE`, in blue and green, respectively, and the unfolded covariance (right). `NuWro` and `GENIE` distributions have been smeared by the output A_c matrix in Figure 7.21. The test performs as expected: The unfolded result agrees with the underlying `NuWro` prediction ($\chi^2/n = 1.5/5$) better than the underlying `GENIE` prediction ($\chi^2/n = 3.4/5$).

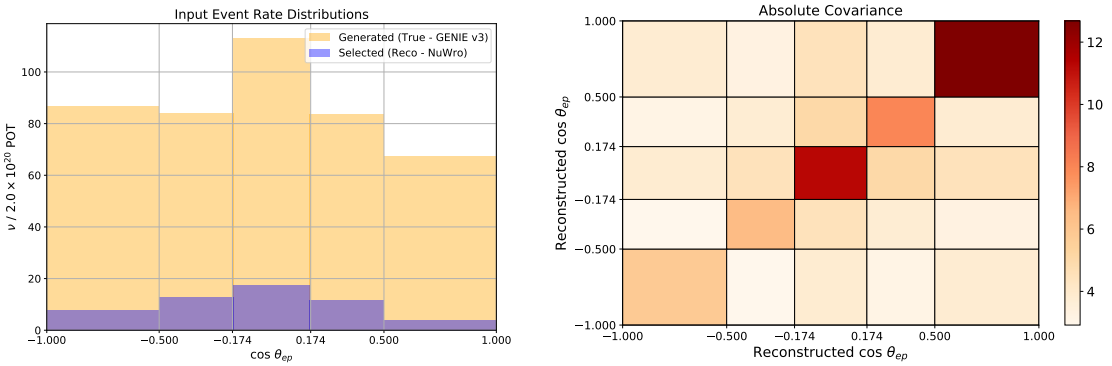


Figure 7.19: The event rates and covariance used as input for the Wiener-SVD fake data study on opening angle. The response matrix from Figure 7.7 is used.

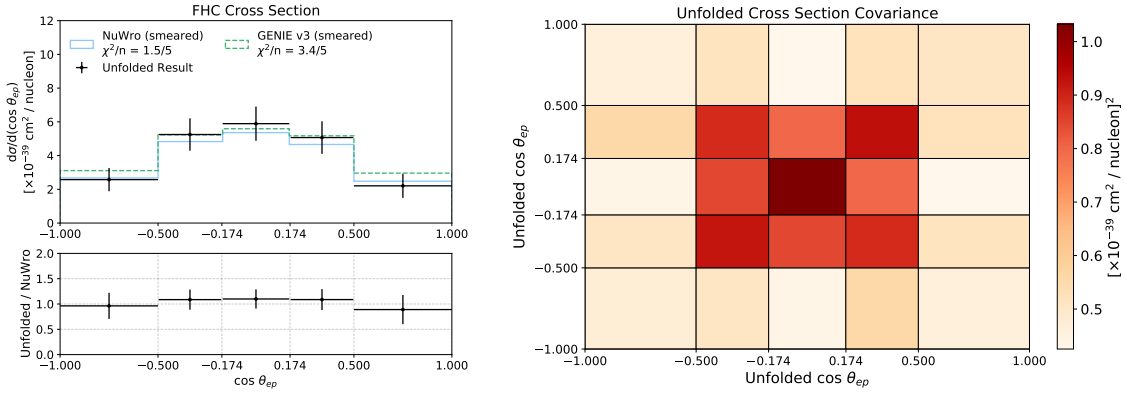


Figure 7.20: Results of the `NuWro` fake data test on opening angle (left), along with the additional smearing matrix (middle), and unfolded covariance (right).

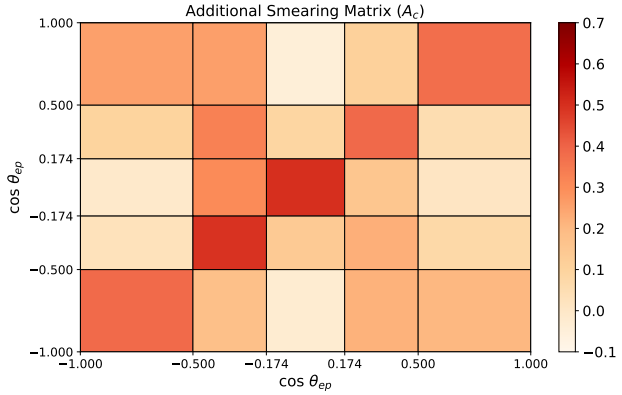


Figure 7.21: The additional smearing matrix A_C produced by the `NuWro` fake data test on opening angle.

7.4 NuMI data results

In this section, the beam-on selected event rates shown in Figure 5.39, subtracted by GENIE-estimated background distributions, are used as the measured event rates input into the Wiener-SVD unfolding algorithm. Final results are presented in comparison with frequently used neutrino event generators.

Electron energy

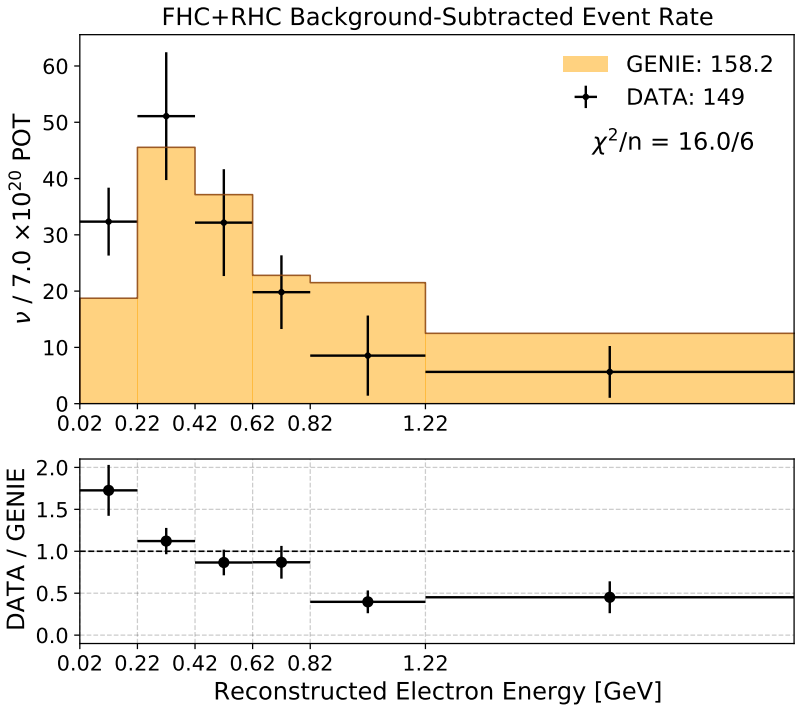


Figure 7.22: The FHC+RHC background-subtracted event rate as a function of reconstructed electron energy.

Figure 7.22 shows the FHC+RHC background-subtracted event rate (black) compared to the GENIE selected signal prediction (orange) as a function of reconstructed electron energy. A total of 149 beam-on events survive the subtraction. Error bars on the data are derived from the covariance matrix shown in Figure 6.47, representative of the total uncertainty with the GENIE systematic contribution from the signal channel turned off. Goodness of fit to the

GENIE prediction is $\chi^2/n = 16.0/6$. This distribution and its covariance are used as the input measured event rate for Wiener-SVD unfolding on electron energy. The same response matrix as in Figure 7.1 also serves as input.

Figure 7.23 displays the unfolded cross section result as a function of electron energy (left), in comparison with the GENIE prediction that has been smeared by the output A_c matrix (right). Unfolded covariance and correlation matrices are shown in Figure 7.24.

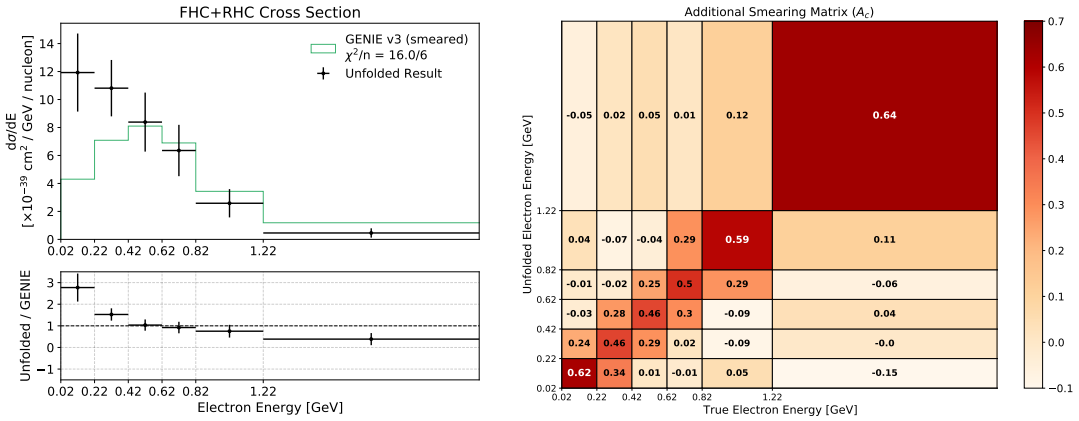


Figure 7.23: The unfolded FHC+RHC cross section using NuMI data as a function of electron energy (left) as compared to the GENIE prediction, which has been smeared by the output A_c matrix (right).

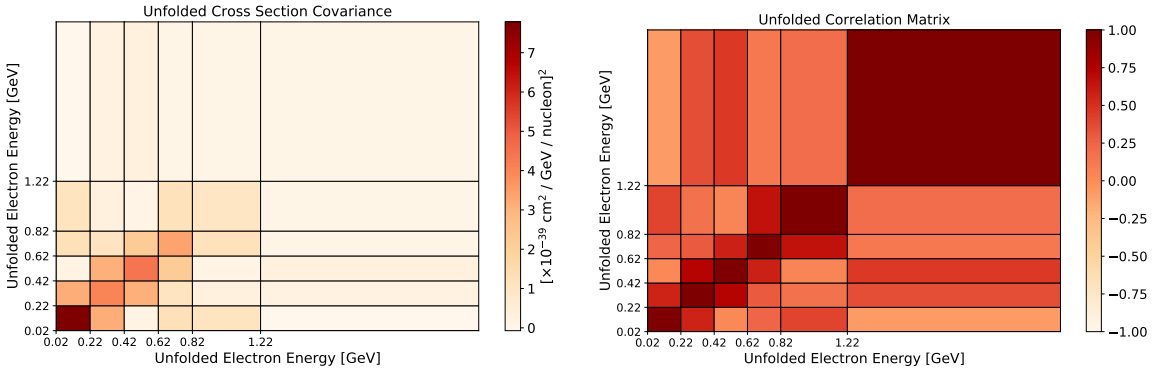


Figure 7.24: Unfolded covariance (left) and correlation (right) matrices of the FHC+RHC cross section using NuMI data as a function of electron energy.

Visible energy

Figure 7.25 shows the FHC+RHC background-subtracted event rate (black) compared to the `GENIE` selected signal prediction (orange) as a function of reconstructed visible energy. Goodness of fit to the `GENIE` prediction is $\chi^2/n = 30.7/7$. This distribution and its covariance are used as the input measured event rate for Wiener-SVD unfolding on visible energy. The same response matrix as in Figure 7.4 also serves as input.

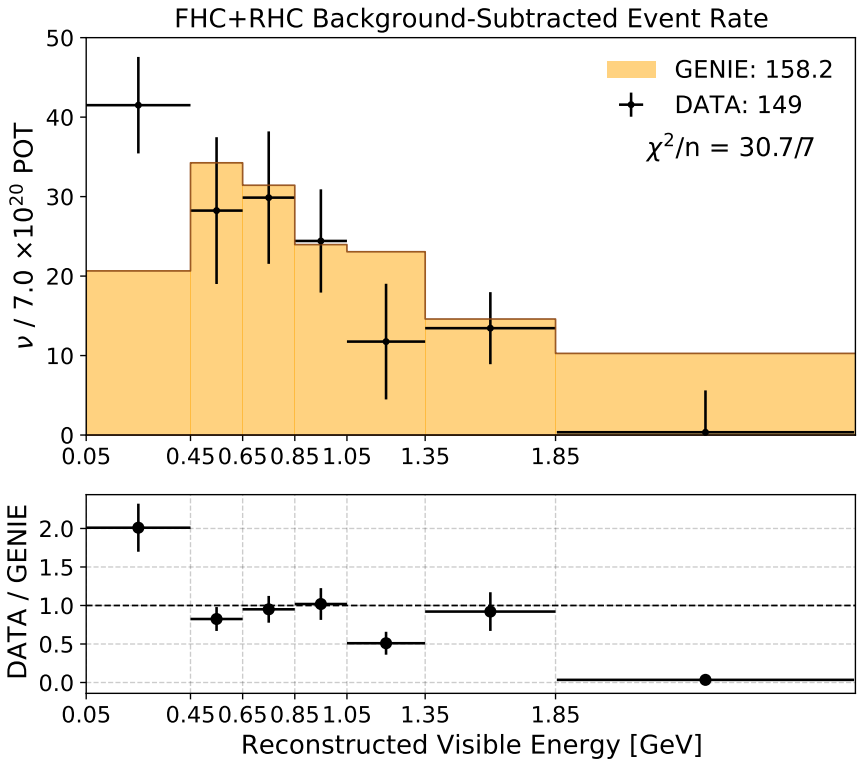


Figure 7.25: The FHC+RHC background-subtracted event rate as a function of reconstructed visible energy.

Figure 7.26 displays the unfolded cross section result as a function of visible energy (left), in comparison with the `GENIE` prediction that has been smeared by the output A_c matrix (right). Unfolded covariance and correlation matrices are shown in Figure 7.27.

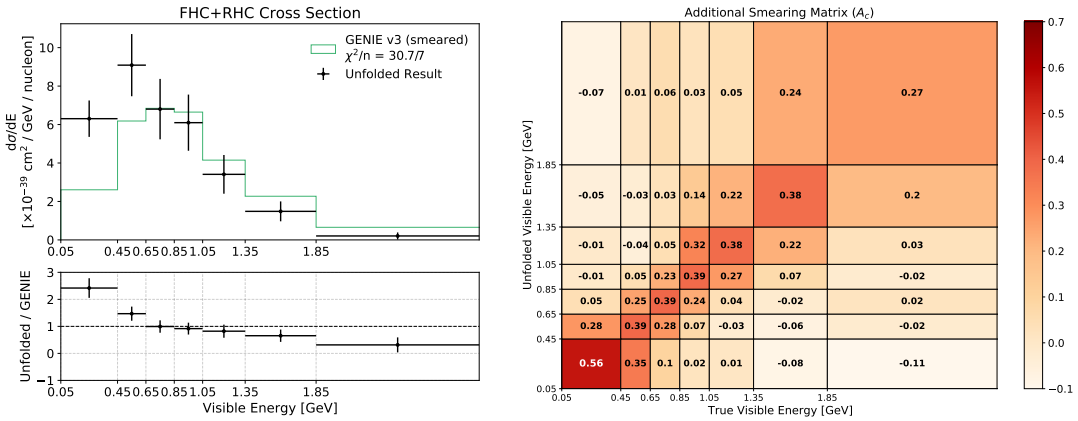


Figure 7.26: The unfolded FHC+RHC cross section using NuMI data as a function of visible energy (left) as compared to the GENIE prediction, which has been smeared by the output A_c matrix (right).

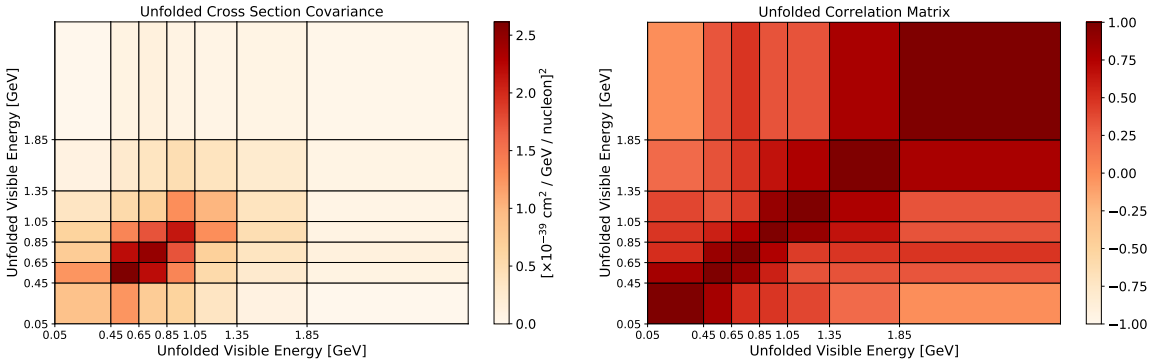


Figure 7.27: Unfolded covariance (left) and correlation (right) matrices of the FHC+RHC cross section using NuMI data as a function of visible energy.

Opening angle

Figure 7.28 shows the FHC+RHC background-subtracted event rate (black) compared to the GENIE selected signal prediction (orange) as a function of reconstructed opening angle. Goodness of fit to the GENIE prediction is $\chi^2/n = 8.1/5$. This distribution and its covariance are used as the input measured event rate for Wiener-SVD unfolding on opening angle. The same response matrix as in Figure 7.7 also serves as input.

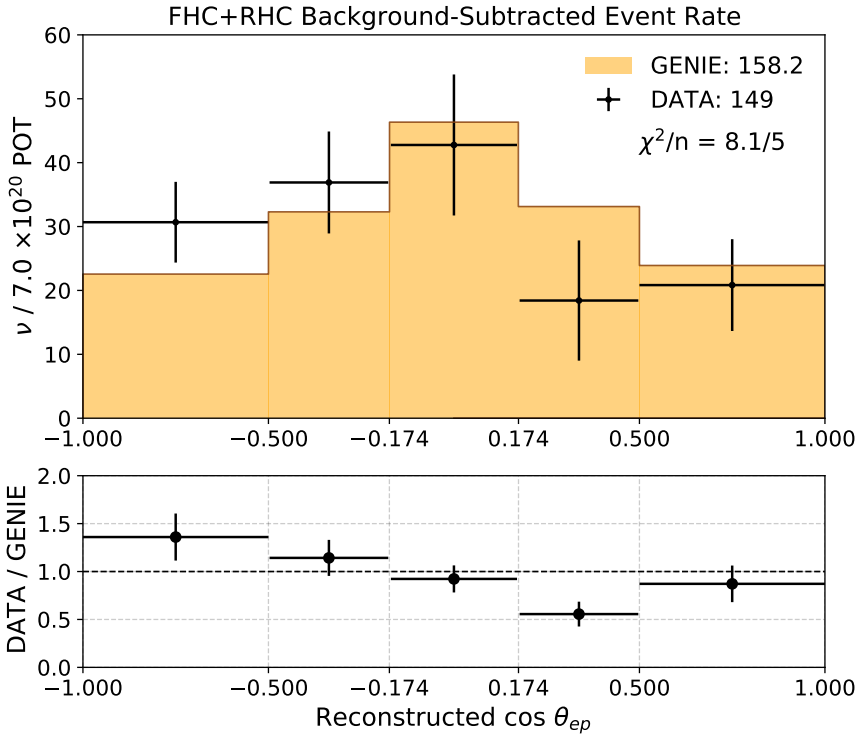


Figure 7.28: The FHC+RHC background-subtracted event rate as a function of reconstructed opening angle.

Figure 7.29 displays the unfolded cross section result as a function of opening angle (left), in comparison with the `GENIE` prediction that has been smeared by the output A_c matrix (right). Unfolded covariance and correlation matrices are shown in Figure 7.30.

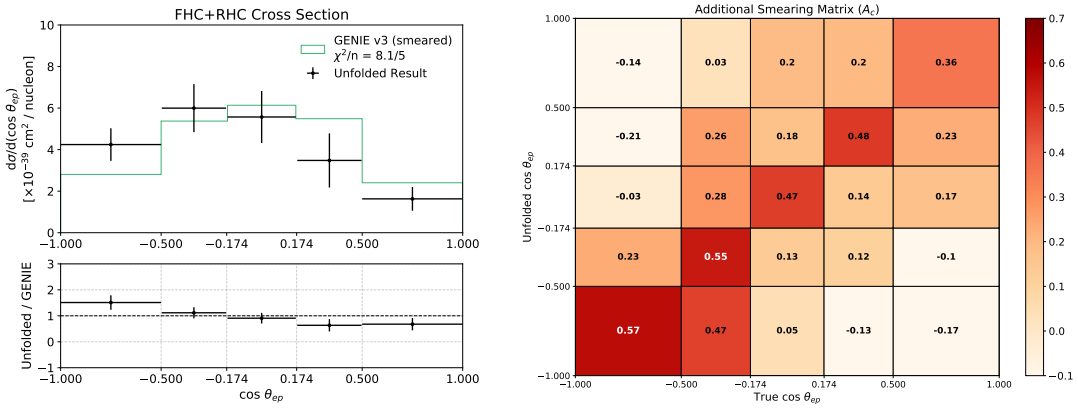


Figure 7.29: The unfolded FHC+RHC cross section using NuMI data as a function of opening angle (left) as compared to the `GENIE` prediction, which has been smeared by the output A_c matrix (right).

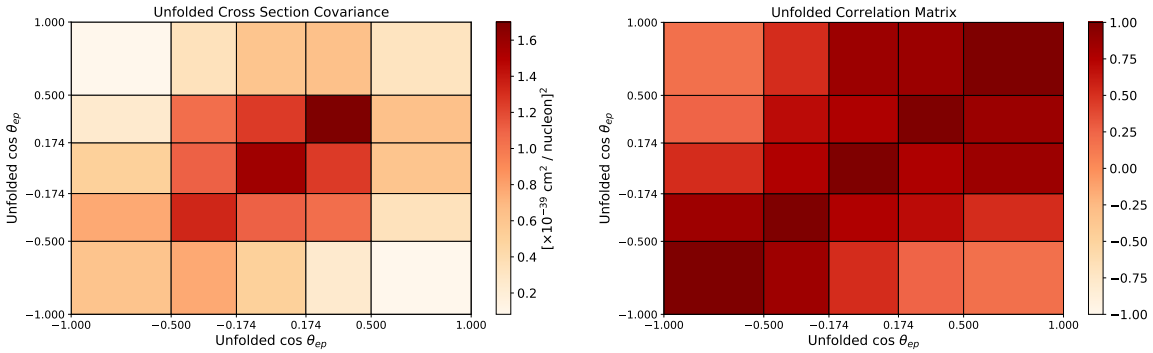


Figure 7.30: Unfolded covariance (left) and correlation (right) matrices of the FHC+RHC cross section using NuMI data as a function of opening angle.

Proton multiplicity

Figure 7.31 shows the FHC+RHC background-subtracted event rate (black) compared to the `GENIE` selected signal prediction (orange) as a function of reconstructed proton multiplicity. Goodness of fit to the `GENIE` prediction is $\chi^2/n = 1.6/3$. The covariance and correlation matrices corresponding to the data distribution are displayed in Figure 7.32.

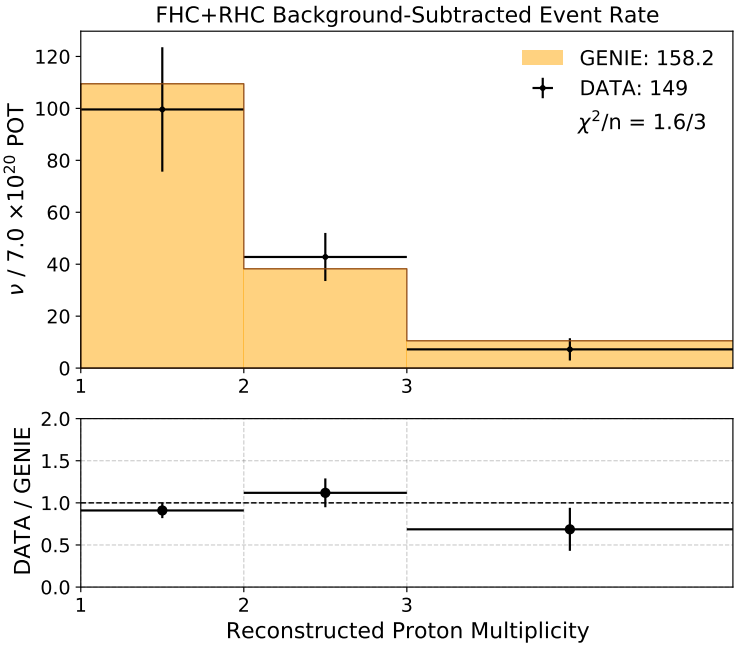


Figure 7.31: The FHC+RHC background-subtracted event rate as a function of reconstructed proton multiplicity.

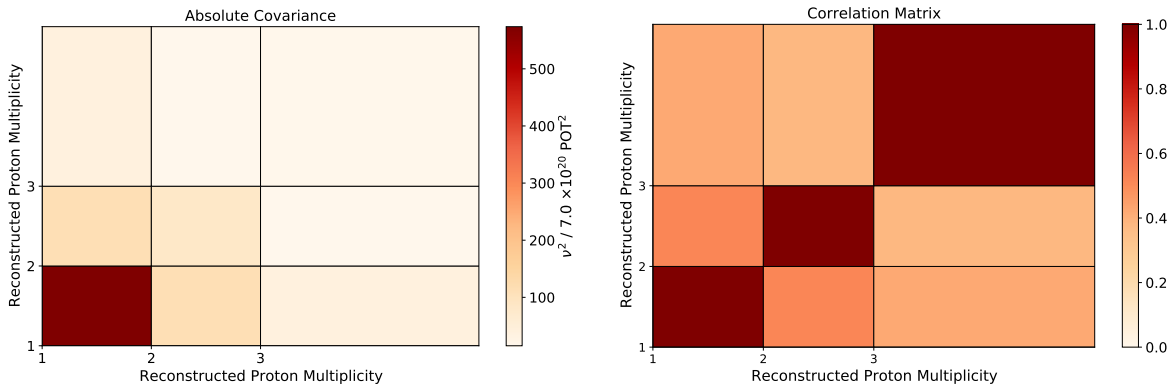


Figure 7.32: Covariance (left) and correlation (right) matrices of the FHC+RHC background-subtracted event rate as a function of the reconstructed proton multiplicity.

7.5 Comparisons to neutrino event generators

Unfolded results are compared to predictions from neutrino event generators commonly employed in the field: `NEUT v5.4.0` [84, 85], `NuWro v.19.02.2` [82, 83], `GiBUU 2021` [86], `GENIE v03.0.6` (untuned) [87], and finally `GENIE v03.0.6` (tuned) [65], the primary simulator used in this work. Each prediction was generated with $O(10^5)$ ν_e events. In precedence with previous MicroBooNE results, uncertainties on the generator predictions are deemed negligible and not included in the goodness of fit assessment. Each prediction uses a NuMI ν_e flux spectrum corrected with `PPFX` as input, and is smeared by A_c for a direct comparison with the unfolded cross section.

Generator comparisons with the unfolded data result are shown in Figures 7.33, 7.34, and 7.35 for electron energy, visible energy, and opening angle, respectively. For opening angle, the data is decently represented by model predictions, with best agreement to `GiBUU 2021` ($\chi^2/n = 3.4/5$), then `NEUT v5.4.0` ($\chi^2/n = 6.1/5$), `GENIE v3.0.6` (tuned, $\chi^2/n = 8.1/5$), `GENIE v3.0.6` (untuned, $\chi^2/n = 8.6/5$), and finally `NuWro v19.02.2` ($\chi^2/n = 10.3/5$).

The data is also in agreement with generators for electron energies ≥ 0.42 GeV and visible energies ≥ 0.65 GeV, though marked disagreement is observed in the lower two bins. Table 7.3 and Table 7.4 show the goodness of fit for these variables, respectively.

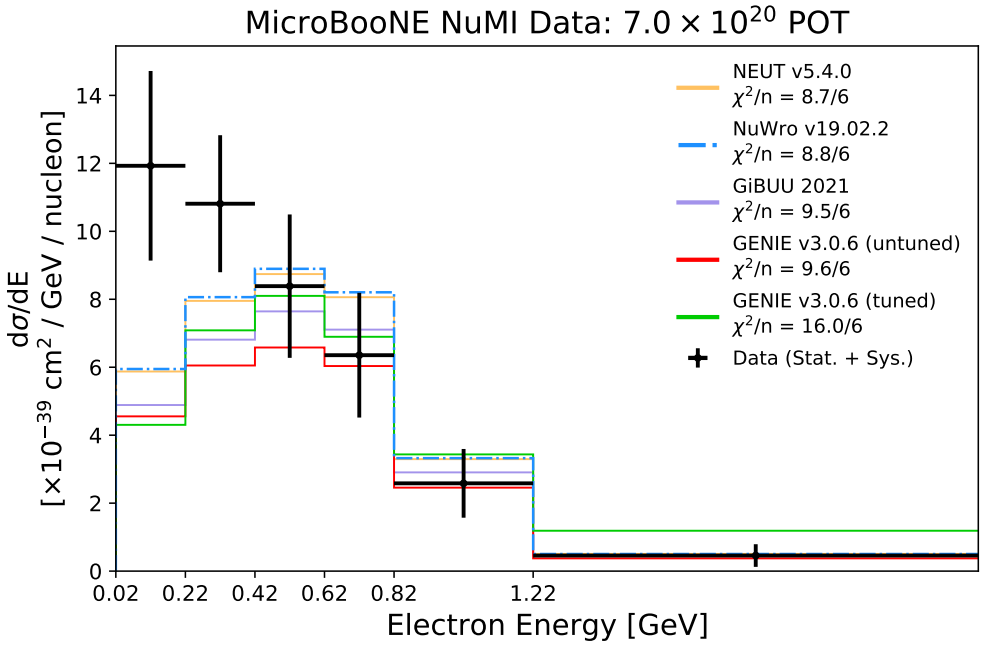


Figure 7.33: The unfolded data result on electron energy, in comparison with predictions from `NEUT v5.4.0` (yellow), `NuWro v.19.02.2` (blue), `GiBUU 2021` (lilac), `GENIE v03.0.6` untuned (red), and `GENIE v03.0.6` tuned (green).

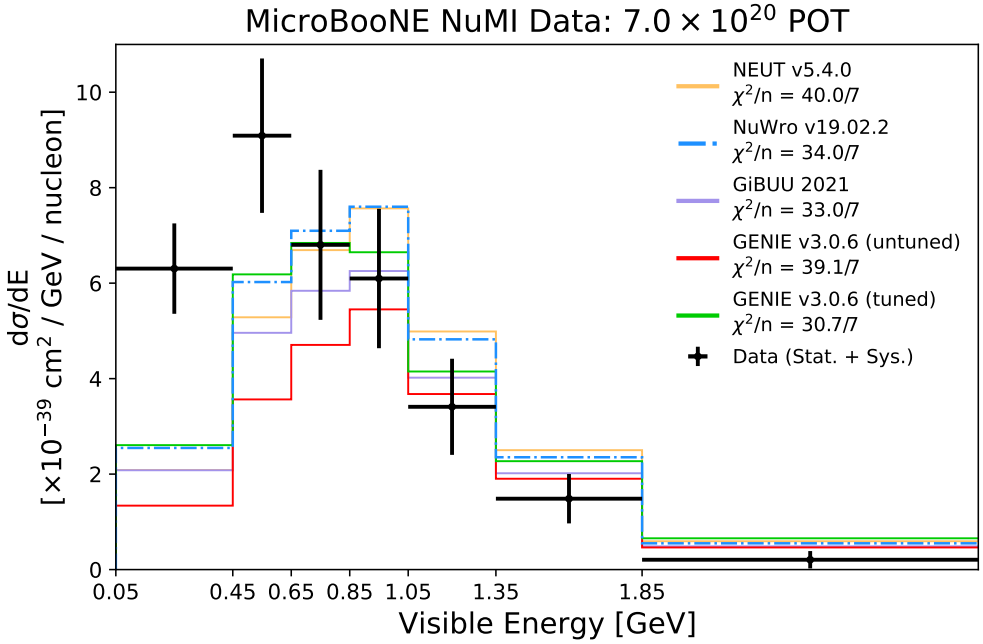


Figure 7.34: The unfolded data result on visible energy, in comparison with predictions from `NEUT v5.4.0` (yellow), `NuWro v.19.02.2` (blue), `GiBUU 2021` (lilac), `GENIE v03.0.6` untuned (red), and `GENIE v03.0.6` tuned (green).

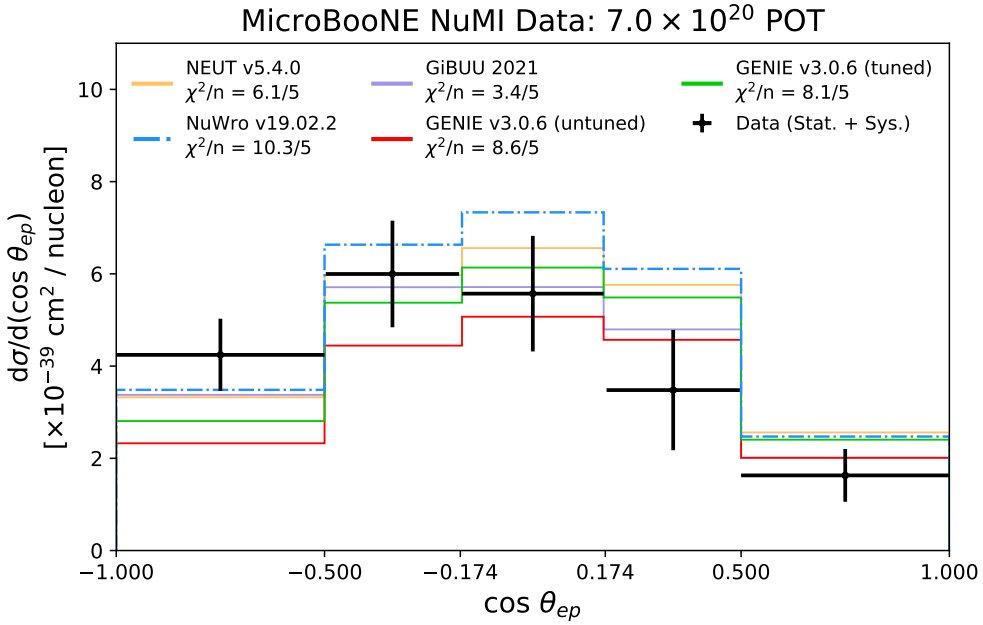


Figure 7.35: The unfolded data result on opening angle, in comparison with predictions from `NEUT v5.4.0` (yellow), `NuWro v.19.02.2` (blue), `GiBUU 2021` (lilac), `GENIE v03.0.6` untuned (red), and `GENIE v03.0.6` tuned (green).

Across the entire phase space, the reduced χ^2 ranges from $8.7/6 = 1.5$ (`NEUT v5.4.0`) to $16.0/6 = 2.7$ (`GENIE v3.0.6`, tuned) for electron energy. Partial χ^2/n values for the first two bins only, computed by inverting the relevant subset of the unfolded covariance matrix, reveal the extent to which the lower energy disagreement impacts the goodness of fit average across the entire phase space: For electron energies ≥ 0.42 GeV, the reduced χ^2 is less than unity for every generator except `GENIE v3.0.6` (tuned, $\chi^2/n = 8.1/4$). For electron energies in the region 0.02-0.42 GeV, however, the reduced χ^2 ranges from $4.6/2 = 2.3$ (`GiBUU 2021`) to $8.1/2 = 4.1$ (`NuWro v19.02.2`).

Similar trends are observed in the χ^2 values for visible energy. Across the entire phase space, the reduced χ^2 ranges from $30.7/7 = 4.4$ (`GENIE v3.0.6`, tuned) to $40.0/7 = 5.7$ (`NEUT v5.4.0`). For visible energies ≥ 0.65 GeV, the partial χ^2/n for all generators is ~ 2. In the visible energy region 0.05-0.65 GeV, however, partial χ^2/n values range from $21.0/2 = 10.5$ (`GENIE v3.0.6`, tuned) to $29.7/2 = 14.9$ (`NuWro v19.02.2`).

Generator	χ^2/n (total)	χ^2/n (0.02 - 0.42 GeV)	χ^2/n ($\geq$ 0.42 GeV)
NEUT v5.4.0	8.7/6	4.8/2	1.3/4
NuWro v19.02.2	9.6/6	8.1/2	1.8/4
GiBUU 2021	8.8/6	4.6/2	1.4/4
GENIE v03.0.6 (untuned)	9.5/6	6.8/2	1.3/4
GENIE v03.0.6 (tuned)	16.0/6	7.6/2	8.1/4

Table 7.3: Reduced χ^2 values for unfolded electron energy in comparison with each of the generators shown in Figure 7.33. The partial χ^2/n is computed by inverting the relevant subset of the full unfolded covariance matrix.

Generator	χ^2/n (total)	χ^2/n (0.05 - 0.65 GeV)	χ^2/n ($\geq$ 0.65 GeV)
NEUT v5.4.0	40.0/7	24.7/2	11.2/5
NuWro v19.02.2	39.1/7	29.7/2	11.2/5
GiBUU 2021	34.0/7	21.1/2	8.4/5
GENIE v03.0.6 (untuned)	33.0/7	23.4/2	8.2/5
GENIE v03.0.6 (tuned)	30.7/7	21.0/2	12.3/5

Table 7.4: Reduced χ^2 values for unfolded visible energy in comparison with each of the generators shown in Figure 7.34. The partial χ^2/n is computed by inverting the relevant subset of the full unfolded covariance matrix.

Notably, the goodness of fit calculation does not account for uncertainties associated with each generator prediction, due to a lack of information to properly assess them, which would indeed diminish these values. In addition, while the Wiener-SVD method preserves χ^2 across all bins before and after unfolding, this statement is not necessarily true for individual bins. Figures 7.22 and 7.25 better illustrate the agreement before bias is introduced via the unfolding algorithm. In the background-subtracted event rates, a discrepancy as compared to the GENIE prediction is observed in the first bin only, quantified as $\chi^2 = 5.1$ ($p = 0.02393 \rightarrow 2.3\sigma$) for reconstructed electron energy and $\chi^2 = 11.8$ ($p = 0.00059 \rightarrow 3.4\sigma$) for reconstructed visible energy, using a one-sided p-value and two-tailed Z score.

To investigate this tension as a low energy excess, signal channel uncertainties must be included, which reduces the significance. Using the event rates in Figure 5.39, goodness of fit in the first bin only is computed as $\chi^2 = 2.4$ ($p = 0.1213 \rightarrow 1.5\sigma$) for reconstructed electron energy and $\chi^2 = 5.8$ ($p = 0.01603 \rightarrow 2.4\sigma$) for reconstructed visible energy.

CHAPTER 8

CONCLUSION & OUTLOOK

This work presents a measurement of $\nu_e + {}^{40}Ar \rightarrow 1e + Np$ interactions using data collected by the MicroBooNE detector from the NuMI neutrino beam at Fermilab. Exclusive differential cross sections are extracted on electron energy, visible energy, and opening angle between the outgoing electron and most energetic proton. Interaction rates and associated uncertainties as a function of proton multiplicity are also reported.

A state-of-the-art NuMI flux prediction is generated via the implementation of `PPFX` into LArSoft, which corrects the raw simulation with data-driven constraints based on a variety of independent hadronic cross section measurements. Using `GENIE` simulations of the FHC Run 1 and RHC Run 3 event rates in MicroBooNE's NuMI dataset, independent, though analogous, selections of electron neutrino interactions with at least 1 proton and no pions present in the final topology are developed. The FHC Run 1 (RHC Run 3) selection algorithm performs with 13.7% (11.4%) efficiency and 77.0% (73.5%) purity. A total of 203 beam-on data events in the FHC+RHC sample passes the selection, and 149 data events survive the background subtraction. Systematic and statistical covariance on the background-subtracted event rate is assessed. At low energies, the largest source of uncertainty is in the hadron production models underlying the flux prediction. Statistical uncertainty dominates at higher energies due to limited events in the beam-on data samples. Across the non-energetic variables, these contributions are comparable.

With the exception of proton multiplicity, closure tests and fake data studies are employed to validate the Wiener-SVD unfolding algorithm, which is then used to derive differential cross sections using the NuMI beam-on dataset. Goodness of fit between the unfolded data result and the primary `GENIE` generator used in this work is reported as $\chi^2/n = 16.0/6$ for electron energy, $\chi^2/n = 30.7/7$ for visible energy, and $\chi^2/n = 8.1/5$ for opening angle. Because the Wiener-SVD approach is not optimized to handle discrete distributions, the

background-subtracted event rate as a function of proton multiplicity is instead reported, with an agreement of $\chi^2/n = 1.6/3$ to `GENIE` prediction.

Tension is observed in the range of 0.02-0.42 GeV electron energy and 0.05-0.65 GeV visible energy. In the background-subtracted event rates (before unfolding), this manifests as a 2.3σ and 3.4σ difference between data and prediction in the lowest reconstructed electron energy and visible energy bin, respectively. To investigate this tension as a low energy excess, cross section uncertainties on the signal channel should be included. This reduces the significance to 1.5σ and 2.4σ in the lowest electron and visible energy bin, respectively. Curiously, a similar systematic excess was observed in the MiniBooNE detector [88].

By far, as is often the case for analyses with exclusive signal definitions, the largest limitation of this work is statistical. The next iteration of this work should include a significant reduction of statistical uncertainty with the addition of NuMI data from MicroBooNE Runs 2, 4, and 5, for which production is currently underway. This will more than double the currently available dataset, enabling future studies to extract individual (FHC-only and RHC-only) cross sections for comparison with the combined result, and potentially probe differences in the rate of the irreducible electron antineutrino background ($\bar{\nu}_e + {}^{40}Ar \rightarrow 1e{+}Np$) contamination.

APPENDIX A

WIENER-SVD SMOOTHING

Wiener-SVD unfolding takes advantage of a technique traditionally used in digital signal processing, which requires deconvolution of the field and electronics response from a measured signal to recover the true signal. For measured time-series distributions, this can be achieved by executing a Fourier transformation of the signal into the frequency domain, then introducing a filter function (often, the Wiener filter) to suppress large fluctuations arising from noise. This approach is adopted in Wiener-SVD unfolding by constructing an additional smearing matrix A_c that is dependent on the Wiener filter function W. More information about this method can be found in [79].

The generalized Wiener-SVD approach involves a free parameter matrix C that sets the derivative between adjacent bins of the measured spectrum to be smoothed in the regularization. That is:

$$\bar{M} = R \cdot C^{-1} \cdot C \cdot \bar{s} \tag{A.1}$$

where $\bar{M}$ is the expectation of the measured signal distribution based on $\bar{s}$, the expectation of the true signal distribution. The introduction of C is effectively a bias that modifies the response matrix R.

Using singular value decomposition, the effective response matrix can be broken down as:

$$R \cdot C^{-1} = U \cdot D \cdot V^T \tag{A.2}$$

where U and V are orthogonal matrices satisfying $U^T U = U U^T = V^T V = V V^T = I$ (the identity matrix) and D is a diagonal matrix with non-negative elements. The final unfolded result can be expressed as:

$$\hat{s} = A_c \cdot (R^T R)^{-1} \cdot R^T \cdot M \tag{A.3}$$

where M is the real measured spectrum and $A_c = C^{-1} \cdot V \cdot W \cdot V^T \cdot C$.

Higher orders of C increase the strength of the regularization applied to the unfolded distribution, but leave the χ^2 between data and prediction consistent. This is illustrated in Figures A.1, A.2, and A.3, which display the unfolded data results and corresponding A_c matrices via direct inversion (no regularization, thereby reducing Equation A.2 to $\bar{M} = R \cdot \bar{s}$) and an applied smoothing of $C = 0$, $C = 1$, and $C = 2$, for electron energy, visible energy, and opening angle, respectively. Note that in the direct inversion result, A_c is an identity matrix, and increasing C causes this shape to migrate away from the diagonals.

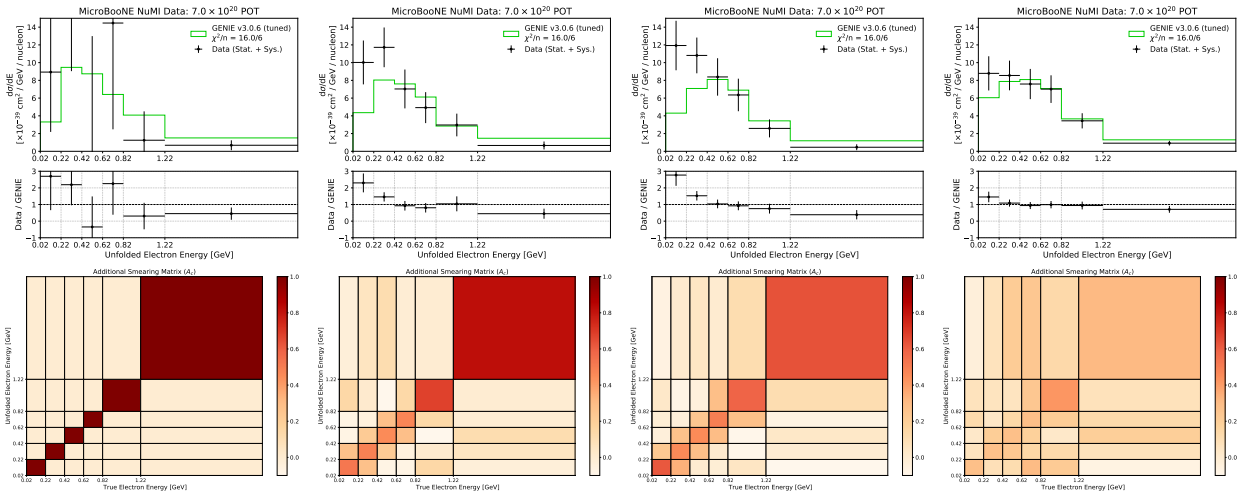


Figure A.1: From left to right, the unfolded data result on electron energy via direct inversion and an applied smoothing of $C = 0$, $C = 1$, and $C = 2$, along with corresponding regularization matrices A_c. Note that the χ^2 is consistent regardless of regularization.

The choice recommended by [79] is $C = 2$, which regularizes the curvature of the unfolded spectrum via a second derivative matrix. However, for visible energy (Figure A.2), the $C = 2$ spectrum suffers from a loss of shape information, indicated by the uniformity of A_c as well as the way the unfolded distribution closes completely onto the prediction at energies < 0.45 GeV as compared to the background-subtracted event rate (Figure 7.25), which displays a discrepancy in the first bin.

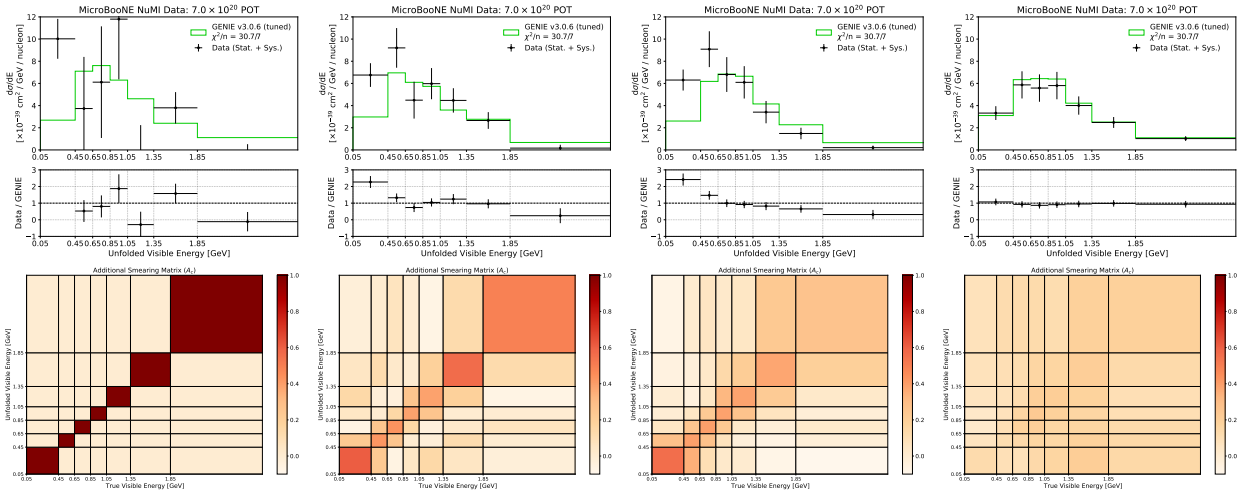


Figure A.2: From left to right, the unfolded data result on visible energy via direct inversion and an applied smoothing of $C = 0$, $C = 1$, and $C = 2$, along with corresponding regularization matrices A_c. Note that the χ^2 is consistent regardless of regularization.

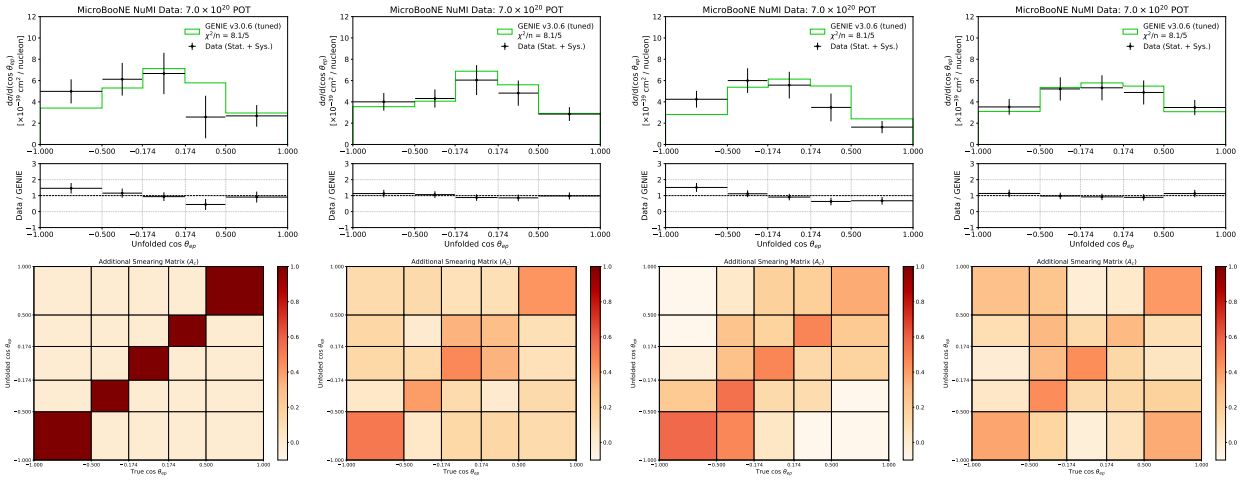


Figure A.3: From left to right, the unfolded data result on opening angle via direct inversion and an applied smoothing of $C = 0$, $C = 1$, and $C = 2$, along with corresponding regularization matrices A_c. Note that the χ^2 is consistent regardless of regularization.

Thus, this work reduces the regularization strength to $C = 1$, which preserves some semblance of the pre-unfolded shape below 0.45 GeV. For consistency, the $C = 1$ choice of regularization is also used for the final unfolded results on electron energy and opening angle.

REFERENCES

[1] F. L. Wilson. Fermi's Theory of Beta Decay. *Am. J. Phys.*, 36(12):1150–1160, 1968.

[2] H. Bethe and R. Peierls. The 'neutrino'. *Nature*, 133:532, 1934.

[3] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the free neutrino: A Confirmation. *Science*, 124:103–104, 1956.

[4] G. Danby, J-M. Gaillard, K. Goulianos, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Phys. Rev. Lett.*, 9:36–44, 1962.

[5] K. Kodama et al. Observation of tau neutrino interactions. *Phys. Lett. B*, 504:218–224, 2001.

[6] B. T. Cleveland, T. Daily, R. Davis, Jr., J. R. Distel, K. Lande, C. K. Lee, P. S. Wildenhain, and J. Ullman. Measurement of the solar electron neutrino flux with the Homestake chlorine detector. *Astrophys. J.*, 496:505–526, 1998.

[7] Y. Fukuda et al. (Kamiokande Collaboration). Solar Neutrino Data Covering Solar Cycle 22. *Phys. Rev. Lett.*, 77:1683–1686, 1996.

[8] Y. Fukuda et al. (Super-Kamiokande Collaboration). Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[9] Q. R. Ahmad et al. Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, 2001.

[10] Z. Maki, M. Nakagawa, and S. Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, 1962.

[11] B. Pontecorvo. Neutrino Experiments and the Problem of Conservation of Leptonic Charge. *Zh. Eksp. Teor. Fiz.*, 53:1717–1725, 1967.

[12] P. A. Zyla et al. (Particle Data Group). Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(083C01), 2020. Includes 2021 update.

[13] B. Abi et al. Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I Introduction to DUNE. *JINST*, 15(08):T08008, 2020.

[14] A. Aguilar et al. Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam. *Phys. Rev. D*, 64:112007, 2001.

[15] 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.

[16] A. de Gouvêa et al. Theory of Neutrino Physics – Snowmass TF11 (NF08) Topical Group Report. 2022.

[17] J. A. Formaggio and G. P. Zeller. From eV to EeV: Neutrino Cross Sections Across Energy Scales. *Rev. Mod. Phys.*, 84:1307–1341, 2012.

[18] M. Day and K. S. McFarland. Differences in Quasi-Elastic Cross-Sections of Muon and Electron Neutrinos. *Phys. Rev. D*, 86:053003, 2012.

[19] K. Abe et al. Measurement of the Inclusive Electron Neutrino Charged Current Cross Section on Carbon with the T2K Near Detector. *Phys. Rev. Lett.*, 113(24):241803, 2014.

[20] M. A. Acero et al. Measurement of the ν_e-Nucleus Charged-Current Double-Differential Cross Section at $\langle E_\nu \rangle = 2.4$ GeV using NOvA. *Phys. Rev. Lett.*, 130(5):051802, 2023.

[21] K. Abe et al. Measurement of the charged-current electron (anti-)neutrino inclusive cross-sections at the T2K off-axis near detector ND280. *JHEP*, 10:114, 2020.

[22] J. Wolcott et al. Measurement of electron neutrino quasielastic and quasielasticlike scattering on hydrocarbon at $\langle E_\nu \rangle = 3.6$ GeV. *Phys. Rev. Lett.*, 116(8):081802, 2016.

[23] R. Acciarri et al. First measurement of electron neutrino scattering cross section on argon. *Phys. Rev. D*, 102(1):011101, 2020.

[24] P. Abratenko et al. Measurement of the flux-averaged inclusive charged-current electron neutrino and antineutrino cross section on argon using the NuMI beam and the MicroBooNE detector. *Phys. Rev. D*, 104(5):052002, 2021.

[25] P. Abratenko et al. First measurement of inclusive electron-neutrino and antineutrino charged current differential cross sections in charged lepton energy on argon in MicroBooNE. *Phys. Rev. D*, 105(5):L051102, 2022.

[26] P. Abratenko et al. Differential cross section measurement of charged current ν_e interactions without final-state pions in MicroBooNE. *Phys. Rev. D*, 106(5):L051102, 2022.

[27] P. Adamson et al. The NuMI Neutrino Beam. *Nucl. Instrum. Meth. A*, 806:279–306, 2016.

[28] B. C. Brown, P. Adamson, D. Capista, W. Chou, I. Kourbanis, D. K. Morris, K. Seiya, G. H. Wu, and M. Yang. The Fermilab Main Injector: high intensity operation and beam loss control. *Phys. Rev. ST Accel. Beams*, 16(7):071001, 2013.

[29] R. Ainsworth, P. Adamson, B. C. Brown, D. Capista, K. Hazelwood, I. Kourbanis, D. K. Morris, M. Xiao, and M. Yang. High intensity operation using proton stacking in the Fermilab Recycler to deliver 700 kW of 120 GeV proton beam. *Phys. Rev. Accel. Beams*, 23(12):121002, 2020.

[30] P. Abratenko et al. Search for long-lived heavy neutral leptons and Higgs portal scalars decaying in the MicroBooNE detector. *Phys. Rev. D*, 106(9):092006, 2022.

[31] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[32] L. Aliaga Soplin. *Neutrino Flux Prediction for the NuMI Beamline*. PhD thesis, William-Mary Coll., 2016.

[33] L. Aliaga Soplin et al. Neutrino Flux Predictions for the NuMI Beam. *Phys. Rev. D*, 94(9):092005, 2016. [Addendum: Phys.Rev.D 95, 039903 (2017)].

[34] C. Alt et al. Inclusive production of charged pions in p+C collisions at 158-GeV/c beam momentum. *Eur. Phys. J. C*, 49:897–917, 2007.

[35] D. S. Barton et al. Experimental Study of the a-Dependence of Inclusive Hadron Fragmentation. *Phys. Rev. D*, 27:2580, 1983.

[36] G. M. Tinti. *Sterile neutrino oscillations in MINOS and hadron production in pC collisions*. PhD thesis, 2010.

[37] A. V. Lebedev. *Ratio of pion kaon production in proton carbon interactions*. PhD thesis, Harvard U., 2007.

[38] R. J. Abrams, R. L. Cool, G. Giacomelli, T. F. Kycia, B. A. Leontic, K. K. Li, and D. N. Michael. Total cross-sections of $K^{\pm}$ mesons and anti-protons on nucleons up to 3.3-GeV/c. *Phys. Rev. D*, 1:1917–1935, 1970.

[39] S. P. Denisov, S. V. Donskov, Yu. P. Gorin, R. N. Krasnokutsky, A. I. Petrukhin, Yu. D. Prokoshkin, and D. A. Stoyanova. Absorption cross-sections for pions, kaons, protons and anti-protons on complex nuclei in the 6-GeV/c to 60-GeV/c momentum range. *Nucl. Phys. B*, 61:62–76, 1973.

[40] A. S. Carroll et al. Absorption cross sections of $\pi^{\pm}$, $K^{\pm}$, p and $\bar{p}$ on nuclei between 60 and 280 GeV/c. 1979. Physics Letters B, Volume 80, Issue 3, Pages 319-322.

[41] J. V. Allaby et al. Total and absorption cross-sections of π^{-}, K^{-} and $\bar{p}$ in the momentum range 20-65 GeV/c. *Yad. Fiz.*, 12:538–557, 1970.

[42] J. W. Cronin, R. Cool, and A. Abashian. Cross Sections of Nuclei for High-Energy Pions. *Phys. Rev.*, 107:1121–1130, 1957.

[43] B. W. Allardyce et al. Pion reaction cross sections and nuclear sizes. 1973. Nuclear Physics A Volume 209, Issue 1, Pages 1-51.

[44] M. J. Longo and B. J. Moyer. Nucleon and Nuclear Cross Sections for Positive Pions and Protons above 1.4 BeV/c. *Phys. Rev.*, 125:701–713, 1962.

[45] V. B. Fedorov, Yu. G. Grishuk, M. V. Kosov, G. A. Leksin, N. A. Pivnyuk, S. V. Shevchenko, V. L. Stolin, A. V. Vlasov, and L. S. Vorobev. Total Inelastic Cross-Sections for π Mesons on Nuclei in the 2-GeV/c to 6-GeV/c Momentum Range. *Sov. J. Nucl. Phys.*, 27:222, 1978.

[46] B. M. Bobchenko et al. Measurement of Total Inelastic Cross-Sections from Proton Interactions with Nuclei in the Momentum Range from 5-GeV to 9-GeV/c and π^- Mesons with Nuclei in the Momentum Range from 1.75-GeV/c to 6.5-GeV/c. *Sov. J. Nucl. Phys.*, 30:805, 1979.

[47] G. Bellettini, G. Cocconi, A. N. Diddens, E. Lillethun, G. Matthiae, J. P. Scanlon, and A. M. Wetherell. Proton-nuclei cross sections at 20 GeV. *Nucl. Phys.*, 79:609–624, 1966.

[48] N. Abgrall et al. Measurements of Cross Sections and Charged Pion Spectra in Proton-Carbon Interactions at 31 GeV/c. *Phys. Rev. C*, 84:034604, 2011.

[49] B. Baatar et al. Inclusive production of protons, anti-protons, neutrons, deuterons and tritons in p+C collisions at 158 GeV/c beam momentum. *Eur. Phys. J. C*, 73(4):2364, 2013.

[50] T. J. Roberts, H. Richard Gustafson, Lawrence W. Jones, M. J. Longo, and M. R. Whalley. Neutron Nucleus Inelestic Cross Sections from 160 to 375 GeV/c. *Nucl. Phys. B*, 159:56–66, 1979.

[51] L. Cremonesi. Cross-section measurements in the NOvA Near Detector, 2020. NOvA Presentation at The XXIX International Conference on Neutrino Physics and Astrophysics.

[52] P. Machado, O. Palamara, and D. Schmitz. The Short-Baseline Neutrino Program at Fermilab. *Ann. Rev. Nucl. Part. Sci.*, 69:363–387, 2019.

[53] A. Aguilar et al. (LSND Collaboration). Evidence for neutrino oscillations from the observation of $\overline{\nu}_e$ appearance in a $\overline{\nu}_\mu$ beam. *Phys. Rev. D*, 64:112007, 2001.

[54] M. Dentler et al. Updated global analysis of neutrino oscillations in the presence of eV-scale sterile neutrinos. *Journal of High Energy Physics*, 2018(8):10, 2018.

[55] S. Gariazzo, C. Giunti, M. Laveder, and Y. F. Li. Updated global 3+1 analysis of short-baseline neutrino oscillations. *Journal of High Energy Physics*, 2017(6):135, 2017.

[56] P. Virtanen et al. SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. *Nature Meth.*, 17:261, 2020.

[57] C. Rubbia. The Liquid Argon Time Projection Chamber: A New Concept for Neutrino Detectors. 1977.

[58] R. Acciarri et al. Design and Construction of the MicroBooNE Detector. *JINST*, 12(02):P02017, 2017.

[59] P. Abratenko et al. Measurement of the longitudinal diffusion of ionization electrons in the MicroBooNE detector. *JINST*, 16(09):P09025, 2021.

[60] M. Del Tutto. *First Measurements of Inclusive Muon Neutrino Charged Current Differential Cross Sections on Argon at 0.8 GeV Average Neutrino Energy with the MicroBooNE Detector*. PhD thesis, Oxford U., 2019.

[61] C. Adams et al. Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation. *JINST*, 13(07):P07006, 2018.

[62] G. F. Knoll. *Radiation Detection and Measurement*. John Wiley and Sons, New York, 3rd edition, 2000.

[63] E. L. Snider and G. Petrillo. LArSoft: Toolkit for Simulation, Reconstruction and Analysis of Liquid Argon TPC Neutrino Detectors. *J. Phys. Conf. Ser.*, 898(4):042057, 2017.

[64] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. *Nucl. Instrum. Meth. A*, 614:87–104, 2010.

[65] P. Abratenko et al. New $CC0\pi$ GENIE model tune for MicroBooNE. *Phys. Rev. D*, 105(7):072001, 2022.

[66] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. CORSIKA: A Monte Carlo code to simulate extensive air showers. 1998.

[67] C. Adams et al. Measurement of the atmospheric muon rate with the MicroBooNE Liquid Argon TPC. *JINST*, 16(04):P04004, 2021.

[68] R. Acciarri et al. The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector. *Eur. Phys. J. C*, 78(1):82, 2018.

[69] R. Acciarri et al. Michel Electron Reconstruction Using Cosmic-Ray Data from the MicroBooNE LArTPC. *JINST*, 12(09):P09014, 2017.

[70] C. Adams et al. Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons. *JINST*, 15(03):P03022, 2020.

[71] E. Shibamura, A. Hitachi, T. Doke, T. Takahashi, S. Kubota, and M. Miyajima. Drift velocities of electrons, saturation characteristics of ionization and W-values for conversion electrons in liquid argon, liquid argon-gas mixtures and liquid xenon. *Nucl. Instrum. Meth.*, 131:249–258, 1975.

[72] R. Acciarri et al. A Study of Electron Recombination Using Highly Ionizing Particles in the ArgoNeuT Liquid Argon TPC. *JINST*, 8:P08005, 2013.

[73] C. Adams et al. A method to determine the electric field of liquid argon time projection chambers using a UV laser system and its application in MicroBooNE. *JINST*, 15(07):P07010, 2020.

[74] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997.

[75] P. Abratenko et al. Calorimetric classification of track-like signatures in liquid argon TPCs using MicroBooNE data. *JHEP*, 12:153, 2021.

[76] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[77] M. Day and K. S. McFarland. Differences in Quasi-Elastic Cross-Sections of Muon and Electron Neutrinos. *Phys. Rev. D*, 86:053003, 2012.

[78] J. Calcutt, C. Thorpe, K. Mahn, and L. Fields. Geant4Reweight: a framework for evaluating and propagating hadronic interaction uncertainties in Geant4. *JINST*, 16(08):P08042, 2021.

[79] W. Tang, X. Li, X. Qian, H. Wei, and C. Zhang. Data Unfolding with Wiener-SVD Method. *JINST*, 12(10):P10002, 2017.

[80] P. Abratenko et al. Multi-Differential Cross Section Measurements of Muon-Neutrino-Argon Quasielastic-like Reactions with the MicroBooNE Detector. 2023.

[81] P. Abratenko et al. First double-differential measurement of kinematic imbalance in neutrino interactions with the MicroBooNE detector. 2023.

[82] T. Golan, J. T. Sobczyk, and J. Zmuda. NuWro: the Wroclaw Monte Carlo Generator of Neutrino Interactions. *Nucl. Phys. B Proc. Suppl.*, 229-232:499–499, 2012.

[83] T. Golan, C. Juszczak, and J. T. Sobczyk. Final State Interactions Effects in Neutrino-Nucleus Interactions. *Phys. Rev. C*, 86:015505, 2012.

[84] Y. Hayato. A neutrino interaction simulation program library NEUT. *Acta Phys. Polon. B*, 40:2477–2489, 2009.

[85] Y. Hayato and L. Pickering. The NEUT neutrino interaction simulation program library. *Eur. Phys. J. ST*, 230(24):4469–4481, 2021.

[86] O. Buss, T. Gaitanos, K. Gallmeister, H. van Hees, M. Kaskulov, O. Lalakulich, A. B. Larionov, T. Leitner, J. Weil, and U. Mosel. Transport-theoretical Description of Nuclear Reactions. *Phys. Rept.*, 512:1–124, 2012.

[87] L. Alvarez-Ruso et al. Recent highlights from GENIE v3. *Eur. Phys. J. ST*, 230(24):4449–4467, 2021.

[88] P. Adamson et al. First Measurement of ν_μ and ν_e Events in an Off-Axis Horn-Focused Neutrino Beam. *Phys. Rev. Lett.*, 102:211801, 2009.