Measurement of Charged Current Inclusive $\bar{\nu}_\mu$ Interactions and 2p2h Contribution Using the NOvA Near Detector

A dissertation submitted by

Saqib Shadman Bashar

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physics

Tufts University

September 2022

Advisor: Prof. W. Anthony Mann

Abstract

More precise measurements of the scattering of antineutrinos on nuclei in the few GeV region of incident energy, E_ν, can reveal new aspects of $\nu/\bar{\nu}$ interactions and of nuclear structure, while providing information crucial to continued progress by neutrino oscillation experiments. This Thesis reports an investigation of charged current $\bar{\nu}_\mu$ interactions in the predominantly hydrocarbon medium of the NOvA Near Detector. The data are obtained with the detector exposed to a narrow-band $\bar{\nu}_\mu$ beam with spectral peak at $E_\nu = 2.0$ GeV, for a total accumulated exposure of $12.5{\times}10^{20}$ protons-on-target. The analysis expresses results in terms of three-momentum transfer, $|\vec{q}|$ and available hadronic energy, E_{avail}. These variables approximate (q_0, $|\vec{q}|$), the variables used in theoretical treatments of lepton scattering on multi-nucleon systems such as in 2-particle-2-hole (2p2h) processes. This Thesis delivers two core measurements: Firstly, the double-differential cross section $d^2\sigma/d|\vec{q}|dE_{avail}$ for inclusive CC $\bar{\nu}_\mu$ interactions in the NOvA detector medium is obtained. The cross section has a roughly gaussian shape in its projection onto $|\vec{q}|$, while being largely concentrated at values of E_{avail} below 100 MeV. The single differential cross sections $d\sigma/d|\vec{q}|$ and $d\sigma/dE_{avail}$ are also obtained, and are compared to predictions based on the GENIE v3.0.6 neutrino generator augmented with four different implementations of 2p2h reactions. Secondly, the contribution of 2p2h reactions to the CC inclusive scattering sample is determined. Its presence is characterized in terms of double-differential and single-differential cross sections. The latter 2p2h distributions are compared to predictions from two experimental data-tune models and from two theory-based models, namely the SuSAv2 model and the Valencia model. Both of the theoretical models underpredict the measured cross-section strength and predict a shape for $d\sigma/d|\vec{q}|$ that is flatter than the measured shape. Thus, this Thesis provides new information about CC antineutrino-nucleus scattering in the vicinity of $E_\nu \sim 2$ GeV, and it highlights improvements that are needed in theoretical treatments of 2p2h reactions.

Acknowledgments

Arriving at the finish line of my PhD has not been easy, and yet here I am. Although the reader may find my name on the document. The reality is that I owe my success to a number of people.

First, I would like to thank Prof. W. Anthony Mann for taking me on as a graduate student and providing me an opportunity to do interesting physics to do with the NOvA experiment, and providing the necessary guidance in the pursuit of this analysis outcome. His attention to detail and deliberate, patient approach to problems is something to emulate. Coming out of my doctoral program, I must confess I am certainly better at documentation and writing. Besides, I have thoroughly enjoyed his classes on classical mechanics, quantum mechanics and particle physics.

Next, I would like to thank my fellow Tufts colleagues from the NOvA collaboration Travis Olson and Jeremy Wolcott. As someone who has avoided coding like the plague prior to graduate school, Travis and Jeremy have been indispensable in the early days of my PhD research. Both have patiently provided me with instructions and helped with debugging code. I don't think I would have been able to do much for NOvA if either were not around.

I must also extend my gratitude to colleagues from the NOvA experiment: Jon Paley and Linda Cremonesi. Jon and Linda have reviewed my work and have provided critical feedback on studies that have been performed in the pursuit of this Thesis. Both have brought in their knowledge from previous cross section analyses and my thesis has certainly benefited from their insights.

During my years at Tufts I have shared office space with Travis, Ozgur, Katie, Josh and Alec. I thank all of them for being great colleagues and for all the fun conversations we've had during our time together, from physics to video games, and many random interesting things.

Besides Tufts, I've had the pleasure of great friends who have made my journey easier and made living in Boston an exciting experience. I thank Mushfiq, Suaida, Afra, Ratib, Tanveer, Shamayla, Intekhab, Lamisa, Najneen, Abedin, Rayyan and Mehnaz for all the great times we've had together. I will dearly miss their presence when I move out of town.

I met my girlfriend Humaira back in 2019 and she has been an immense source of patience

throughout the years, especially during lockdown during the covid era. My thanks to her for all the support she has provided during the grueling last half of my PhD journey.

Finally, I would like to thank my family. Getting accepted to Tufts University would certainly not have been possible had my Mother not helped me with graduate school applications. I am indebted to her for getting accepted into a doctoral program and for the career I am about to embark on. My gratitude to my Father for taking care of the logistics of moving to the United States. Moving across the world to a new country is not easy, and my Father's organization has certainly helped the transition into graduate school much easier. My gratitude to both my parents for providing support, both moral and material, throughout the years. I must thank my sister, who is going to be a doctor soon, for being a constant source of support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Multi-Nucleon $\nu/\bar{\nu}$ Interactions: A New Phenomenon

Previous measurements of inclusive neutrino and antineutrino charged current (CC) cross sections have revealed an excess event rate that is not predicted by conventional neutrino-nucleus or antineutrino-nucleus phenomenology [1][2][3]. The extra event rate occurs in CC interaction topologies that are quasielastic-like, that is, the final states contain a $\mu^{\pm}$ plus nucleon(s) but are devoid of produced mesons. These events are believed to be caused by 2-particle-2-hole (2p2h) with meson exchange currents (MEC) interactions, in which a single incoming lepton scatters off a di-nucleon pair. The nucleons of the initial state di-nucleon are held together by the strong force, involving the exchange of virtual pions, heavier mesons [4][5], and possibly even gluons. Two-particle-two-hole interactions (hereafter 2p2h) contribute significantly to the $\overline{\nu}_{\mu}$-nucleus CC interactions at sub-GeV to few GeV beam energies [3]. Understanding the physics of these reactions is important for experiments that seek to determine the neutrino mass hierarchy and the size of CP violation in the neutrino sector using neutrino and antineutrino long baseline oscillations. That the measurement of oscillation parameters for neutrinos and antineutrinos is sensitive to the physics of 2p2h interactions and related nuclear physics phenomena is acknowledged in a recent publication by the T2K collaboration: "The poor (anti)neutrino-nucleus interaction modeling highlighted by this analysis is a limiting factor for the future neutrino oscillation ex-

periments that have as primary goal the measurement of the CP violation, calling for a deeper understanding of the underlying processes involved in (anti)neutrino-nucleus interactions and for new cross-section analyses with larger statistics and improved systematic uncertainties." [7]. In addition to facilitating neutrino oscillation measurements, a better understanding of 2p2h interaction may provide new perspectives concerning nucleon-nucleon interactions and its possible connection to the quark and gluon substructure of nuclei.

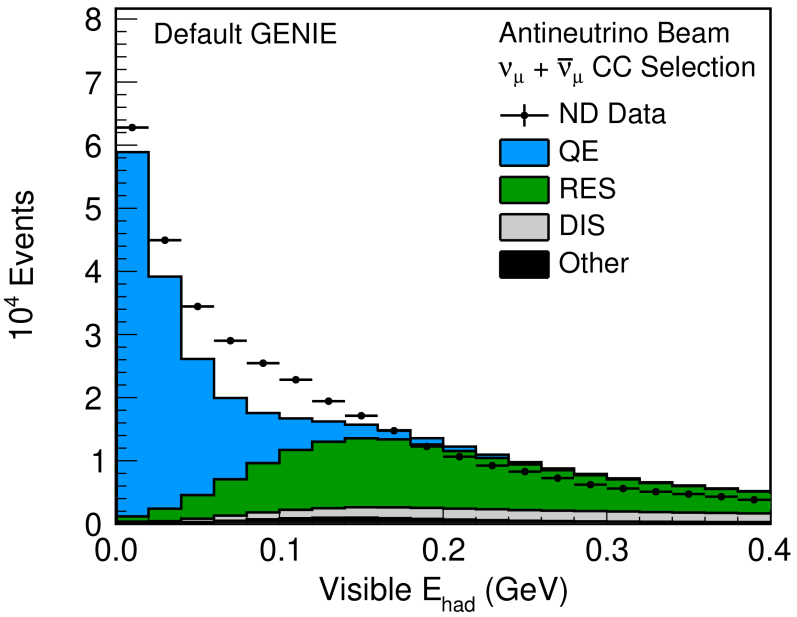

Figure 1.1: Distribution of visible hadronic energy in events of the NOvA ND data (solid circles) compared to the prediction (sum of histograms) based on estimated contributions from $\bar{\nu}_\mu$+single nucleon scattering in nuclei modeled as a Fermi gas. The total predicted event rate underpredicts the data, indicating the presence of reaction(s) that are not accounted for [6].

In the NOvA experiment, such an excess has been observed in the data, as shown in Fig. 1.1. The figure shows the distribution of CC events in visible hadronic energy observed in the NOvA Near Detector data. Quasi-elastic (QE) events (blue histogram in Fig. 1.1) have low visible hadronic energy, lying almost entirely below 200 MeV. This is because the only hadron that is produced from QE interactions is a neutron, which by itself produces zero ionization. Occasionally, the neutron rescatters off of other nuclei, resulting in visible hadronic energy. Baryon resonance production (RES) and subsequent resonance decays ($N^* \rightarrow N\pi$) gives events with final-state pions. Since charged pions give ionizing tracks and neutral pions decay into a pair of photons,

RES events have a higher visible energy. As indicated by the green histogram in Fig. 1.1, RES events tend to populate the region $E_{had} > 100$ MeV. Deep inelastic scattering (DIS) events are topologically similar to RES interactions, with the hadronic final state consisting of pions and nucleons. DIS events (grey histogram) occur at a relatively lower rate in NOvA data and have a flat event distribution across E_{had}. However, Fig. 1.1 shows a large data excess in the QE-region of $E_{had} \leq 200$ MeV, This unpredicted excess is believed to come from 2p2h events. Because of the substantial contribution of 2p2h events in the sample, understanding this process is important for neutrino oscillation experiments. Without knowledge of the 2p2h interaction cross section, the total event rate in the detectors cannot be predicted correctly, resulting in erroneous measurement of neutrino mixing angle and CP violation.

For muon antineutrino CC interactions, the incoming $\overline{\nu}_\mu$ changes into a μ^+ via the electroweak interaction, and a virtual W^- gauge boson is emitted. In $\bar{\nu}_\mu$ 2p2h-MEC interactions, the virtual W^- boson is believed to interact with either a neutron-proton (np) or a proton-proton (pp) pair, yielding a quasielastic-like final state that is constrained by electric charge and baryon number conservation. The two proposed 2p2h channels are listed in Eqs. (1.1) and (1.2). In both cases, an initial-state proton is converted to a final-state neutron. A Feynman-like schematic of the 2p2h process is given in Fig. 1.2.

$$\overline{\nu}_\mu + np \rightarrow \mu^+ + n + n \tag{1.1}$$

$$\overline{\nu}_\mu + pp \rightarrow \mu^+ + n + p \tag{1.2}$$

The two 2p2h channels defined above may not have equal strength. Rather, the larger contribution likely arises from the incoming antineutrino interacting with a np initial-state dinucleon, as in Eq. (1.1). This initial state is favored because the binding energy between an np pair is expected to be higher than that of a pp (or nn) dinucleon pair [8], as evidenced by the existence of the deuteron. Consequently, an antineutrino is more likely to interact with an np pair and give rise to a two-neutron hadronic system, a final-state configuration that yields almost zero scintillator light in the NOvA detector medium [9, 10].

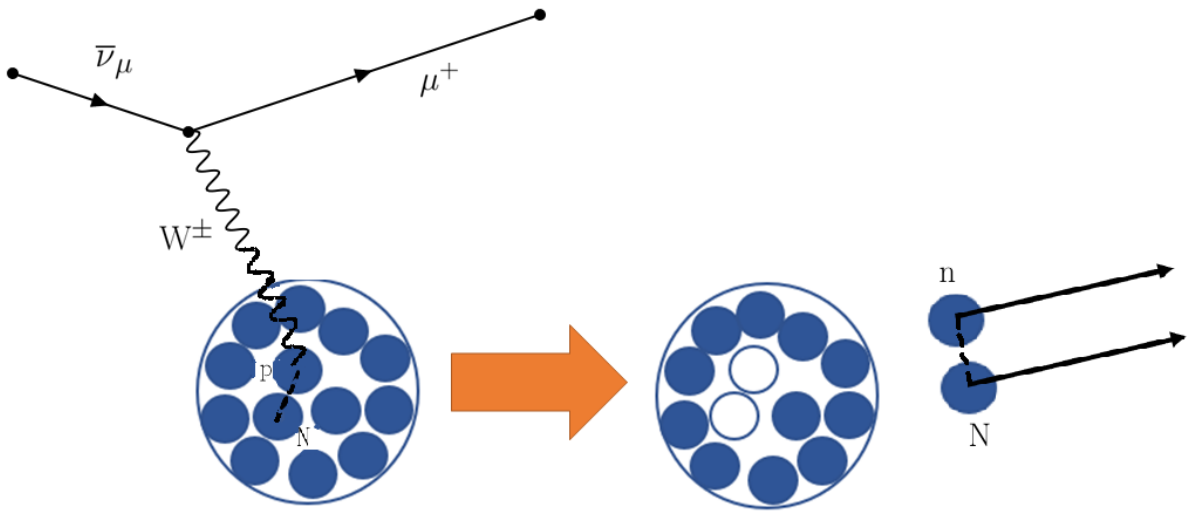

Figure 1.2: Schematic of the $\overline{\nu}_\mu$ CC 2p2h-MEC interaction. The incoming antineutrino interacts with a dinucleon pair in the nucleus, converting a/the proton in the pair, knocking out the two nucleons and leaving two "holes" in the Fermi sea of the struck nucleus.

As is elaborated in Chapter 2, several theoretical models have been proposed to explain 2p2h-MEC interactions. While the models have been partially successful in describing neutrino cross-section measurements, there are shortcomings [8]. Currently the only experiment to have reported a detailed measurement of antineutrino 2p2h-MEC reaction is the MINERvA experiment [2]. The measurement undertaken in the work reported here is carried out using data recorded by the NOvA Near Detector. The analysis presented here probes the physics of 2p2h-MEC interactions at lower incident neutrino-energies and with much higher event statistics than the MINERvA study.

This Thesis proceeds in two stages. In the first stage, CC events in the NOvA Near Detector are isolated and the double differential cross section is measured using the variables three-momentum transfer, $|\vec{q}|$, and available hadronic energy, E_{avail}. In the second stage, the 2p2h contribution to the $\bar{\nu}_\mu$ CC inclusive sample is isolated and the double differential cross section for 2p2h is measured using the same $|\vec{q}|$ and E_{avail} variables. The double and single differential cross sections obtained for the CC inclusive sample and for the 2p2h sample are compared to predictions based on current data tunes and theory models for the 2p2h-MEC process.

Chapter 2

Observations and Theory for $\bar{\nu}_\mu$ 2p2h Reactions

2.1 Measurement of $\bar{\nu}_\mu$ 2p2h in MINERvA

An excess in the neutrino inclusive cross section was first reported by the MiniBooNE experiment in 2010 [1]. The MINERvA experiment, which is a high-resolution cross-section experiment, was the first to scrutinize CC antineutrino interactions and obtain a characterization of the 2p2h excess [3]. Results from the MINERvA cross-section measurement are shown in Fig. 2.1. Model predictions are superimposed in the plots for the purpose of illuminating the physics. For the two top plots on the figure, the model prediction consists of CCQE interactions built from the Llwellyn-Smith formalism, $\Delta(1232)$ resonance cross sections based on the Rein-Sehgal model, Deep Inelastic Scattering (DIS) reactions and CC coherent reactions. Without a 2p2h model, MINERvA observes an apparent excess in the data in the region 20 - 200 MeV of E_{avail}. However, with the addition of a (tuned) 2p2h model and the application of an RPA correction to the CCQE model, the total inclusive cross section prediction gives a much better representation of the data. In a subsequent measurement of quasielastic $\bar{\nu}_\mu$+hydrocarbon scattering detailed in [11], the MINERvA experiment examined distributions in event vertex energy. The relevant plots are reproduced in Fig. 2.2. A comparison of the MC without 2p2h compared to the data reveals a

shortfall across the range of vertex energy values. However when a 2p2h component is included, a much better agreement between the data and the MC is obtained.

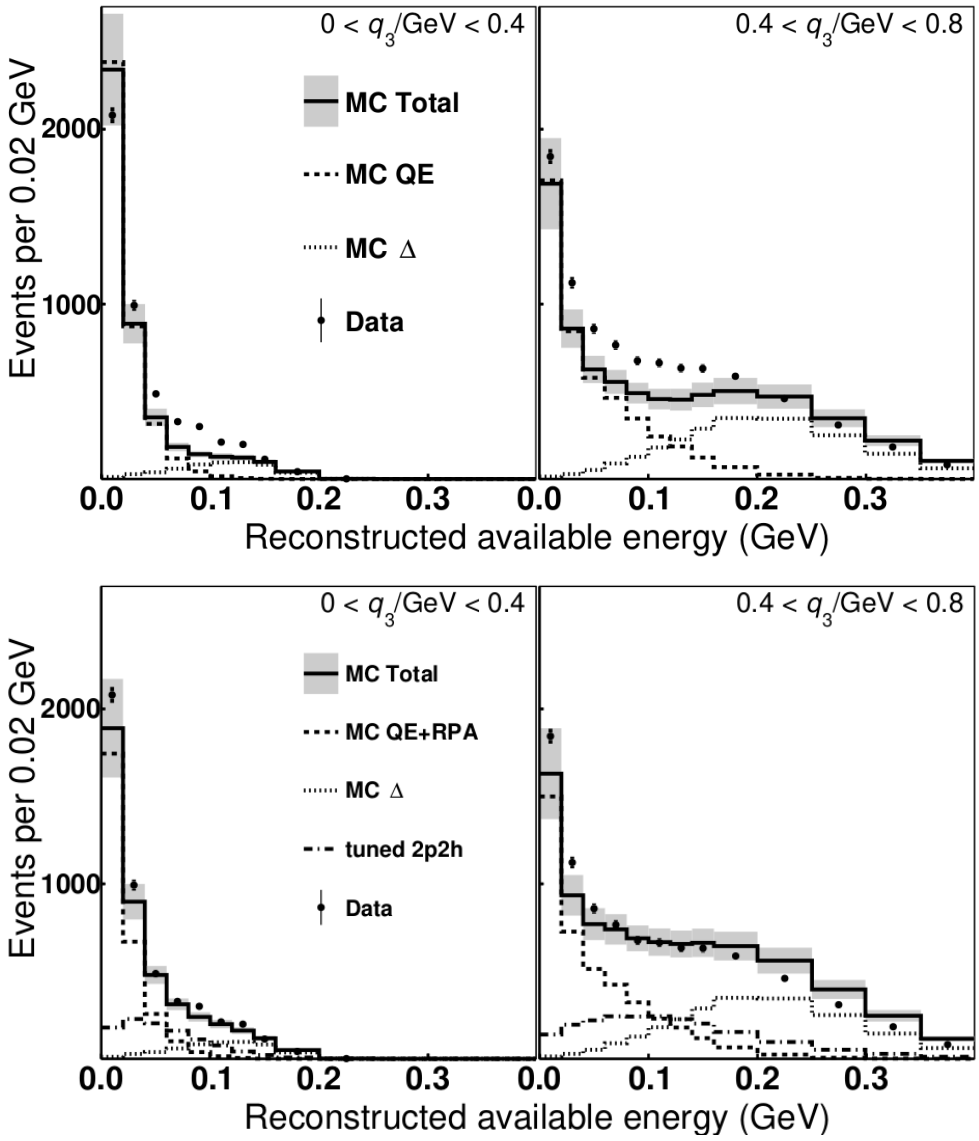

Figure 2.1: .
]Results from the inclusive $\bar{\nu}_\mu$-CC cross section measurement by the MINERvA experiment [3]]. Conventional antineutrino single-nucleon interactions do not predict the excess observed in the region 20 - 200 MeV of E_{avail} (upper plots). However with the addition of a data-tuned 2p2h model, an extra event rate is predicted which accounts for the apparent excess in the data (lower plot).

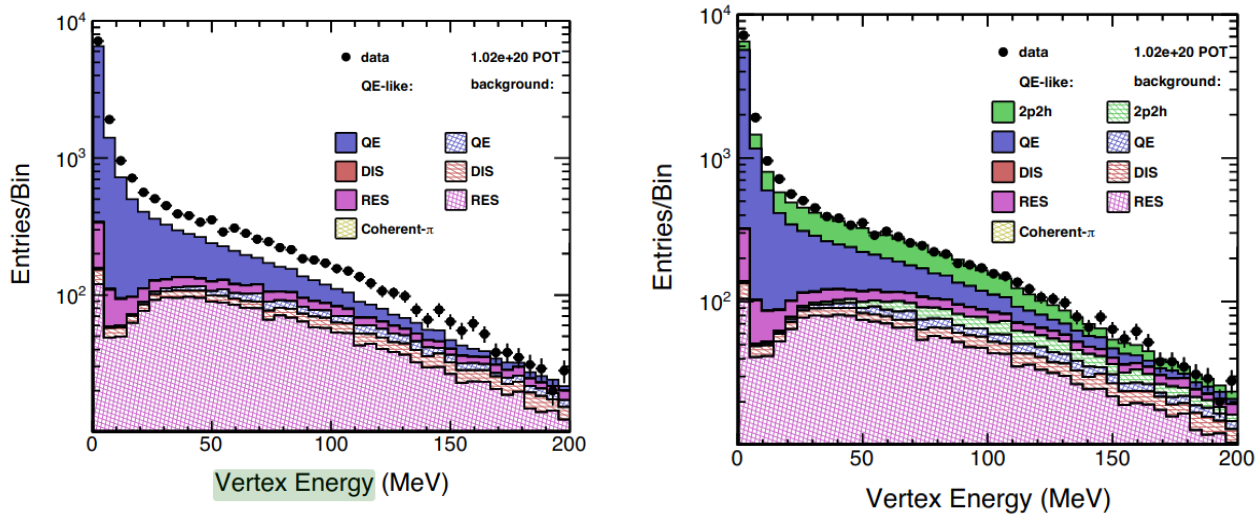

Figure 2.2: .
]Results from the MINERvA measurement reported in [11]]. The event distributions are plotted in terms of vertex energy. Since vertex energy is a measure of activity near the interaction vertex, it probes the energy that is carried by the final-state hadronic system. The excess in the data is apparent over the range of energies being displayed.

2.2 Models of 2p2h interactions with antineutrinos

There are a number of models in literature that attempt to describe muon antineutrino 2p2h-MEC interactions. For this analysis, the theory models considered are the Valencia 2p2h model [4, 12] and the Superscaling (SuSA-v2) 2p2h-MEC [5, 13] model. It is observed that both of the models tend to underpredict the observed data rates. Consequently experiments often tune (reweight) theory 2p2h models in order to better represent their data. The data tunes available for this analysis are the MINERvA tune [2, 3] and the NOvA tune.

Valencia 2p2h-MEC model

This model has been developed by theorists J. Nieves and collaborators at Valencia to predict the inclusive CC $\nu_\mu/\bar{\nu}_\mu$ scattering cross sections. In this model, the hadronic tensor for the amplitude calculations is derived by examining the possible virtual interactions of the W boson in nuclear matter. A perturbative expansion of the hadronic tensor gives a number of terms, with each term representing a possible way in which the W boson interacts with nuclear matter. The term

which is suggested to give the dominant amplitude involves the W boson creating a single proton hole in the nucleus and a virtual pion [14]. The virtual pion in turn creates a second nucleon hole. Upon applying a "Cutkosky cut" horizontally through the center of the diagram shown in Fig. 2.3, one infers that two nucleons have been converted into two holes.

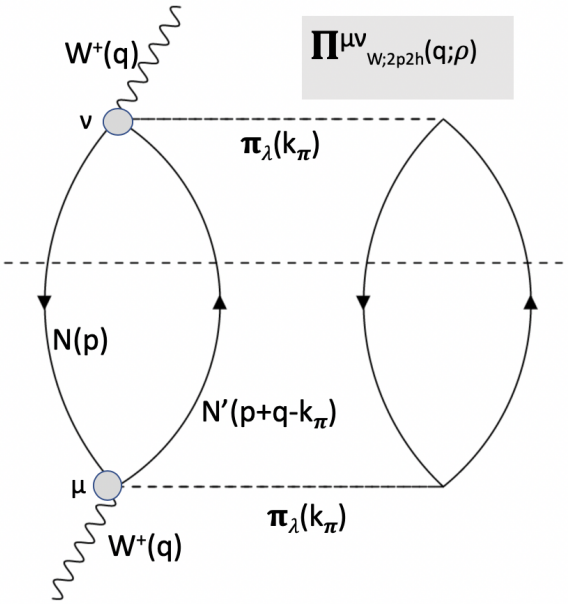

Figure 2.3: Diagram showing one of the ways in which the W boson interacts with the nuclear matter and contributes to the 2p2h excess. The W boson interacts with an nucleon and creates a hole in the nuclear medium. The process also creates a virtual pion which in turn dislodges another nucleon and creates a second hole.

The Valencia prediction of the double differential cross section $d\sigma/d|\vec{q}|dq_0$ for $\bar{\nu}_\mu$ is calculated in Ref. 21. The cross section for $\bar{\nu}_\mu$-carbon scattering for $E_{\bar{\nu}} = 3$ GeV is reproduced in Fig. 2.4. The Valencia cross section as predicted for the NOvA near detector, in which the neutrinos have the energies ranging from 0 - 5 GeV, with modal energy around 2 GeV, is shown in Fig. 2.5. The cross section is characterized by event production in two regions. There is a population of events around $q_0 \approx 0.5$ GeV with $|\vec{q}|$ from 0.3 to 0.6 GeV. There is a second enhancement in the distribution at q_0 of 0.15 GeV and $|\vec{q}|$ from 0.1 to 0.5 GeV. The model has a cutoff at $|\vec{q}| = 1.2$ GeV/c which is clearly visible in Fig. 2.5.

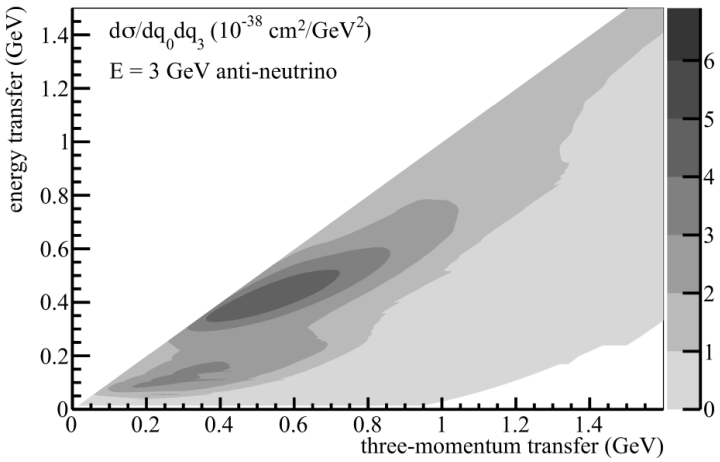

Figure 2.4: Muon antineutrino CC Valencia 2p2h cross section in truth variables $|\vec{q}|$ and q_0 for antineutrinos with fixed energy, $E_{\bar{\nu}} = 3$ GeV. This model predicts two, somewhat distinct, event populations.

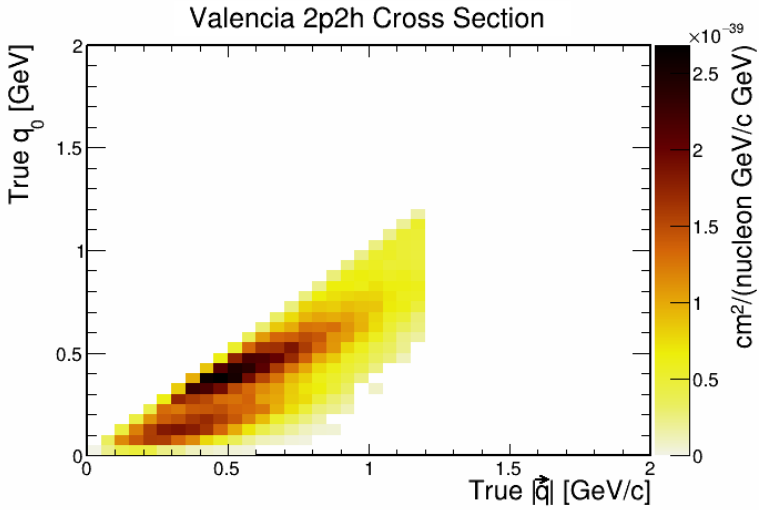

Figure 2.5: Muon antineutrino CC Valencia 2p2h cross section in truth variables $|\vec{q}|$ and q_0 for the NOvA ND beam flux. As in Fig. 2.4, the model predicts two somewhat separate event populations.

SuSA-v2 2p2h-MEC model

In this phenomenological model, a scaling function, $f(\psi)$, is developed as a function of a variable ψ, which itself is a function of q_0 and $|\vec{q}|$ [14]. Typically, an inclusive semileptonic CC cross section is a function of q_0 and $|\vec{q}|$. However, at high $|\vec{q}|$, it is observed in electron-nucleus scattering that the inclusive cross section in terms of ψ (and only in terms of ψ) is a product of the

electron plus nucleon cross section times the scaling function, $f(\psi)$. Previously, inclusive electron scattering data was examined with this superscaling approach, where the scaling function was extracted based on the longitudinal response of quasi-elastic electron-nucleus interactions. It was found that this scaling function, in conjunction with the electron-single nucleus interaction cross section, describes the electron-nucleus scattering data quite well. Assuming that the superscaling approximation applies to neutrino-nucleus interactions, the same scaling function can be used in conjunction with the neutrino quasi-elastic cross section to obtain a superscaled 2p2h-MEC model for neutrinos and antineutrinos [15].

Unlike the Valencia model, the reach of the SuSA model (Fig. 2.6) is not limited to 1.2 GeV in $|\vec{q}|$. This model predicts more 2p2h-MEC events compared to the Valencia 2p2h model. However, the events populate a single contiguous region with a smaller spread in q_0. The population of 2p2h events extends to $|\vec{q}|$ values beyond 1.2 GeV, however the cross section in this region is quite small.

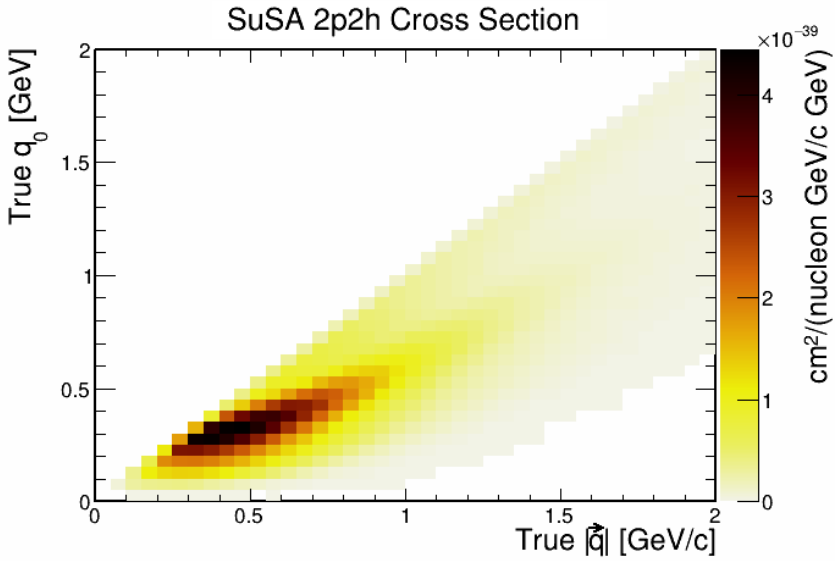

Figure 2.6: Muon antineutrino SuSA 2p2h-MEC cross section in "true" variables $|\vec{q}|$ and q_0. The model predicts more 2p2h-MEC events than does the Valencia model. However, unlike Valencia, there is only a single contiguous population of events for q_0 in the vicinity of 0.3 GeV.

A comparison of 2p2h cross sections as a function of (anti)neutrino energy for different theory models is provided in Ref. [21] and the relevant figure is reproduced in Fig. 12.11. The antineu-

trino cross sections are represented by the dashed distributions. The SuSA-v2 model predicts a higher cross section than that of the Valencia model, especially for $\nu_\mu/\bar{\nu}_\mu$ energies above 1 GeV.

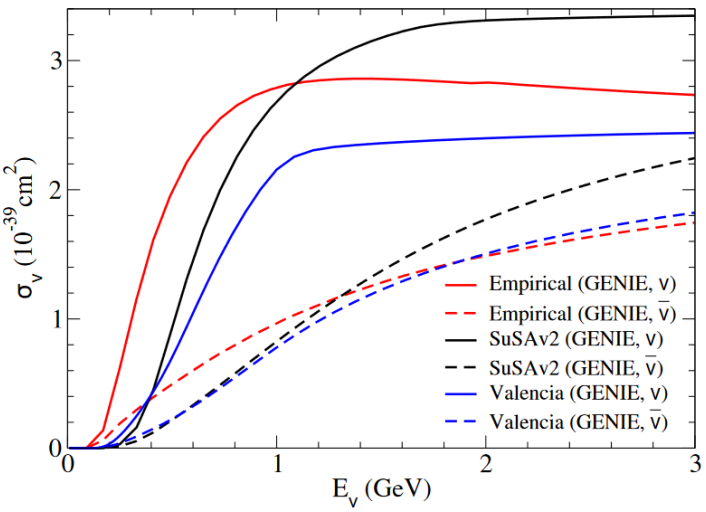

Figure 2.7: Cross sections of the GENIE empirical, SuSA-v2 and Valencia 2p2h-MEC models versus (anti)neutrino energy, E_ν. The extent to which the cross sections of the SuSA and GENIE empirical models exceed that of the Valencia model is more pronounced for neutrino interactions (solid-line cross sections) than for antineutrino interactions (dashed-line cross sections).

MINERvA 2p2h-MEC data tune

The MINERvA collaboration has reported an excess of observed $\overline{\nu}_\mu$ events when compared to the experiment's GENIE-based reference model [3]. An empirical 2p2h model was developed by augmenting the interaction rate of the base Valencia model. The augmentation doubles the strength of 2p2h between the $\Delta(1232)$ baryon resonance peak and the quasi-elastic peak [4].

The MINERvA 2p2h-MEC distributions (Figs. 2.8 and 2.10) are obtained by weighting the Valencia model. Consequently, these "tunes" retain the $|\vec{q}| = 1.2$ GeV cutoff. The significant enhancement in the event rate is apparent when the MINERvA distributions are compared to the Valencia 2p2h-MEC cross section. Unlike the Valencia model, the MINERvA tune has a single, narrow peak centered at $q_0 = 0.25$ GeV with momentum transfers concentrated in the interval from 0.3 GeV to 0.6 GeV. The event distribution is also significantly changed, with the two-peak structure of the model replaced by a single, concentrated population of events.

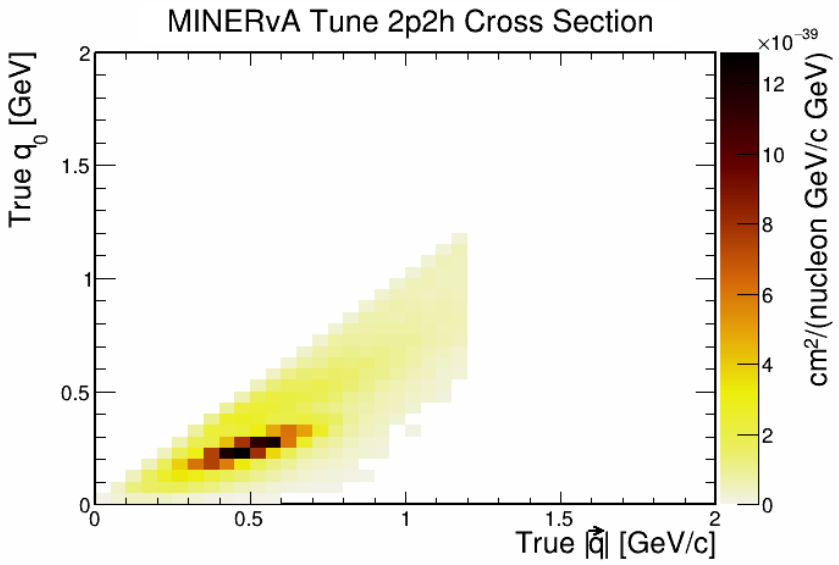

Figure 2.8: Differential cross section of the MINERvA tune of the Valencia 2p2h-MEC model shown in Fig. 2.5. In order to match MINERvA data, the interaction rate of the Valencia model has been enhanced by a factor of 1.5.

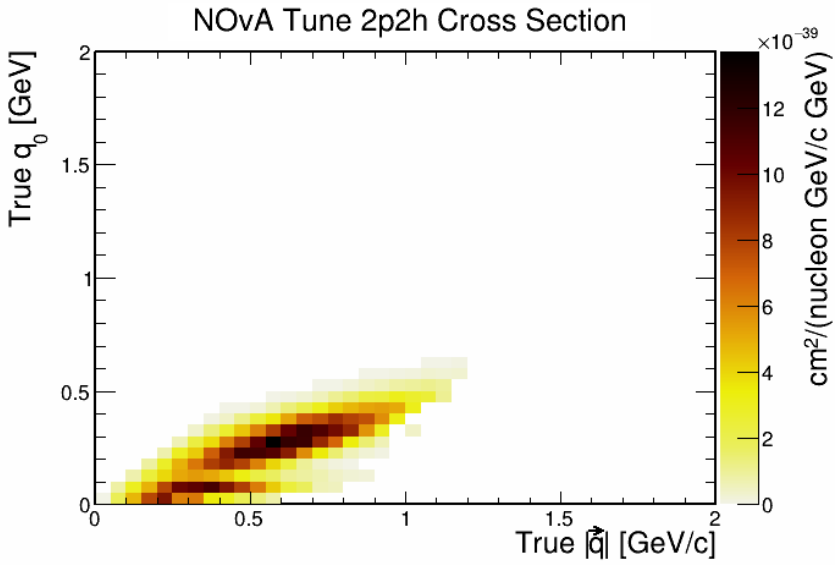

Figure 2.9: Differential cross section for the NOvA tune of the Valencia 2p2h cross section shown in Fig. 2.5. This data tune also has two event populations, however their locations differ somewhat from those predicted by Valencia.

NOvA 2p2h-MEC data tune

The NOvA experiment has its own data tune to the Valencia 2p2h model [16], shown in Fig. 2.9.

The tuning results in a significant enhancement of the event populations predicted by the Valencia

model, namely the event rate predicted by the NOvA tune is nearly six times that predicted by Valencia. However, the phase space coverage of the tune is less than that of the base model, especially in q_0 where no observable cross section is predicted above 0.7 GeV.

MINERvA tunes for np and pp initial states

In addition to the default enhancement, the MINERvA experiment developed two alternate tunes. In one tune, only the np initial-state nucleon-pair events, which transform into final-state nn state events, receive a rate enhancement. In the other tune, the pp initial-state events, which transform into final-state np pair events, are produced at an enhanced rate. These latter tunes were developed in order to assign a systematic uncertainty for the final-state dinucleon content. The two tunes are displayed in Fig. 2.10. The MINERvA np tune involves only enhancing the np $\rightarrow$ nn component of the 2p2h interactions. Likewise, the MINERvA pp tune involves only enhancing the pp $\rightarrow$ np component. Compared to the default tune shown in Fig. 2.8, these alternate tunes have smaller cross sections overall. However, they have larger spreads in q_0.

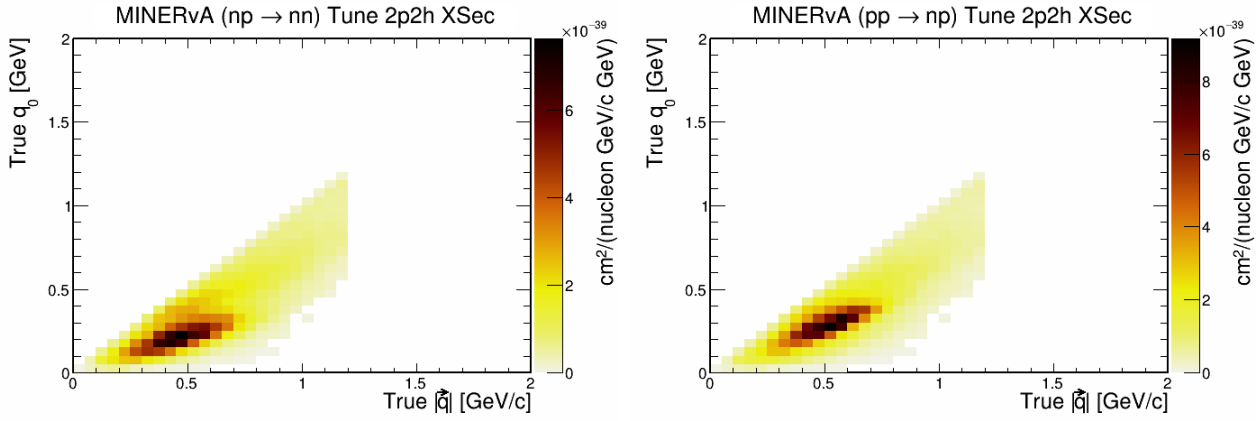

Figure 2.10: The MINERvA np-only tune (left) and pp-only tune (right) based on the Valencia model. For the np-only tune, only the subset of interactions wherein the incoming $\bar{\nu}_\mu$ interacts with a np dinucleon is enhanced. For the pp-only tune, only the subset of interactions where the $\bar{\nu}_\mu$ interacts with a pp dinucleon is enhanced.

2.3 Measurements of this Thesis

The antineutrino event spectrum at the NOvA Near Detector has a relatively narrow $E_{\bar{\nu}}$ distribution that peaks at 2 GeV. This peak is below the $E_{\bar{\nu}}$ peak for the event sample examined by MINERvA, which is at 3 GeV. On the other hand, the NOvA data is in an $E_{\bar{\nu}}$ range that lies above the sub-GeV interactions studied by the T2K experiment. The NOvA measurement undertaken by this Thesis has hadronic energy resolutions that are comparable but not equal to those obtained by MINERvA, however the NOvA data sample has higher statistics. The measurement presented in Ref. [11] uses an exposure of 1.02×10^{20} Protons-on-Target (POT), whereas the data exposure for this analysis is 12.5×10^{20} POT. Furthermore, the theory models have been improved since the publication of the MINERvA measurements. Thus the measurements of this Thesis shed new light on the physics of anti-neutrino induced 2p2h-MEC interactions.

Chapter 3

The NOvA experiment

The NuMI Off-Axis ν_e Appearance (NOvA) experiment is a long-baseline experiment designed to study neutrino and antineutrino oscillations. Specifically, it is designed to measure the rate of electron neutrino/antineutrino ($\nu_e/\bar{\nu}_e$) appearance ($\nu_\mu \to \nu_e$) and of muon neutrino/antineutrino ($\nu_\mu/\bar{\nu}_\mu$) disappearance ($\nu_\mu \to \nu_\mu$). The experiment consists of two detectors, designated the Near Detector and the Far Detector. Two detectors are required to monitor the time evolution and oscillation of the neutrino flavor states. The detectors are constructed of individual modules, which are comprised of PVC-cells filled with scintillation liquid. The liquid enables the detection of the passage of charged particles. The Near Detector is placed 1 km away from the start of the NuMI beam. Its dimensions are 3.8m×3.8m×15.9m and it has a total mass of 300 tons. The NOvA Far Detector is located at a distance $L = 810$ km away from the Near Detector in Ash River, Minnesota. At this distance, the probabilities for ν_e flavor appearance and for ν_μ flavor disappearance, which depend on the ratio of the propagation length to the neutrino energy, L/E_ν, are maximal. The Far Detector mass is 14 kilotons and its dimensions are 15.9m×15.9m×54.9m. The locations of the two detectors on the surface of the Earth are shown in Fig. 3.1.

The Neutrino Main Injector (NuMI) provides the neutrinos or antineutrinos whose interactions and properties are to be studied. The NuMI beam can operate in two separate modes: The Forward Horn Current (FHC) mode provides a neutrino beam, and the Reverse Horn Current (RHC) mode provides an antineutrino beam.

Figure 3.1: Map showing the geographic locations of the two NOvA detectors. The Near Detector is in Batavia, Illinois, while The Far Detector is situated 810 km away at Ash River, Minnesota. The long baseline neutrino beam enables the study of neutrino appearance and disappearance.

3.1 The neutrino main injector

The Fermilab Main Injector provides highly accelerated protons which can be used to initiate neutrino beams [17]. Figure 3.2 shows a diagram of the accelerator complex at Fermilab. A source provides hydrogen ions (protons), which are injected into the Booster. The Booster accelerates these protons to 8 GeV. The protons are then fed into a recycler ring where they are grouped into batches to obtain a focused, more energetic beam. The proton batches then enter the Main Injector where they are accelerated to 120 GeV. These highly energetic protons are then directed onto a carbon target, where the protons interact to form $\pi^{\pm}$ and $K^{\pm}$ mesons, which in turn decay to provide neutrinos and antineutrinos.

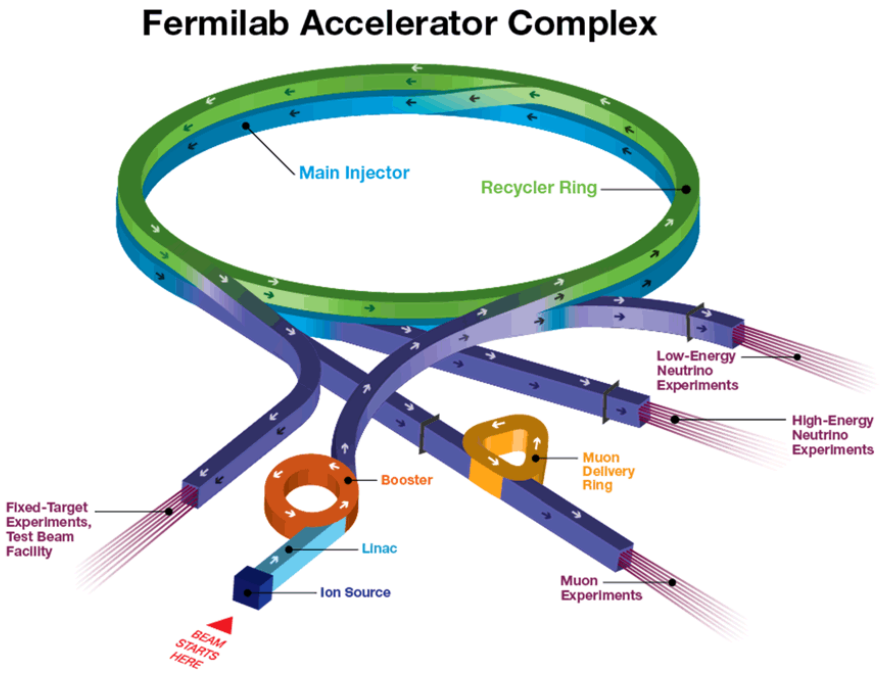

Figure 3.2: Diagram of the Fermilab accelerator complex. Protons are produced by ionizing hydrogen atoms. The ions are accelerated by the Booster (red ring) and then fed into the recycler ring (green ring), where the protons are bunched together. The bunched protons are then accelerated by the Main Injector (blue ring) to 120 GeV before being directed onto a carbon target for neutrino production (blue straight section).

3.2 NuMI neutrino/antineutrino beam

A diagram of the NuMI beam is shown in Fig. 3.3 [17]. There are six components: the target, two magnetic focusing horns, the decay pipe, the hadron monitor, the absorber and the muon monitors. The protons from the Main Injector are directed onto a graphite target in batches of approximately 4.8×10^{13} protons. The window of time during which a proton bunch is directed onto the target, known as the spill, is 10 μs. The time duration between consecutive spills is 1.3 seconds. Producing neutrinos in spills enables timing of the neutrino signals, which is highly useful for reducing cosmic ray backgrounds which are usually out-of-time with respect to spills. The proton bunches interact with the graphite target, which is 1.2 m in length, to create pions

and kaons. These pions and kaons are then focused by parabolic shaped magnetic horns into an intense beam. There are two horns, with the second horn focusing charged particles that are poorly focused by the first horn. The polarity of the horn electric current can be adjusted based on whether the beam required is a neutrino or an antineutrino beam.

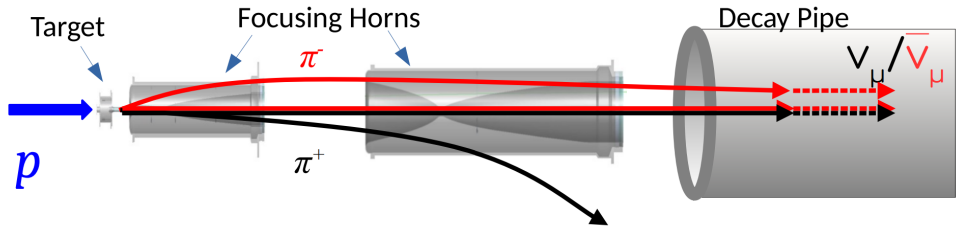

Figure 3.3: Diagram of the NuMI neutrino beam. Accelerated protons from the Main Injector interact with graphite to produced charged mesons. Depending on whether neutrinos or antineutrinos are required, the horns focus either the positively charged or the negatively charged mesons.

The focused π^- mesons (for antineutrino production) are then directed into a 675 m long decay pipe filled with helium. Inside the decay pipe, the charged mesons decay via the following modes:

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \tag{3.1}$$

$$K^- \rightarrow \mu^- + \bar{\nu}_\mu \tag{3.2}$$

Once past the decay pipe, the remnant beam traverses the hadron monitor, the absorber and the muon monitors. The beam is then directed into 240 m of rock. The purpose of passing the beam through the absorber and the rock is to remove the remaining hadrons and muons in the beam, leaving only neutrinos and antineutrinos. The energy spectrum of different neutrino flavors in the RHC mode of operation is shown in Fig. 3.4. The beam consists mostly of antineutrinos, with an energy spectrum that peaks at 2 GeV. The next largest flux in the RHC beam consists of muon neutrinos. This small ν_μ flux is present because not all π^+ particles are sufficiently deflected by the horns. There also exists a very small contribution of ν_e and $\bar{\nu}_e$ in the beam.

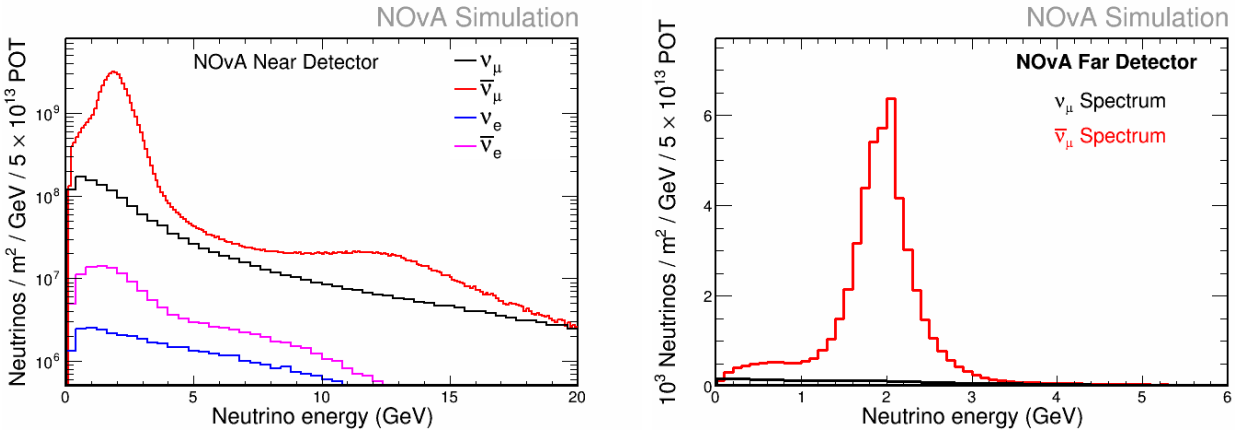

Figure 3.4: Energy distributions of the neutrino and antineutrino spectra in the RHC beam. The leading contribution is the $\bar{\nu}_\mu$ component, with a smaller contribution from ν_μ arising from defocused π^+ mesons. There is also a very small amount of ν_e and $\bar{\nu}_e$ contamination in the beam.

3.3 Off-axis neutrinos

The axis of the NOvA detectors is 14 mrad from the direction of the NuMI beam. By placing the detectors off-axis, the decay kinematics of the mesons is utilized to produce antineutrinos whose energy spectrum peaks sharply around 2 GeV. However, this positioning of the detectors results in a lower beam flux. The following equations specify the neutrino energy and flux at small angles :

$$E_{\bar{\nu}} = \frac{0.43 E_\pi}{1+\gamma^2\theta^2}, \tag{3.3}$$

$$\Phi = \left(\frac{2\gamma}{1+\gamma^2\theta^2}\right)\frac{A}{2\pi L}. \tag{3.4}$$

Here, $E_{\bar{\nu}}$ is the antineutrino energy, E_π is the energy of the π^- meson, γ is the ratio, E_π/m_π, of the meson total energy to its rest mass, θ is the angle between the detector and the NuMI beam axis, and Φ is the (anti)neutrino beam flux. The cross-sectional area of the detector is A, and the distance from the beam source to the detector is L. Figure 3.5 shows how the flux and the antineutrino energies are related to the pion energy at different off-axis angles. Figure 3.7 shows the Far Detector flux spectra for different off-axis angles. To get a relatively high statistics

sample of events in which the peak neutrino energy is 2 GeV, a 14 mrad off-axis orientation is required.

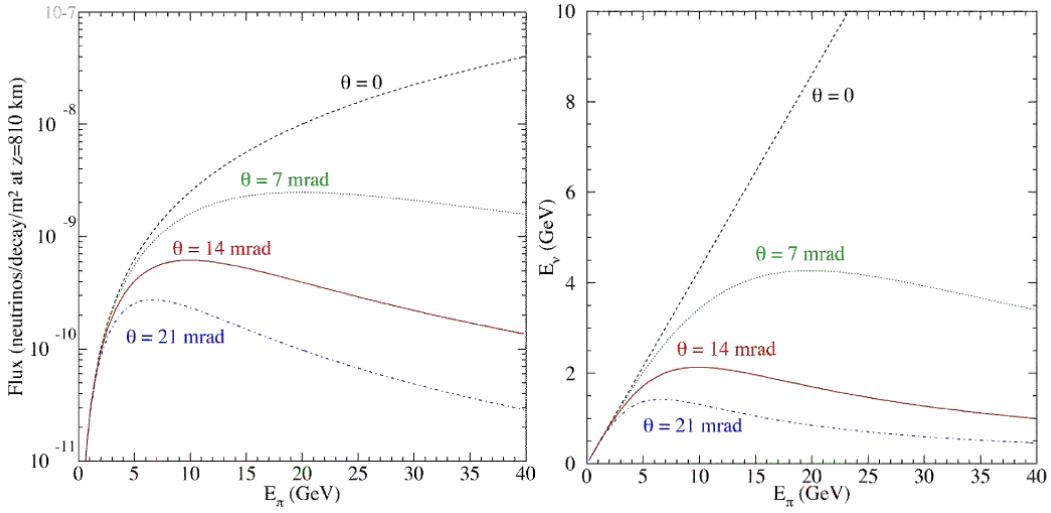

Figure 3.5: The energy of the decaying meson, E_π, affects the neutrino flux (left) and the neutrino energy (right) at different off-angle directions.

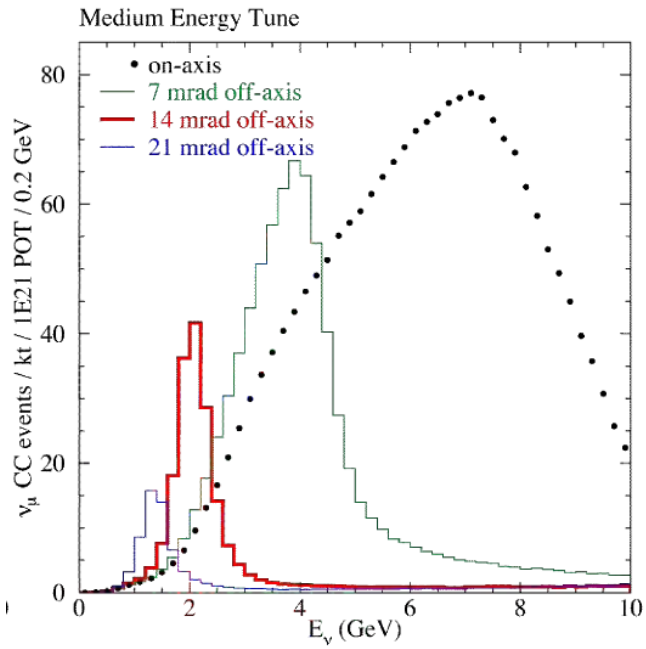

Figure 3.6: Distributions of neutrino energy predicted for different off-axis angles. When the detectors are aligned with the NuMI beam, there is a broad distribution of neutrino energies, peaking at $\sim$8 GeV. With increasing off-axis angle, the mean of neutrino energy values becomes smaller and the distribution becomes narrower. At 14 mrad, the distribution peaks at 2 GeV of neutrino energy.

3.4 The NOvA detectors

As previously stated, The NOvA experiment has two detectors, a Near Detector and a much larger Far Detector. The Near and Far Detectors are built from modules which are functionally identical, however the volume of the Far Detector is 75 times that of the Near Detector. The relative sizes of the two detectors is indicated by the sketch in Fig. 3.7.

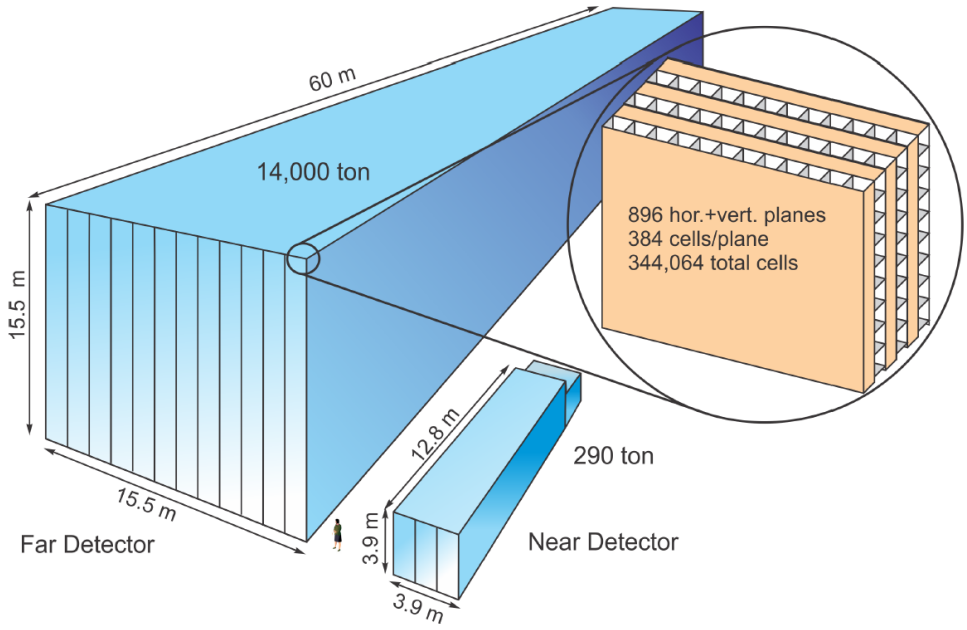

Figure 3.7: The two detectors of the NOvA experiment. The Far Detector has a cross sectional area that is 15 times larger and a length 5 times longer than that of the Near Detector. The detectors are constructed from modules of identical design.

3.4.1 Detector modules

Detector modules are assembled as a collection of planes. The planes are built from collections of cells which are filled with liquid scintillator. Each cell contains a wavelength shifting fiber and is read out by an avalanche photodiode (APD). The resulting electronic signals are then used to reconstruct events occurring in the detectors. In both detectors, a section of 16 cells are attached side by side, as shown in Fig, 3.8. Two such sections are stacked together. An end plate, side seal manifold cover, and an electronics bow are attached to this two-section extrusion. Together they form a detector module as shown in Fig. 3.9. The manifold cover contains the wiring from the individual cells to the APD, which converts light signals to electronic signals.

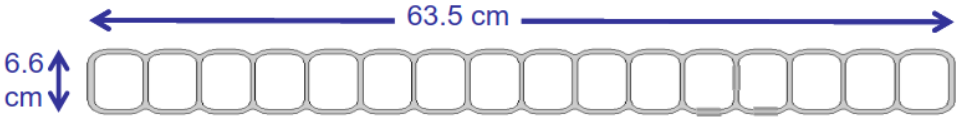

Figure 3.8: A collection of sixteen cells placed side-by-side. Two such sections are glued together to form a module.

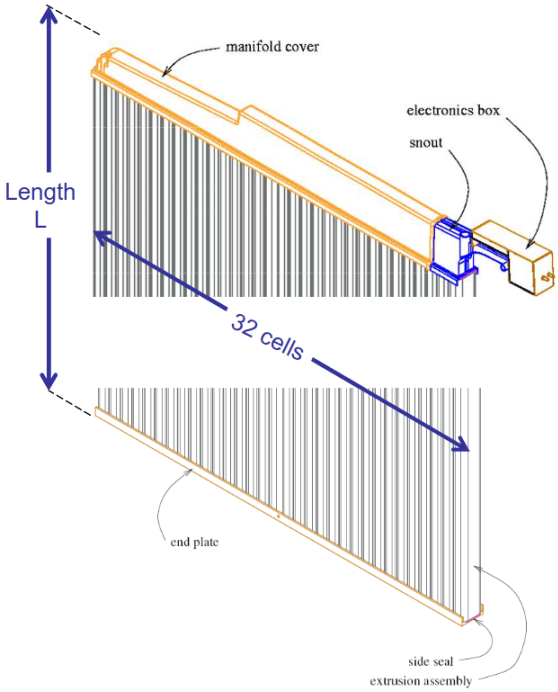

Figure 3.9: Diagram of a detector module. The module consists of 32 cells with the wiring from the cells being directed to a single avalanche photodiode in the electronics box.

The module lengths are 3.9 m and 16.2 m for the Near and Far Detectors respectively. In order for each detector to have a square cross section, a collection of 3 (12) modules are glued together side-by-side to form detector planes for the Near (Far) Detector. These are alternately placed vertically and horizontally behind one another. This alternating orientation of planes enables 3-dimensional reconstruction of the charged particles passing through the detector. A collection of 64 planes is referred to as a diblock. The Near and Far Detectors are made up of 3 and 14 diblocks respectively.

3.4.2 Detector module components

PVC cells

The building blocks of the detector modules are hollow cells composed of a reflective dioxide rigid polyvinylchloride (PVC). The cross section of a cell has dimensions of 3.9 cm×6.6 cm. Each cell is filled with liquid scintillator and contains a wavelength shifting fiber.

Liquid scintillator

The cells of the NOvA detectors are filled with liquid scintillator. When charged particles pass through, the ionization of the scintillator liquid produces light with wavelength 360 - 390 nm, which is collected by the wavelength shifting fiber. The average ΔE value for a minimum ionizing particle (MIP), which is a particle with $\beta \approx 1$ that gives minimal ionization as it traverses the cell, is 12.9 MeV. The relation between ionization and particle speed is given by the Bethe-Bloch equation [20]. Roughly, 10% of the energy of the particle is lost in the walls. A schematic of the process is shown in Fig. 3.10. The liquid scintillator is predominantly composed mineral oil, with pseudocumene being the scintillant. PPO (2,5-diphenyloxazole) and bis-MSB(1,4-di(methylstyryl)benzene) are also present in small amounts in order to shift the wavelength of the light emitted to 400 - 450 nm [18]. The cell walls are coated with an anti-static agent to prevent charge buildup when the cells are being filled.

Wavelength-shifting fiber

Each cell contains a loop of wavelength shifting fiber [19] that is 33.5 m long. The fiber captures light of wavelengths between 400 - 450 nm and shifts it to 490 to 500 nm, wavelengths at which the APDs have a more efficient response. The fiber is composed of polystyrene and has trace amounts of R27 dye as a wave-shifter.

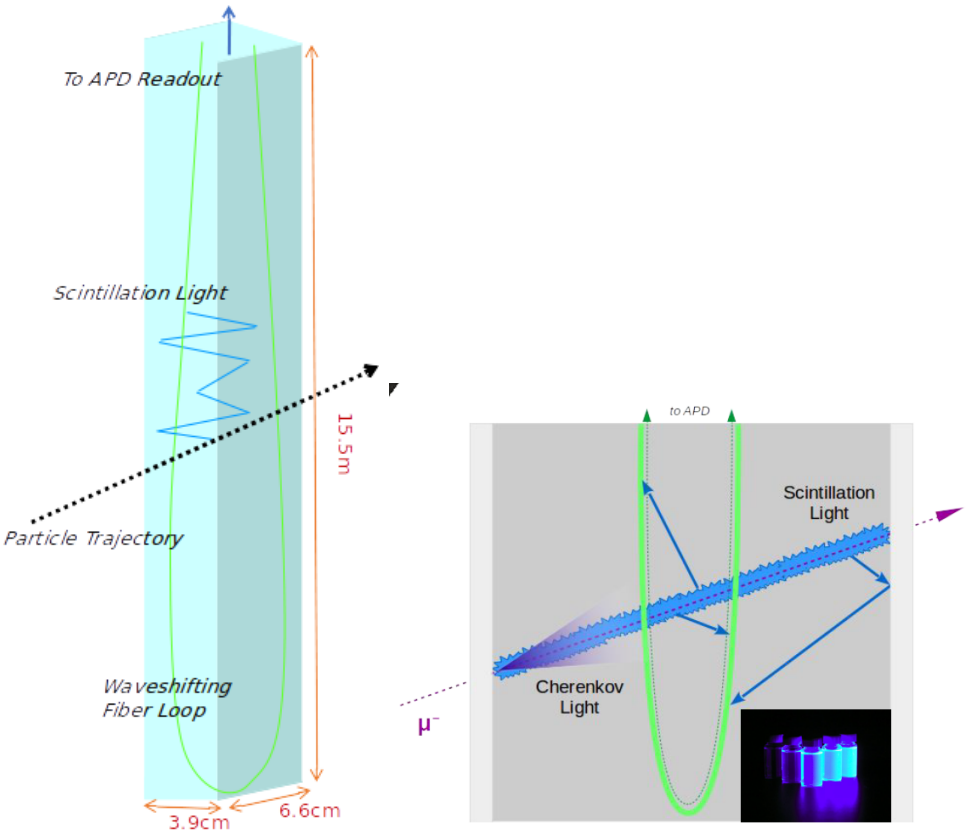

Figure 3.10: Scintillator light collection in the NOvA detectors. A charged particle passing through a cell causes the scintillator to emit light (left). This light is picked up by the wavelength-shifting fiber for processing by the photodiode (right).

3.4.2.1 Avalanche photodiode

Avalanche photodiodes(APDs) are used by the detector modules to convert light signals into electrical signals [19]. Figure 3.11 shows how a cell fiber connects to an APD. Both ends of the fiber are received into a single pixel/port. Each APD has 32 pixels and receives the signals from 32 cells. The APD used by NOvA has a quantum efficiency of 85% for the 500 - 520 nm light coming from the fibers. Light from the fibers excites electron-hole pairs. A strong electric field accelerates this pair. In the process, the primary electrons scatter on other electrons and eject them from the atoms of the APD material. This process cascades, resulting in an avalanche of electrons.

For optimal performance, the APD is cooled down to a temperature of -15^o Celsius. The cooling is done with a thermoelectric cooler. Cold water is constantly pumped in to and out of the

thermoelectric cooler to remove heat. A dry gas distribution system prevents humidity from condensing on the APDs. The operating voltage for the APD are 425 V and the current gain is 100. The signal from the APDs is fed into a Front End Board (FEB) circuit. The FEB circuit shapes the pulse so that it has a fast rise time and a slow fall time. An analog-to-digital converter digitizes the signal. The output from many FEBs is used for data acquisition (DAQ).

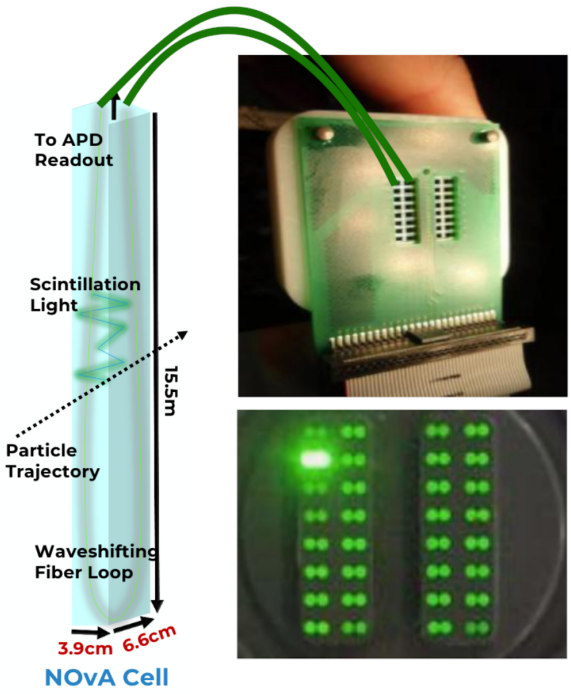

Figure 3.11: Connection of the wavelength-shifting fiber of a cell to an APD. Each APD has 32 ports (pixels) which collect the light signals coming from 32 cells of the detector module.

3.4.2.2 Data acquisition

The data from FEBs is collected by Data Concentrator Modules (DCM). Each DCM is connected to 64 FEBs, and it processes the data collected from a unique region in the detector. The Near and Far Detectors have 14 and 168 DCMs respectively. The DCM aggregates the data from the FEBs into batches, which are then used as inputs for trigger algorithms to decide whether a segment of data collected by the detectors of interest.

NOvA DAQ also includes a timing system. The Near Detector, because of the number of neutrinos that pass through it, often has events with tracks that overlap each other. The timing system separates hits in time and thus alleviates signal pile-up. Both detectors are also synchronized

to an external reference clock. The synchronization allows the NOvA DAQ to know when spill occurs, so that neutrino beam data can be recorded.

Most of the light signals collected by the NOvA detectors are not associated with events of interest. The various triggers installed on both detectors serve to identify the data of potential interest. There are two types of triggers, namely data-driven triggers and external triggers. Data-driven triggers identify events that satisfy pre-determined criteria, and then keep track of the targeted events. Externally driven triggers are activated when an outside party reports an interesting physics event. An example of the latter is the supernova trigger installed on each of the detectors.

3.4.3 Detector capabilities

A simulated event display from a beam spill into the Near Detector is shown in Fig. 3.12. The display shows two orthogonal views. The top view shows the X-Z projection, and the bottom view shows the Y-Z projection. The information from the two views can be combined to reconstruct the three-dimensional kinematics of the outgoing particles. The color depicts relative ionizations, with a 'hotter' color (the range from yellow to red) showing a higher energy deposition and a 'cooler' color representing low energy deposition (from green to cyan to dark blue). The beam spill is 1μs long, whereas the hits from a single event occur within a few nanoseconds. Because the time needed for all the hits to occur and be read out is much smaller than the duration of a beam spill, hits of an event can be grouped together in time. The energy due to various particles in an event can be reconstructed from the energy depositions left during traversal. The Near Detector is especially good at tracking and stopping muon tracks. These capabilities have been used extensively in the kinematic reconstructions required by the Thesis analysis.

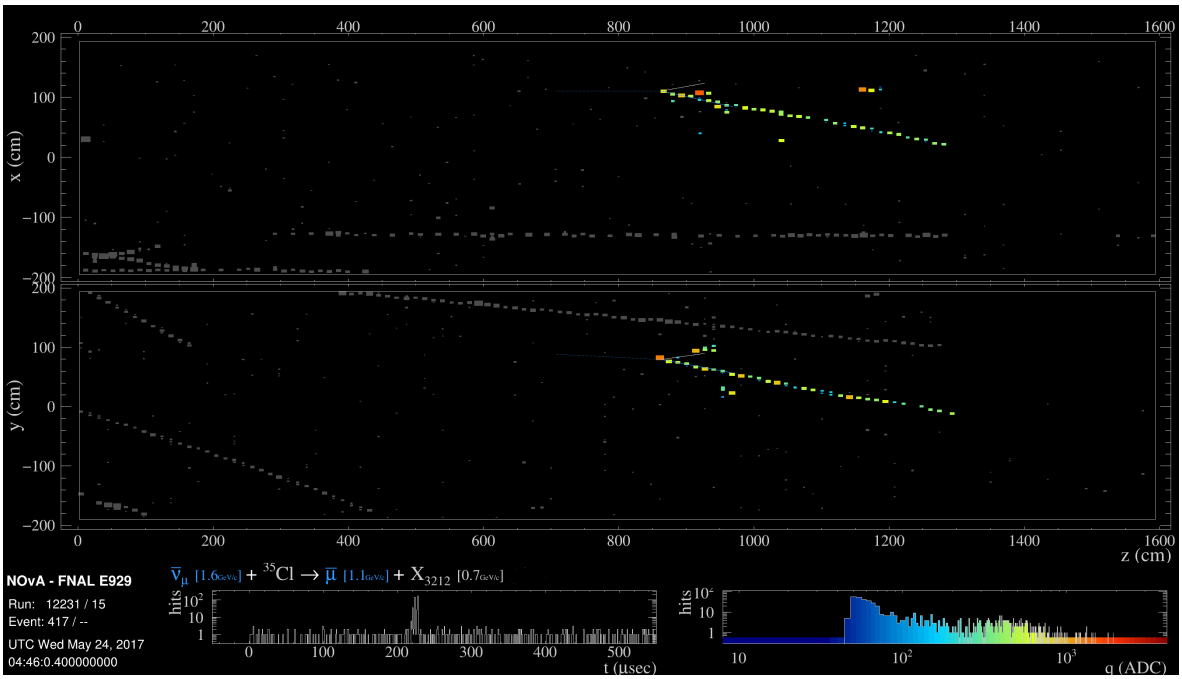

Figure 3.12: Simulated event display in the NOvA Near Detector. There are two views. The upper view shows the X versus Z projection while the lower view shows the Y-Z projection. The colored hits and track arise from a $\bar{\nu}_\mu$-CC event, with the track created by the outgoing (anti)muon. The remote clusters are caused by a Σ^0 hyperon that decays, producing photons which then convert into e^+e^- pairs: $\bar{\nu}_\mu(p) \to \mu^+ \Sigma^0$, $\Sigma^0 \to \Lambda^0 \gamma$, $\Lambda^0 \to n \pi^0$, $\pi^0 \to \gamma \gamma$.

Chapter 4

Data, Monte Carlo, and Analysis Variables

4.1 Data and Monte Carlo

For this Thesis, the fifth analysis NOvA ND dataset and Monte Carlo (MC) have been utilized. The RHC data has been collected between June 2016 and October 2018. The typical power in the NuMI beam ranged between 300 - 700 kW during the period.

For the simulated predictions, the event generator GENIE version v3.00.00 [22] has been used. The total MC exposure is 5.06×10^{21} protons-on-target.

4.1.1 Monte Carlo central value weights

The predictions made with GENIE event generator are weighted to reflect the current knowledge of neutrino interactions and the NOvA ND flux. The 2020 cross section weights, referred to as *kXSecCVWgt2020* in CAFAna [16], make two major changes to the base GENIE predictions. The first is the tuning of pion kinematic and reinteraction parameters. In the region of interest to NOvA, which is $T_\pi \leq 350$ MeV, the mean free path of the pions is reduced by 40%, the pion absorption is increased by 40%, charge exhange is reduced by 30% and rescattering likelihood is increased by 10%. The other major change that the 2020 tuning makes to the base GENIE prediction is the enhancement of the 2p2h-MEC interaction rate. The weighting scheme consists of two two-dimensional gaussians.

4.2 Analysis variables

The analysis is carried out in variables $|\vec{q}|$ and E_{avail}. Previously considered variables were muon energy E_μ, muon production angle (θ_μ), four-momentum transfer (Q^2) and invariant mass (W). The latter two variables did have enough information but were not very well reconstructed from observables and had poor resolution, or did not separate the 2p2h-MEC intractions from other interaction types in the phase space. Three-momentum transfer contains sufficient information on the kinematics of the outgoing muon.

Typically, the reconstructed hadronic energy is used to describe hadronic kinematics. This variable, unfortunately is too model-dependent to be useful for a cross section analysis. There is also the visible hadronic energy, which is too detector specific and not very useful to audiences outside the NOvA collaboration. Instead, the variable E_{avail} is used, which accurately describes the observed hadronic kinematics with little model dependence. The E_{avail} variable has been used by the MINERvA collaboration for cross section measurements [2].

4.2.1 Three-momentum transfer

The $|\vec{q}\,|$ variable characterizes how much three-momentum the leptonic system transfers to the hadronic part in the interaction. Calculating $|\vec{q}|$ requires information about the antineutrino energy, $E_{\bar{\nu}}$, the muon energy, E_{μ^+}, and the muon production angle, θ_{μ^+}, and m_{μ^+}. The relationship between the variables can be understood from the interaction displayed in Fig. 4.1.

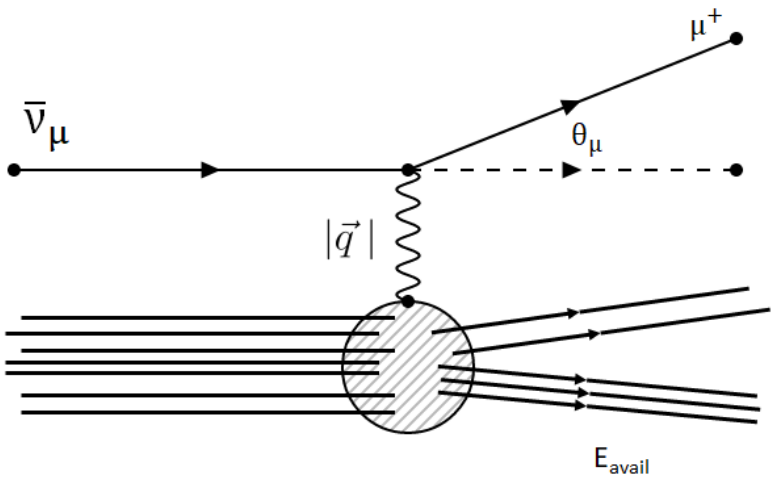

Figure 4.1: Feynman diagram showing a antineutrino - nucleus reaction. The diagram goes from left to right. The W-boson transfers three-momentum, $|\vec{q}|$ to the hadronic system. The E_{avail} variable represents how much of the outgoing hadronic energy is visible to the detector.

The neutrino energy of an event can be reconstructed by summing the reconstructed muon energy with the estimated hadronic energy of the system:

$$E_{\bar{\nu}_\mu} \;=\; E_{had} + E_{\mu^+} \tag{4.1}$$

Because charged current $\bar{\nu}_\mu$ events produce neutrons, the hadronic energy E_{had} must often be corrected to fully account for the energy due to the hadronic events. To obtain the value of $|\vec{q}|$, it is convenient to calculate the four-momentum transfer, Q^2, as follows:

$$Q^2 \;=\; 2E_{\bar{\nu}}(E_{\mu^+} - p_{\mu^+}\cos(\theta_{\mu^+})) - m^2_{\mu+} \tag{4.2}$$

Once Q^2 has been obtained, the $|\vec{q}|$ value can be calculated:

$$|\vec{q}| \;=\; \sqrt{Q^2 + (E_{\bar{\nu}} - E_{\mu^+})} \tag{4.3}$$

The relationship between true $|\vec{q}|$, which is the actual three-momentum transfer of events, and reconstructed $|\vec{q}|$, which is three-momentrum transfer calculated using E_{μ^+}, p_{μ^+}, $cos(\theta_{\mu^+})$ and E_{had}, is shown in Fig. 4.2.

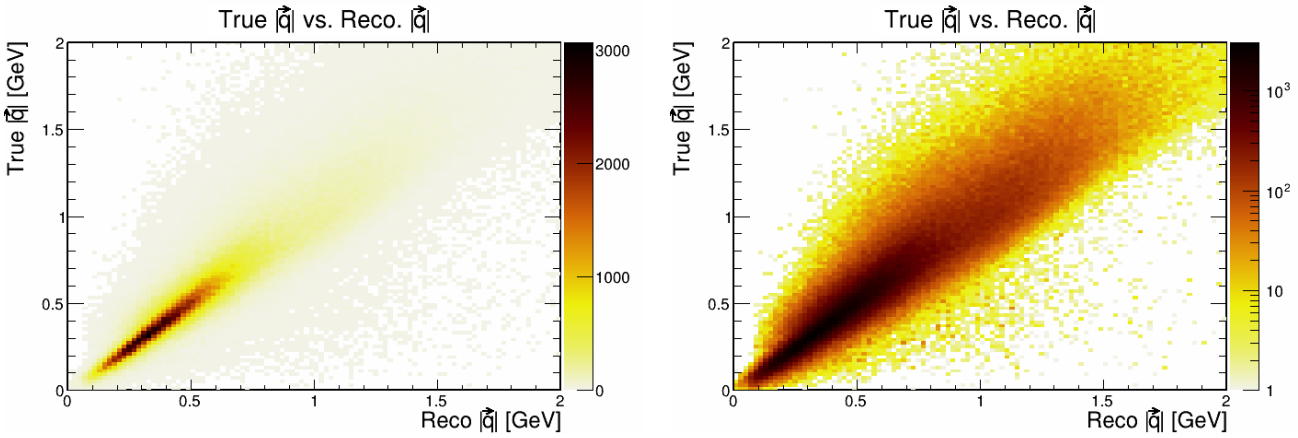

Figure 4.2: Distribution of true $|\vec{q}|$ vs. reconstructed $|\vec{q}|$ of RHC events in linear z-axis scale (left) and log z-axis scale (right).

The distribution of the mean true $|\vec{q}|$ value for each reco $|\vec{q}|$ bin is shown in Fig. 4.3. The reconstructed $|\vec{q}|$ has a mostly linear relationship with True E_{avail} up to 1.5 GeV, above which reco $|\vec{q}|$ tends to underestimate the actual $|\vec{q}|$.

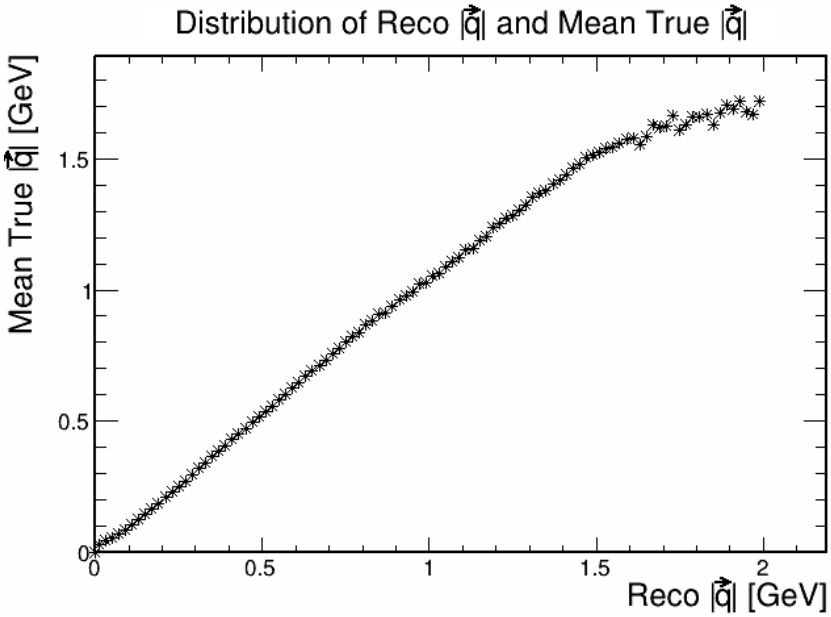

Figure 4.3: Distribution of reco $|\vec{q}|$ versus mean true $|\vec{q}|$ of events in the RHC CC inclusive sample for reco $|\vec{q}|$ of 0 - 2 GeV. The mean true $|\vec{q}$ has a linear relationship with reco $|\vec{q}|$ upto 1.4 GeV/c.

For this analysis, it is necessary to correct the reco $|\vec{q}|$ variable. The correction is done by fitting a quadratic function over the points shown in the above figure. With ROOT, the best fit quadratic function is

$$y = 0.007\,x^2 + 1.061\,x - 0.003. \tag{4.4}$$

The function is superimposed on the profile points in Fig. 4.3 is shown in Fig. 4.4.

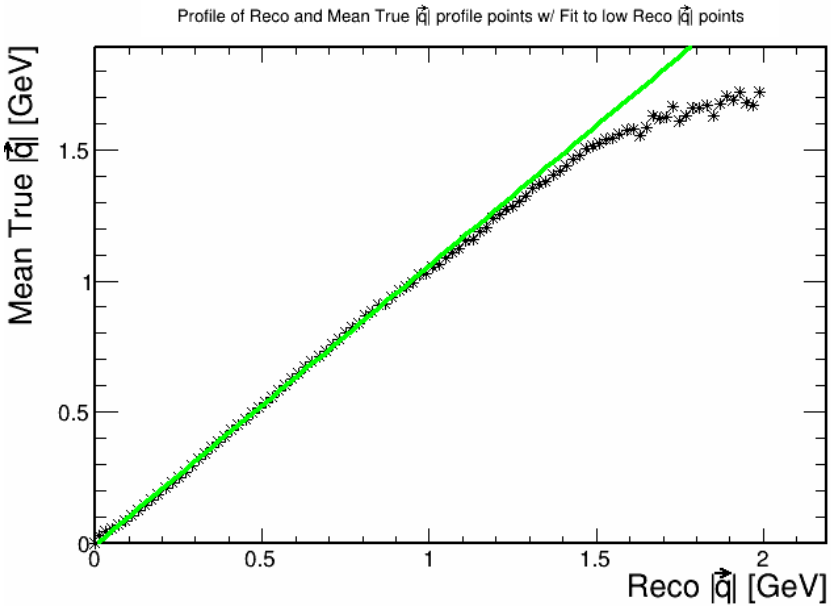

Figure 4.4: Reco $|\vec{q}|$ versus mean true $|\vec{q}|$ of events in the RHC CC inclusive sample. The profiling function shown in Eq. (4.4) is also superimposed. The profile accurately calculates the mean true $|\vec{q}|$ value upto 1.3 GeV/c, beyond which the values get overestimated.

This function corrects the reconstructed three-momentum transfer obtained from experimental observables. The expression then, for converting a reconstructed $|\vec{q}|$ to a corrected value is

$$Corrected\ Reco\ |\vec{q}| = 0.007(Reco\ |\vec{q}|)^2| + 1.061(Reco\ |\vec{q}|) - 0.003. \tag{4.5}$$

The mapping function superimposed on the reconstructed $|\vec{q}|$ vs. true $|\vec{q}|$ is shown in Fig. 4.5. This corrected reconstructed $|\vec{q}|$ variable is the one used throughtout the analysis.

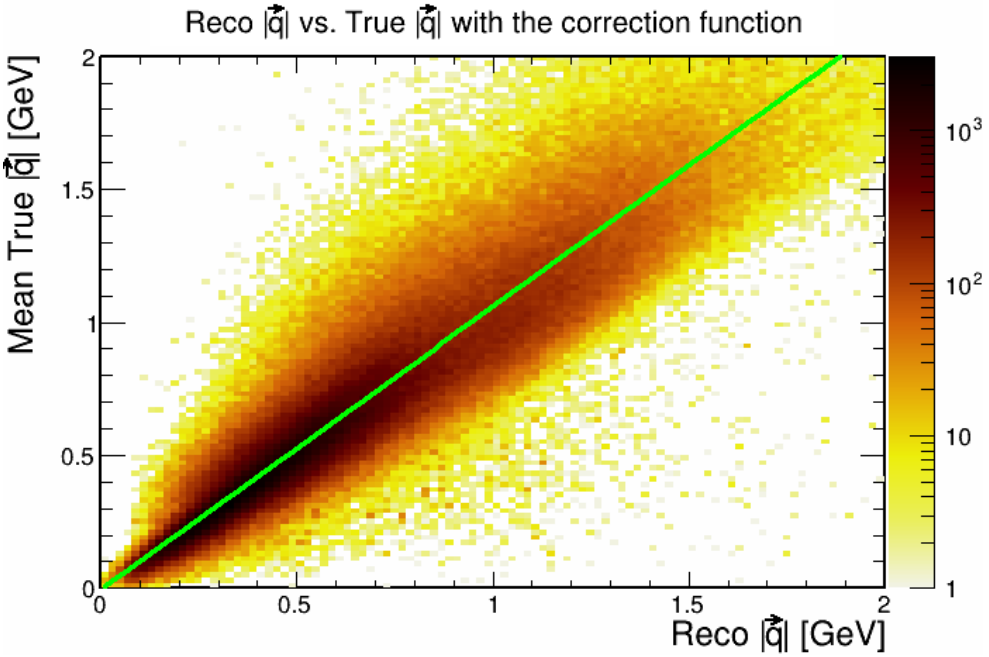

Figure 4.5: Event distribution in variables reco $|\vec{q}|$ versus true $|\vec{q}|$. The profile function lies in the middle of the population of events for each slice in reco $|\vec{q}|$. There is an even spread of events on both sides of the function.

Two useful quantities that are worth examining are the *absolute residual* and the *fractional residual*. These quantities help determine a suitable binning scheme for the analysis, which will be discussed in a later section. The aforementioned quantities for an event are defined in the following way:

$$Absolute\ residual\ =\ True\ |\vec{q}|\ -\ Reco\ |\vec{q}|, \tag{4.6}$$

$$Fractional\ residual\ =\ \frac{True\ |\vec{q}|\ -\ Reco\ |\vec{q}|}{True\ |\vec{q}|}. \tag{4.7}$$

The absolute and fractional residuals for a sample of $\bar{\nu}_\mu$ CC events are shown in Fig. 4.6. The distributions of both quantities are observed to be centered around a value of 0.0, showing that the reconstructed $|\vec{q}|$ is capable of correctly estimating the momentum transfer for events. The overall absolute resolution of the sample is the Full-Width Half Maximum of the absolute resolution distribution, which is 0.2 GeV.

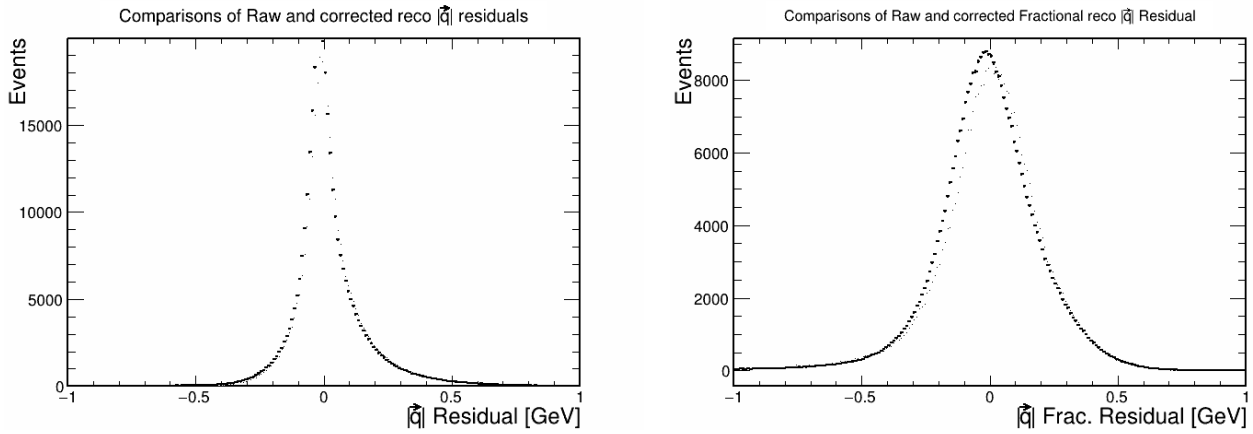

Figure 4.6: Absolute (left) and fractional residual (right) of the $|\vec{q}|$ for $\bar{\nu}_\mu$ CC events in the NOvA Near Detector.

The absolute and fractional residuals of $|\vec{q}|$ has also been examined across different ranges of $|\vec{q}|$: 0.0 GeV$\leq$ $|\vec{q}| \leq$ 0.5 GeV, 0.5 GeV $< |\vec{q}| \leq$ 1.0 GeV, 1.0 GeV $< |\vec{q}| \leq$ 1.5 GeV, 1.5 GeV$< |\vec{q}| \leq$ 2.0 GeV. The plots for these are shown in Figs. 4.7 and 4.8. It can be seen that at the higher ranges of three-momentum transfer, the absolute residual distribution becomes broader and the central value peak moves towards higher residual values. This bias indicates that the reconstructed $|\vec{q}|$ tends to underestimate the true $|\vec{q}|$ at higher values.

The biases at higher values of $|\vec{q}|$ can be understood from looking at Fig. 4.5. It can be seen that, from 1.5 GeV of reco $|\vec{q}|$ upwards, more of the events lie above the mapping function, where true $|\vec{q}|$ is greater than reco $|\vec{q}|$. There are not as many events below the mapping line. Such an imbalance of events results in the biases observed in figs. 4.7 and 4.8.

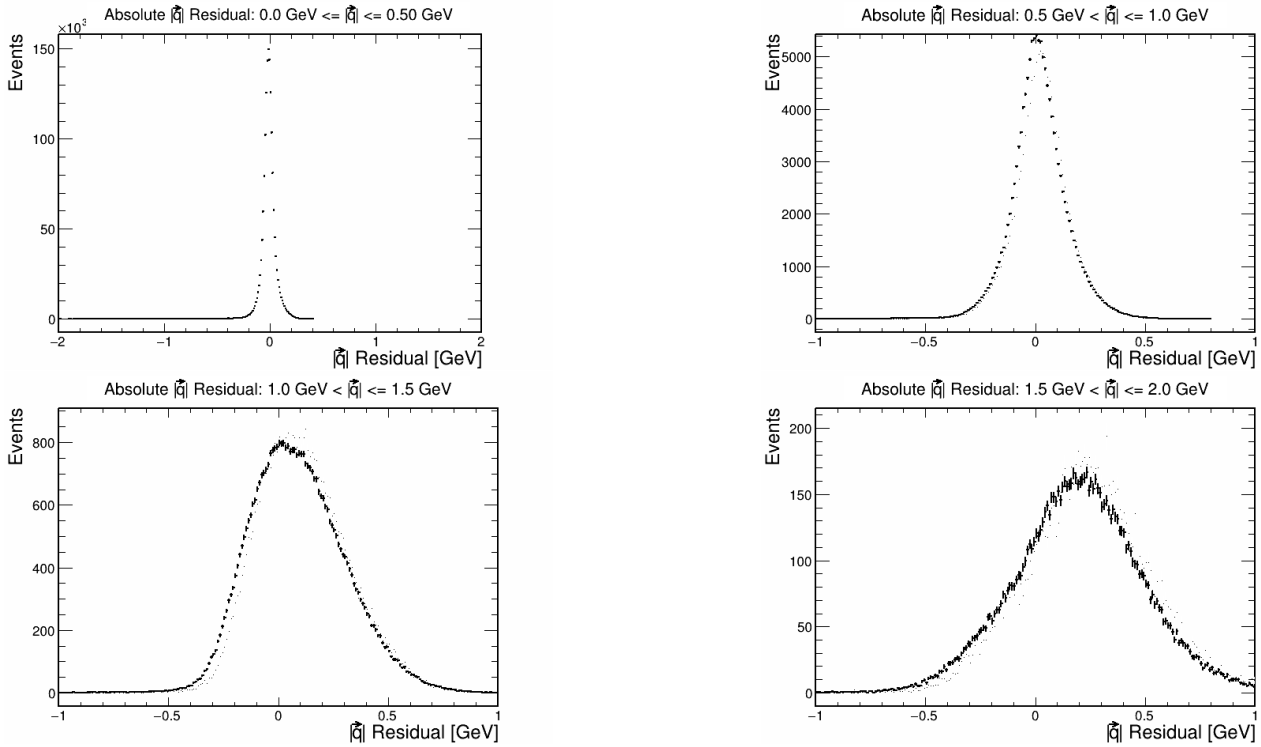

Figure 4.7: Absolute resolution of the $|\vec{q}|$ variable displayed for different ranges of $|\vec{q}|$. It can be seen that with increasing $|\vec{q}|$, the distributions become broader.

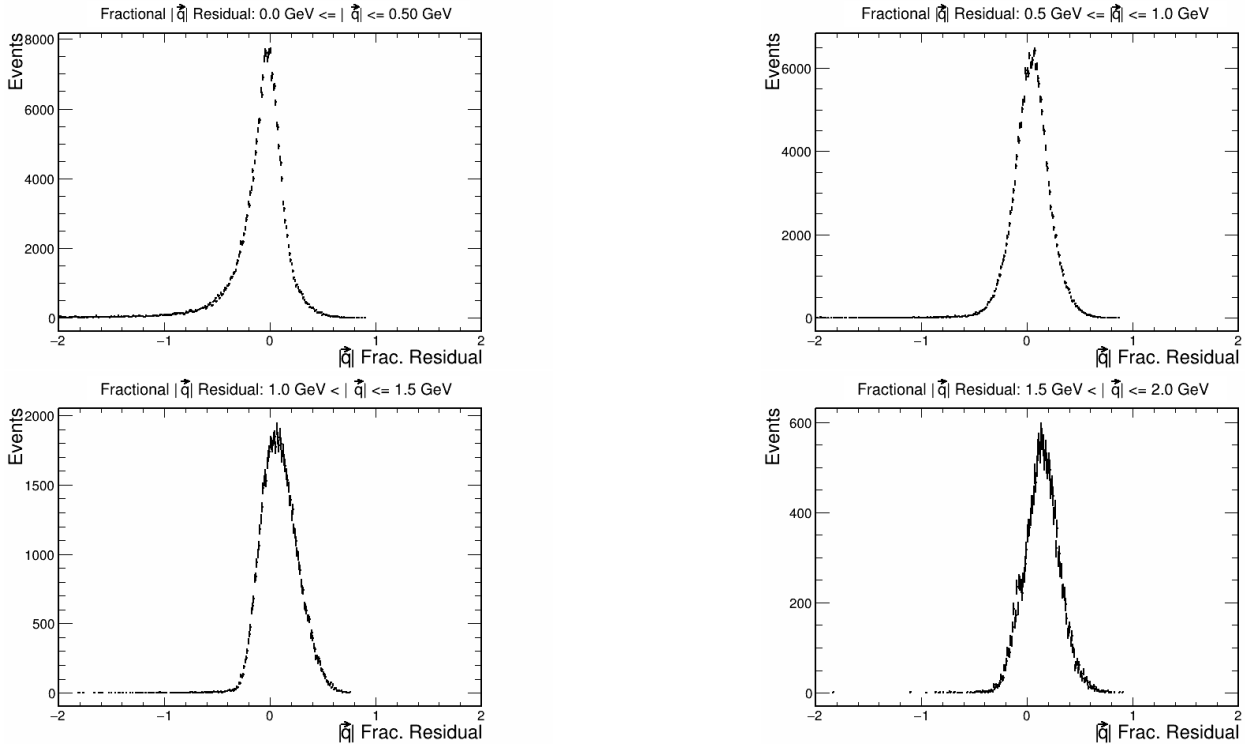

Figure 4.8: Fractional residual of the $|\vec{q}|$ variable for different ranges of $|\vec{q}|$.

4.2.2 Available hadronic energy

The available energy, E_{avail}, describes the total hadronic energy that is expected to be visible from events. This variable is the sum of the following ionization energies:

1. **Kinetic Energy of protons and charged pions**. The momenta of these particles give rise to ionization tracks inside the detector.

2. **Energy from electrons, photons and neutral pions**. Electrons and photons initiate EM showers. Neutral pions decay into a pair of photons ($\pi^0 \rightarrow \gamma + \gamma$), and the photons subsequently convert ($\gamma \rightarrow e^+ + e^-$).

3. **For hyperons, the total energy minus the nucleon mass**. The charged decay products of hyperons give ionization tracks, e.g. ($\Lambda^0 \rightarrow p + \pi^-$ and $\Sigma \rightarrow p + \pi^0$).

4. **For anti-nucleons, the total energy minus the nucleon mass**. For nucleon annihilation, the resultant products are charged/neutral pions and photons, all of which give visible ionization signatures.

Intentionally and very importantly, ionization energy released by neutron scatters is not included in E_{avail}. In individual events, these energies are generally small, typically from few to a few tens of MeV. However, their simulation has large uncertainties. The neglect of neutron induced kinetic energy in E_{avail} is accounted for by assigning systematic uncertainties that provide appropriate coverage.

4.2.2.1 Estimator for E_{avail}

Measurement of available hadronic energy involves establishing a relationship between visible hadronic energy and the true available energy. Visible hadronic energy is defined as sum of the energies from all visible cell hits not belonging to the muon track. For higher values of E_{had}^{vis}, this works well because charged hadrons (e.g protons/pions) are produced in the process, which leave behind ionization tracks in the detector. A plot of E_{had}^{vis} versus true E_{avail}, as shown in Fig. 4.9,

shows a clear linear relationship between the two, enabling a nearly linear map to be constructed between the two variables.

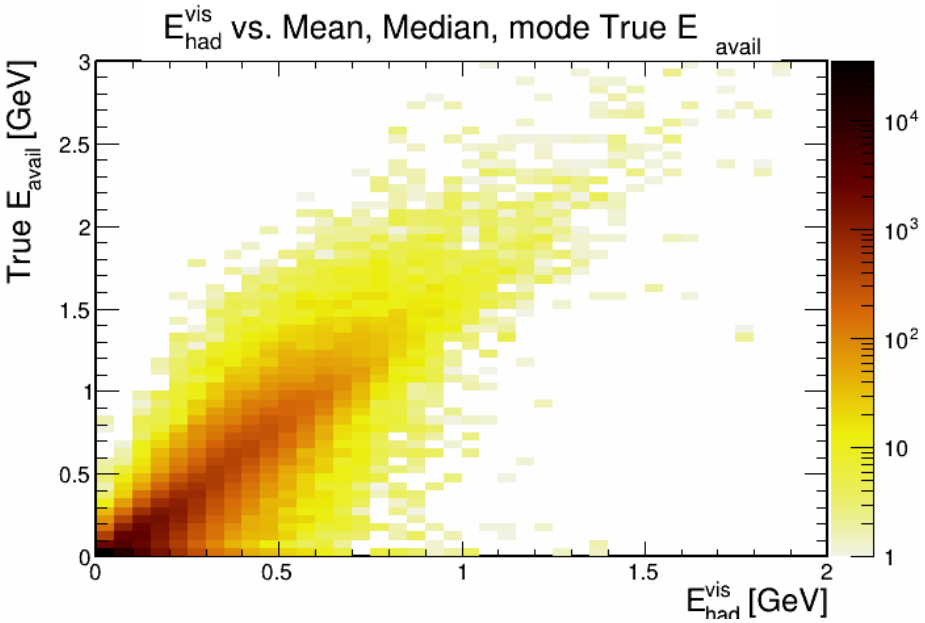

Figure 4.9: Distribution of E_{had}^{vis} vs. E_{avail} of selected $\bar{\nu}_\mu$ CC events in the NOvA ND.

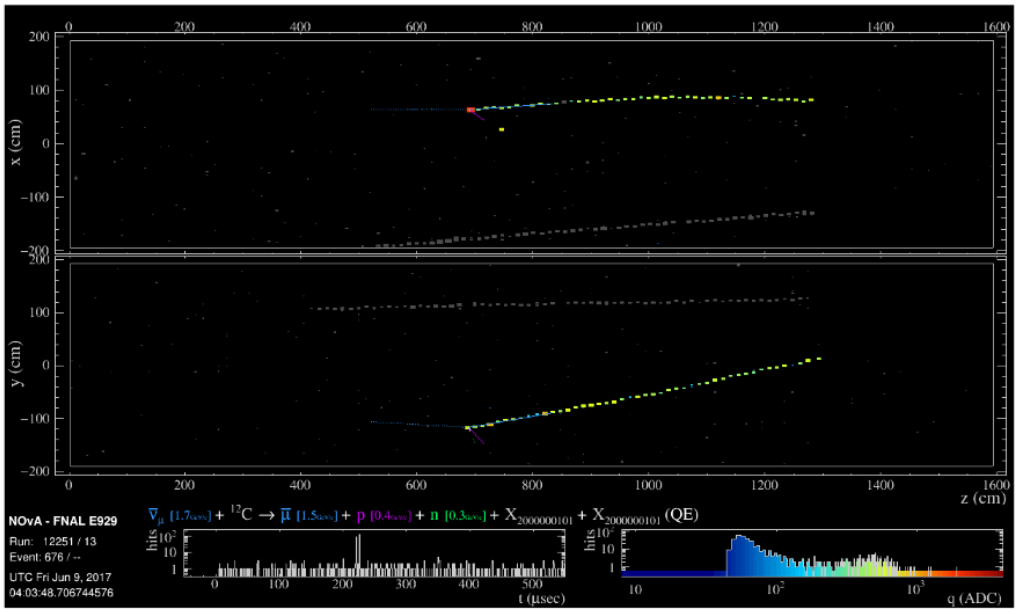

Figure 4.10: Event display showing a $\bar{\nu}_\mu$ CC QE reaction. A heavy hit is visible in the x-view. The true E_{avail} value of the event is 100 MeV. The visible hadronic energy is 57.5 MeV.

However, for low values of visible hadronic energy, that is not the case. The hadronic content of events in this region consists of neutrons and low energy charged hadrons. Such particles do

not create tracks. They are visible only when they cause spallations of nuclei, or when they scatter off another charged particle with significant momentum transfer. The event display in Fig. 4.10 gives an example of an event with low hadronic energy. Unfortunately, there is no clear relationship between the visible energy produced by these hadrons (or the daughters of the hadrons) and the true available hadronic energy for events with very low visible hadronic activity. Because of this, the true available energy of the events with low visible hadronic energy must be characterized separately.

The mapping of E_{had}^{vis} to E_{avail} consists of two parts, one for low E_{had}^{vis} and one for higher E_{had}^{vis} values. This requires a threshold value for E_{had}^{vis} to be established. Above the threshold, there is a linear relationship between the two variables. The threshold is established by observing the behavior of the neutron kinetic energy value as visible hadronic energy rises. Figure 4.11 shows that for lower values of E_{had}^{vis}, most of the visible hadronic energy comes from the neutron kinetic energy. However, from Fig. 4.12, which shows the relationship between visible hadronic energy and average true neutron kinetic energy, it is seen that for E_{had}^{vis} values of 85 MeV and above, the neutron kinetic energy contribution diminishes as E_{had}^{vis} increases. The threshold therefore is set at 85 MeV. A linear fit is used to map E_{avail} values for visible hadronic energy values above 85 MeV. Below this value, an average E_{avail} value is assigned.

4.2.2.2 E_{avail} for events with $E_{had}^{vis} > 85$ MeV

The estimator for E_{avail} is constructed as follows:

1. From a diplot of the visible hadronic energy vs. the true available energy, the mean true E_{avail} value for each bin of E_{had}^{vis} is determined and plotted, as is shown in Fig. 4.13.

2. A line is fitted through the profile points for E_{had}^{vis} except for the very first point. It was found, after many fitting trials, that a three-part, piecewise fit, describes the relationship between E_{had}^{vis} and mean true E_{avail} well for E_{had}^{vis} upto 1.5 GeV. Beyond that, given the spread of points, it is difficult to draw meaningful relationships due to the lack of statistics.

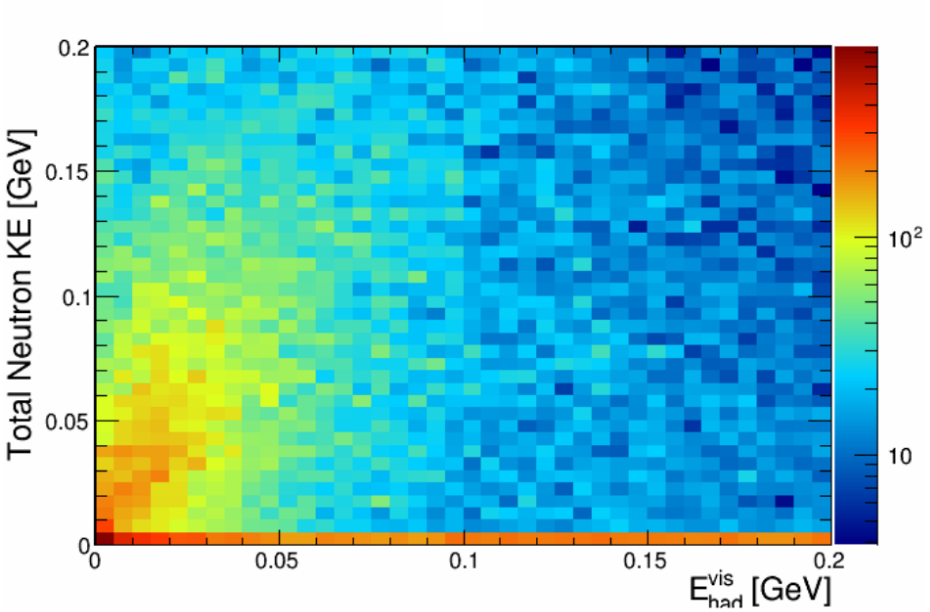

Figure 4.11: Visible hadronic energy versus true neutron kinetic energy for events with $E_{vis}^{had} < 200$ MeV.

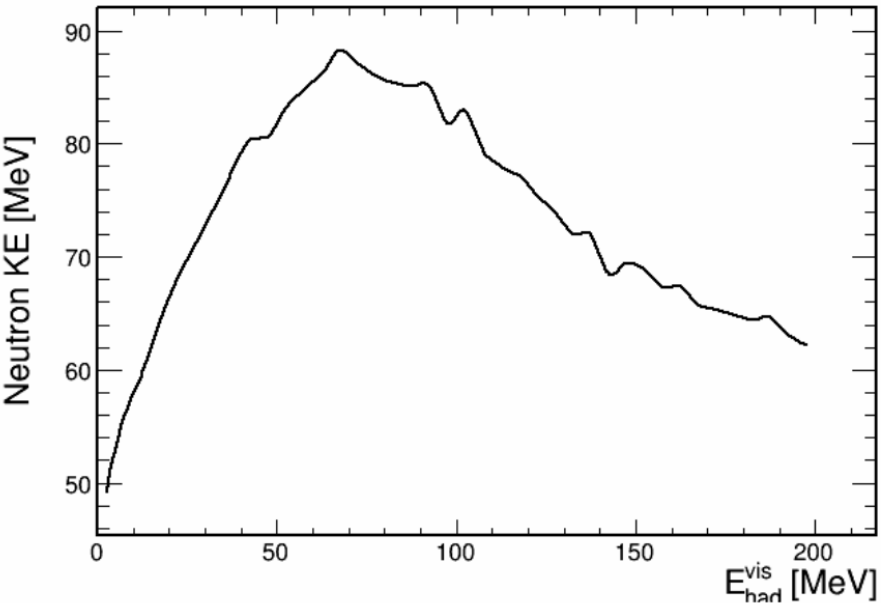

Figure 4.12: Visible hadronic energy versus Average True Neutron Kinetic energy for events with $E_{vis}^{had} < 200$ MeV.

3. A straight line mapping between 0.085 GeV to 0.2 GeV of E_{had}^{vis} characterizes the relation between visible hadronic energy and available energy. Above that, quadratic corrections

are required. The mapping function superimposed on the mean profiling points, and is shown in Fig. 4.13.

$$E_{avail} = (1.797 \times E_{vis}^{had}) - 0.1403, \ for\ 0.085\ GeV < E_{vis}^{had} \leq 0.2\ GeV, \quad (4.8)$$

$$E_{avail} = (0.002 \times (E_{vis}^{had})^2) + (1.79635 \times E_{vis}^{had}) - 0.1403, \ for\ 0.2\ GeV < E_{vis}^{had} \leq 0.8\ GeV,$$

$$E_{avail} = (-0.04 \times (E_{vis}^{had})^2) + (1.65 \times E_{vis}^{had}), \ otherwise.$$

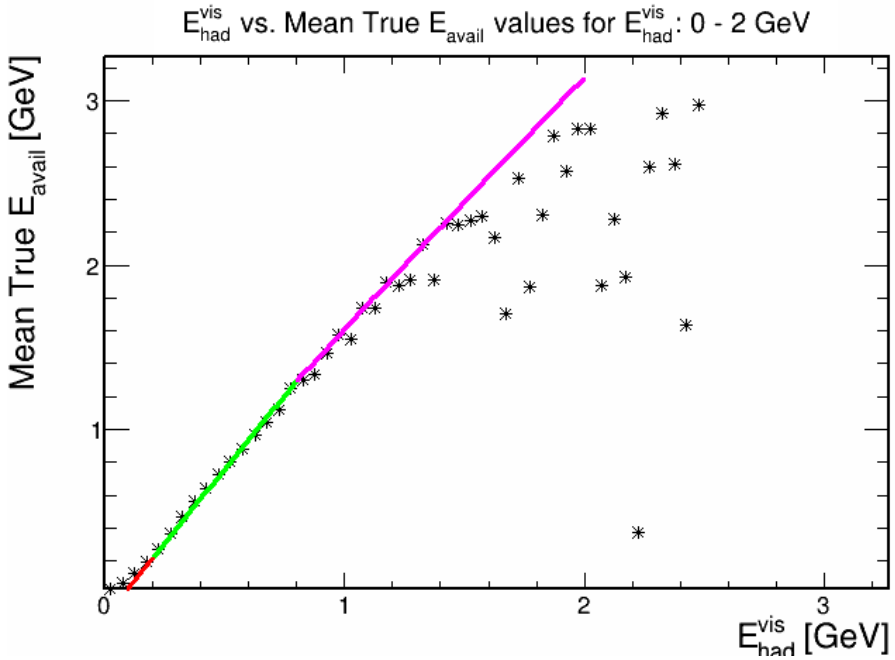

Figure 4.13: Visible hadronic energy versus mean true available energy for selected events. The mapping function, as shown in Eq. 4.8 has been superimposed

4.2.2.3 E_{avail} for events with $E_{had}^{vis} \leq 85$ MeV

As mentioned above, the hadronic content of events in this region is dominated by neutrons which, by definition, have zero available energy. Nevertheless, many neutrons give visible low energy ionizations. A plot of E_{had}^{vis} versus true E_{avail} for $E_{had}^{vis} \leq 85$ MeV events is presented in Fig. 4.14.

Because of the concentration of events in the lowest true E_{avail} bin, as seen in Fig. 4.15, an average true E_{avail} is calculated for E_{had}^{vis} for this particular range ($0 \leq E_{had}^{vis} \leq 85$ MeV). Figure

4.15 shows a histogram of the true E_{avail} projection. One can roughly surmise from this figure what the average E_{avail} value is. The average true E_{avail} value is calculated to be 12.4 MeV. We assign this average E_{avail} value to all CC events having $E_{had}^{vis} < 85$ MeV.

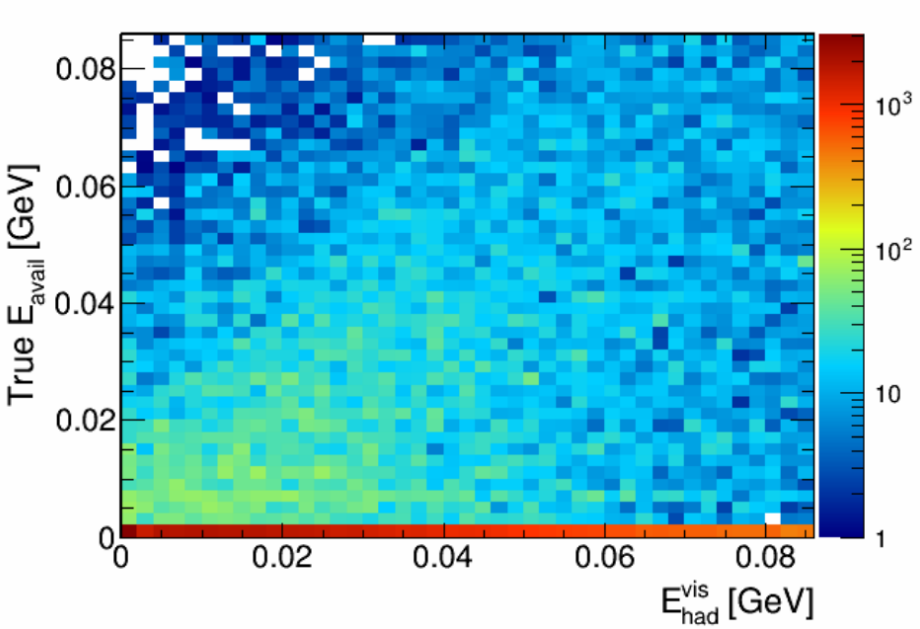

Figure 4.14: Distribution of visible hadronic energy versus true available energy of events with $E_{had}^{vis} < 85$ MeV.

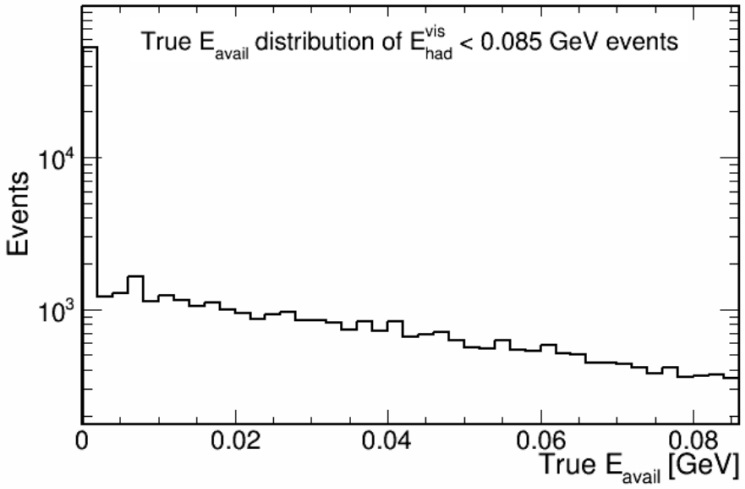

Figure 4.15: True E_{avail} distribution for $E_{had}^{vis} < 85$ MeV events. The hadronic content of most events consist solely of neutrons, with zero available energy. This leads to the concentration of events in the lowest true E_{avail} bin.

The mapping of E_{had}^{vis} to true E_{avail} for CC $\bar{\nu}_\mu$ events of the RHC sample is necessarily bimodal,

for the reasons indicated in Secs. 2.2.1 - 2.2.3. The full mapping is defined by the mapping in Sec. 2.2.2 for $E_{had}^{vis} > 85$ MeV together with the assignment of $E_{avail} = 12.4$ MeV for $E_{had}^{vis} \leq 85$ MeV detailed in Sec. 2.2.3.

4.2.2.4 Residual of E_{avail}

The absolute and fractional residuals for E_{avail} are defined as follows:

$$Absolute\ residual\ =\ True\ E_{avail}\ -\ Reco.\ E_{avail}, \tag{4.9}$$

$$Fractional\ residual\ = \frac{True\ E_{avail}\ -\ Reco.\ E_{avail}}{True\ E_{avail}}. \tag{4.10}$$

In the case of E_{avail}, it must be noted that a vast majority of the events of the RHC sample have zero True $E_{avail}\ = 0.0$ and little, if any, E_{had}^{vis}. Such events get assigned a reconstructed E_{avail} value of 12.4 MeV. For such events, the quantity of residuals is meaningless, because the residual values will always be biased. As such, absolute and fractional residuals of events will be displayed for events with reco $E_{avail} > 100$ MeV. The absolute and fractional residual distributions for the E_{avail} variable are shown in Fig. 4.16. It can be seen that both the absolute and fractional residual distributions have gaussian distributions, with a peak and a relatively even spread on both sides of the peak. However, the residuals demonstrate that there is a bias in the available energy estimator, indicating that our estimator in general slight underestimates the actual available energy value.

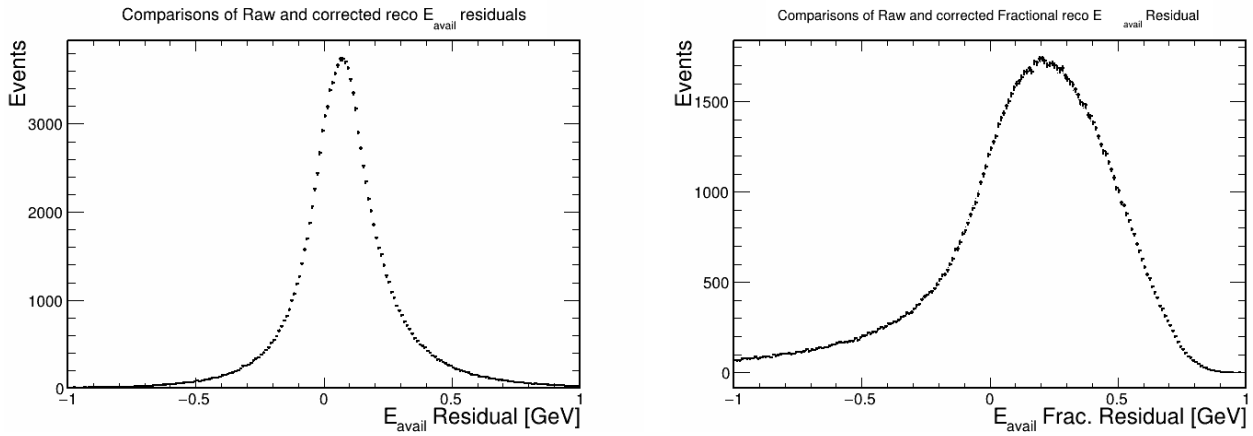

Figure 4.16: Absolute (left) and fractional residual (right) for available energy of $\bar{\nu}_\mu$ CC events.

The absolute and fractional residuals are also broken out in the following ranges 0.1 GeV$\leq$ $E_{avail} \leq$ 0.5 GeV, 0.5 GeV $< E_{avail} \leq$ 1.0 GeV, 1.0 GeV $< E_{avail} \leq$ 1.5 GeV, 1.5 GeV$< E_{avail} \leq$ 2.0 GeV and displayed in Figs. 4.17 and 4.18. The residual plots for the range show that most of the overall distributions are driven by the kinematics of the events in the range 0.0 GeV$\leq$ $E_{avail} \leq$ 0.5 GeV, and in particular, from events that have very little available hadronic energy. A suitable binning scheme, discussed in a later section, is required to absorb the differences in true and reconstructed available energy so that the analysis is performed correctly. At higher E_{avail}, the bias in the estimator is particularly noticable. The lack of statistics in this region makes it difficult to form a more accurate mapping of visible hadronic energy values. Such biases are taken care of by the unfolding procedure, which is discussed in a later section.

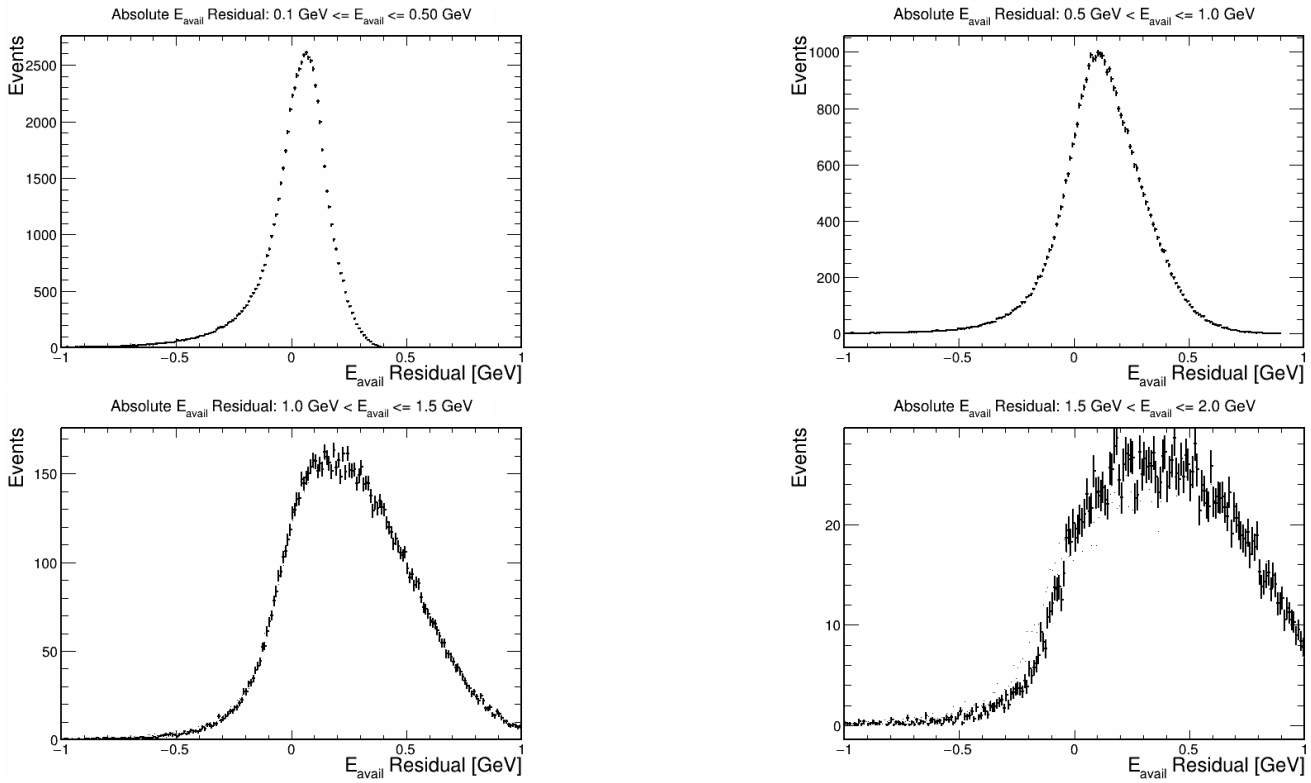

Figure 4.17: Absolute resolution of the E_{avail} variable broken out in different ranges of E_{avail}. With increasing E_{avail}, the distributions become broader and biased.

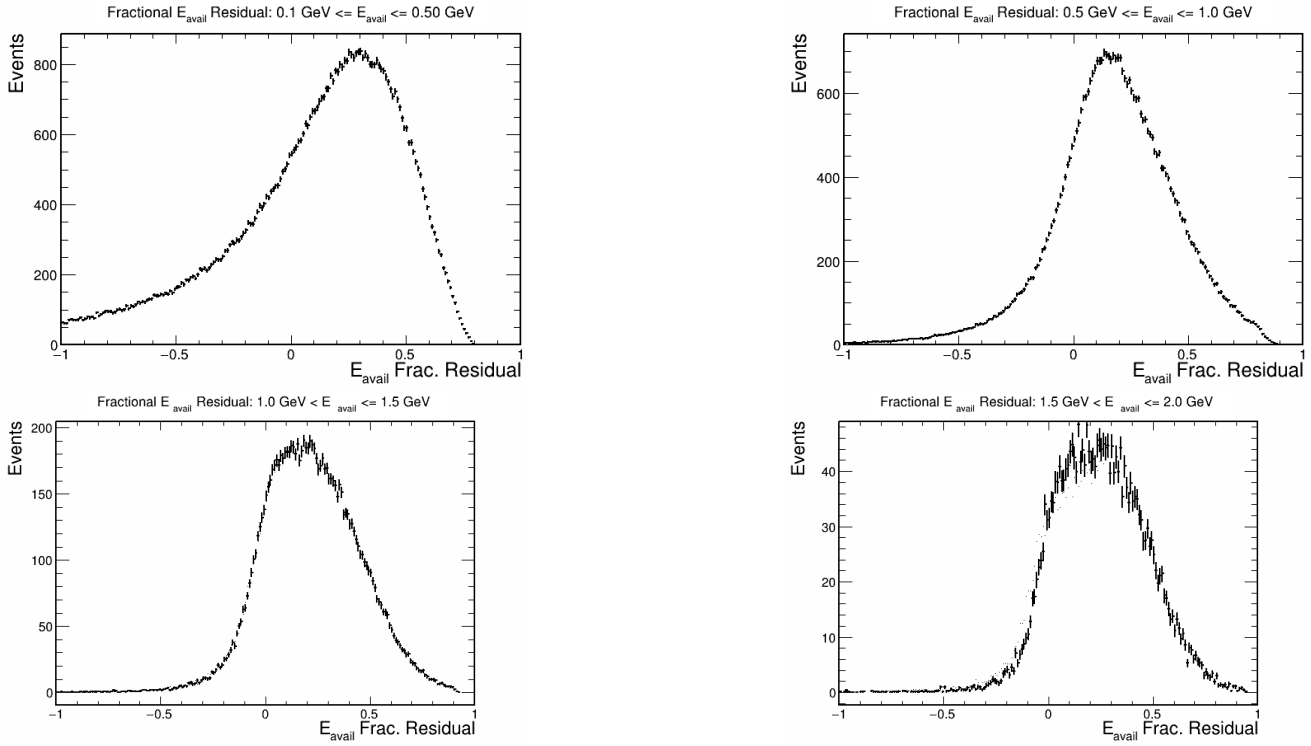

Figure 4.18: Fractional residual of E_{avail} broken out by different ranges of E_{avail}.

4.3 Selection of events

4.3.1 Signal event definition

The signal definition for this analysis is modeled after the one made for the NOvA ν_μ-CC inclusive analysis [24]. True signal events for this analysis must satisfy the following criteria:

1. The event is a true $\bar{\nu}_\mu$-CC event.

2. The true primary vertex of the event must be within the fiducial volume. The coordinates of the Near Detector fiducial volume are (-130 cm, +140 cm) for X and Y, and (100 cm, 1000 cm) for Z.

3. The truth kinematics of the muon must pass the muon phase space cuts, which is a combination of muon production angle and muon kinetic energy requirements. In general, the true muon kinetic energy must be between 0.5 - 2.5 GeV and the muon production angle must be less than 60 degrees. The detailed kinetic energy and production angle combinations are listed in Table 4.1.

4.3.2 $\bar{\nu}_\mu$ CC event selection criteria

The CC-$\bar{\nu}_\mu$ inclusive cross section measurement uses the same selections as for the ν_μ CC inclusive cross-section measurement [24]. In brief, the selection requirements are as follows:

1. There must be at least one reconstructed "Kalman track". In Kalman tracking, cell hits are clustered together by time. Hits that occur in a narrow window of time belong in the same cluster. If these clusters occur one after the other in a line, they are assumed to belong to a track. A fit is performed over these clusters, forming a Kalman track that encompasses all these hits. Details of the Kalman tracking method can be found in [25].

   In addition, The event must have at least 20 cell hits and the hits must occur in at least four consecutive planes. These criteria ensure that selected events have sufficient activity to enable good reconstruction.

2. There must be a Kalman track that is identified as a muon. For muon identification, a boosted decision tree was trained on muon tracks and non-muon tracks [24][26]. The identifier utilizes the dE/dx log-likelihood and the scattering log-likelihood of muons and pions and the average dE/dx in the last 10 cm and 40 cm of the tracks to assign a MuonID score to the track. Any event in which the highest MuonID score among all tracks is 0.27 or above is likely to contain a muon.

3. All tracks must be contained within the active detector volume.

4. The muon track is identified first. The start of this track is the event vertex. The vertex of the interaction must be within the fiducial volume of the detector.

5. The muon produced from the event must pass certain kinematic constraints. These are designed to reject events in which the muon is likely to escape the detector. The details of the muon phase space requirement are summarized in Table 1. The muon in the event must satisfy one of the conditions listed in the table. In general, the muon kinetic energy must be 0.5-2.5 GeV and the production angle must be less than 60 degrees.

Muon Production Angle, $\cos(\theta_{\mu^+})$	Muon Kinetic Energy, KE_{μ^+} (GeV)
0.5 - 0.56	< 1.1
0.56 - 0.62	<1.2
0.62 - 0.68	<1.3
0.68 - 0.74	<1.4
0.74 - 0.80	<1.4
0.80 - 0.85	<1.4
0.85 - 0.88	<1.8
0.88 - 0.91	<2.9
0.91 - 0.94	<2.4
0.94 - 1.00	<2.5

Table 4.1: Summary of the muon phase space kinematic cuts. The reconstructed muon in the event must satisfy one of the listed conditions.

Chapter 5

Formulation of the Cross Section

The likelihood of an interaction between two particles is characterized with *cross sections* [20]. In a simple case, suppose beam of particles of type a with flux ϕ_a (rate at which particles pass through) crosses through a region of space that consists of particles of type b. There are n_b particles of type b per unit volume. The interaction rate of type a particles on a single particle of type b is $\mathcal{N}_{ab}$. The relationship between interaction rate per particle of type b and flux is given by

$$\mathcal{N}_{ab} \;=\; \sigma \times \phi_a. \tag{5.1}$$

The quantity σ in Eq. (5.1) is the cross section. This quantity contains information about the underlying interaction between particles a and b [23]. Consider a particle a is moving with velocity v_a through the volume containing particles b, in which the illuminated area of the target volume is A. In a certain time δt, a crosses a region containing $\delta N \;=\; n_b\, v_a\, A\, \delta t$ particles of type b. The probability of the two particle types interacting is given by the effective cross sectional area of the δN particles divided by the incident area A.

$$\delta P = \frac{\delta N\, \sigma}{A} = \frac{n_b\, v_a\, A\, \delta t}{A}\, \sigma = n_b\, v_a\, \sigma\, \delta t. \tag{5.2}$$

Dividing both sides by δt yields the interaction probability of a single particle a interacting with a target in the region:

$$\frac{\delta P}{\delta t} = n_b \, v_a \, \sigma. \tag{5.3}$$

The total interaction rate of a beam of particles a moving through the volume of particles b is then

$$\mathcal{N}_{ab} = \frac{\delta P}{\delta t} \times n_a \, V \tag{5.4}$$

where V is the volume of the target region and n_a is the number density of the a particles moving through the space. This equation can be rearranged into:

$$\mathcal{N}_{ab} \; = \; (n_b \, v\sigma) \, n_a \, V \; = (n_a \, v)(n_b \, V) \, \sigma \; = \phi_a \, N_b \, \sigma. \tag{5.5}$$

Equation (14.1) can be rearranged to yield the formula for the cross section:

$$\sigma \; = \; \frac{\mathcal{N}_{ab}}{\phi_a \times N_b}. \tag{5.6}$$

It is of interest to express the cross section differentially as a function of various kinematic variables. In this analysis, the variables of interest are $|\vec{q}\,|$ and E_{avail}. However, knowing how to calculate the differential cross section in terms of antineutrino energy is important for closure tests, and will be discussed first.

5.1 Calculation of differential cross section in $E_{\bar{\nu}}$

The first step is to obtain a distribution of signal events in $E_{\bar{\nu}}$. This is done by subtracting the total distribution of data events with an estimate of the background:

$$N^{Reco}_{Signal,\alpha} = (N_{data,\alpha} - N_{bkg,\alpha}). \tag{5.7}$$

Here, the subscript α represents bins of reconstructed antineutrino energy. For comparison against theoretical models, this event distribution has to be unfolded so that one obtains true antineutrino energy distribution:

$$N^{True}_{Signal,i} = \sum_{\alpha} U_{\alpha,i}(N_{data,\alpha} - N_{bkg,\alpha}). \tag{5.8}$$

In Eq. 5.8, i is the bin in true antineutrino energy. The unfolding matrix, $U_{\alpha,i}$, converts distributions from reco $E_{\bar{\nu}}$ to true $E_{\bar{\nu}}$. The number of signal events in a particular bin i of true $E_{\bar{\nu}}$ is represented by $N^{True}_{Signal,i}$.

The event distribution obtained in eq. (5.8) only consists of events that have passed selection cuts. There are other true signal events that do not pass the selection criteria. These events must also be accounted for to obtain the correct cross section. To do this, we must know the efficiency, ϵ, for selecting true signal events:

$$\epsilon = \frac{Signal\ events\ that\ pass\ CC\ inclusive\ selections}{All\ signal\ events}. \tag{5.9}$$

The number of selected signal events is then efficiency-corrected by dividing the true signal distribution by the efficiency distribution. The efficiency-corrected true signal distribution is the distribution of all signal events that occur in the detector. A cross section for $\bar{\nu}_{\mu}$ CC Dytman MEC on carbon-12 in neutrino energy is calculated by dividing this efficiency corrected distribution by the flux distribution in true $E_{\bar{\nu}}$, and by the number of nucleons from all carbon-12 nuclei in the fiducial volume:

$$\sigma(E_{\nu})_i = \frac{N^{True}_{Signal,i}}{T\,\epsilon_i\,\phi_i}. \tag{5.10}$$

5.2 Double-differential cross section in $|\vec{q}\,|$ and E_{avail}

As previously, the first step is to calculate the distribution of signal events in the phase space of $|\vec{q}|$ and E_{avail} by subtracting the background events from the sample:

$$N^{True}_{Signal,ij} = \sum_{\beta}\sum_{\alpha} U_{\alpha\beta,ij}(N_{data,\alpha\beta} - N_{bkg,\alpha\beta}) \tag{5.11}$$

Each bin now has two indices as two variables are involved. Here, α and β designate bins of reconstructed $|\vec{q}|$ and E_{avail} respectively. Bins of true $|\vec{q}|$ and E_{avail} are designated using i and j. This signal distribution must be efficiency corrected. The unfolded, efficiency corrected signal distribution is given by the bin-by-bin ratio of unfolded selected signal events to the selection efficiency. The ratio for the bin (i, j) is $(N^{True}_{Signal,ij})/\epsilon_{ij}$.

To get to a flux-averaged cross section, The unfolded, efficiency corrected signal distribution is divided by the bin area, the product of the bin widths $\Delta|\vec{q}|$ and ΔE_{avail}, the integral of the $\bar{\nu}_\mu$ flux from 0.0 to 5.0 GeV and the number of target nucleons, T:

$$(\frac{d\sigma}{d|\vec{q}|dE_{avail}})_{ij} = \frac{N^{True}_{Signal,ij}}{T\ \epsilon_{ij}\ (\Delta|\vec{q}|)_i\ (\Delta E_{avail})_j\ \int \phi(E_\nu)\ dE_\nu}. \tag{5.12}$$

The cross section expressions 5.12 with 5.11 is used in the analysis to follow.

Chapter 6

Resolution Binning

In cross-section measurements, the accuracy of measurement is limited by the resolution of the detector. The resolution of the analysis variables is not constant over the kinematic range of the variables. Furthermore, the phase space distribution of the events can introduce large statistical errors into the cross section measurement. However, certain features of the measurement are important and therefore there are cases where finer binning is required. With all of these factors in mind, it is necessary to define a variable binning scheme in both $|\vec{q}|$ and E_{avail}. The binning scheme is constructed according to the resolution of the reconstructed $|\vec{q}|$ and E_{avail} variables. The resolution of a variable is the RMS spread of the absolute residual of the variable about the mean value. The binning scheme developed for the two variables is discussed below.

6.1 Bin assignments for $|\vec{q}|$

The distribution of $|\vec{q}|$ residuals as a function of $|\vec{q}|$ is shown in Fig. 6.1 (left). The corresponding resolution of the $|\vec{q}|$ variable is displayed on Fig. 6.1 (right). It can be seen that the resolution of the variable increases roughly linearly with increasing $|\vec{q}|$. Note that the number of events with higher $|\vec{q}|$ are fewer in number compared to those with lower momentum transfer.

To construct the binning scheme, the resolution at the smallest $|\vec{q}|$ value, which is 0.0 GeV/c, is examined; it is approximately 20 MeV/c. The width of the first bin is defined by this value,

with 0.0 GeV/c being the low edge of this bin and with the high edge of the bin 20 MeV/c. The resolution of the $|\vec{q}|$ variable is then examined at 20 MeV/c, which is about 30 MeV/c. The 30 MeV/c value defines the width of the second bin, with 20 MeV/c being the low edge of the bin. The procedure is continued is continued until the range of the analysis variable is covered. The ranges of the bins in $|\vec{q}|$ are shown in Table 6.1.

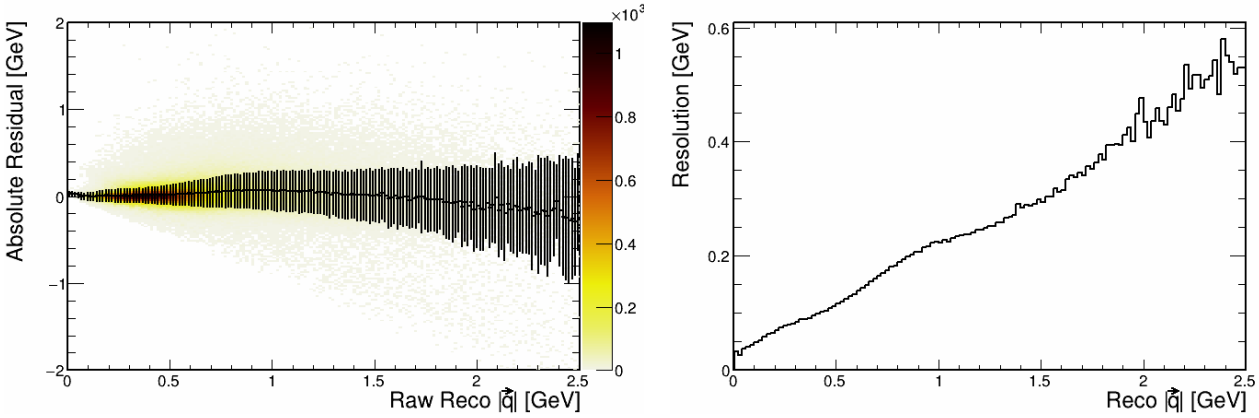

Figure 6.1: Distribution of the absolute $|\vec{q}|$ residual versus reco $|\vec{q}|$ with the resolution overlaid (left); the resolution as a function of reconstructed $|\vec{q}|$(right).

$|\vec{q}|$ Bins	0 - 0.02	0.02 - 0.05	0.05 - 0.08	0.08 - 0.12	0.12 - 0.16	0.16 - 0.23
	0.23 - 0.30	0.30 - 0.40	0.40 - 0.48	0.48 - 0.58	0.58 - 0.7	0.7 - 0.85
	0.85 - 1.05	1.05 - 1.30	1.30 - 2.0			

Table 6.1: Ranges of bins for the $|\vec{q}|$ variable. Values are in GeV/c.

6.2 Bin assignments for E_{avail}

The absolute residual for E_{avail} as a function of E_{avail} is shown in Fig. 6.2 (left) and the resolution of E_{avail} is plotted in Fig. 6.2 (right). As with $|\vec{q}|$, it can be seen that the resolution variables increases roughly linearly with increasing available energy. A binning scheme is constructed using the same procedure in the previous subsection. At zero E_{avail}, the resolution is 60 MeV. Thus the first bin range is chosen to be from 0 to 60 MeV. Next, the resolution value at 60 MeV E_{avail} is 120 MeV. This defines the second bin to have the range 60 - 180 MeV, with 120 MeV being the width of the bin.

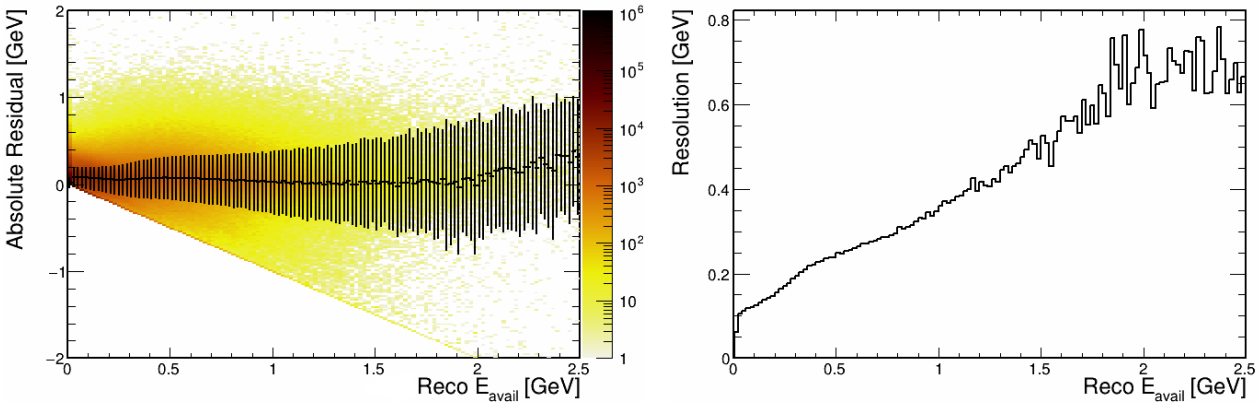

Figure 6.2: Distribution of the absolute E_{avail} residual versus reco E_{avail} (left). The resolution of the variable is also overlain on the same distribution. The resolution values are plotted on the figure to the right.

The ranges of the bins in E_{avail} are listed in Table 6.2. A plot of MC simulated events in the ND passing the CC inclusive selection criteria is shown in Fig. 6.3.

E_{avail} Bins	0.00 - 0.06	0.06 - 0.18	0.18 - 0.33	0.33 - 0.52
	0.52 - 0.75	0.75 - 1.05	1.05 - 1.45	1.45 - 2.0

Table 6.2: Ranges of bins for the E_{avail} variable. Values are in GeV.

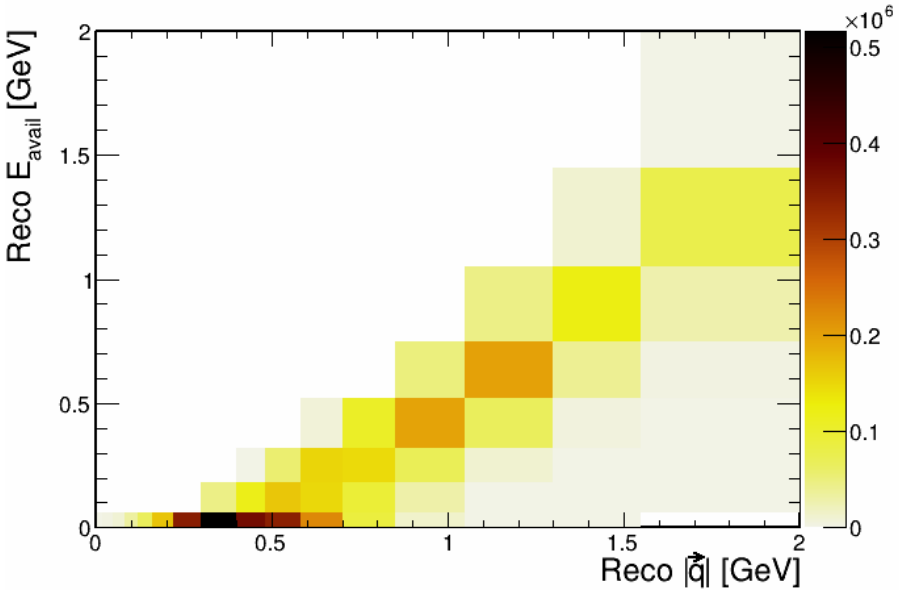

Figure 6.3: Distribution of simulated ND events that pass the CC inclusive selection criteria plotted in analysis variables $|\vec{q}|$ and E_{avail} with the new binning scheme.

The viability of this binning scheme is checked by examining the statistical error in each bin. The statistical error in in bin $(i,\ j)$, $\delta N_{i,j}$ is given by the square root of the number of events in the bin, $\sqrt{N_{i,j}}$. The fractional statistical error in bin (i, j) is given by the following:

$$\frac{\delta N_{i,j}}{N_{i,j}} = \frac{\sqrt{N_{i,j}}}{N_{i,j}} = \frac{1}{\sqrt{N_{i,j}}}. \tag{6.1}$$

It is highly useful to ensure that no bins contain very few events. The highest two bins in $|\vec{q}|$, raging from 1.3 - 1.55 GeV and 1.55 - 2.0 GeV are merged together, yielding one $|\vec{q}|$ bin going from 1.3 - 2.0 GeV. The fractional statistical error of the event distribution with the last two momentum-transfer bins merged is displayed in Fig. 6.4. This merging finalizes the binning scheme for the analysis.

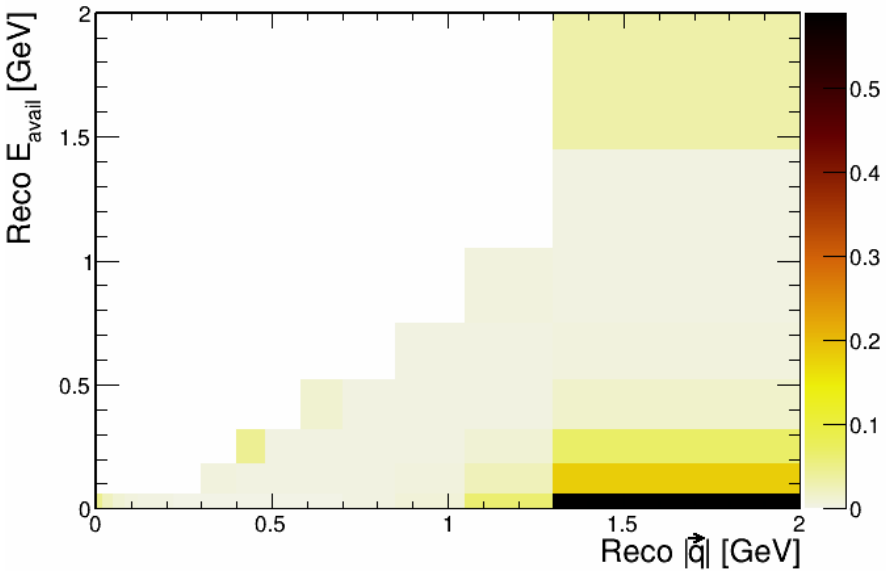

Figure 6.4: Fractional statistical error of simulated ND events that have passed the CC inclusive selections. The region consisting of $|\vec{q}|$ 1.3 - 2.0 GeV and 0 - 330 MeV of E_{avail} has low statistics, consequently it is a region of high fractional statistical error.

Chapter 7

Selection of Signal Events

A pure sample of signal events, namely CC-$\bar{\nu}_\mu$ events, is required for this analysis. Signal events consist of an incoming muon antineutrino interacting with nucleon(s) via a $W^{\pm}$ boson exchange. In the process, the antineutrino turns into a μ^+ particle and the nucleon turns into a hadronic shower, which may consist of a single nucleon, or a nucleon plus pion(s). The process is portrayed by the Feynman diagram shown in Fig. 7.1. Some event displays of true signal events are shown in Fig. 7.2.

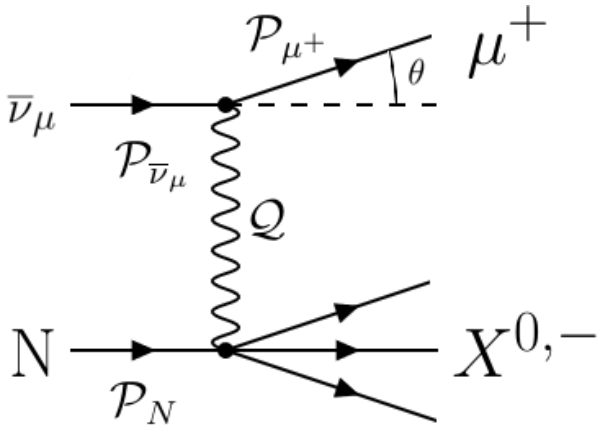

Figure 7.1: Feynman diagram illustrating a generic CC antineutrino - nucleon interaction. The target nucleon, N, is usually bound within a nucleus, $\mathcal{A}$.

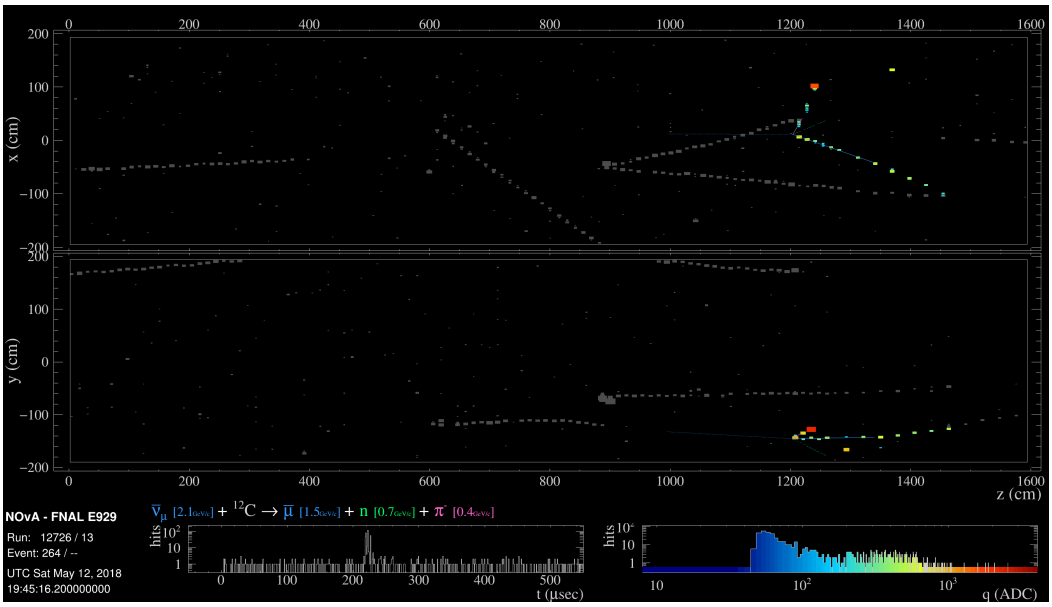

Figure 7.2: Event display of a simulated $\bar{\nu}_\mu$ CC event in the NOvA ND that passed the inclusive selection criteria. The event is a CC DIS event that produced a π^- particle.

The signal selection criteria are based on the visible signatures possessed by true signal events. Selected events are required to have a reconstructed muon track, among other criteria. However, the CC inclusive selections can allow ν_μ CC and neutral current (NC) events to be selected. The former so-called wrong-sign events are selected because such interactions also produce a muon track. NC interactions can produce a long pion track which may be misidentified as a muon track, thereby enabling them to pass the selections.

7.1 Background events

The analysis requires a relatively pure sample of CC-$\bar{\nu}_\mu$ events, however there are other event types that also pass the CC inclusive selections. The major background interaction type consists of CC-ν_μ events. There are magnetic horns to separate the π^+ and π^- particles that decay to produce antineutrinos and neutrinos respectively. The presence of wrong-sign events (ν_μ CC events) indicates that the separation is not complete.

A second interaction category consists of NC events, in which a neutrino or antineutrino interacts with a nucleon via a Z boson exchange. The resulting final state contains an invisible neutrino

together with a recoiling hadronic system.

Electron neutrinos are also present in the sample. And finally, there is a small sample of events where the true event vertex lies outside the detector. However, the vertex of these events are misreconstructed to be within the fiducial volume, and therefore pass the selection. All nonfiducial events, including those generated by $\bar{\nu}_\mu$ CC interactions, are treated as background because these events cannot be properly reconstructed.

7.2 Events in the NOvA near detector - no selections

Most of the observed tracks, showers and hits come from interactions (about 87% of the total sample) that occur outside the detector. This point is illustrated by the true neutrino/antineutrino energy distribution in Fig. 7.3. The CC-$\bar{\nu}_\mu$ events originating within the detector make up about 8% of the total number of observed interactions.

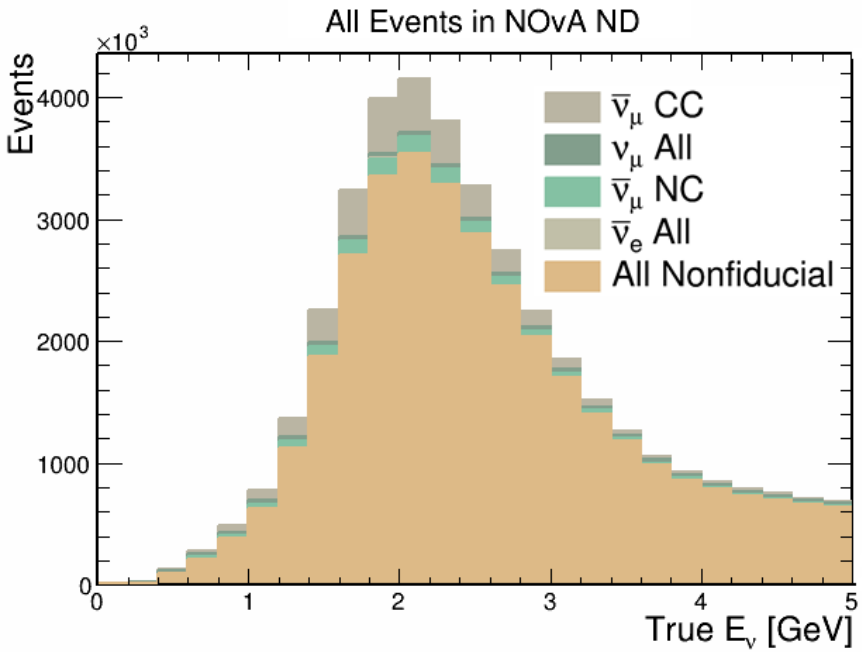

Figure 7.3: True Antineutrino/neutrino energy distribution of the NOvA RHC MC sample in the Near Detector. The nonfiducial events (bottom histogram) make up most of the activity inside the detector, while CC-$\bar{\nu}_\mu$ interactions (top histogram) make up most of the events that originate inside the detector.

7.3 Selection of signal events

While the vast majority of events observed in the ND originate outside the detector, there are significant numbers of background events that originate inside the detector together with the signal events. The CC inclusive selection consists of a number of criteria that are effective in isolating a relatively pure sample of true $\bar{\nu}_\mu$ CC events. The reduction of the raw sample to an analyzable sample is shown in the sequence of distributions in Fig. 7.4.

In Fig. 7.4(a), the requirement of a candidate muon removes many NC events. The application of the fiducial cut eliminates almost all events that originate outside the detector, as can be seen in Fig. 7.4(b). The quality cut does not affect the event distribution as much, merely removing a small number of very low energy events that do not create many hits. The containment cut rejects events in which particles exit the detector, and generally reduces the count of higher energy events. This still leaves a fair amount of NC and ν_μ-CC events, as shown by Fig. 7.4(d). Finally, the muon phase space cut eliminates almost all of the remaining NC events, leaving behind only antineutrino CC events and a much smaller number of neutrino CC backgrounds, as shown in Fig. 7.4(e). In Fig. 7.4(e), it can be seen that the selection criteria almost completely eliminates electron neutrino/antineutrino and nonfiducial events, and minimizes the presence of NC events, leaving only true signal events and a portion of wrong sign events. The percentages of signal and background event types in the CC inclusive selected sample are shown in Table 7.1.

Event Type	Percentage contribution
$\bar{\nu}_\mu$ CC signal events	77.3%
$\bar{\nu}_\mu$ CC background events	8.09%
ν_μ CC interactions	10.6%
NC events	2.60%
$\bar{\nu}_e$ and ν_e CC events	0.16%
Nonfiducial events	1.25%

Table 7.1: List of background event types along with percentage contribution to observed events in the NOvA ND that pass the CC inclusive selection. The dominant background ν_μ arises from wrong sign events (10.6%).

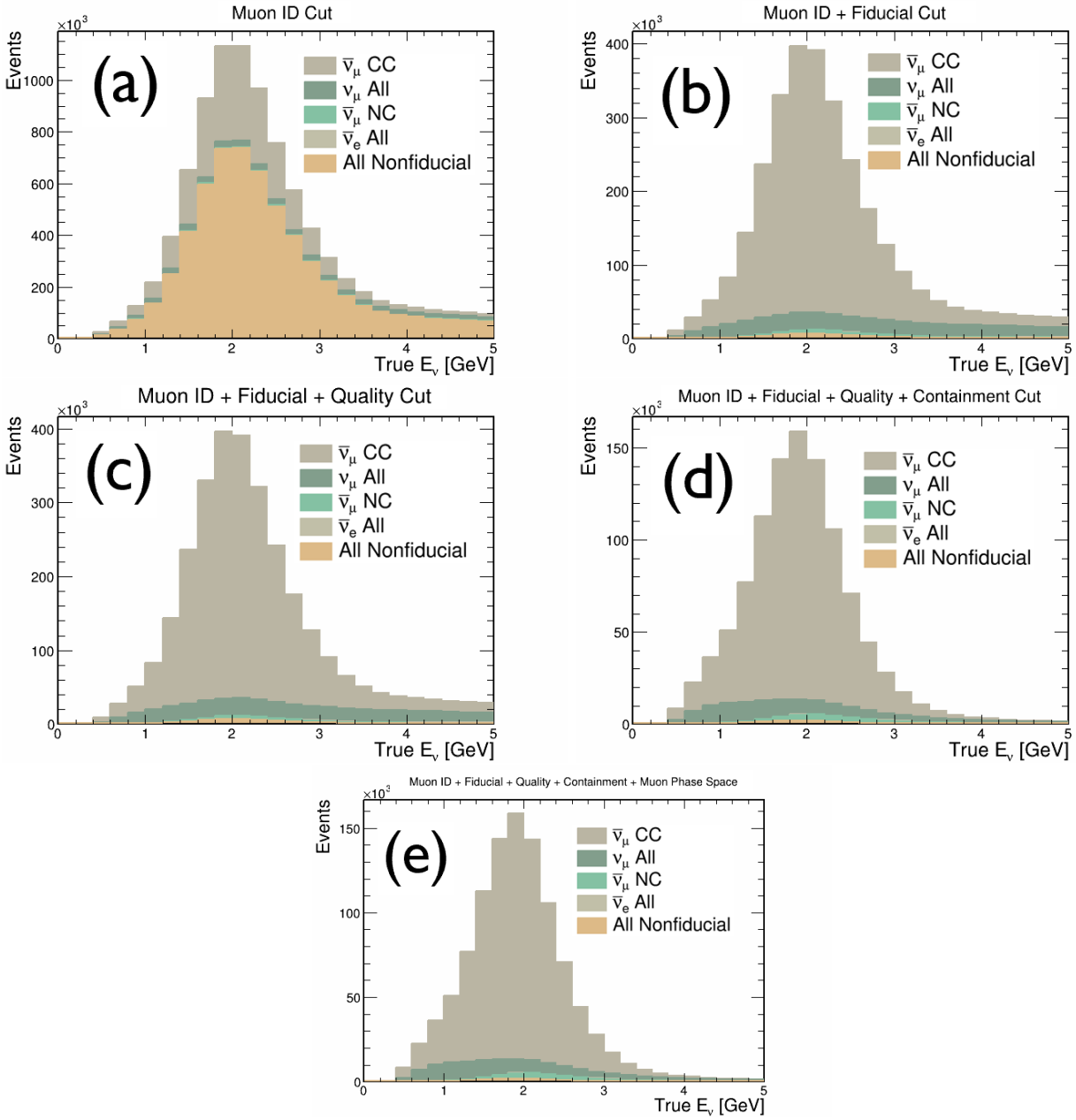

Figure 7.4: Effect of the individual selection criteria on the analysis event sample. The sequence of cuts yields a sample that is rich in $\bar{\nu}_\mu$ CC events (85.4%) plus a certain amount (14.6%) of background events.

7.4 $\bar{\nu}_\mu$ signal and ν_μ background distributions

The CC inclusive selections eliminate most backgrounds. However, there remains a fraction of ν_μ-CC interactions contributing to the signal event count. The antineutrino signal and neutrino background distributions in variables $|\vec{q}|$ and E_{avail} are displayed in Figs. 7.5(a) and 7.5(b)

respectively. It can be seen that the signal (Fig. 7.5(a)) and the contamination events (Fig. 7.5(b)) are concentrated in different parts of the phase space. This happens because CC ν_μ events often produce a leading proton, which is visible and therefore contributes to the available energy. On the other hand, signal CC $\bar{\nu}_\mu$ events often produce a neutron, which is not visible in the detector and does not contribute to E_{avail}. This accounts for the concentration of events in the lowest E_{avail} slice of Fig. 7.5(a), which is dominated by CCQE and 2p2h events.

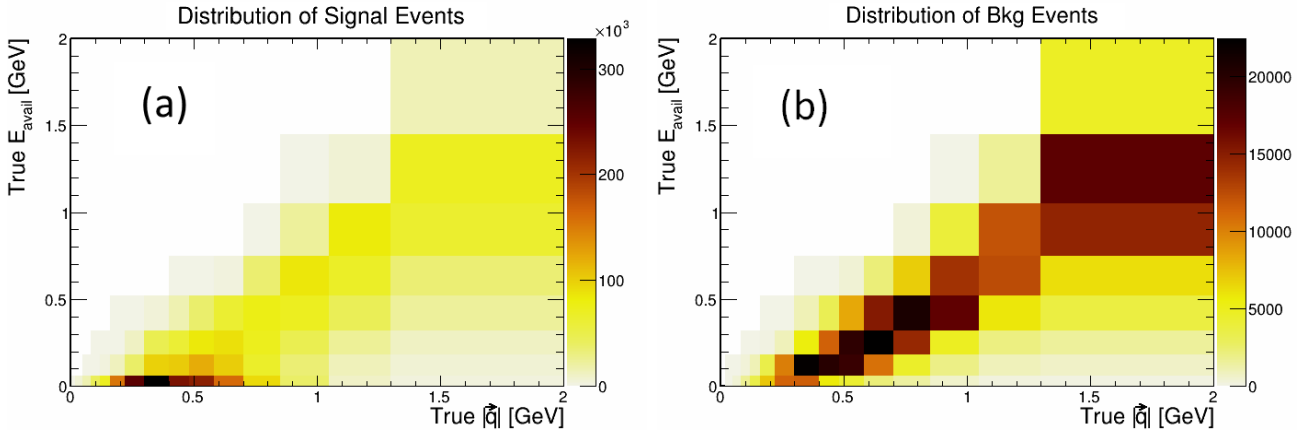

Figure 7.5: Distribution of signal and background events in variables $|\vec{q}|$ and E_{avail}. The signal is observed at very low E_{avail} with $|\vec{q}| < 0.7$ GeV/c, while the background extends to higher E_{avail} values.

7.4.1 Separation of background events from signal

7.4.1.1 Separation using Bjorken y

A variable that can in principle produce a separation between muon neutrino and antineutrino events is Bjorken y [20], which is approximately the fraction of the incident neutrino or antineutrino energy that is transferred to the target (see Eq. (7.4)). For the formalism summarized here, it is helpful to refer to Fig. 7.1. The Bjorken y variable is defined by the following equation:

$$y = \frac{\mathcal{P}_N \cdot q}{\mathcal{P}_N \cdot \mathcal{P}_{\bar{\nu}_\mu}}. \tag{7.1}$$

Here, $\mathcal{P}_{\bar{\nu}_\mu}$ is the 4-momentum of the incoming neutrino, $\mathcal{P}_N$ is the 4-momentum of the target nucleon and q^λ is the 4-momentum transfer. The following 4-momentum convention is utilized

for this formalism: $q^\mu = (q^0, \vec{q})$ and $q_\mu = (q^0, -\vec{q})$.

For a stationary nucleon, the initial state 4-momenta expressions can be written:

$$\mathcal{P}^\lambda_{\bar{\nu}_\mu} = (E_{\bar{\nu}_\mu}, 0, 0, E_{\bar{\nu}_\mu}) \text{ and } \mathcal{P}^\lambda_N = (m_N, 0, 0, 0). \tag{7.2}$$

Taking the plane defined by the lepton momenta to be the x-z plane, we can write:

$$\mathcal{P}^\lambda_{\mu^+} = (E_{\mu^+}, |\vec{p}_{\mu^+}| \sin(\theta), 0, |\vec{p}_{\mu^+}| \cos(\theta)) \text{ and } q^\lambda = (E_{\bar{\nu}_\mu} - E_{\mu^+}, \vec{p}_{\bar{\nu}_\mu} - \vec{p}_{\mu^+}). \tag{7.3}$$

Assuming the initial-state nucleon is stationary, Eq. (7.1) can be simplified [20]:

$$y \;=\; \frac{m_N(E_{\bar{\nu}_\mu} - E_{\mu^+})}{m_N E_{\bar{\nu}_\mu}} \;=\; 1 - \frac{E_{\mu^+}}{E_{\bar{\nu}_\mu}} \;=\; \frac{E_{hadronic}}{E_{\bar{\nu}_\mu}}. \tag{7.4}$$

Thus Bjorken y represents the fraction of the incident neutrino energy that is transferred to the hadronic system. The relevant four-momenta are illustrated for CC $\bar{\nu}_\mu N$ scattering by the Feynman diagram of Fig. 7.1.

The Bjorken y distributions for CC ν_μ versus CC $\bar{\nu}_\mu$ interactions are different. To elicit the difference, we consider the idealized case of neutrino scattering on pointlike fermions, such as neutrino - quark scattering. The scattering S-matrix element for a neutrino interacting with a point-like spin-1/2 particle is

$$S = \frac{g^2}{2 \cdot M_W^2}. \tag{7.5}$$

The Fermi coupling and electroweak coupling are related in the following way:

$$\frac{G}{\sqrt{2}} \;=\; \frac{g^2}{8M_W^2}. \tag{7.6}$$

According the Fermi's Golden Rule #2, the differential cross section can be written [20]

$$\frac{d\sigma}{dq^2} = \frac{2\pi}{v_{rel}}|S|^2 p^2 \frac{dp}{dE_f}\frac{d\Omega}{dq^2}\frac{2}{(2\pi)^3}\ , \tag{7.7}$$

where p is the momentum of a final state particle in center-of-momentum system (cm) frame and E_f is the total energy of the final state particles in the cm frame. Since the leptonic particles move at relativistic speeds, the relative velocity, $v_{rel} = c$ and $dp/dE_f = 1/2$. A neutrino is left-handed and has the $j_z = -1/2$ helicity. But the final state muon can have either $j_z = -1/2$ or $j_z = +1/2$ helicities. Furthermore, $q^2 = 2p^2(1 - cos(\theta))$ and $d\Omega/dq^2 = \pi/p^2$. Plugging these expressions into Eq. (7.7) gives

$$\frac{d\sigma}{dq^2} = \frac{G^2}{2\pi}. \tag{7.8}$$

To convert this into an expression for $d\sigma/dy$, we must first find a relation between q^2 and y. This is provided by Eqs. (5.12) and (5.15) of Ref. [20], which are reproduced below:

$$q^2 = 2p^2(1 - cos(\theta)),\ \text{and}\ \ y = \frac{1}{2}(1 - cos(\theta)). \tag{7.9}$$

Using the fact that $d\sigma/dy = d\sigma/dq^2 \times dq^2/dy$, we obtain:

$$\frac{d\sigma}{dy}(\nu N \to \mu N^*) = 4p^2\frac{G^2}{\pi}. \tag{7.10}$$

Antineutrinos are right-handed and so the total initial state angular momentum is $J = 1$, with a polarization of $J_z = +1$. This must be conserved in the final state. It is not necessary to recalculate the antineutrino cross section. Rather, the following Wigner d-function can be utilized in conjunction with the cross section for neutrinos to obtain the cross section for the antineutrino case:

$$d^1_{+1,+1} = \frac{1}{2}(1 + cos(\theta)). \tag{7.11}$$

Multiplying the square of this function with Eq. (7.10) gives the antineutrino differential cross section in terms of y:

$$\frac{d\sigma}{dy}(\bar{\nu}N \to \bar{\mu}N^*) = 4p^2\frac{G^2}{\pi}(1-y)^2. \tag{7.12}$$

Since there is the factor of $(1-y)^2$ in the antineutrino cross section, but not in the neutrino one, the event distributions for the two different particles in the Bjorken y variable will differ. These cross sections are in the center-of-mass frame. Nonetheless, the calculations indicate why the event distributions of neutrinos and antineutrinos differ in y in the LAB frame.

The Bjorken y distributions of the different event types passing the CC inclusive selections are shown in Fig. 7.6. The left plot of Fig. 7.6 shows the true y distribution and the right plot is the reconstructed y distribution. As seen from the figures, the neutrino component (red, lower histogram) has a relatively flat distribution, slowly decreasing to zero as Bjorken y increases. On the other hand, the antineutrino component (blue, upper histogram) has a peak at low y but then decreases rapidly with increasing y. This steeper falloff for the antineutrino component is reflected by the y-dependence of the cross section as evaluated in Eq. (7.10). The distribution in both true and reconstructed quantities is similar. However, there is a depletion in low reconstructed y, compared to true y, for true $\bar{\nu}_\mu$-CC events. This is because the energy of the outgoing neutron cannot be measured directly.

There still exists a small neutral current component in the sample. The reconstructed y calculator uses the energy of misidentified hadrons to calculate Bjorken y. This calculation is wrong and the Reco y distribution for the NC sample does not represent the true y distribution. However, the NC contribution is very small and so these considerations do not have impact.

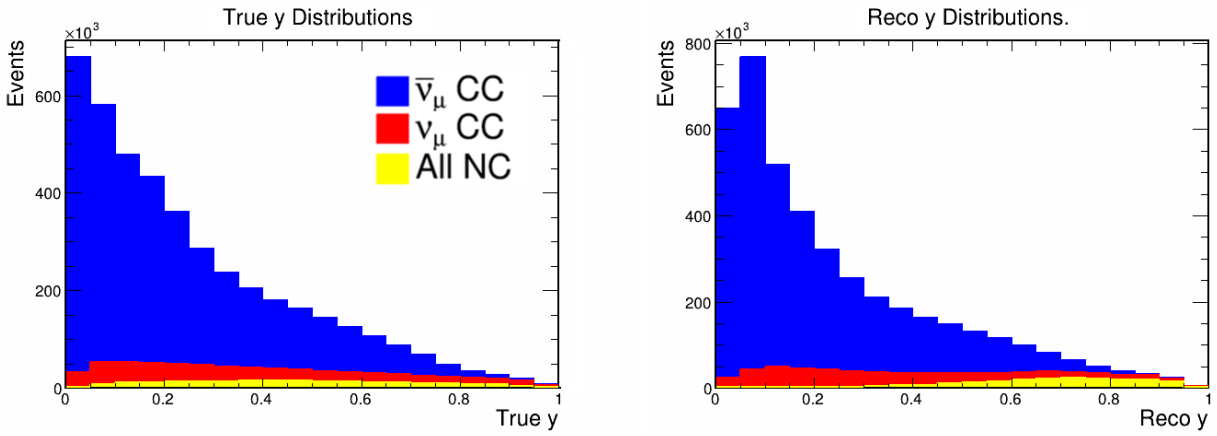

Figure 7.6: Bjorken y distributions for CC-$\bar{\nu}_\mu$, CC-ν_μ and NC components (blue, red and yellow histograms respectively) of the selected event sample. The left (right) plot shows the distribution in true (reconstructed) y.

A plot of the absolute resolution (or residual) for Bjorken y is shown in Fig. 7.7. The residual of an event is defined as: $(True\ y - Reconstructed\ y)$ for that event. The residuals are mostly centered around 0. However, there is a bias towards positive y, indicating that the reconstructed y often underestimates the actual y. This bias may arise from neglecting the initial Fermi momentum of the target nucleon.

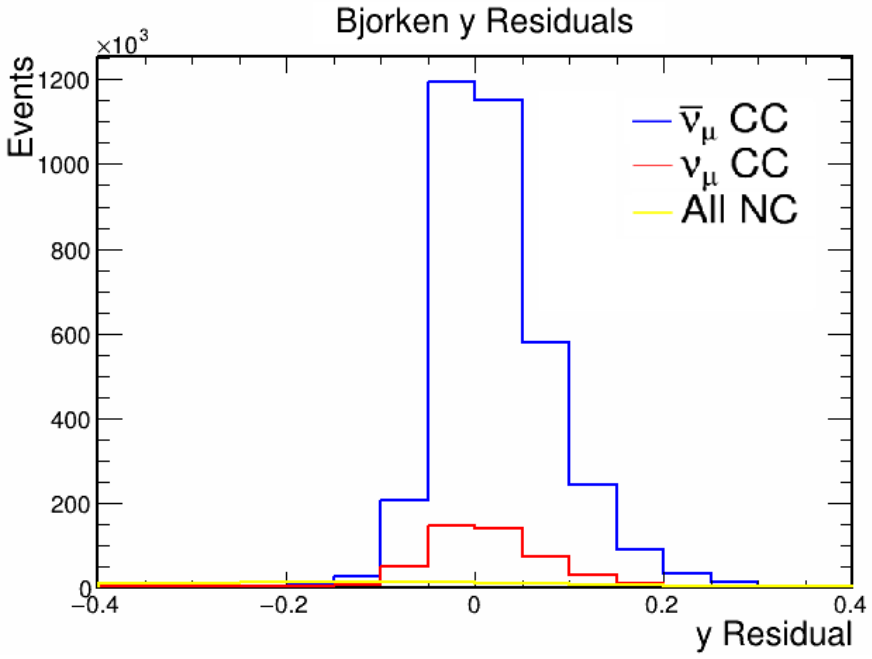

Figure 7.7: Residual of the reconstructed y variable for the antineutrino and neutrino components in the selected event sample. The full-width-half-maximum for the CC-$\bar{\nu}_\mu$ distribution is 0.10.

A method for obtaining a data-driven constraint for the background has been explored previously; details are presented in Ref. [27]. The procedure involves utilizing the following function to develop a prediction of the event distribution:

$$N_P(y_i) \;=\; fN_D^{Total}B(y_i) + (1-f)N_D^{Total}S(y_i). \tag{7.13}$$

Here, $N_p(y_i)$ is the predicted number of events in the i-th bin in Bjorken y, N_D^{Total} is the total number of data events that are being analyzed, f is the fraction of the total number of events in the sample that are background events; $B(y_i)$ and $S(y_i)$ are the values of the unit-normalized predictions of the background (consisting of CC-ν_μ, NC, $\nu_e/\bar{\nu}_e$ and nonfiducial events) and the signal (all CC-$\bar{\nu}_\mu$ events) respectively. The nominal background is estimated to be around 10%. To fit the prediction $N_p(y_i)$ to the data, the variable f is varied from 0.5 to 1.5 in increments of 0.1. The closeness of the prediction to the data is measured via χ^2, with a smaller value representing a better fit. The value of f that yields the lowest χ^2 is retained, as represents the best estimate of the fraction of the total events that constitutes the background.

The procedure was found to consistently overestimate the value of f, and therefore has not been used to estimate the fraction of background in the selected sample of events.

7.4.1.2 Separation using muon longitudinal momentum

Another variable with the potential to quantify the contribution of wrong-sign, NC, and Other background events from signal events is the muon longitudinal momentum, $P_{||}$, the component of the muon in the direction of the incoming antineutrino/neutrino. The "Other" category is comprised of nonfiducial events and $\nu_e/\bar{\nu}_e$ interactions. The event distribution of $\bar{\nu}_\mu$-CC events, ν_μ-CC events and NC/Other events in the detector are shown in Fig. 7.8 (left). The event distribution of true signal events peaks at 1.50 GeV/c, whereas the event distributions of wrong-sign events and NC/Other events peak in the vicinity of 0.65 GeV/c.

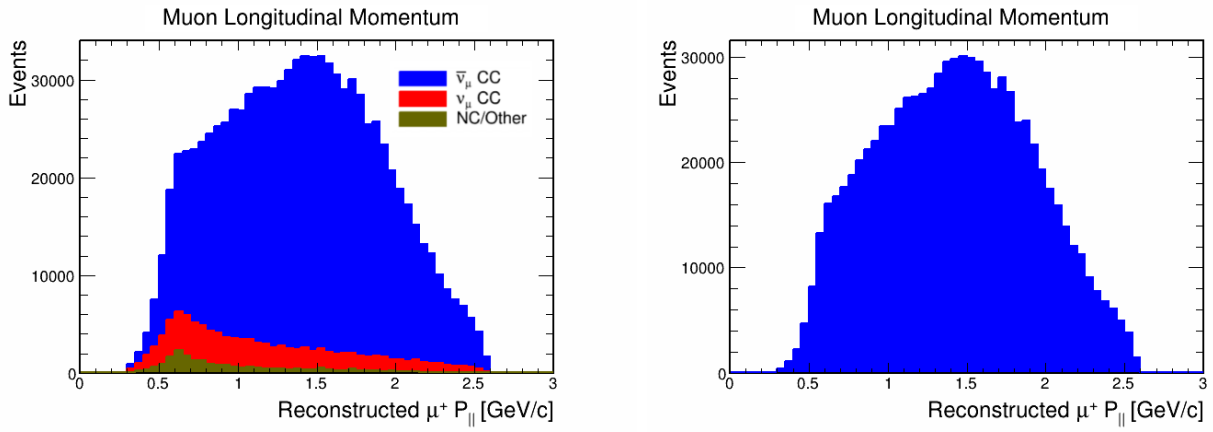

Figure 7.8: Event distributions for $\bar{\nu}_\mu$-CC, ν_μ-CC and NC components (blue, red and yellow histograms respectively) of the selected event sample muon longitudinal momentum. On the left, the muon longitudinal momentum distributions of the different reaction categories are stacked. The plot on the right shows the CC-$\bar{\nu}_\mu$ event distribution by itself.

The 'low-shoulder' in $P_{||}$ that arises from NC/Other events is even more pronounced when the sample is restricted to events with $E_{avail} > 100$ MeV, as shown in Fig. 7.9. These observations motivate an approach wherein the NC/Other normalization is adjusted to account for the shoulder observed at low $P_{||}$, and then the normalization for the wrong-sign and signal events are subsequently adjusted to match the total selected sample distribution shown in Fig. 7.8.

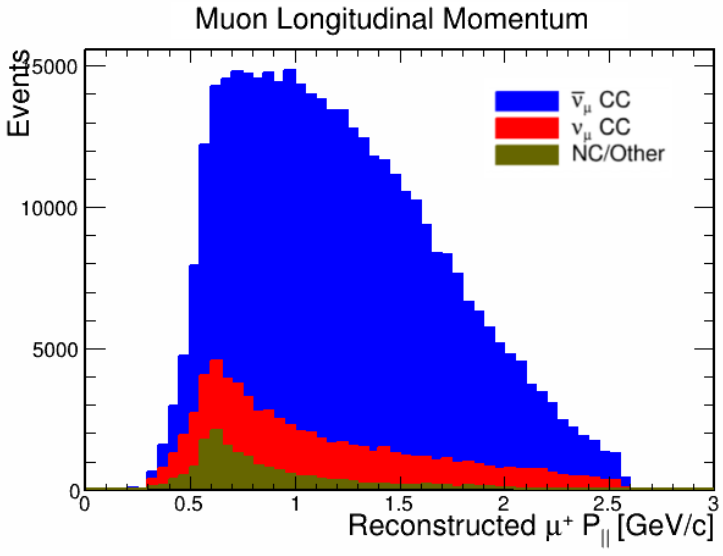

Figure 7.9: Event distributions of $\bar{\nu}_\mu$-CC + ν_μ-CC events and of NC/Other interactions in muon longitudinal momentum. The events are required to satisfy $E_{avail} > 100$ MeV.

The procedure for obtaining a data-driven constraint on the NC/Other and wrong-sign contri-

butions consists of two iterations of two fitting stages. A number of fitting approaches were explored and the procedure presented below is the one that has been found to provide the best constraints on the wrong-sign and NC/Other backgrounds.

The first stage consists of a two-template fit. For events with reco $E_{avail} > 100$ MeV, a combined ν_μ CC + $\bar{\nu}_\mu$ CC template and an NC/Other template are fit to the data in the $P_{||}$ variable with MINUIT [28]. MINUIT varies the normalizations of the two templates until the total prediction best matches the data. An example of this fit is shown in Fig. 7.10. The normalizations of the two templates are retained for the next stage.

A metric that shows how well the MC prediction matches the data event distribution is the χ^2/DoF (chi-square per degrees of freedom). The χ^2 per bin is the square of the difference between the MC prediction and the data event counts divided by the sum of the squares of the errors associated with the prediction and data. The error considered is the statistical error only, which is the square root of the event count in each bin. The χ^2 across the full range of $P_{||}$ in Fig. 7.10, is the sum of the chi-square value from each bin that has events in it. The degrees of freedom are the number of bins minus the number of parameters being floated, which is two parameters for both stages of the fit.

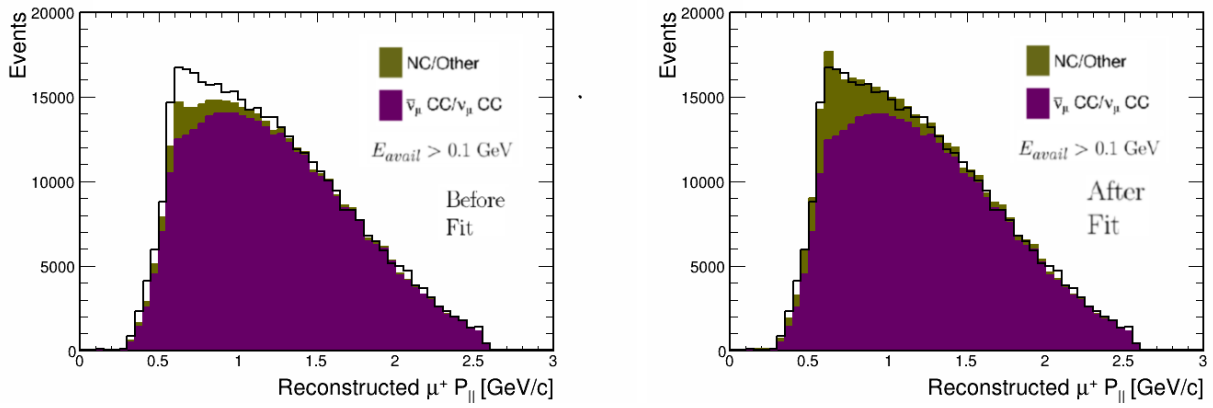

Figure 7.10: Demonstration of the first stage of a two-stage fit of ν_μ + $\bar{\nu}_\mu$ CC templates and NC/Other templates to data in the variable $P_{||}$. MINUIT adjusts the event rate normalizations of the two interactions until the prediction best matches the data.

A smaller χ^2/DoF usually indicates a better match. To confirm that MINUIT correctly performs the fitting, the χ^2/DoF values of the MC with respect to the data before and after the fit were

examined for 100 systematically shifted universes. A significant reduction in the χ^2/DoF is observed after the fit is carried out, indicating MINUIT does fit the MC to the data correctly.

A two-parameter fit is carried out in the second stage, consisting of the templates of the ν_μ-CC, $\bar{\nu}_\mu$-CC and NC/Other event distributions, of events that pass the CC inclusive selections. The procedure is illustrated in the plots of Fig. 7.11. First, the normalization of the ν_μ-CC + $\bar{\nu}_\mu$-CC template from stage 1 is applied to the $P_{||}$ event distributions of the ν_μ-CC and $\bar{\nu}_\mu$-CC interactions in stage 2. Likewise, the NC/Other template normalization from the previous fit is applied to NC/Other template. For this stage, the NC/Other and $\bar{\nu}_\mu$ CC templates are allowed to float independently and the ν_μ CC template is fixed. MINUIT adjusts the normalizations of the signal and NC/Other templates until the prediction best matches the data, as shown in Fig. 7.11. The normalizations of the templates are retained.

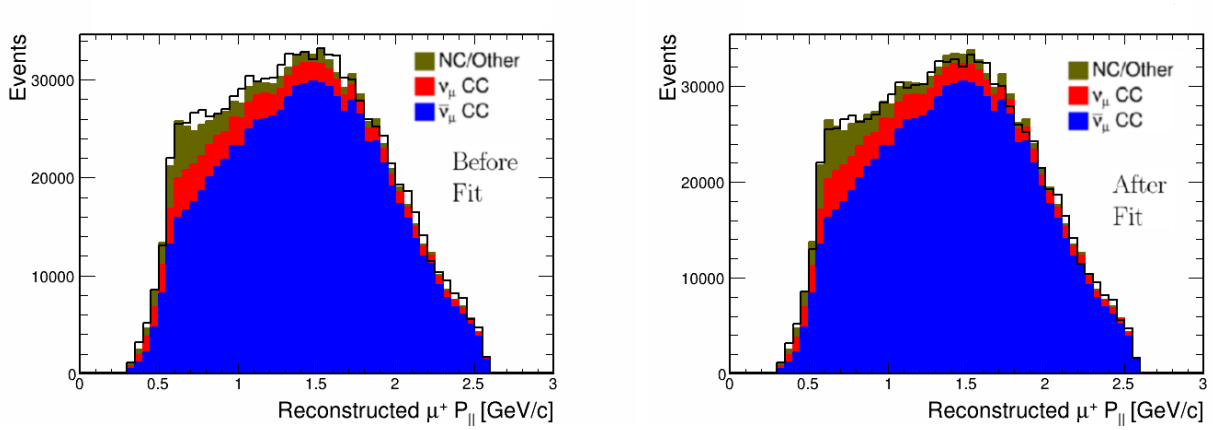

Figure 7.11: Demonstration of the revised second stage of a two-stage fit of ν_μ CC, $\bar{\nu}_\mu$ CC templates and NC/Other templates of events that have passed the CC inclusive selections to data in the variable $P_{||}$. MINUIT adjusts the event rate normalizations of the two interactions until the prediction best matches the data.

The above two stages are then repeated for a second iteration. The normalizations of the templates from stage 2 are applied as starting $P_{||}$ predictions for $\bar{\nu}_\mu$-CC, ν_μ-CC and NC/Other templates of event with $E_{avail} > 100$ MeV. The $\bar{\nu}_\mu$-CC and ν_μ-CC distributions are consolidated, and then the first-stage two parameter fit is carried out. MINUIT adjusts the normalizations of the two templates until the prediction matches the data. This process is illustrated in Fig. 7.12.

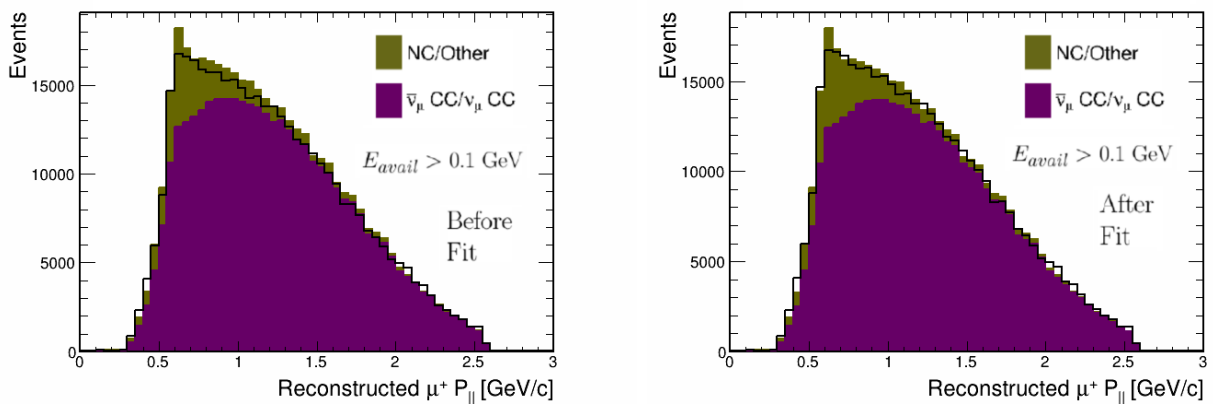

Figure 7.12: Demonstration of the second iteration first stage of a two-stage fit of ν_{mu} + $\bar{\nu}_\mu$ CC templates and NC/Other templates to data in the variable $P_{||}$. The normalization results from the first iteration of fit stage 2 are applied first; then MINUIT adjusts the event rate normalizations of the two interactions until the prediction best matches the data.

As previously, the normalization of the $\bar{\nu}_\mu$-CC + ν_μ-CC template is applied to the $P_{||}$ event distribution of true signal and wrong sign events passing the CC inclusive selections. The NC/Other event distribution is given the normalization of its counterpart from the previous fit. The second stage of the fit is repeated, with the $\bar{\nu}_\mu$-CC and NC/Other templates allowed to float. An example of this fit is shown in Fig. 7.13.

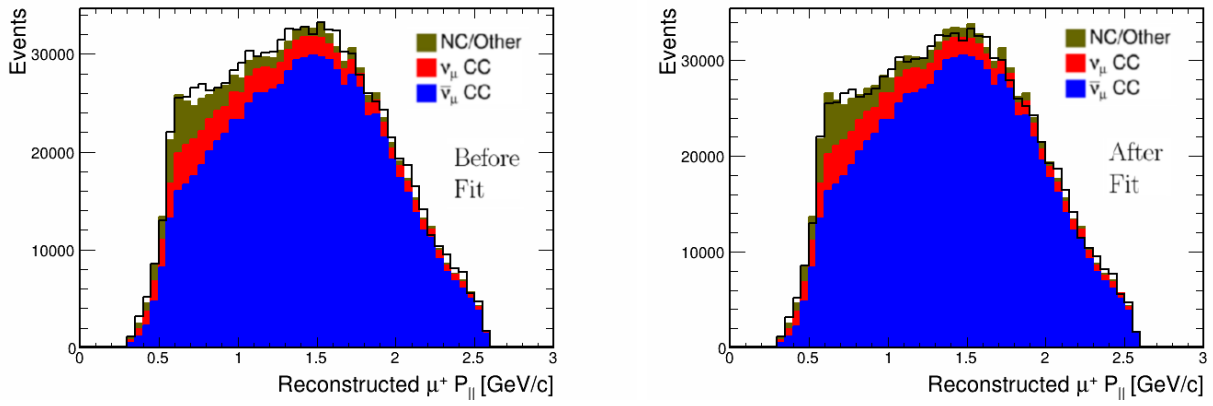

Figure 7.13: Illustration of the second stage, second iteration fitting process. The results from the first stage are applied and MINUIT is then allowed to float the $\bar{\nu}_\mu$-CC and ν_μ-CC distributions to match the prediction to the data.

The final normalizations of the $\bar{\nu}_\mu$-CC, ν_μ-CC, and NC/Other templates are applied to their counterparts in $|\vec{q}|$ and E_{avail}. This provides a data-driven constraint on the background estimates. The application of normalizations changes the estimated event rates of the background

interactions, as can be seen in Fig. 7.14. The projections of the background event distributions onto each analysis variable are presented in Figs. 7.15 and 7.16. A noticable reduction in the NC/Other background is obtained, particularly in the region 0.7 GeV/c$<$ $|\vec{q}| \leq$ 2.0 GeV/c. The reduction in the wrong-sign event rate is event rate is modest by comparison.

From the projection plots in Figs. 7.15 and 7.16, it can be seen that there is some improvement in the background event estimates. For the NC/Other interaction, a noticeable improvement is observed across all $|\vec{q}|$ bins. However, there is a noticeable overestimation of events in the lowest E_{avail} bin. The same pattern is observed for the wrong-sign event estimation, except that the event rate in the lowest available energy bin is further underestimated.

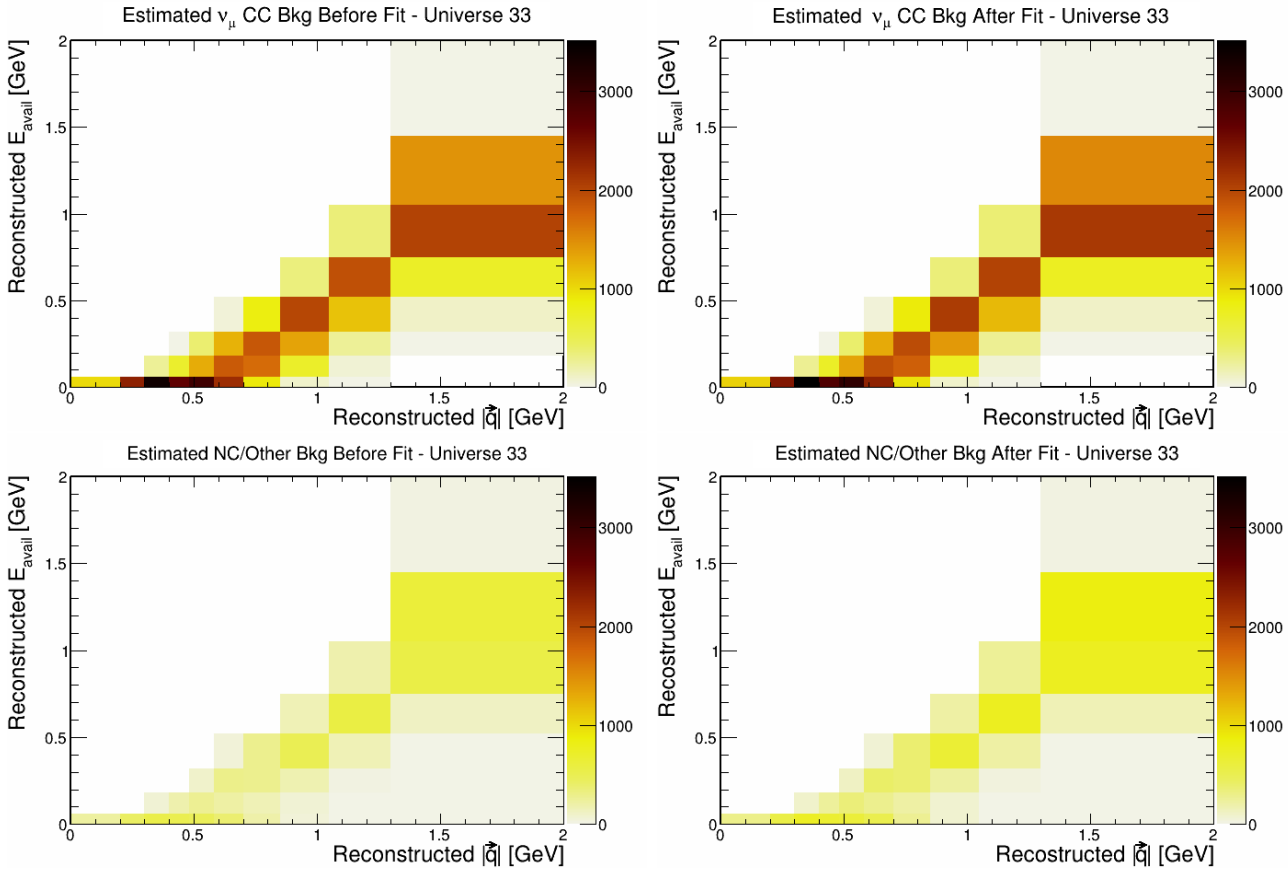

Figure 7.14: Estimated $|\vec{q}|$ and E_{avail} distributions of the ν_μ CC and NC/Other background interactions (top, bottom) before and after the two-stage two-iteration fit (left, right). The fit changes the normalizations of the templates.

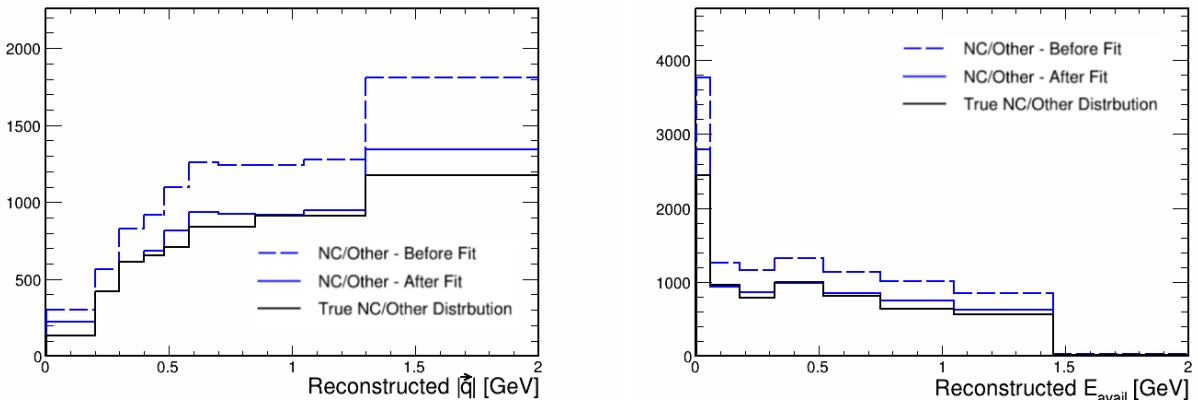

Figure 7.15: Estimated $|\vec{q}|$ and E_{avail} distributions of the NC/Other background interactions before and after two iterations of the fit.

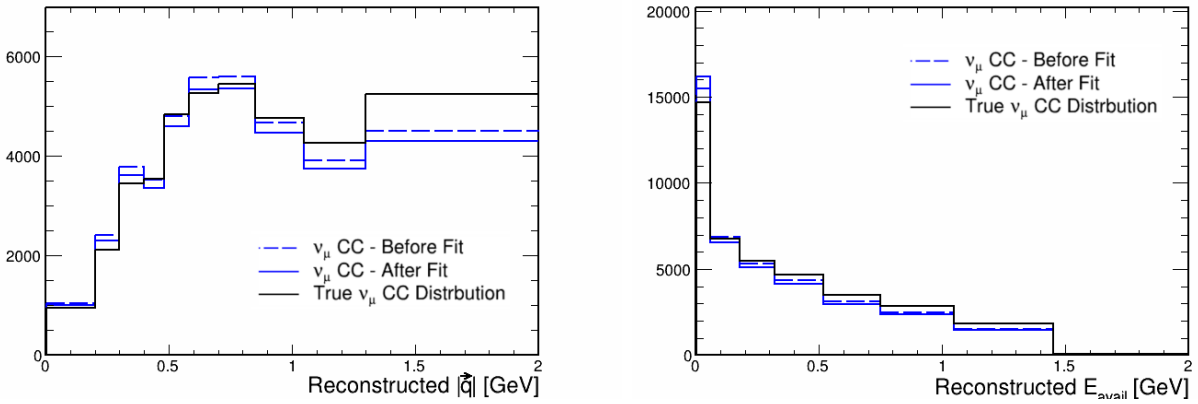

Figure 7.16: Estimated $|\vec{q}|$ and E_{avail} distributions of the ν_μ CC background interactions before and after two iterations of the fit. The fit improves the estimate of the template, only further underestimating the event rate in the lowest E_{avail} bin.

The $\bar{\nu}_\mu$ CC event distribution in $|\vec{q}|$ and E_{avail} is obtained from the data by subtracting the adjusted background event distributions. The $|\vec{q}|$ and E_{avail} projections of the signal estimate obtained for this universe before and after fits, as well as the true signal event distribution are shown Fig. 7.17. In this universe, the fitting procedure modestly improves the overall signal estimation.

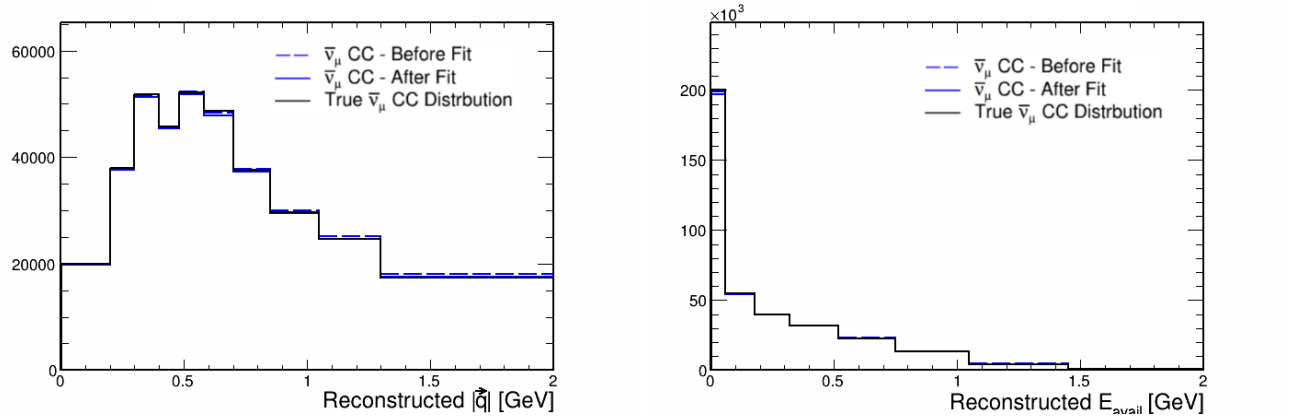

Figure 7.17: The $|\vec{q}|$ and E_{avail} projections of the estimated signal event distributions before and after fit (dashed blue and solid blue lines respectively). The true signal event distribution is shown with the solid black line. In this universe, the fitting procedure modestly improves the estimates of the backgrounds while the signal estimate remains nearly the same.

To examine how well the fitting sequence works in yielding the correct background estimates, the fitting procedure is applied to fake data samples in 200 universes. The residual of the estimated NC/Other event count versus that of the true NC/Other event count is calculated before and after the fit in each universe. The residual comparison distributions are shown in Fig. 7.18. Good estimates bring the residual value closer to zero, and that is what is observed in the plots of Fig. 7.18. The residual values obtained after the fit have a mean value of -0.03 GeV, compared to 0.18 GeV before the fit. The same metric was also applied to the wrong-sign estimation, and the results are displayed in Fig. 7.19. For the wrong-sign component, the residual distributions span almost the same range of values. The mean of the residual values for the ν_μ-CC wrong-sign estimation after the fit is -0.007 GeV, compared to 0.018 GeV before the fit, indicating a modest improvement.

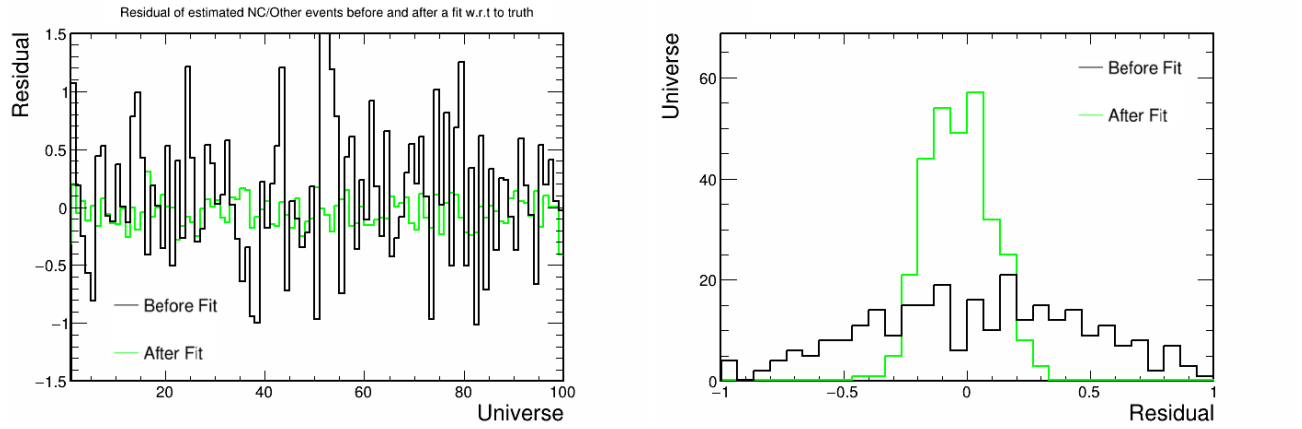

Figure 7.18: Residual of the estimated NC/Other event count with respect to truth before (black) and after (green) the interaction templates in $P_{||}$ have been constrained. The left plot shows a universe by universe comparison and the right plot compares the residual distributions before versus after the fitting procedure.

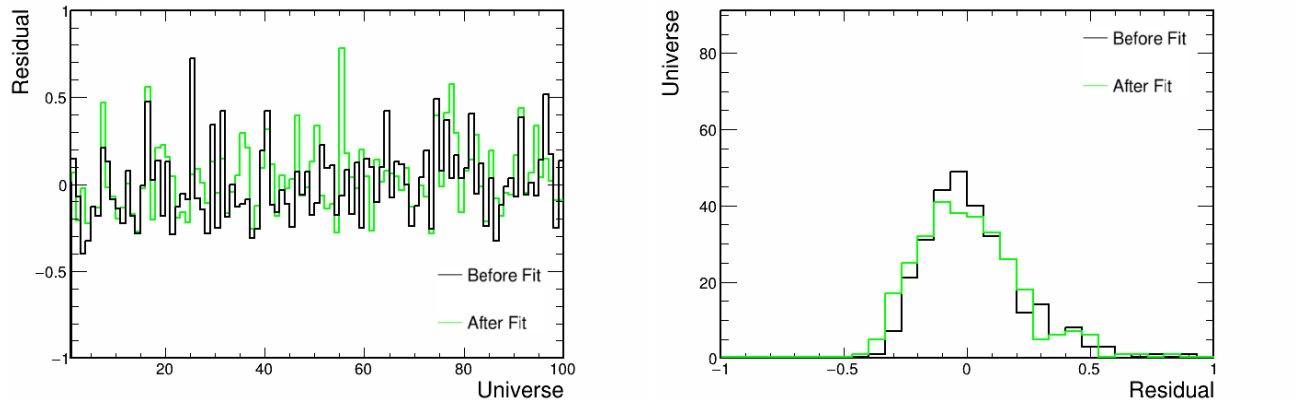

Figure 7.19: Residual of the estimated ν_μ-CC with respect to truth before (black) and after (green) the interaction templates in $P_{||}$ have been constrained. The distribution of residuals before and after fit, shown on the right, indicate that the fitting procedure modestly improves the estimate, with a mean value that is closer to zero.

For each universe, the residual of the estimated signal event count with respect to true signal event count before and after the fitting procedure was calculated. The results are shown in Fig. 7.20. The mean value of the residuals after the fit is -0.008, compared to 0.005 before the fit. However, the residual values after the fit have a tighter spread about 0, with a full-width half maximum (FWHM) of 0.03. The FWHM of the residual distributions before the fit is 0.05. This indicates that the procedure provides a modest constraint on the background estimation.

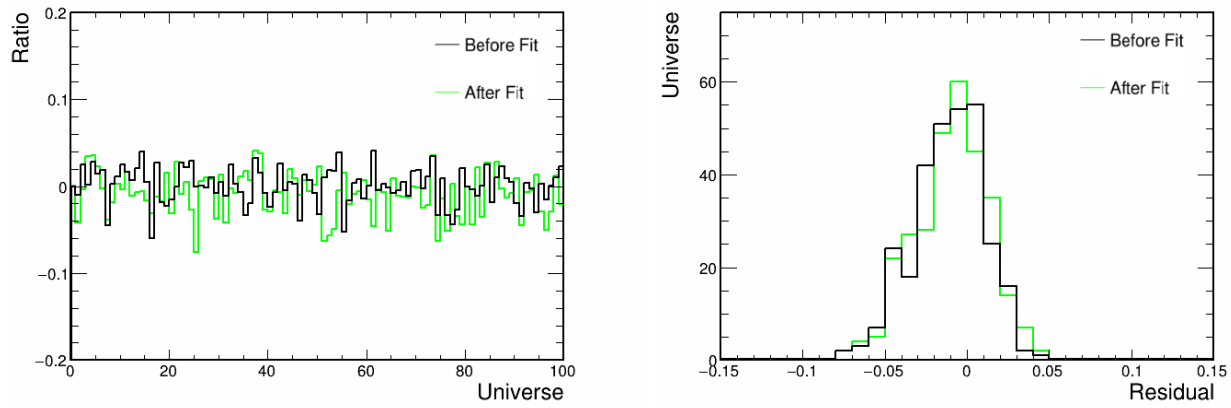

Figure 7.20: Residual of the estimated signal with respect to truth before (black) and after (green) the interaction templates in $P_{||}$ have been constrained. The right plot, which summarizes the residual values before and after the fit, show that the fit improves the signal estimate, where the residual values after the fit distribute more tightly around zero.

The residual plots in Figs. 7.18, 7.19 and 7.20 show that our two-stage, two-iteration fit in the $P_{||}$ variable gives on average a better result than simply subtracting the GENIE nominal predictions of the background event distributions from the data. The fit provides constraints on the two leading backgrounds, namely ν_μ-CC wrong-sign and NC/Other events. The average gain for extracting true signal is quite modest, but having a constraint on backgrounds is useful because it reduces the systematic errors arising from the background estimation. With this method, observed features of the data have been utilized to derive better estimates of the backgrounds.

Chapter 8

Selection Efficiency and Sample Purity

The CC inclusive selections mentioned in Sec. 4 do not capture all signal events because the reconstruction fails for certain events. For example, when the event is initiated by a very low energy neutrino, there may not be enough hits for the interaction to be properly isolated. Or, when a high energy muon or charged hadron from a CC event escapes the detector, the event cannot be fully reconstructed. However, the cross section measurement must still account for these events, making an efficiency correction necessary. The efficiency of the signal detection and selection is a measure of how well the analysis is capable of isolating signal events. The CC inclusive selection criteria allow some NC and wrong-sign neutrino events to enter the candidate signal sample. They do, however, almost completely reject all electron neutrino events. The purity of the selection criteria is the proportion of true signal events relative to the total number of events, including NC and ν_μ-CC events.

8.1 Selection efficiency

The efficiency is defined according to

$$Efficiency = \frac{Selected\,True\,Signal\,Events}{All\,True\,Signal\,Events}. \tag{8.1}$$

True signal events are defined as true $\bar{\nu}_\mu$-CC events, excluding $\bar{\nu}_\mu + e^-$ events, with the true vertex of the event occurring within the fiducial volume and with the muon truth kinematics satisfying the muon phase space cuts listed in Table 1. There are true $\bar{\nu}_\mu$-CC events that fail the true vertex fiducial volume or phase space requirements, and hence get categorized as background.

The overall selection efficiency of signal events is shown in Fig. 8.1. Above a boundary, there is a region that has zero efficiency. This region above the boundary is occupied by a few antineutrino-on-electron events. In general, the efficiency is around 20% across the phase space, with a higher efficiency (40% - 70%) close to the boundary region where numerical values of $|\vec{q}|$ are approximately equal to the numerical values of E_{avail}. The efficiency distribution as a function of each analysis variable is displayed in Fig. 8.2. It can be seen that the efficiency is very high for events with true $|\vec{q}| < 500$ MeV/c. Above this 500 MeV/c threshold, the efficiency is relatively flat around 20%. In terms of E_{avail}, the efficiency is relatively constant at 30%. With a higher efficiency at around 35% for $E_{avail} < 200$ MeV.

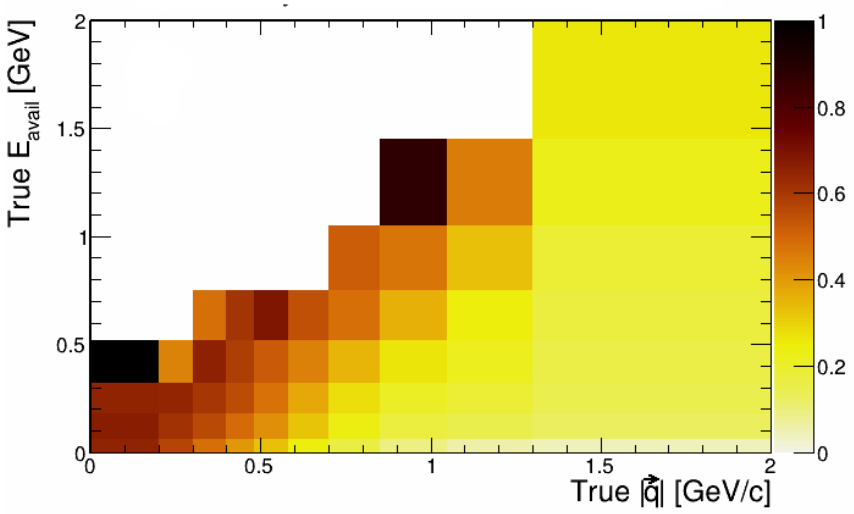

Figure 8.1: Selection efficiency of true signal events. The efficiency is constant at 20% across most of the phase space. The efficiency is higher around the boundary region.

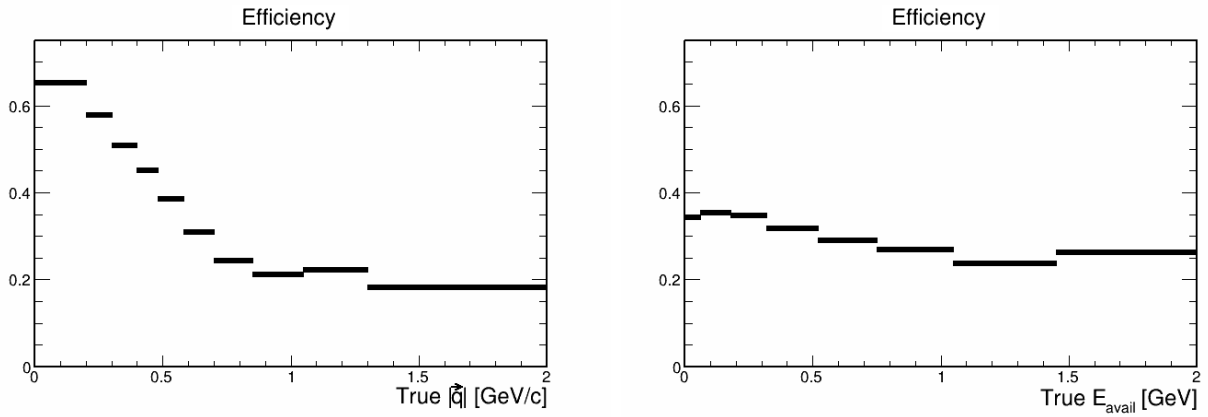

Figure 8.2: Selection efficiency as a function of true $|\vec{q}|$ (left) and of true E_{avail} (right).

The CC inclusive selections is composed of a number of criteria. To better understand the overall selection efficiency, it is useful to examine how the efficiency is affected as each successive criteria is applied; this is shown in Fig. 8.3. The efficiency distribution for selecting signal events is very similar to that observed for ν_μ-CC events from the FHC sample [29], with a higher efficiency near the border region, and a relatively flat efficiency distribution across the remaining phase space.

The quality cut and the track reconstruction have no observable effects on the efficiency, as shown by Fig. 8.3(a). The muon identification requirement (Fig. 8.3(b)) reduces the efficiency to 83.6%. The efficiency is reduced further to 81.8% with the application of the vertex fiducial volume requirement (Fig. 8.3(c)). In both cases, the change in efficiency is not uniform across the phase space. Rather, the efficiency drop is higher at higher values of $|\vec{q}|$ and E_{avail}. The muon containment criterion (Fig. 8.3(d)) causes the efficiency to drop to 37.8%; it is the cut that has the largest effect. Again, the effect is not uniform; the efficiency is much higher in the vicinity of the boundary region. The successive application of muon efficiency and track/shower containment cuts (Figs. 8.3(e) and (f)) reduce the overall efficiency to 36.2% and 33.1%. The effect of successively applying the selection criteria one-by-one is summarized in Table 8.1.

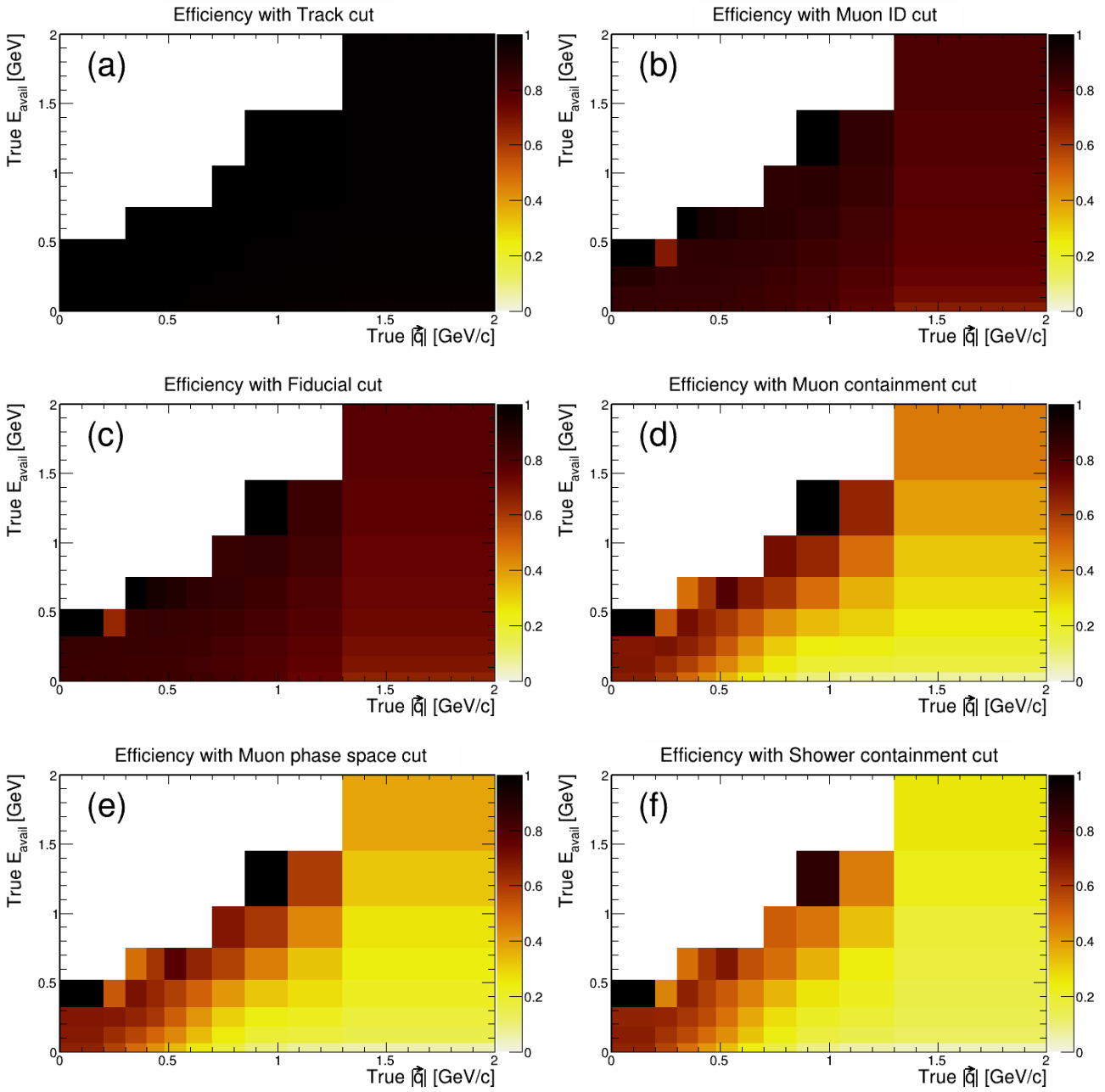

Figure 8.3: Effect of selection criteria on efficiency. The track fiducial (a), the muon ID (b), and fiducial cut (c) reduce the efficiency at higher $|\vec{q}|$ and E_{avail} values. The muon containment cut (d) affects efficiency the most. The muon phase space and shower containment cuts (e) and (f) reduce the efficiency at the boundary region.

Selection Cut	Selected Signal Events	Efficiency
No Cut	2,560,540	100%
Quality	2,555,200	99.8%
Track	2,540,870	99.4%
Muon ID	2,140,340	83.6%
Fiducial	2,095,690	81.8%
Muon Containment	968,303	37.8%
Muon Phase Space	927,617	36.2%
Track/Shower Containment	847,821	33.1%

Table 8.1: Effect of selection criteria on the selection efficiency for the signal $\bar{\nu}_\mu$-CC sample. The largest effect is observed when the muon containment cut is applied. The selection criteria reduce the number of true signal events in the sample from over 2.5 million events to just under 847k events.

8.2 Sample purity

The purity of the selected sample is the proportion of true signal events relative to the total number of events, including NC and neutrino event backgrounds.

In equation form, purity is

$$Purity \; = \; \frac{Selected\, True\, Signal\, Events}{All\, Selected\, Events}. \tag{8.2}$$

The purity distribution of $\bar{\nu}_\mu$-CC events in the selected sample as a function of true $|\vec{q}|$ and E_{avail} is shown in Fig. 8.4.

The sample purity is greatest in the lowest E_{avail} slices. This observation contrasts with what is observed for ν_μ-CC events in the FHC sample, which has a more nearly constant purity across the phase space [29]. The low-E_{avail} region is mostly occupied by $\bar{\nu}_\mu$-CCQE and $\bar{\nu}_\mu$-CC 2p2h-MEC events. The sample purity decreases with increasing E_{avail}. This occurs because ν_μ-CC wrong-sign and NC events occupy these regions. The purity is observed to be the lowest close to the boundary region with $|\vec{q}|$ values 200 MeV/c - 1 GeV/c and E_{avail} values 300 - 750 MeV. In this low purity region, $\bar{\nu}_\mu$-CC signal events make up 63% of the population. Background $\bar{\nu}_\mu$-CC comprise 10% of the events here. The remaining 27% of events is made up of wrong-sign events.

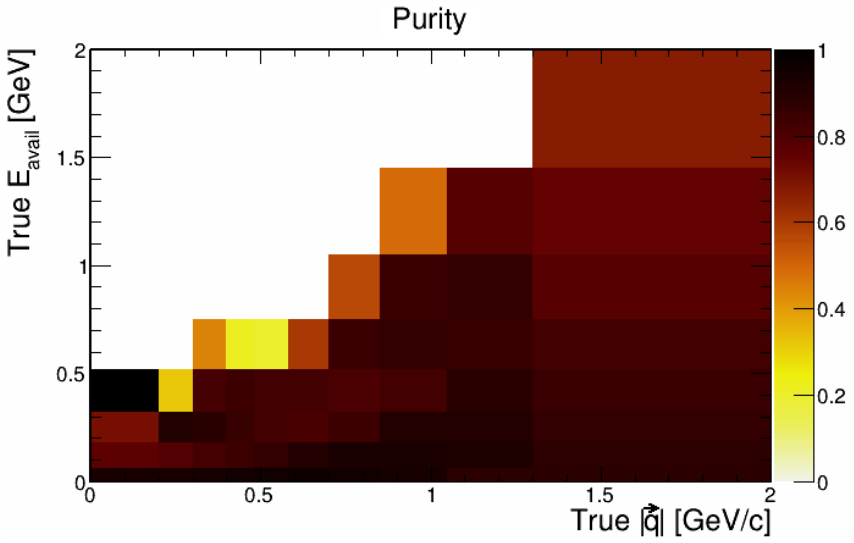

Figure 8.4: Purity distribution of true signal events. The sample purity is relatively flat, with a lower purity at higher values of $|\vec{q}|$, and E_{avail} and around the boundary region. These regions are also occupied by wrong-sign and NC/Other events.

The purity distribution in terms of individual variables is shown in Fig. 8.5. The purity is relatively flat between 0 - 1.3 GeV/c, with a lower purity in the ighest $|\vec{q}|$ bin. On the other hand, the purity is 95% on the lowest E_{avail} bin, but then averages to about 80% for higher E_{avail} values.

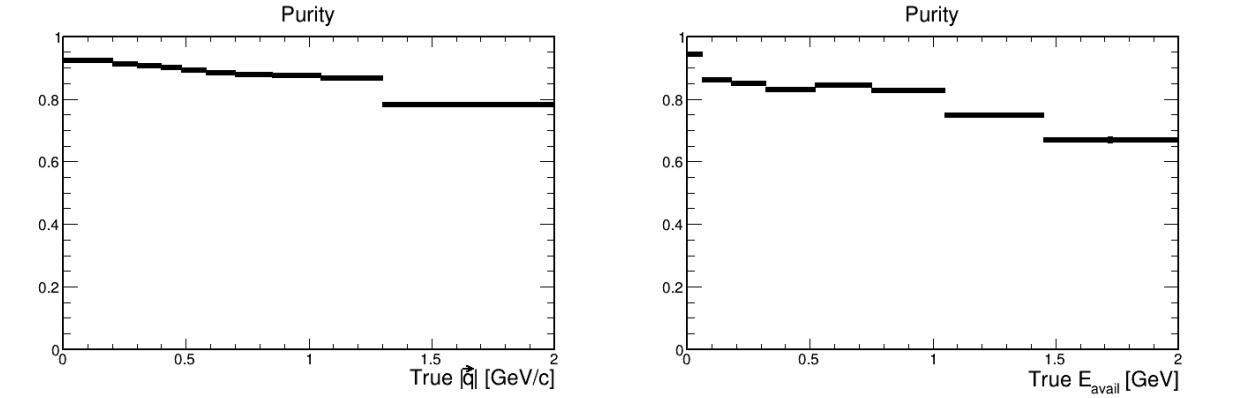

Figure 8.5: Sample purity as a function of true $|\vec{q}|$ (left) and of true E_{avail} (right).

The effect of individual selection cuts on the overall sample purity is shown in Fig. 8.6. The track and the muon ID cuts (Figs. 8.6(a) and (b) respectively) do not improve the purity of the sample, with the overall purity being at 5%. However, the application of the fiducial cut (Fig. 8.6(c))

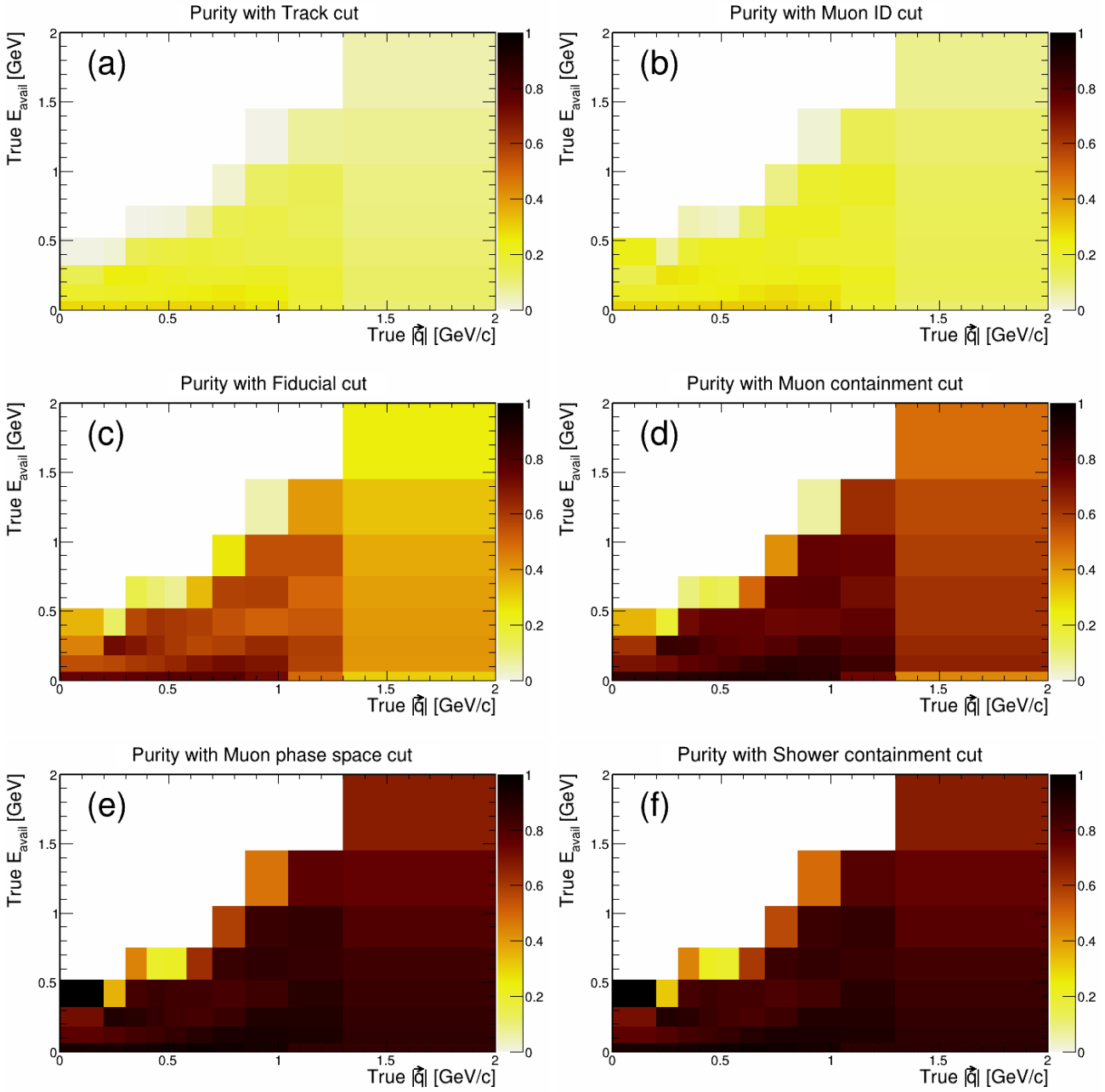

Figure 8.6: Effect of selection criteria on purity. The track, (a), and muon ID cut, (b) have little effect on the sample purity. The vertex fiducial volume improves purity at lower E_{avail}. The muon containment cut improves purity at higher $|\vec{q}|$. The most significant improvement comes from applying the phase space cut (e). There does not seem to be further observable improvement from the application of the shower containment cut (f).

gives visible improvement at some of the lower E_{avail} bins, and raises the overall purity to 8%. The muon containment cut (Fig. 8.6(d)) further improves the purity, especially in the region 0 GeV/c $< |\vec{q}| \leq$ GeV/c and E_{avail} between 0.0 - 1.5 GeV. The muon phase space cut (Fig. 8.6(e))

brings about the most significant improvement in purity across the phase space, save for some bins near the boundary region. The final overall sample purity is 88.57%. The effect of the cuts on the purity of the sample is summarized in Table 8.2.

Selection Cut	Selected Signal Events	Selected Events	Purity
No Cut	2,560,540	48,195,000	5.31%
Quality	2,555,200	31,469,400	8.12%
Track	2,545,870	13,182,400	19.31%
Muon ID	2,140,340	9,265,950	23.10%
Fiducial	2,095,690	3,425,580	61.18%
Muon Containment	968,303	1,214,920	79.70%
Muon Phase Space	927,617	1,047,660	88.54%
Track/Shower Containment	847,821	957,224	88.57%

Table 8.2: Effect of the selection criteria on the sample purity of signal $\bar{\nu}_\mu$-CC. The largest effect is observed when the muon phase space cut is applied.

8.3 Effect of removing individual criterion

It is useful to examine the effect of removing individual cuts from the inclusive selection criteria. The details are reported in Table 8.3. It can be seen that the muon containment requirement rejects the most signal events, namely 954,638 events. This is followed by the Muon ID criterion, which rejects 100,340 events. The the third leading cut for rejecting events is the track/shower containment requirement. Other cuts have much smaller effects on the selected event count.

Selection Cut	Selected Signal Events	Signal Events Removed
All Selected Events	847,821	-
Quality	847,821	0
Track	847,821	0
Muon ID	948,160	100,340
Fiducial	865,054	17,233
Muon Containment	1,802,460	954,638
Muon Phase Space	882,993	35,172
Track/Shower Containment	927,617	79,796

Table 8.3: Table showing the effect of removing individual cuts from the CC inclusive selection criteria. The total number of selected signal events is shown on the first row. The rows below show how many true signal events are selected if the cut shown in the left column is removed, as well as the number of events that do not pass when the cut is added back in.

Chapter 9

Data Unfolding

We seek to obtain measurements that represent reality as closely as possible. Such measurements enable meaningful comparisons to theoretical predictions. However, instrumentation distortions of detectors, in conjunction with statistical fluctuations affect our measured data [30]. It is necessary to correct these effects before the data is analyzed. Unfolding is a procedure designed to correct for detector distortions based upon knowledge of instrumentation responses to data inputs. For this analysis, the D'Agostini unfolding is utilized [31] and is implemented in ROOT using RooUnfold [32]. An important property of this unfolding method is the retention of the total event count during unfolding. The method is iterative, and the number of unfolding iterations can be adjusted as required to achieve reliable results.

9.1 Constructing the unfolding matrix

The unfolding matrix of this analysis is a 4-dimensional object. However, there is a way to display this 4-D object in two-dimensional form, as shown in Fig. 9.1. In this 2D unfolded matrix, the horizontal axis displays bin numbers for event distributions in the reconstructed three-momentum transfer, $|\vec{q}|$ and available hadronic energy, E_{avail}. The vertical axis is the truth counterpart to the bins of the aforementioned variables. On each axis, the $|\vec{q}|$ variable is combined with the E_{avail} variable to express the 2D binning scheme in 1 dimension. There are

12 bins for the $|\vec{q}|$ variable and 10 bins for the E_{avail} variable, including overflow and underflow bins. The bin ranges have been listed in Tables 1 and 2 in Sec. 5. In each axis of Fig. 9.1, each block of 10 bins on either axis represents all the E_{avail} bins in one slice of $|\vec{q}|$, and there are 12 such blocks on both axes.

While it is difficult to glean information from the unfolded two-dimensional matrix, the unfolding matrices for the individual variables can be extracted from this object. These are shown in Fig. 9.2. Ideally, the matrices should be diagonal, indicating that migration of events between different bins is rare. This situation is seen in the $|\vec{q}|$ unfolding matrix shown in the left-side plot of Fig. 9.2.

At low $|\vec{q}|$, there is very little migration between different E_{avail} or $|\vec{q}|$ values. This is because the three-momentum transfer limits how much available energy the hadronic system has. Very rarely does an event with low reconstructed $|\vec{q}|$ migrate to high $|\vec{q}|$. As $|\vec{q}|$ increases, there are noticably more migrations between the various E_{avail} bins. This migration in E_{avail} results from a wider range of energies being available for the primary neutron coming from CC interations. A neutron hit contributes to the reconstructed available energy, but not

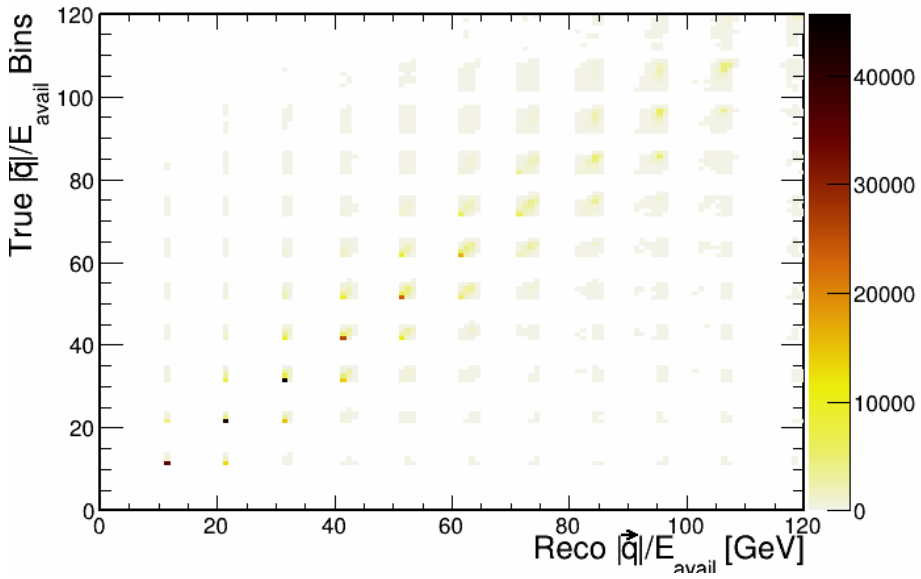

Figure 9.1: Unfolding matrix for variables $|\vec{q}|$ and E_{avail}. On each axis, block of 10 bins represents all the E_{avail} bins in one slice of $|\vec{q}|$.

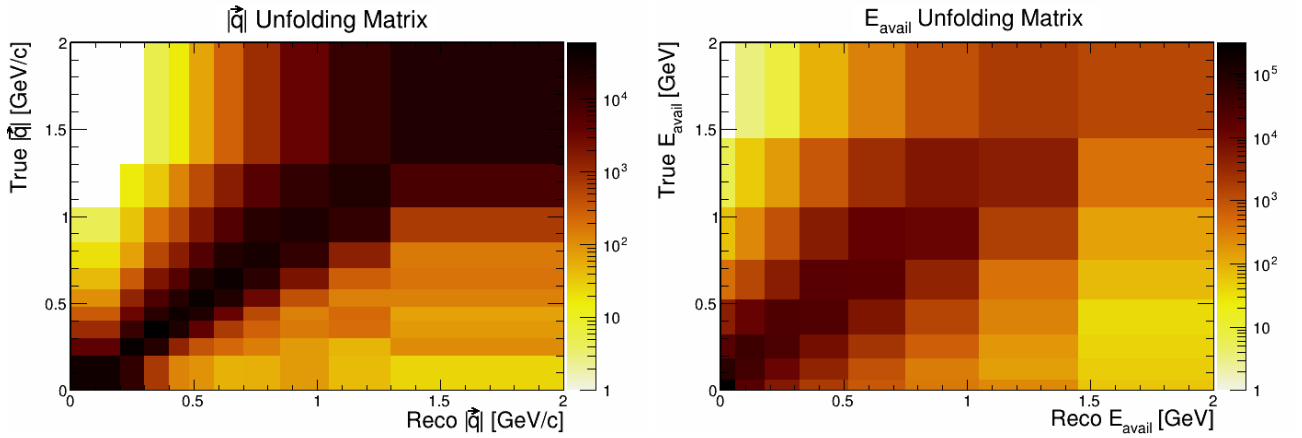

Figure 9.2: Unfolding matrix for each variable. The $|\vec{q}|$ is on the left and the E_{avail} unfolding matrix is on the right.

the true available energy. This results in overestimation or underestimation of the available energy, resulting in E_{avail} migration. There is also noticably more migration from very high $|\vec{q}|$ (1.5 GeV and higher) to mid-range $|\vec{q}|$ values (around 1 GeV), as evinced by the increased spread of off-diagonal elements with increasing bin number. Many of these migrations in the $|\vec{q}|$ variable result from nuclear breakup, with a high three-momentum transfer that is not observable because of the low energy of the exiting nucleons.

9.2 Testing the unfolding procedure

The raw signals collected from the detector (as estimated by realistic simulation) may be smeared by instrumentation effects and these will lead to event reconstruction that is distorted relative to the actual truth value. In general, a single unfolding iteration will not completely convert a reconstructed distribution to the truth, and it is necessary to carry out multiple unfolding iterations. The number of unfolding iterations should be kept to a minimum, however, in order to avoid biasing the data by the MC.

The performance of the unfolding matrix and the procedure can be examined by applying them to distributions from multiple universes. For this study, 500 universes were generated with randomized values assigned to GENIE cross-section parameters and to parameters that characterize

the flux uncertainties. This number is sufficient to capture the full spread of universes allowed by the systematics.

9.3 Unfolding in a typical universe

From a single universe, event distributions in true and reconstructed variables can be made; Fig. 9.3 shows an example. The horizontal axis in Fig. 9.3 represents a combination of the two analysis variables. Each block of 10 bins represents all the E_{avail} bins in one slice of $|\vec{q}|$. Lower values of $|\vec{q}|$ are on the left side of the plot, and the left-side bins in each block represent lower values of E_{avail}. The figure shows that the reconstructed event distribution (Reco, red histogram) approximates the MC truth distribution (True, black histogram), however there are some clear differences. As mentioned before, this smearing arises due detector effects plus statistical fluctuations.

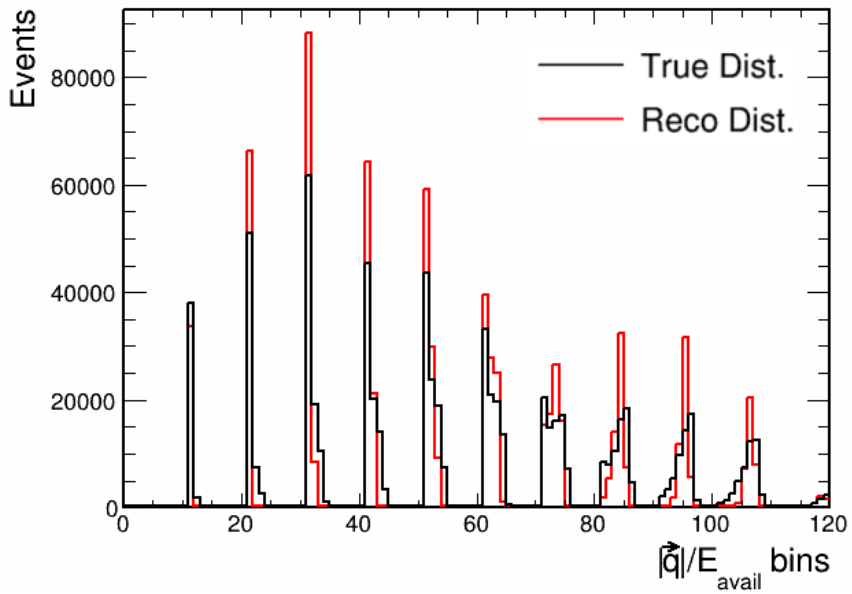

Figure 9.3: Unfolding matrix for variables $|\vec{q}|$ and E_{avail}. On each axis, block of 10 bins represents all the E_{avail} bins in one slice of $|\vec{q}|$.

The purpose of the unfolding procedure is to undo the smearing effects and yield the truth distribution underlying the measurement. In Fig. 9.4, the effect of the unfolding procedure on the reconstructed event distribution is shown for 4 unfolding iterations. Going from upper

left to upper right, to lower left and finally lower right, the resulting distribution coming from an additional unfolding iteration is shown. It can be seen that the first unfolding iteration significantly improves the event distribution estimate. Subsequent unfolding iterations improve the estimate in some bins but worsen it in others.

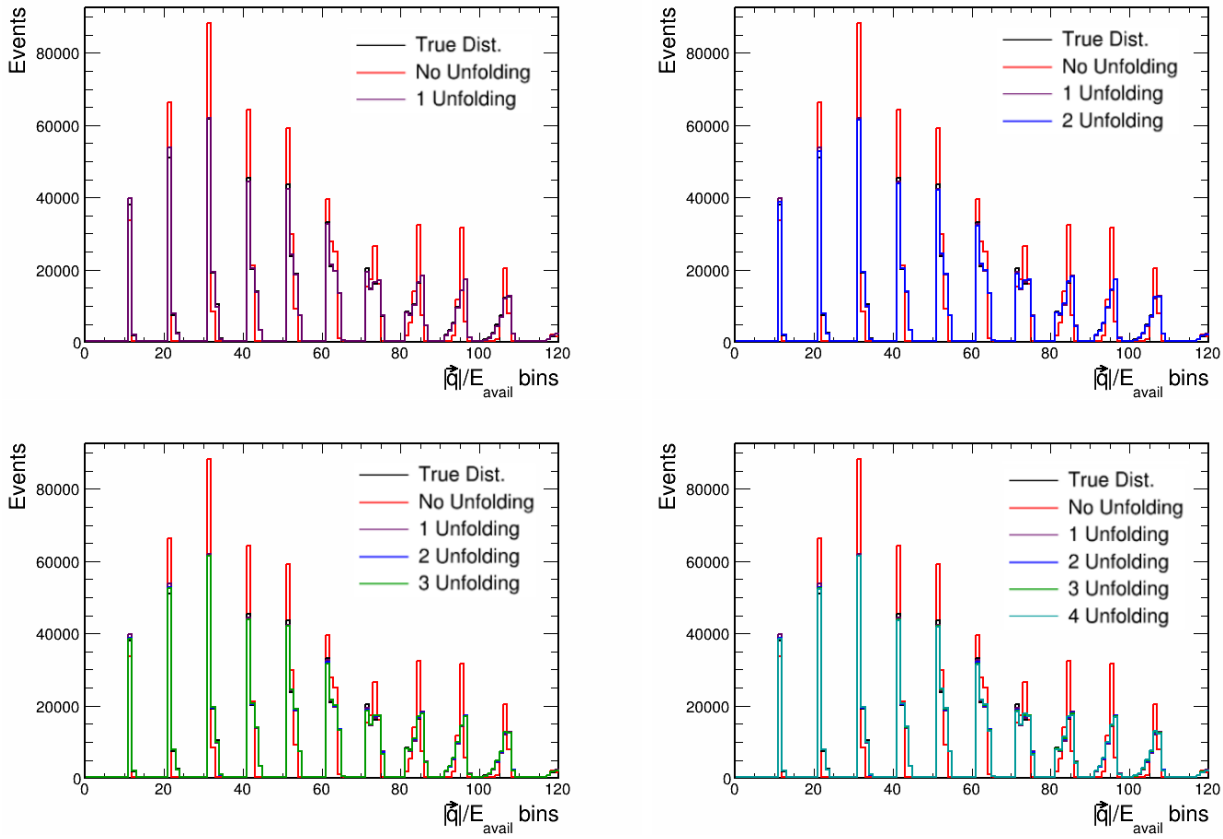

Figure 9.4: Effect of Unfolding iterations on the reconstructed event distribution in a typical universe. The first unfolding iteration significantly improves the event distribution estimate. Subsequent unfolding iterations make very small improvements to the estimated event distribution.

The overall difference between the true and reconstructed event distributions can be quantified with Mean Squared Error (MSE), which is defined as:

$$Mean\,Squared\,Error\ = \sum_{j=1}^{Bins} \frac{(Unfold_j\ -\ True_j)^2}{True_j}. \tag{9.1}$$

Here, the subscript j represents the bin number. The MSE changes with the number of unfolding iterations carried out, which is illustrated by Fig. 9.5. The first unfolding iteration significantly

undoes the smearing effects of the detector. However, more than a single unfolding iteration is required to minimize the MSE.

To determine the optimal number of unfolding iterations, the average MSE as a function of unfolding iteration is plotted for 500 universes according to

$$Avg\,MSE = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{Bins}\frac{(Unfold_{j,i} - True_{j,i})^2}{True_{j,i}}. \tag{9.2}$$

Here, the subscript i represents the index of the universe. A plot of the average MSE as a function of unfolding iterations is displayed in Fig. 9.6. The first unfolding iteration reduces the differences between true and reco distributions. However, it takes two or three unfolding iterations to minimize the average MSE.

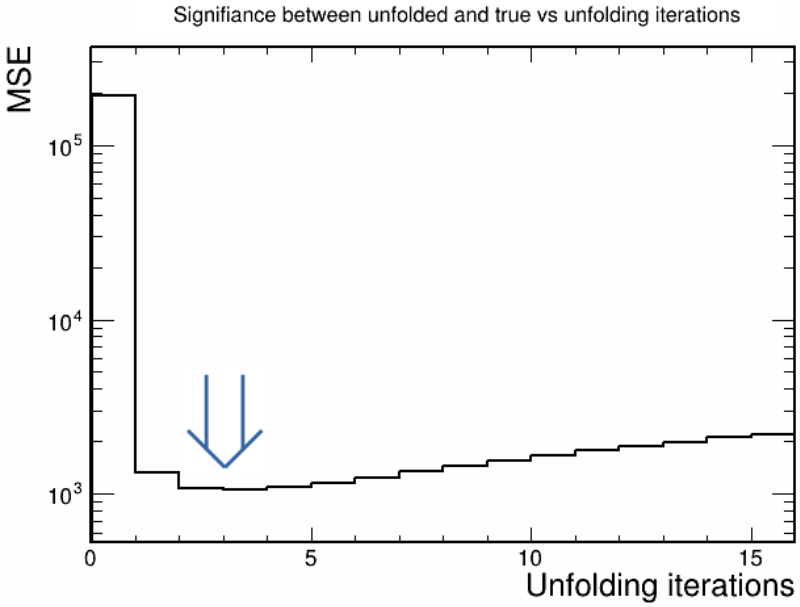

Figure 9.5: Mean Squared Error of the reconstructed unfolded distribution with respect to truth as a function of unfolding iteration for a typical universe. The first unfolding iteration significantly removes the smearing effects due to detector imperfections.

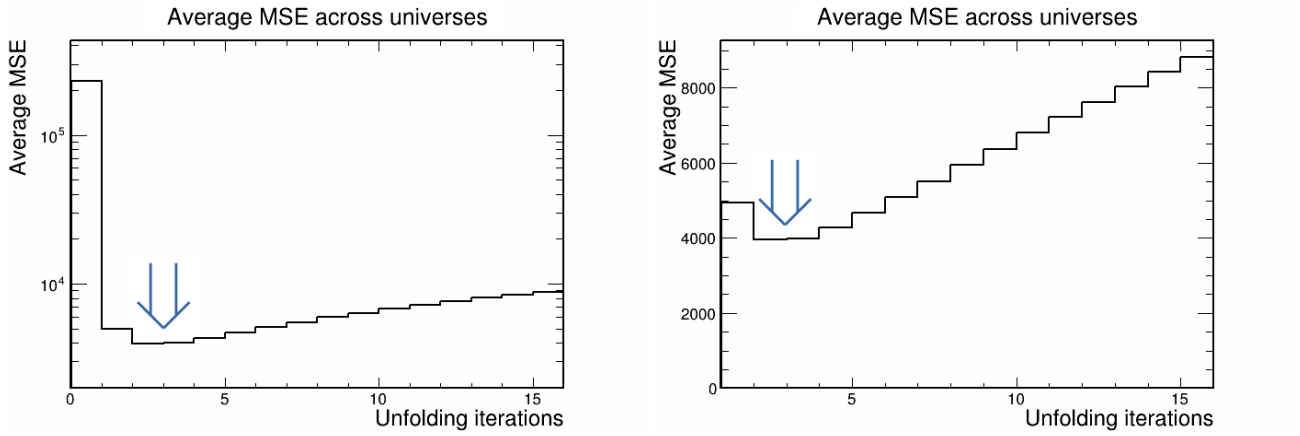

Figure 9.6: Average MSE across 500 universes versus unfolding iteration. On the left plot, the first bin shows the average MSE between true and reconstructed distributions without any unfolding. The first unfolding iteration significantly minimizes differences. The right plot shows the same information, but with the first bin from the left plot taken out so that the minimum can be better observed.

The number of unfolding iterations over which the MSE between unfolded and truth distributions is minimized was examined for each universe. The results of the examination are summarized in Fig. 9.7, where the x-axis is the number of unfolding iterations and the y-axis is the number of universes where the MSE minimizes for a given number of unfolding iterations. Out of 500 universes, the distance between truth and unfolded distribution minimizes at 3 unfolding iterations for 170 universes.

In addition to MSE, the unfolding procedure has been evaluated with χ^2/DoF and with a significance metric. The significance is defined as follows:

$$Significance = \sum_{j=1}^{Bins} \frac{n_{reco,j} - K n_{true,j}}{\sqrt{n_{reco,j} + K^2 n_{true,j}}}. \tag{9.3}$$

Essentially, this is the sum of the difference between the true and reconstructed histograms, weighted by the number of true and reco events in each bin. The constant $K = \frac{n_{true}}{n_{reco}}$ normalizes the number of true and reconstructed events. The average χ^2/DoF and Significance across universes is displayed in Figs. 9.8 and 9.8. For both metrics, the average metric value minimizes after two unfolding iterations. However, Figs. 9.6, 9.8 and 9.8 show that the MSE is very close to the minimum for three unfolding iterations.

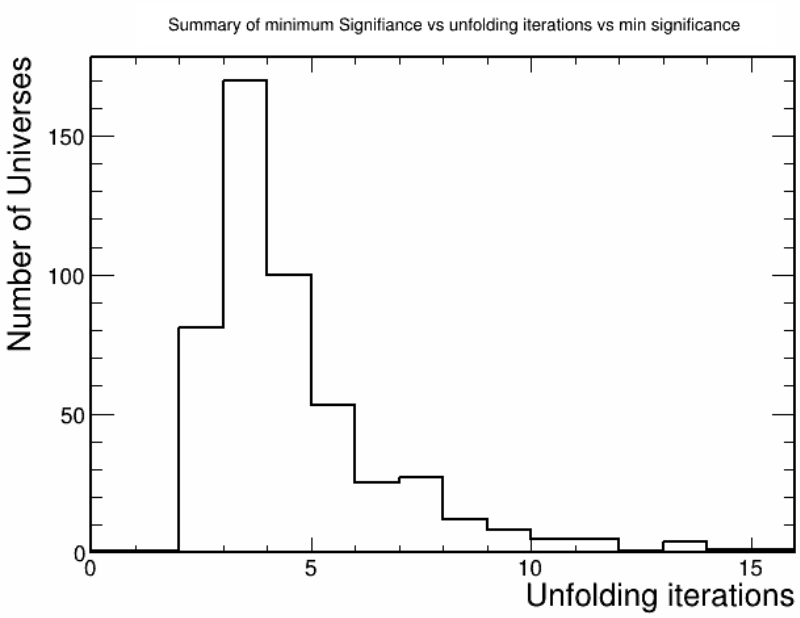

Figure 9.7: Mean Squared Error of the reconstructed unfolded distribution with respect to truth as a function of unfolding iteration for a typical universe. The first unfolding iteration significantly undoes the smearing effects due to detector imperfections.

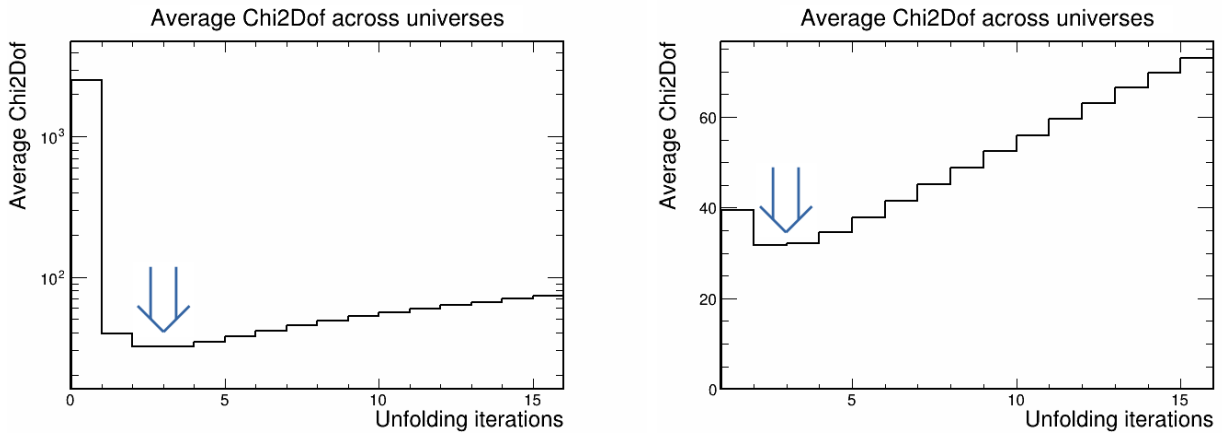

Figure 9.8: Average χ^2/DoF across 500 universes versus unfolding iteration. On the left plot, the first bin shows the average difference between true and reconstructed distributions without any unfolding. The right plot shows the same information, but with the first bin from the left plot taken out so that the minimum can be better observed.

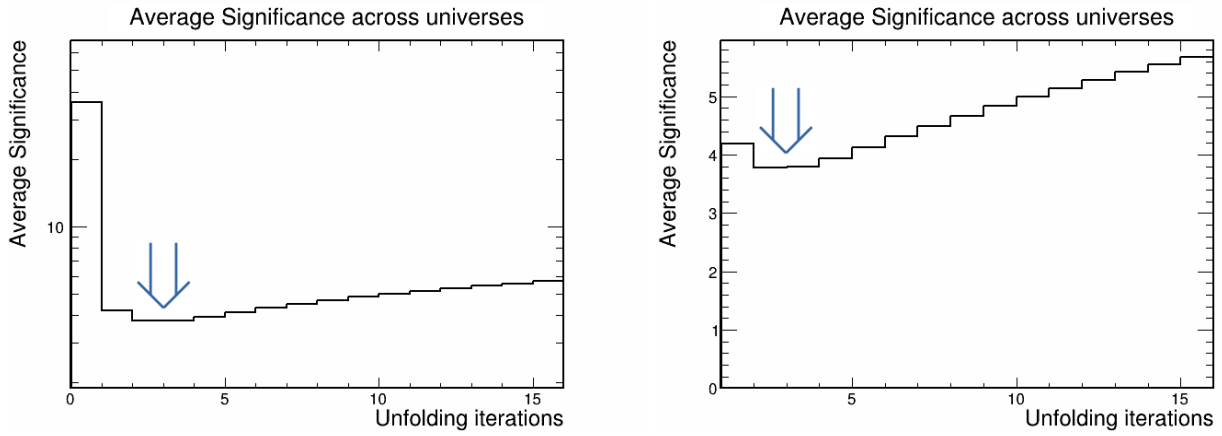

Figure 9.9: Average Significance across 500 universes versus unfolding iteration. On the left plot, the first bin shows the average significance between true and reconstructed distributions without any unfolding. The right plot shows the same information, but with the first bin from the left plot taken out so that the minimum can be better observed.

One item of interest is to observe how the metric values rating the unfolding performance changes as a function of unfolding iterations in the special systematic samples: light, calibration and Cherenkov radiation. The MSE, χ^2/DoF and Significance were calculated. For each sample, the metric values have been observed to minimize at the same number of unfolding iterations. Therefore, only the change in the MSE as a function of unfolding iterations will be displayed for these systematic samples.

The MSE as a function of unfolding iterations is shown in Fig. 9.10 for light level samples: light up and light down. Unlike the metric plots in Figs. 9.6, 9.8 and 9.8, it is observed that the MSE minimizes at just one unfolding iteration. The MSE arising from a second iteration is almost close to the MSE of the first iteration. The MSE as a function of unfolding iterations is shown in Fig. 9.11 for calibration samples: calib up, calib down and calib shape. It is observed that the MSE minimizes at just one unfolding iteration. The MSE as a function of unfolding iterations for the two Cherenkov radiation samples: Cherenkov up and Cherenkov down is shown in Fig. 9.12. The metric plot for the Cherenkov systematic shifted samples shows that the diferences between truth and reco distributions minimize at 1 unfolding iteration.

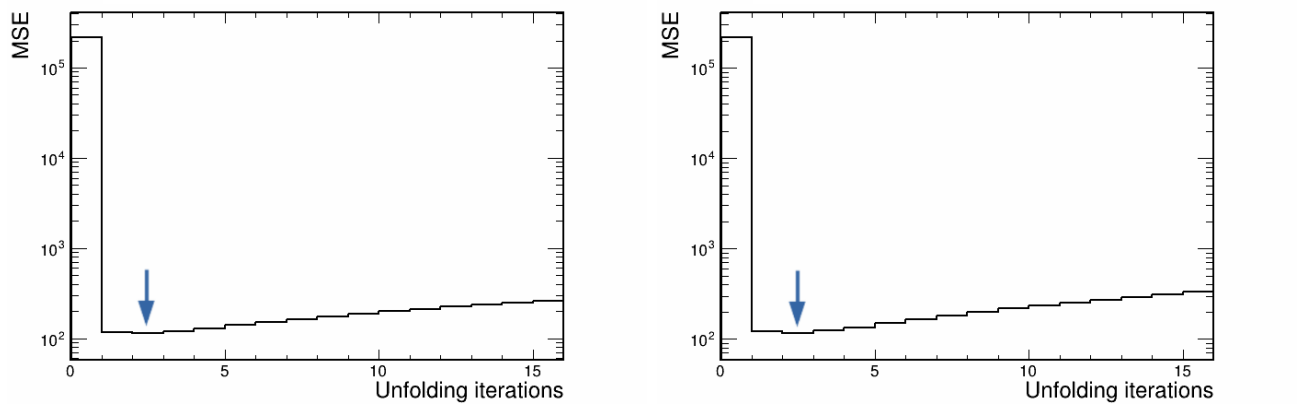

Figure 9.10: MSE of the reconstructed event distribution with respect to truth as a function of unfolding iterations for the light level systematic samples. The plots are made from the light up (left), light down (right) samples.

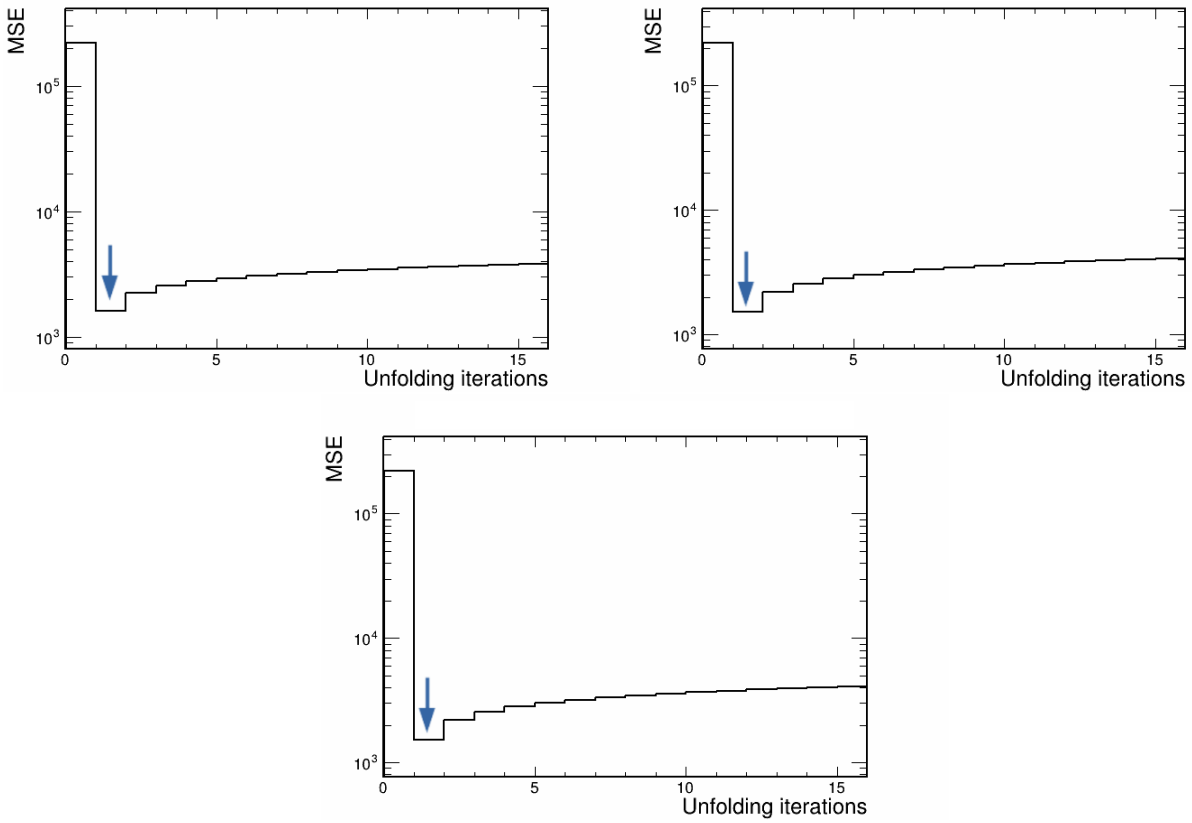

Figure 9.11: MSE of the reconstructed event distribution with respect to truth as a function of unfolding iterations for the calibration systematic samples: calib up, down, and shape.

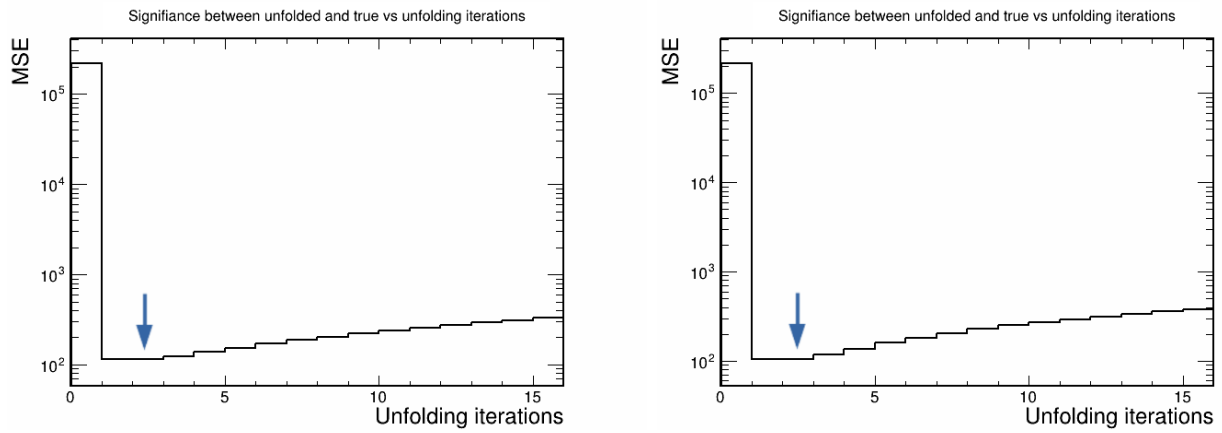

Figure 9.12: MSE of the reconstructed event distribution with respect to truth as a function of unfolding iterations for the Cherenkov systematic samples. The plots are made from the Cherenkov up (left), Cherenkov down (right) samples.

As shown in Fig. 9.13, the number of unfolding iterations for which the χ^2/DoF and significance minimize in each universe have also been examined. In both cases, a similar conclusion is reached. In over 150 universes, the difference between reconstructed and truth distribution is minimized after 3 unfolding iterations. However, the number of universes in which the metric value minimizes is not far behind for 2 or 4 iterations. Few universes require more than five unfolding iterations to resolve the differences between the estimated and true event distribution.

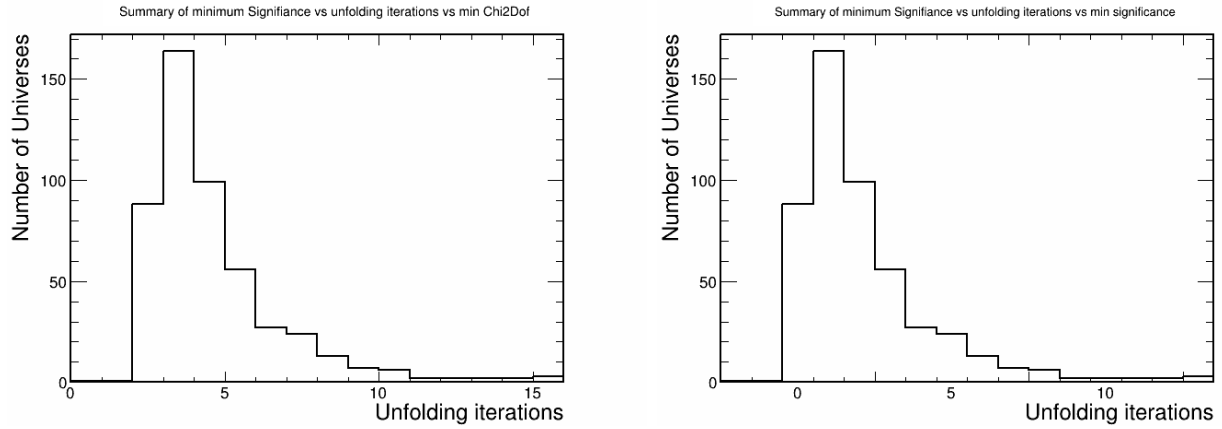

Figure 9.13: Distribution of the number of unfolding iterations it takes to minimize χ^2/DoF (left) and significance (right). The distances between true and recon distributions minimize for more than 160 universes after three unfolding iterations.

The average metric values in Figs. 9.6, 9.8 and 9.9 indicate that the differences minimize for two unfolding iterations. However there is very little difference between unfolded distributions

and truth for two or three unfolding iterations, and that distribution yielded from two or three unfolding iterations is practically the same. However, the modal distributions of the unfolding iterations from Figs. 9.7 and 9.13 show that the optimal number of unfolding iterations is three. The combination of metric plots indicate that either two or three unfolding iterations will do. For this analysis, two unfolding iterations are performed to extract the true event distribution from the data sample.

9.4 Unfolding performance

The performance of the unfolding procedure can be examined using the reconstructed event distribution of individual universes. The true and reconstructed event distributions of an illustrative universe are displayed in the upper plots of Fig. 9.14. The bin-by-bin differences between the two distributions are shown on the lower plots. The true and reco distributions are somewhat different from each other. In particular, the reconstructed distribution has more events concentrated along a ridge close to the boundary region. Whereas the event distribution of the same events in true $|\vec{q}|$ and true E_{avail} is more smeared out along the same ridge.

The reconstructed event distribution is unfolded three times. The result of the procedure is displayed in Fig. 9.15. It can be seen that unfolding undoes the smearing effects and transforms the event distribution into one that is much closer to the truth distribution. The ratio of the unfolded reco event count to true event count is closer to 1 on average across all bins.

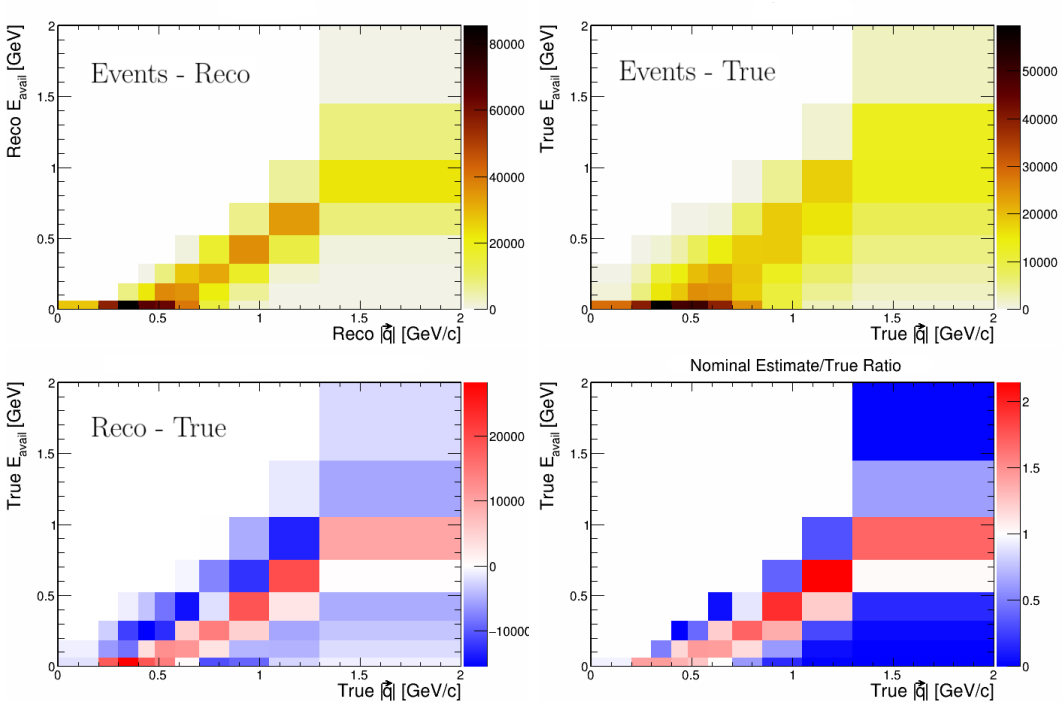

Figure 9.14: Nominal reconstructed (upper left) and truth (upper right) event distributions in $|\vec{q}|$ and E_{avail}. The reco - true difference and reco/true ratios are on the lower left and lower right respectively.

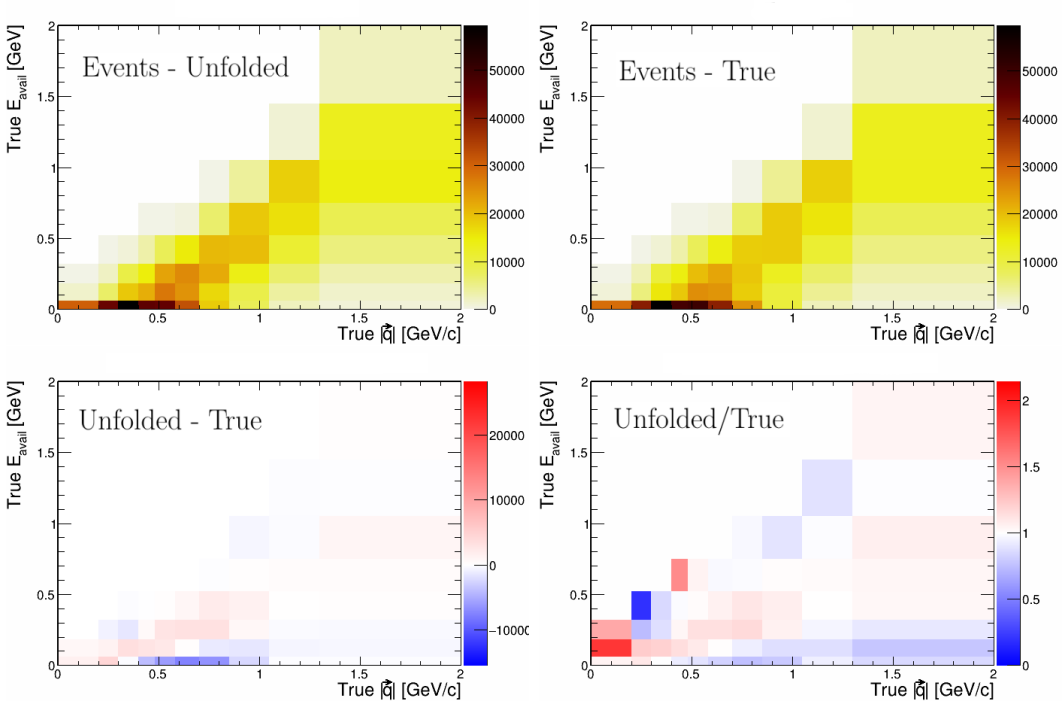

Figure 9.15: Unfolded reconstructed (upper left) and truth (upper right) event distributions in $|\vec{q}|$ and E_{avail}. The reco - true difference and reco/true ratios are on the lower left and lower right respectively.

The nominal reconstructed, true reconstructed and truth event distributions are compared for each analysis variable in Figs. 9.16 and 9.17. The event count is underestimated at higher $|\vec{q}|$ but overestimated at lower values. This is corrected by unfolding the distribution three times.

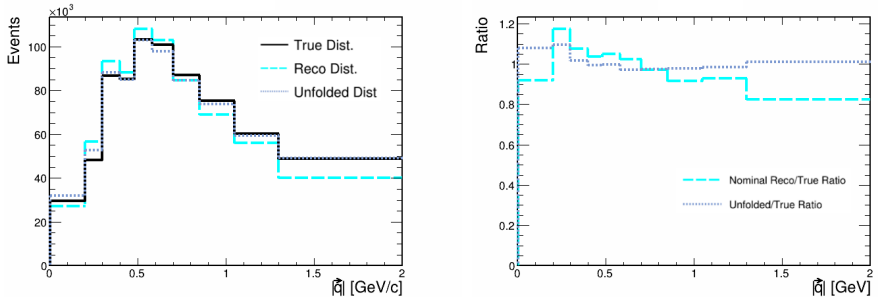

Figure 9.16: Examination of reconstructed event distribution before and after unfolding in $|\vec{q}|$. The left plot shows how the reconstructed event count compares with the truth distribution before and after unfolding. The right plot shows the ratio of the reconstructed distribution with respect to the true distribution.

The improvement in the estimated event distribution is evident upon looking in at the ratio plot on Fig. 9.17 (right). The ratio of estimated/true event count is closer to one on average across the bins. In E_{avail}, the reconstructed distribution has more events in the lowest bin and less in the remaining bins compared to the truth. The underestimation becomes significant with increasing value of E_{avail}. Again the ratio of the unfolded reconstructed to truth is closer to 1 across all bins in E_{avail}, demonstrating that the unfolding procedure using three iterations is robust.

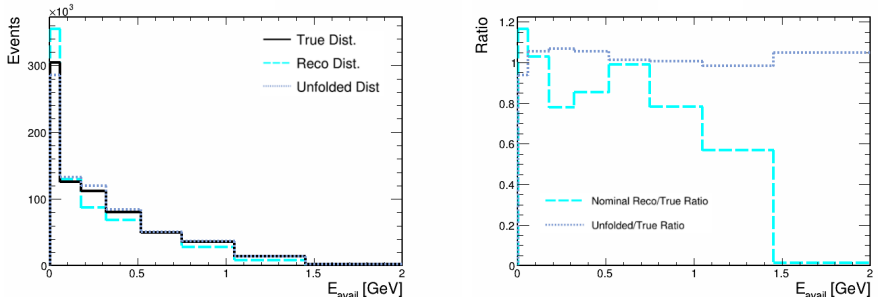

Figure 9.17: Examination of reconstructed event distribution before and after unfolding in E_{avail}. The left plot shows how the reconstructed event count compares with the truth distribution before and after unfolding. The right plot shows the ratio of the reconstructed distribution with respect to the true distribution.

Chapter 10

Closure Tests for Cross-Section Calculation

This Section documents the closure tests that have been carried out to validate the analysis software package. These closure tests - referred to as In-Out tests - examine whether cross sections determined by the package are identical to those that underwrite the input information. If the cross section of a process in some variable(s) is known and this cross section is used to generate an event distribution in the same variable(s) (i.e. what goes in), a cross-section calculation made from the same event distribution (i.e. what comes out) must yield the cross section that was originally used to make the same distribution.

Closure tests for the methods used to constrain background processes are also discussed in this Section. In the latter tests, the nominal background estimates are correct from the beginning. The fitting methods should not adjust the normalizations of the background templates for a successful closure test. For the closure tests, the software package is used to calculate the cross sections of true $\bar{\nu}_\mu$-CC inclusive and true $\bar{\nu}_\mu$-CC 2p2h-MEC events on carbon-12. Signal events must satisfy the criteria listed in Sec. 4.1, with the exclusion of the muon phase space cuts (point 3 of Sec. 4.1) and the added requirement that the interaction occurs on a carbon-12 nucleus. The omission of the phase space cuts is necessary so that the cross sections yielded by the packages can be compared with GENIE splines to confirm that the correct cross sections are being generated.

10.1 Closure test in E_ν

The identification of background events versus $E_{\bar{\nu}}$ is determined with MC truth. The background event distribution consists of events that do not satisfy the true signal definition for signal events; this includes $\bar{\nu}_\mu$-CC events and $\bar{\nu}_\mu$-CC 2p2h-MEC events that do not interact on carbon-12 nuclei. The signal contribution in the data is obtained by subtracting the estimated background from the data. The signal distributions and the background to $\bar{\nu}_\mu$-CC events and $\bar{\nu}_\mu$-CC 2p2h-MEC events are shown in Figs. 10.1 and 10.2 respectively. The signal and background estimates have also been plotted from the MC and superimposed for comparison. The distributions used by the cross-section package for bookkeeping coincide with the distributions made from the MC, indicating the package does the necessary bookkeeping correctly for the cross-section measurement. The distributions of the signal events in Figs. 10.1 and 10.2 peak around 2 GeV antineutrino energy, which is close to the peak of the antineutrino flux.

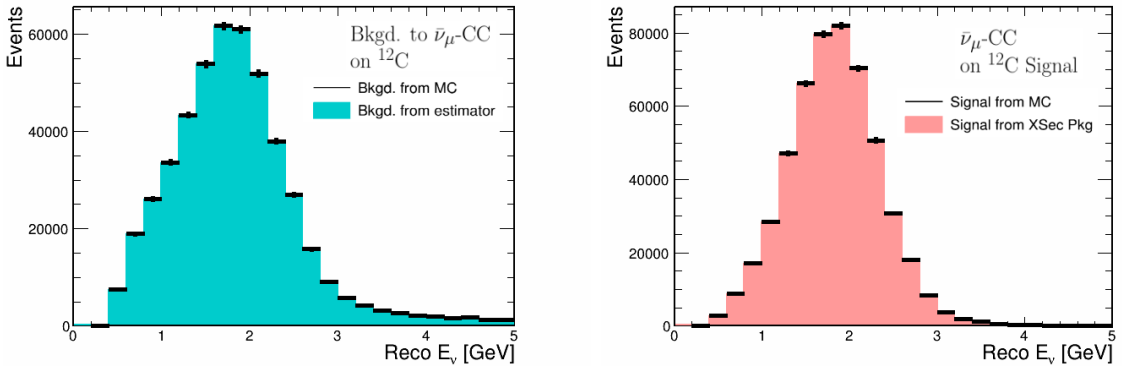

Figure 10.1: Distributions of the background and $\bar{\nu}_\mu$-CC signal in reconstructed antineutrino/neutrino energy. The signal and background generated by the cross-section package, represented by the colored histograms, coincide with the same distributions made directly from the MC, represented by the solid black points, confirming that the cross section infrastructure does the bookkeeping correctly.

The extracted signal event distributions are then unfolded, a process which undoes the smearing effects introduced by the detector and converts the distributions to true E_ν. The results of unfolding the signal estimates are shown in Fig. 10.3. Again, true E_ν distributions made directly from the MC are superimposed, and it can be seen that the unfolding done by the cross-section package is correct.

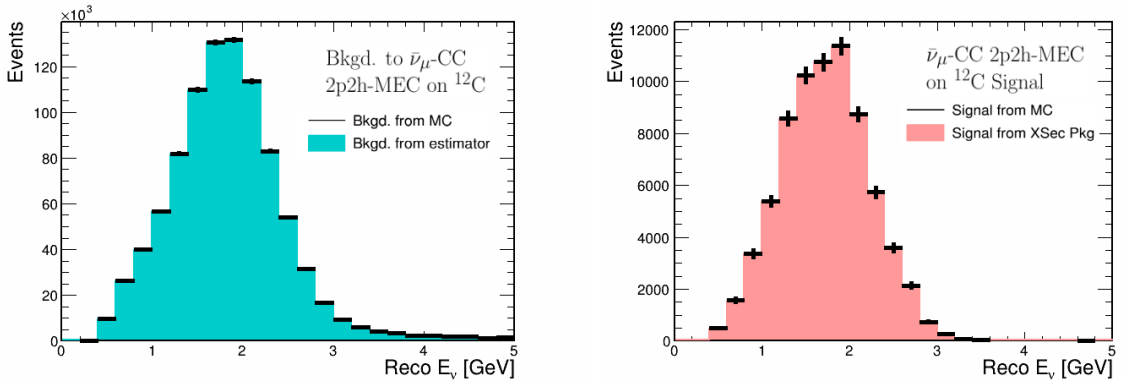

Figure 10.2: Distributions of the background and $\bar{\nu}_\mu$-CC 2p2h-MEC signal in reconstructed antineutrino/neutrino energy. Again, the event distributions utilized by the cross-section package (colored histograms) coincide with the same distributions made from the MC sample (solid black points).

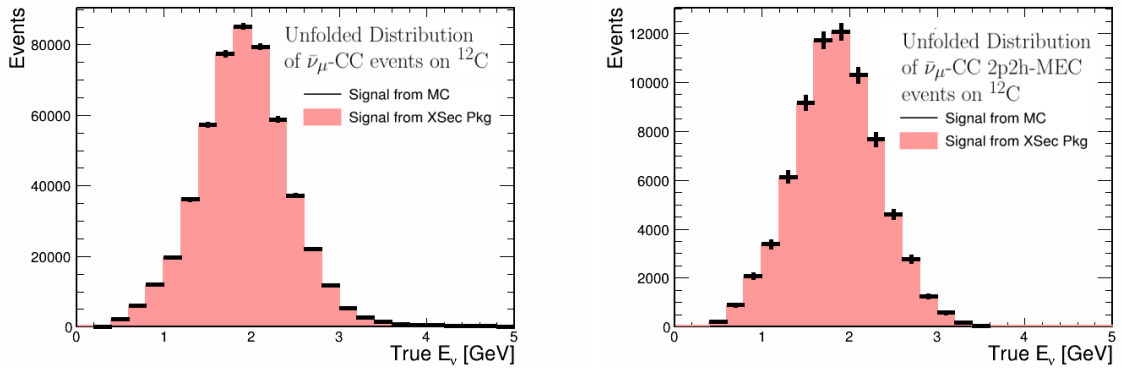

Figure 10.3: Unfolded distribution of the 'signal' for the $\bar{\nu}_\mu$-CC inclusive analysis (left). Unfolded distribution for the 'signal' 2p2h-MEC analysis (right) in true antineutrino energy. The shaded histograms in this figure are obtained by unfolding the signal distributions in Figs. 68 and 69.

The criteria used to select candidate events for analysis do not pick up all signal events that occur in the the Near Detector. Therefore, this unfolded event distribution must be efficiency-corrected to obtain the true distribution of signal events in the detector. The selection efficiencies versus $E_{\bar{\nu}}$ for the two interaction categories are shown in Fig. 10.4. Note that these efficiencies are specifically for inclusive $\bar{\nu}_\mu$-CC events and for $\bar{\nu}_\mu$-CC 2p2h-MEC events that occur on carbon-12 nuclei.

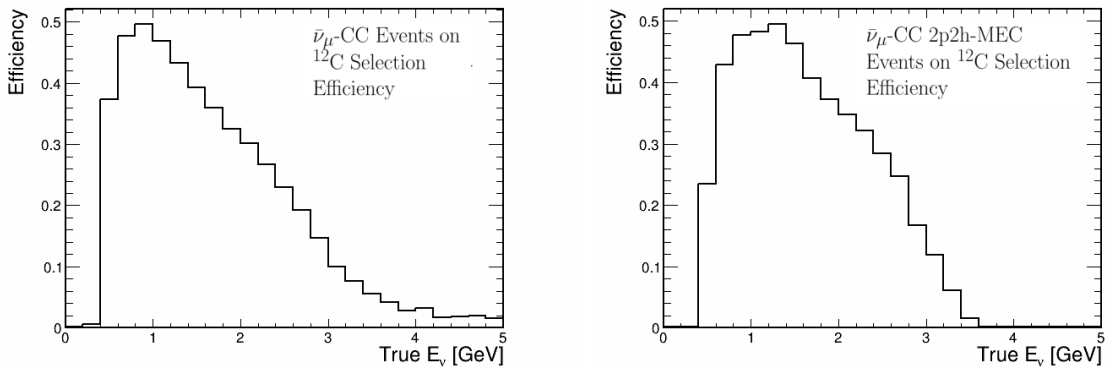

Figure 10.4: Efficiency distributions for $\bar{\nu}_\mu$-CC events and $\bar{\nu}_\mu$-CC 2p2h-MEC events on carbon-12 nuclei.

The unfolded distributions are then divided bin-by-bin with the antineutrino flux distribution of the RHC beam shown in Fig. 10.5.

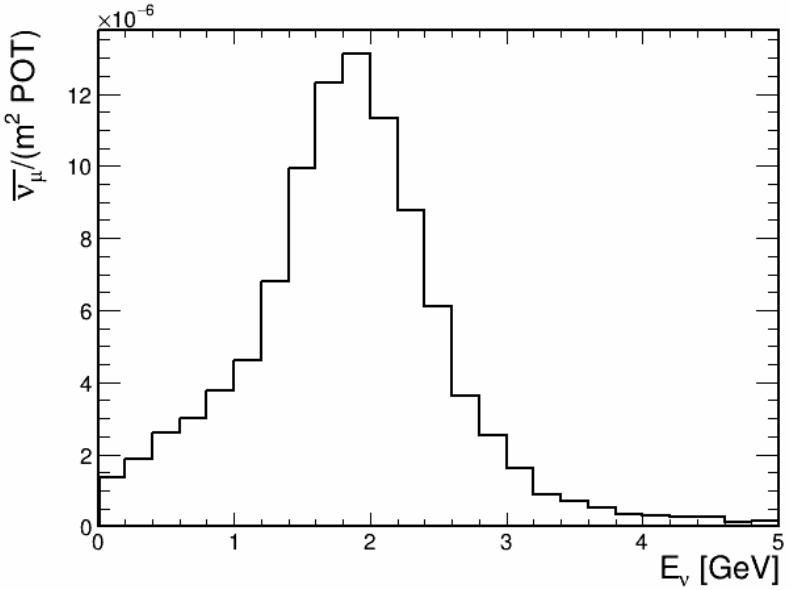

Figure 10.5: Flux of incoming muon antineutrinos in the RHC beam. The flux peaks at 1.86 GeV and extends to nearly 5 GeV in antineutrino energy.

The final step is to divide the resulting distribution with the total number of targets, which is the total number nucleons of carbon-12 nuclei contained in the fiducial volume of the detector. The cross sections of the inclusive sample and of the 2p2h-MEC interactions, superimposed with the GENIE splines, are shown in Fig. 10.6, expressed in units of 10^{-38} cm^2/ nucleon/GeV. Based

on the agreement between the splines and the histograms, the cross section infrastructure yields the correct cross section in $E_{\bar{\nu}}$.

Thus the cross-section package, given a correct signal estimate, unfolding and efficiency correction, is capable of yielding the correct cross section distribution in E_{ν}. In the next subsection, the package is tested to see whether it yields the correct cross section in $|\vec{q}|$ and E_{avail}.

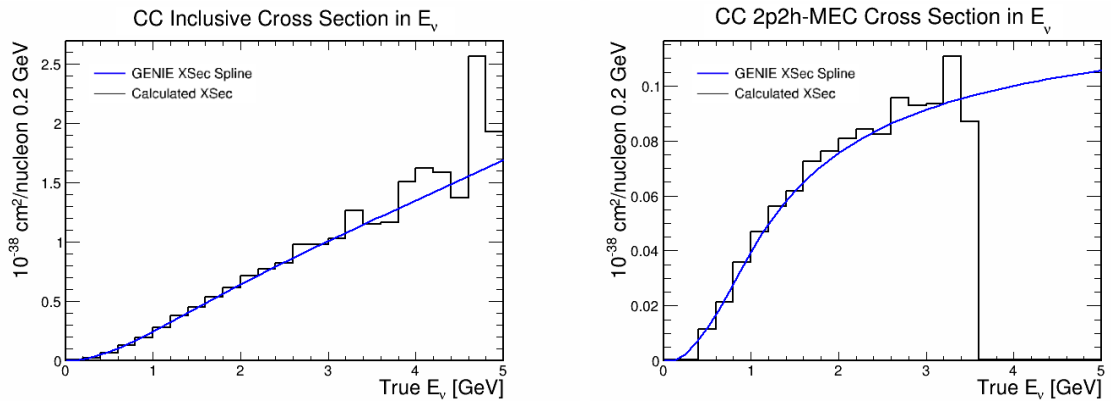

Figure 10.6: Single differential cross section distributions in E_{ν} for $\bar{\nu}_{\mu}$-CC events (left) and $\bar{\nu}_{\mu}$-CC 2p2h-MEC events (right) on carbon-12. The relevant GENIE splines have also been superimposed for comparison.

10.2 Closure test for cross sections in $|\vec{q}|$ and E_{avail}

The procedure for cross-section calculation in $|\vec{q}|$ and E_{avail} is similar. The first step is to estimate the background contribution in the sample in the two analysis variables, which is then subtracted from the data to yield the distribution of signal events. The background and signal distribution from the calculation of cross sections of the $\bar{\nu}_{\mu}$-CC inclusive and CC 2p2h-MEC samples are shown in Figs. 10.7 and 10.8 respectively.

The distributions are then unfolded to convert the reconstructed signal distributions to true $|\vec{q}|$ and E_{avail}. The unfolded distributions for $\bar{\nu}_{\mu}$-CC inclusive and for $\bar{\nu}_{\mu}$-CC 2p2h-MEC are shown in Fig. 10.9. In the case of 2p2h-MEC events, there are no events observed above 1.6 GeV of $|\vec{q}|$, reflecting the $|\vec{q}|$ cutoff at 1.2 GeV that occurs in the Valencia MEC model.

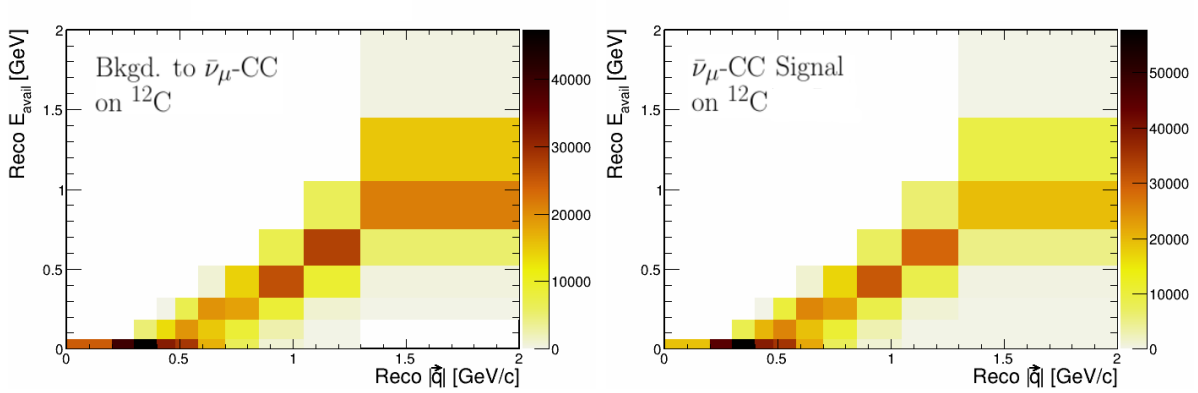

Figure 10.7: Event distributions of the background and $\bar{\nu}_\mu$-CC inclusive signal in reconstructed $|\vec{q}|$ and E_{avail}.

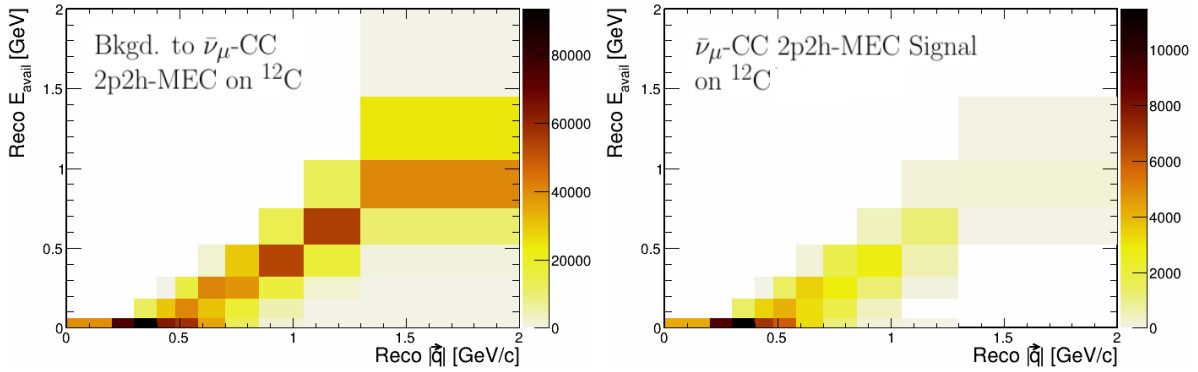

Figure 10.8: Event distributions of the background and $\bar{\nu}_\mu$-CC 2p2h-MEC signal in reconstructed $|\vec{q}|$ and E_{avail}. The MEC signal tends to lie in the region of low $|\vec{q}|$ and low E_{avail}.

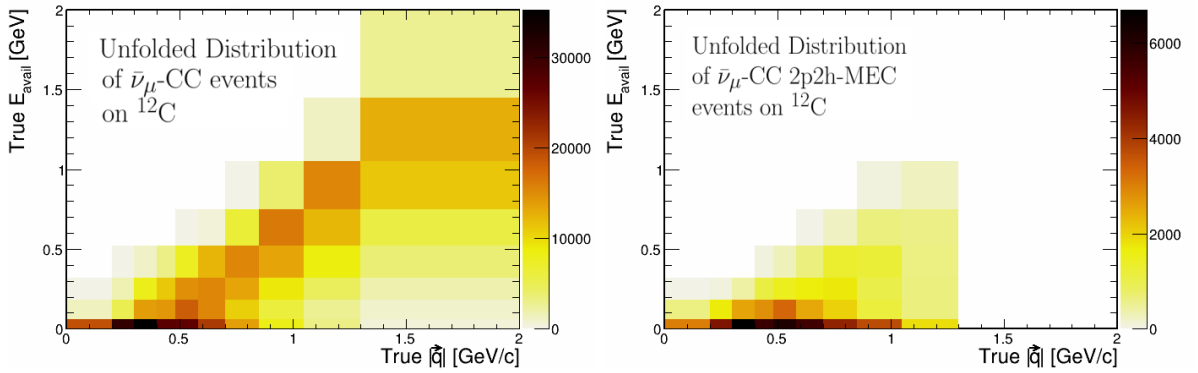

Figure 10.9: Unfolded distributions of the signal interaction event distributions: $\bar{\nu}_\mu$-CC inclusive and of the $\bar{\nu}_\mu$-CC 2p2h-MEC in $|\vec{q}|$ and E_{avail}.

The unfolded signal distribution must now be efficiency-corrected to account for all the true signal events not passing the CC inclusive selection. This is done by dividing the signal distributions in

Fig. 10.9 with the relevant selection efficiency distributions, which are displayed in Fig. 10.10. In both cases the efficiency is higher around the boundary region where $|\vec{q}|$ and E_{avail} are numerically approximately equal.

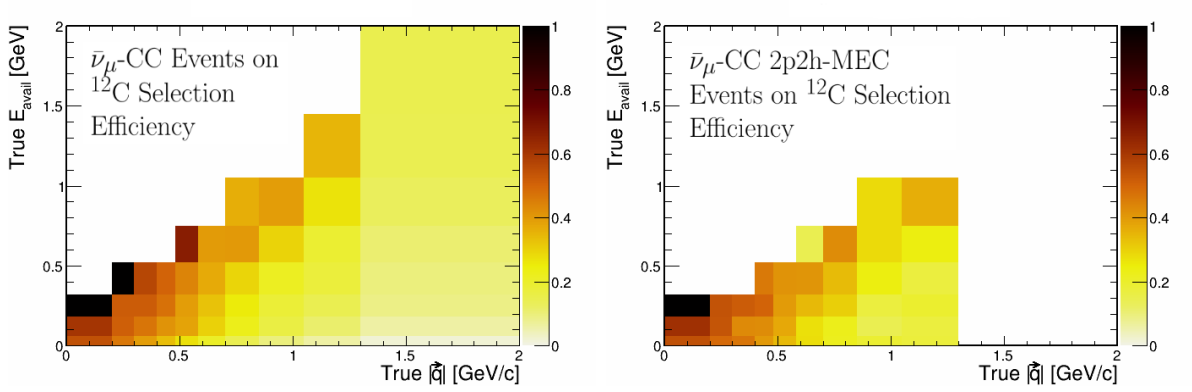

Figure 10.10: Selection efficiency of $\bar{\nu}_\mu$-CC Inclusive events (left) and $\bar{\nu}_\mu$-CC 2p2h-MEC (right). In both distributions, the efficiency is relatively higher in the boundary region.

The double-differential cross sections that are to be measured are flux-integrated, meaning that the efficiency-corrected distributions are divided by the integral of the flux, shown in Fig. 10.5. The final step is to divide the result with the total number of nucleons of the target carbon-12 nuclei in the detector fiducial volume, yielding the double differential cross sections. These are shown in Fig. 10.11.

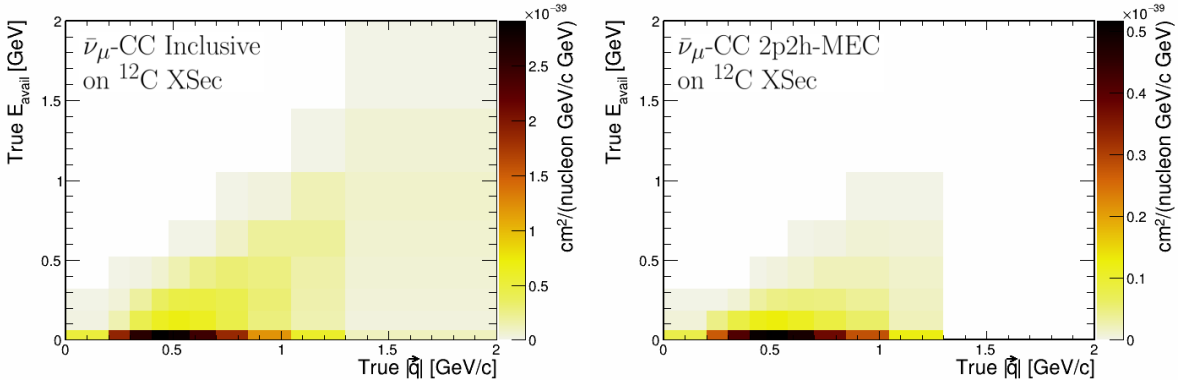

Figure 10.11: Double-differential cross sections of the $\bar{\nu}_\mu$-CC inclusive signal (left) and the $\bar{\nu}_\mu$-CC 2p2h-MEC signal (right) in $|\vec{q}|$ and E_{avail}. The units of the cross section are cm^2/(nucleon GeV / GeV/c).

To confirm that the cross sections obtained are correct, the total cross sections can be calculated from these distributions and compared with total cross sections calculated from the distributions

in Fig. 10.6. Details about obtaining a total cross section from both of these cases are given in Ref. [33]. Initially, it was observed that the total cross sections calculated for each interaction differ. This occurs because the effective kinematic cuts for the two cross sections are not the same. To resolve this discrepancy, the cross sections are recalculated, but only with events that satisfy the following kinematic cuts: True $E_\nu \leq 3.5$ GeV/c, true $|\vec{q}| < 2$ GeV and true $E_{avail} \leq 2$ GeV. The total cross sections thereby obtained are presented in Table 10.1. With the kinematic restrictions in place, the total cross sections calculated from two different sources agree with each other.

Interaction Type	Total Cross Section from E_ν	Total Cross Section from $\|\vec{q}\|$ vs. E_{avail}
$\bar{\nu}_\mu$-CC Inclusive	0.53×10^{-38} cm^2/nucleon	0.53×10^{-38} cm^2/nucleon
$\bar{\nu}_\mu$-CC 2p2h-MEC	0.64×10^{-39} cm^2/nucleon	0.64×10^{-39} cm^2/nucleon

Table 10.1: Total cross section calculated from the single differential cross section in E_ν and double differential cross section in $|\vec{q}|$ versus E_{avail}. The values are in units of cm^2/nucleon of carbon nuclei.

10.3 Closure test with reduced signal

It is of interest to see what distribution the cross-section package yields when the event rate of the signal is down-weighted. For this test, the event rates of the two signal interactions, $\bar{\nu}_\mu$-CC inclusive and $\bar{\nu}_\mu$-CC 2p2h-MEC, are reduced by 25%. It is expected that the extracted cross sections should be 75% of the nominal cross sections displayed in Fig. 10.6. The cross sections that are obtained by this 25% event rate reduction are displayed in Fig. 10.12. The nominal GENIE splines have also been reduced by the same factor and superimposed for comparison. To better display how these cross section compare to the nominal cross sections (i.e. ones without down-weighting), the nominal cross sections are also superimposed on the same figure.

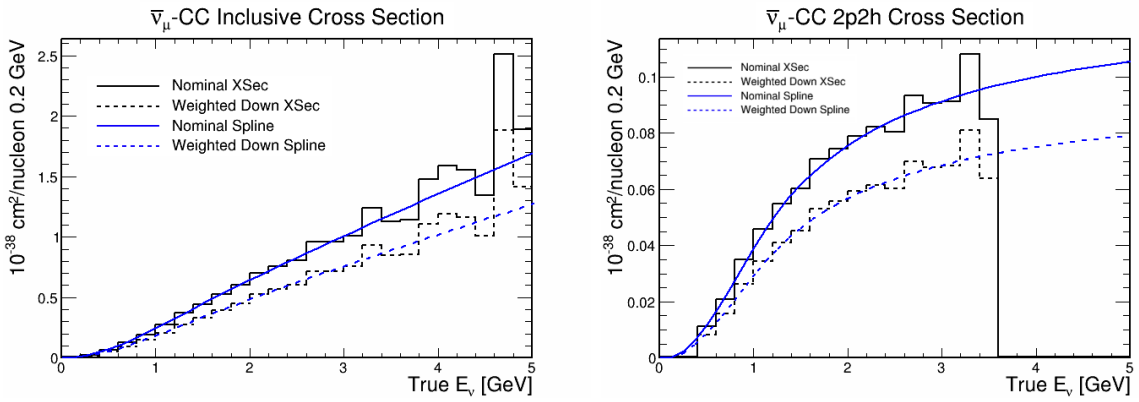

Figure 10.12: Cross sections of $\bar{\nu}_\mu$-CC inclusive events (left) and of CC 2p2h-MEC events (right) that are unweighted (solid lines) and weighted down by 25% (dashed lines).

10.4 Closure test for cross section in E_ν on NOvA medium

It is necessary to perform a closure test with the cross section of true signal events as defined by the criteria in Sec. 4.1 on the NOvA medium. For this test, a sample of true signal events, i.e. $\bar{\nu}_\mu$-CC events in the NOvA medium, is used to calculate a cross section in E_ν. The cross-section package is then used to reconstruct another cross section from the same sample of true signal events that have passed the selections in Sec 4.2. The aforementioned two cross sections are observed to coincide as shown in Fig. 10.13.

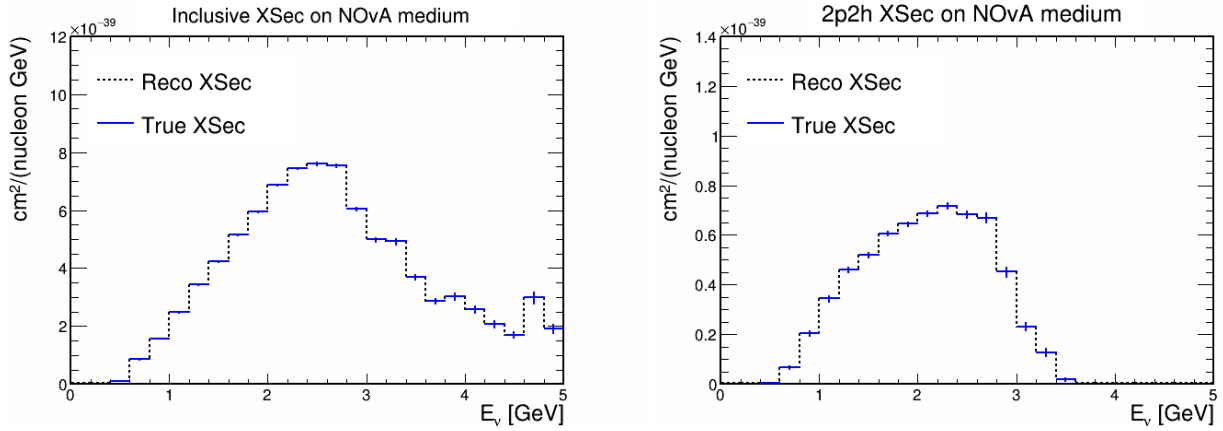

Figure 10.13: Cross section of $\bar{\nu}_\mu$-CC inclusive events and $\bar{\nu}_\mu$-CC 2p2h-MEC events that interact in the NOvA medium and pass the muon kinematic cuts. The dashed black histograms represent the cross sections 'measured' from a sample of true signal events passing the CC inclusive selection criteria, while the blue line histograms represent the cross sections used to generate the true signal events.

10.5 Closure test with background estimator

The background estimators discussed in Sec 6.5 has been integrated with the cross-section calculator. It is therefore necessary to validate the performance of this software configuration. For this closure test, the fake data and MC samples are drawn from statistically independent samples. This is to check that the background estimator behaves reasonably when the MC does not perfectly describe the data. It is expected that the normalizations of the templates coming out of the fit should be close to 1.0.

Plots displaying the fits are presented in this subsection to show that the fitting procedure does not change the normalizations of the background contributions. The first stage of the first iteration of the fit is shown in Fig. 10.14. The fitting procedure does not alter the normalizations of the two underlying templates in the prediction.

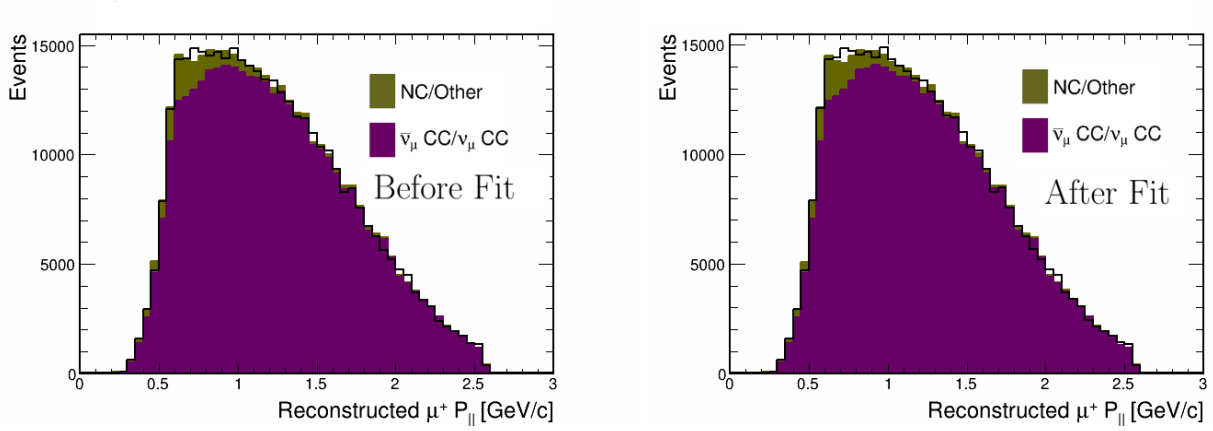

Figure 10.14: First iteration of the first stage of the fitting procedure. As seen from the plots before (left) and after (right) the fit, the normalizations of the two templates are virtually unchanged.

The fitting carried out in the second stage of the first iteration, as well as both iterations of the next stage are displayed in Fig. 10.15, 10.16 and 10.17 respectively. It can be seen that the normalizations of the different templates involved in the fit remain virtually unchanged. A summary of all the normalizations applied to the various background templates at each stage of the fit, and the overall extrapolated normalization is shown in Table 10.2. It can be seen that the normalizations applied to the templates are close to 1, which is in line with the expectations from

the fitting procedure. The results of the fit are extrapolated to $|\vec{q}|$ and E_{avail}. These constrained backgrounds are then used to extract the signal from the data, which can be then used for a cross-section measurement.

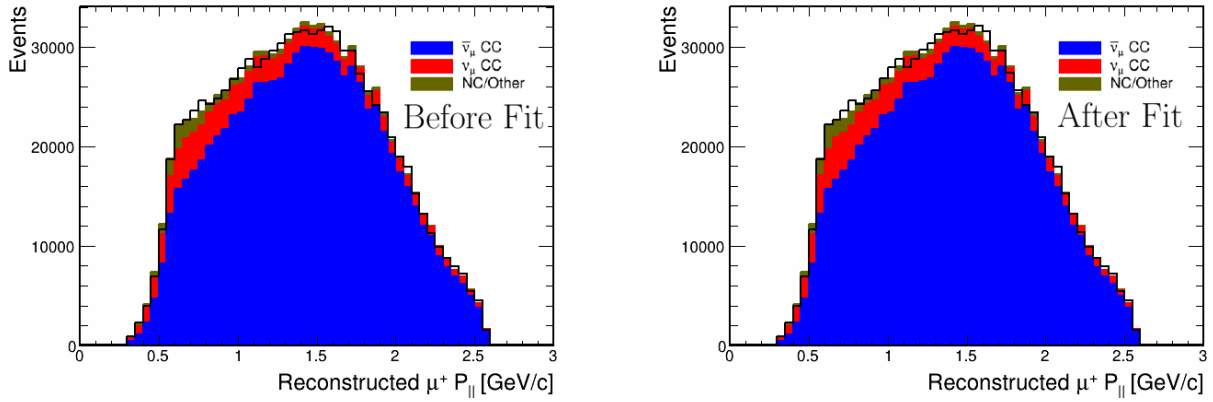

Figure 10.15: First iteration of the second stage of the fitting procedure. The left versus right plots show the data-MC comparison before and after the fit.

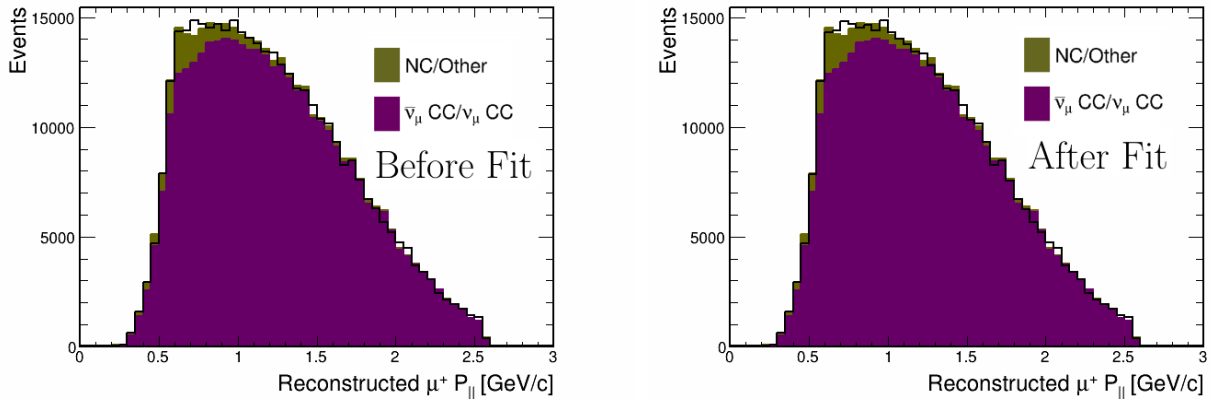

Figure 10.16: Second iteration of first stage of the fitting procedure. The left versus right plots show the data-MC comparison before and after the fit.

A comparison of the cross section calculated with this data-driven background versus one with nominal background prediction is shown in Fig. 10.18. Any difference between the two cross sections is three-orders of magnitude smaller than the actual values of the cross sections in a bin. The small differences arise because the fitting procedure does slightly alter the normalizations of the background templates.

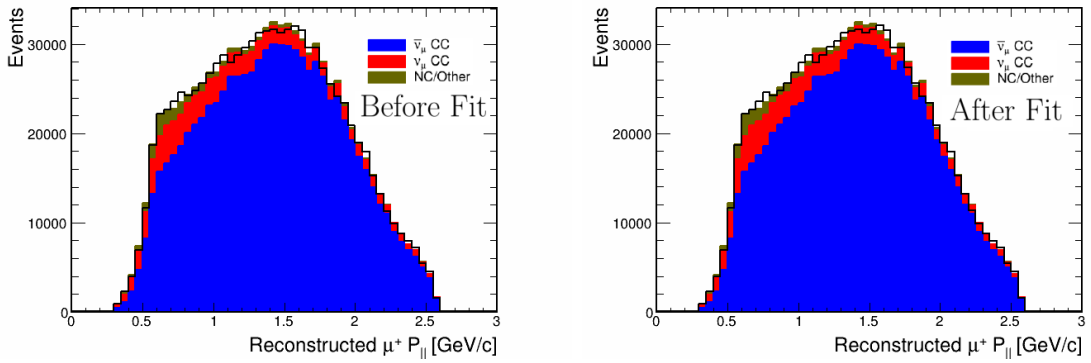

Figure 10.17: Second iteration of second stage of the fitting procedure. The left versus right plots show the data-MC comparison before and after the fit.

Bkg. Int. Type	Stage 1	Stage 2	Stage 3	Stage 4	Overall
$\bar{\nu}_\mu$-CC Bkg.	0.997	1.002	0.998	1.002	0.999
ν_μ-CC	0.998	1.00	0.998	1.00	0.995
NC/Other	0.996	1.01	0.998	1.004	1.007

Table 10.2: Normalizations applied by the fitting method for background estimation on the various background templates.

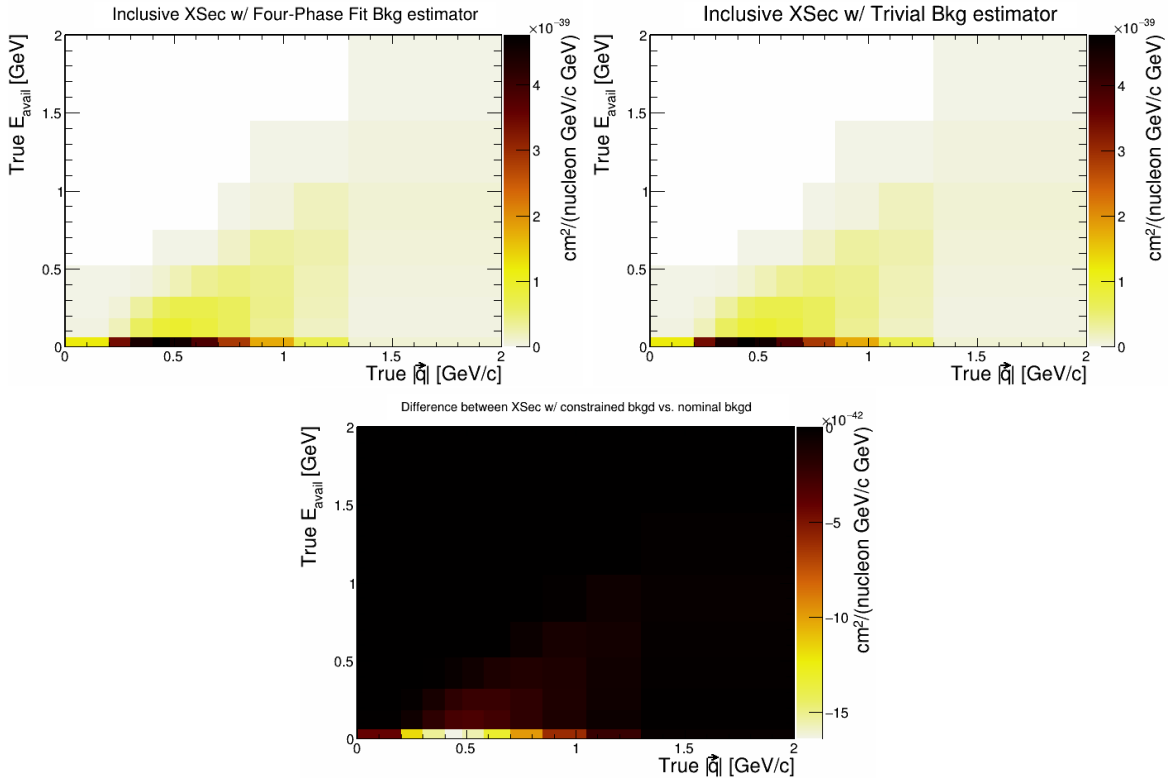

Figure 10.18: Cross section of $\bar{\nu}_\mu-$CC inclusive events with the background estimate returned by the fitting package. The nominal cross section and the difference with respect to the nominal cross section is provided for comparison.

Chapter 11

Warp Tests Using Weighted Backgrounds

Another type of test involves taking a certain background interaction component in the data, weighting it up and down, carrying out the fitting, and observing whether the procedure yields the correct normalization of the component. This is known as a warp test. Three warp tests have been carried out: One with the NC/Other background component in the data weighted down by 25%, another with the same component weighted up by 25%, and a third test with the wrong-sign component increased in the pseudo-data by 25%.

11.1 Warping the NC/Other background

Firstly, The NC/Other component in the data is weighted down by 25%, while the other components in the data are untouched. It is expected that the fitting procedure should yield a normalization of 0.75 for the NC/Other component and 1.0 for the $\bar{\nu}_\mu$-CC and ν_μ-CC components.

Table 10 summarizes the normalizations obtained by MINUIT from the four fitting stages, as well as the overall normalization applied to each component involved in the fit. It can be seen that the overall normalization assigned to NC/Other events by the fit is 0.749, which matches the weight that was applied to the pseudodata. The weights on the other two components, $\bar{\nu}_\mu$-CC and ν_μ-CC are both 1.0.

Bkg. Int. Type	Stage 1	Stage 2	Stage 3	Stage 4	Overall
$\bar{\nu}_\mu$-CC Bkg.	1.000	0.999	1.000	1.005	1.005
ν_μ-CC	1.000	1.000	1.000	1.000	1.000
NC/Other	0.751	1.01	0.999	0.999	0.749

Table 11.1: Normalizations applied by the fitting method for background estimation from the warping test where NC/Other is weighted down by 25%.

Secondly, the NC/Other component in the data is weighted up by 25% while the other components are left at their nominal rates. It is expected that the fitting procedure should yield a normalization of 1.25 for the NC/Other component and 1.0 for the $\bar{\nu}_\mu$-CC and ν_μ-CC components. The normalizations yielded by the fitting procedure are presented in Table 11.2. The normalization applied to the NC/Other component in the sample is 1.251. The normalization of the $\bar{\nu}_\mu$-CC is 0.99 and the normalization of the ν_μ-CC component is 1.0.

Bkg. Int. Type	Stage 1	Stage 2	Stage 3	Stage 4	Overall
$\bar{\nu}_\mu$-CC Bkg.	1.000	1.000	1.000	0.991	0.991
ν_μ-CC	1.000	1.000	1.000	1.000	1.000
NC/Other	1.250	1.001	0.999	1.001	1.251

Table 11.2: Normalizations applied by the fitting method for background estimation from the warping test where NC/Other is weighted up by 25%.

11.2 Warping the wrong-sign ν_μ-CC background

It is important to observe the results of a warp test in which the wrong-sign component in the pseudo-data sample is weighted, to determine whether the fitting procedure yields the correct normalization on the wrong-sign component. The normalizations yielded by the process for this warp test are shown in Table 12.

Bkg. Int. Type	Stage 1	Stage 2	Stage 3	Stage 4	Overall
$\bar{\nu}_\mu$-CC Bkg.	1.027	0.988	1.011	0.958	0.982
ν_μ-CC	1.027	1.000	1.011	1.000	1.038
NC/Other	1.215	1.139	0.870	1.000	1.205

Table 11.3: Normalizations applied by the fitting method for background estimation from the warping test where wrong-sign ν_μ-CC is weighted up by 25%.

Unfortunately, the procedure increases the NC/Other component in the MC to compensate for the increased rate of wrong-sign events in the pseudo-data. In the normal fitting sequence, the ν_μ-CC and $\bar{\nu}_\mu$ are combined in the first stage of the fit, and the NC/Other template is free to move. The relative freedom is used by MINUIT to make up for the increased rate of wrong-sign events in the pseudo-data. In both iterations, the default fitting procedure floats the NC/Other template and the combined $\bar{\nu}_\mu$-CC and ν_μ-CC templates in stage 1 and letting the $\bar{\nu}_\mu$-CC and NC/Other templates float in stage 2.

Two alterations to the fitting procedure were made to see whether the normalizations can be correctly estimated. The first modification tried was to let NC/Other be assigned half the shift in normalization, freeze the $\bar{\nu}_\mu$-CC template in the second stage, and let the ν_μ-CC template float instead. This enables the fitting to move the wrong-sign template independently in two of the subsequent fitting stages. It was expected that by letting the wrong-sign template have an opportunity to move around, the correct normalizations whould be obtained. These changes did not improve the results: it was observed that the procedure further reduces the normalization of the wrong-sign template while increasing the NC/Other prediction.

A final variation of the fitting procedure involved simply freezing the NC/Other template in the second stages of both iterations. This meant that the NC/Other template could only float in stage 1. This variation on the fitting stage could not yield the correct normalizations on the different event rates. The failure of multiple different combinations indicate that the shapes of the templates that are fit are not distinctive enough to yield reliable estimates of the backgrounds in the data sample.

However, the latter warp test indicates that significant renormalizations imposed by the fitting sequence is evidence of one or more of the templates not predicting the data well. Thus this method can still be utilized to examine how well the NOvA ND MC describes the data collected by the detector.

Chapter 12

Treatment of Systematic Uncertainties

A cross-section measurement involves the reconstruction of kinematic distributions and the isolation of signal contributions. The latter requires prior knowledge/measurement of the cross sections of other processes. Uncertainties are invariably associated with these prior measurements. Furthermore, event reconstruction is an imperfect process, which in turn introduces more uncertainties into the measurements. As a result, the final cross section that is measured contains errors propagated from these two sources. A full accounting of the sources of uncertainty and their effect on the measurement is required. This Section describes how the systematic uncertainties associated with $\bar{\nu}_\mu$ charged current (CC) events in the NOvA Near Detector can be estimated.

In short, estimating an uncertainty for the measurement involves simulating how the cross section distribution might differ from the measurement across a large number of universes in which various systematic parameters have been randomly shifted from their nominal central values. This method has been used for the determination of errors on prior cross-section measurements in NOvA[34] [35]. The differences are used to calculate a covariance matrix [36]. The elements of this matrix are used to assign errors in the $|\vec{q}|$ and E_{avail} bins.

12.1 How multiverses are generated

The multi-universe technique involves the generation of measurements from multiple simulated universes. There is one nominal universe and multiple alternate universes. The nominal universe reflects physics described by the current knowledge, with processes quantified by various parameters known through experimental procedures. In alternate universes, the universe is arranged such that some, many or all physics parameters deviate from the accepted central values. In short, the physics mechanisms in these alternate universes is different from the nominal. Therefore cross sections generated in these alternate universes, for the same interaction, are different.

As a demonstration of multi-universe generation, suppose there is a physics process that is parametrized by the function $y(x) = a(x+60)^3 + b(x+60)$, a and b are the systematic parameters that control the process. Suppose that the nominal central values are $a = 3$ and $b = 1$. A plot of the function $y(x)$ is shown in Fig. 12.1.

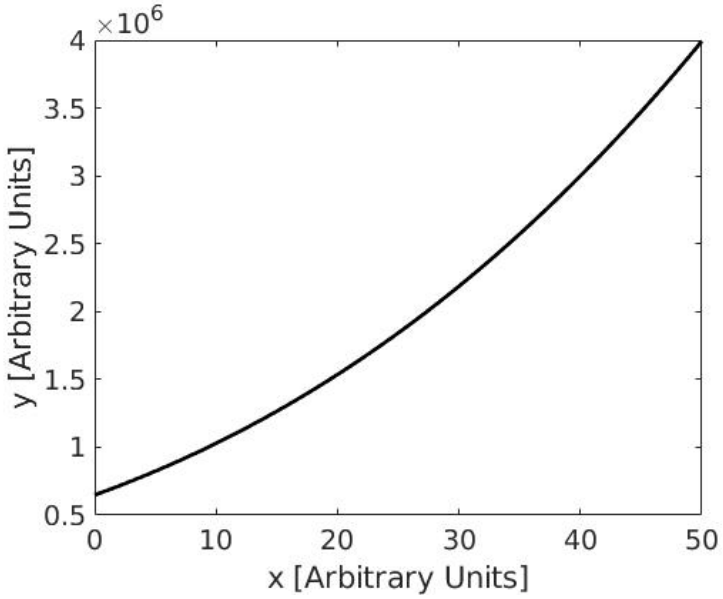

Figure 12.1: Illustration of $y(x) = a(x+60)^3 + b(x+60)$ for the nominal universe with parameter central values of $a = 3$ and $b = 1$.

The first step of this multi-universe technique is to alter the values of the parameters that affect the process. In this example, these are a and b. The variations are done by letting a and b be gaussian random variables with their respective central values and with a standard deviation of

one.

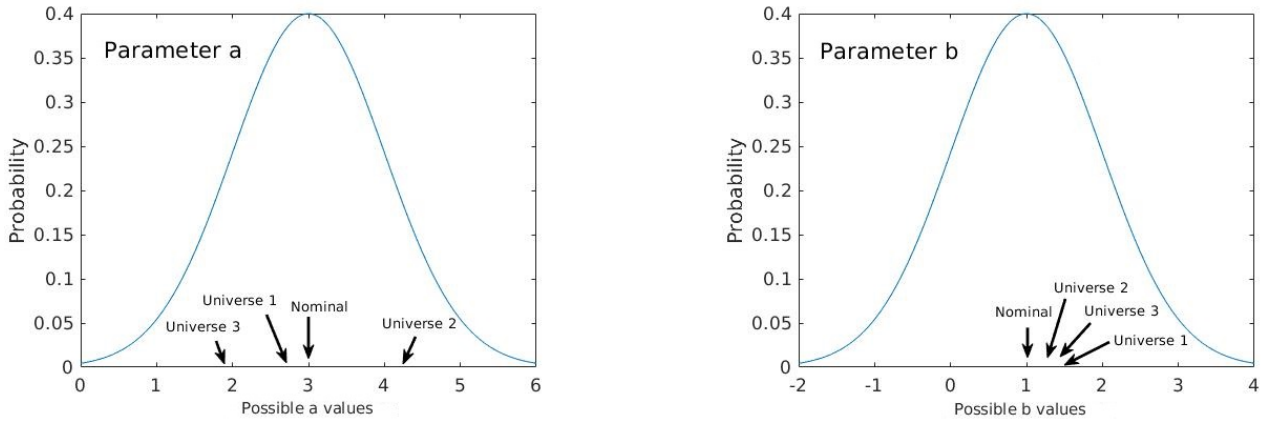

Figure 12.2: Probability distributions of a (on the left) and b (on the right). For the multi-technique, for each universe, a and b take on values randomly and independently based on the the probability distributions. a couple of (a, b) values have been picked out and their effect is demonstrated in Fig. 87.

For each pair of random values of a and b that are chosen, a different y versus x relationship is obtained. This represents the process in one universe. Next, another pair of random values of a and b are picked and the another different relationship is obtained. The different relationship represents the process in another universe, and so on. The distribution of this y-x process in different universes is shown in Fig. 12.3.

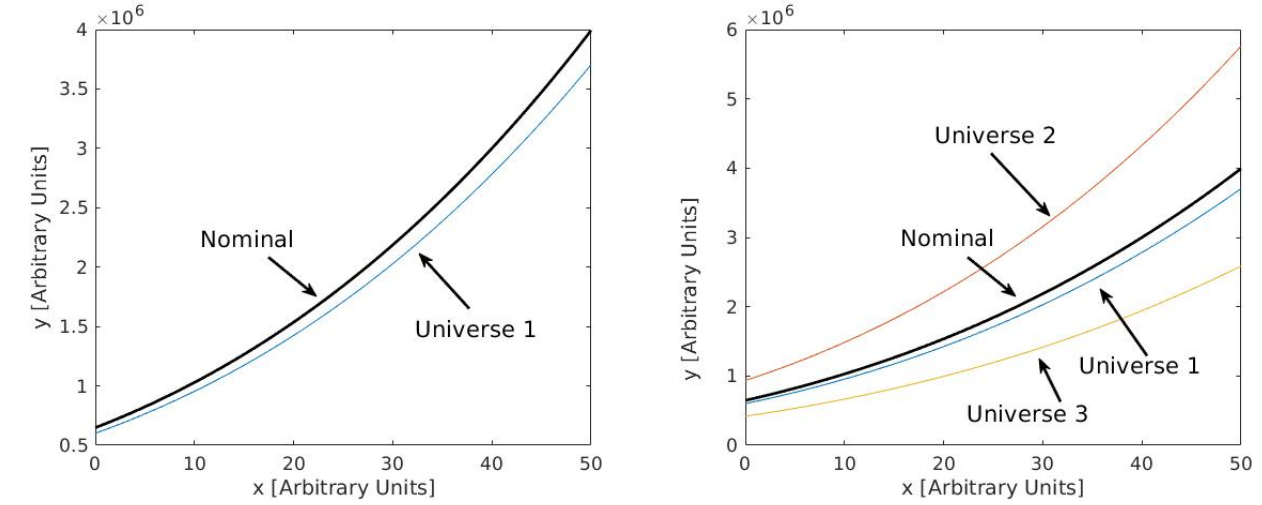

Figure 12.3: Graphs of $y - x$ from multiple universes. The solid black line represents the process from the nominal universe with $a = 3$ and $b = 1$. The left plot shows the graph from one universe, with altered values of a and b, besides the nominal. The figure on the right shows the distribution for three universes.

Following this prescription, the process is repeated for multiple universes. As universes are generated, a spread is created around the nominal distribution. The upper (lower) error band can be obtained by calculating the rms (root mean square) of the deviation of the y values from the nominal curve:

$$\text{Upper (lower) error on y(x)} \;=\; \text{rms}[y_i(x) - y_0(x)], \text{ for all } y_i(x) > (<)\, y_0(x). \tag{12.1}$$

Here, $y_0(x)$ is the nominal value and $y_i(x)$ is the value from the ith alternate universe.

The rms of the spread obtained changes as more and more universes are added. However, the change in the spread diminishes, i.e. the error band converges, when a large sample of universes is generated. The rate of convergence is important, as generating more universes requires a longer computation time. It is necessary to find a number, N, of universes generated where the error band has converged sufficiently that any further additions of universes produces negligible changes.

Consider an error band that is formed by an ensemble of n universes, denoted *error band*(n). Adding another universe to n shifts the rms of the spread and produces another error band, *error band*$(n+1)$. Adding another universe to the ensemble produces another error band, *error band*$(n+2)$. Now, the shift in the error bands upon generation of more universes is usually unequal, i.e. the difference between *error band*$(n+1)$ and *error band*$(n+2)$ may not be the same as *error band*(n) and *error band*$(n+1)$. However, the average change in the error bands, for adding universes one by one, for a certain number of universes, n_1, can be calculated as follows:

$$\delta\epsilon_{n+n_1,n} = \frac{\sum_{i=n}^{n_1} \text{e-band}(i+1) - \text{e-band}(i)}{n_1}. \tag{12.2}$$

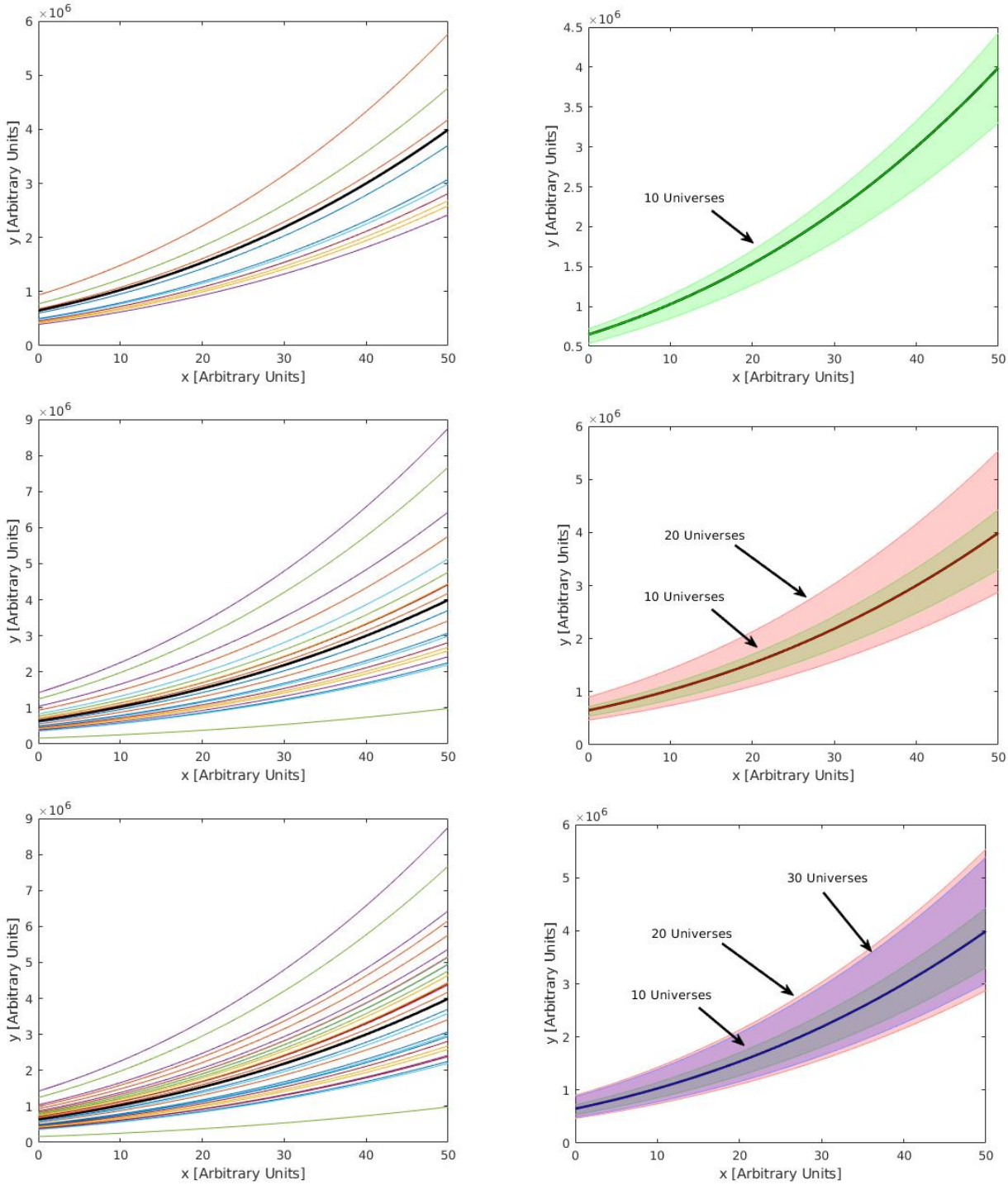

Figure 12.4: The spread of universes obtained from generation of multiple universes. On the right side, from top to bottom, is the result of generation ten, twenty and thirty universes. On the right side, from top to bottom, is the RMS of the spread (above and below the nominal) from generating ten, twenty and thirty universes. The RMS spreads are superimposed for the middle-right and lower-right figure to demonstrate how this changes with the addition of universes.

In Eq. 12.2, e-band(i) represents the error band obtained from generating i universes. The

numerator expression yields the sum of the changes in the error band as successive universes are added, for n_1 universes. The mean change in the error bands for adding n_1 universes to the first n universes is $\delta\epsilon_{n+n_1,n}$. For the example shown in this section, n_1= 10.

The fractional mean change in the error bands can also be calculated:

$$\text{Fractional mean change in error bands, } F = \frac{\delta\epsilon_{n+n_1,n}}{\textit{e-band}(n+n_1)}. \tag{12.3}$$

This fractional mean change is a quantity that converges as universes are added to the ensemble. This convergence trend is shown in Fig. 12.5. This metric is suitable for determining the number of universes that are required to get proper error coverage. The first bin of the plot represents how the rms spread changes, as a fraction of the new rms, when 10 universes are added on top of the first 10. The second bin represents the same quantity for 10 universes added to the first 20, and so on. A respectable degree of convergence is achieved for F $<$ 1%. In this case, the appropriate location is seemingly 270. However, upon further addition of universes, the value of F spikes when the total number of universes is 480. Therefore, the appropriate number of alternate universes to generate the error band is 480 in this case.

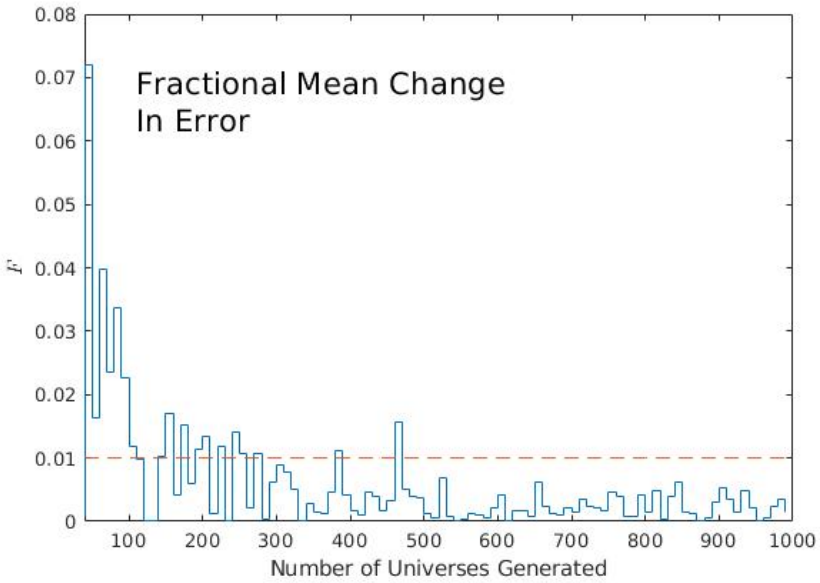

Figure 12.5: Fractional change in the mean of the total spread of y(x). The dashed line denotes the value of F = 1%.

12.1.1 Sources of systematic uncertainty

12.1.1.1 Uncertainty categories

There are 97 sources that affect the error range of the cross section measurement. Among them, the effect of the 93 systematic sources are examined by applying systematic shifts to the MC. The remaining 4 systematic uncertainties are examined via alternate MC samples. Each source falls under one of the four following categories:

1. Modeling parameters that govern the simulation of the neutrino-nucleus interactions.
2. Neutrino flux parameters.
3. Parameters that affect the energy depositions of particles in the detector.
4. Parameters that affect detector calibrations and reconstruction.

Some key features of each of these four categories are discussed below:

Neutrino-nucleus interaction parameters

An MC sample is used estimate the distribution of signal events from the measured data. Furthermore, this MC sample is also used for unfolding and efficiency correcting the estimated signal distributions. Consequently, any uncertainties in the distributions made from MC samples will propagate to the final measurement.

The MC sample reflects current knowledge of neutrino interaction physics. Event distributions generated from the MC depend on the cross sections of various channels and on intranuclear hadron interactions. These are modeled using a number of different parameters such as the axial mass of the axial-vector dipole form factor and nucleon mean free path. There are 80 systematic parameters in this category.

Neutrino flux parameters

Reconstruction of the neutrino flux in the near detector requires information about the beam focus and transport, flux model and PPFX (Package to Predict the Flux), all of which have

associated uncertainties. Furthermore, the RHC beam has a significant neutrino component, which is estimated to be around 10% [37]. An uncertainty of the flux reflects the fact that the actual beam flux can be higher or lower, which can result in a measurement that is higher or lower compared to the actual cross section.

Energy scale systematics

These systematics represent the uncertainties involved in reconstructing the energy deposited by final state particles in the detector. For the $\bar{\nu}_\mu$ CC inclusive cross section analysis, an important systematic is the neutron energy deposition. This is because most $\bar{\nu}_\mu$ CC events have an outgoing primary neutron. Two of the systematics in this category are related to the muon energy deposition. The muon deposits energy in the actual ND and the muon catcher of the ND, and there is a systematic error uncertainty associated with the energy deposited in each region. Furthermore, another systematic affects the correlated energy deposition in the two near detector regions. A fifth systematic reflects the uncertainty in the deposited muon energy arising from the overlapping of tracks. The sixth uncertainty covers the uncertainty involved in reconstructing the muon momentum.

Calibration and light level systematics

Energy reconstruction of particles in the detector require the collection of scintillation and Cherenkov light. And the detector cells must be properly calibrated so that the energy reconstruction is accurate. There are uncertainties associated with how much scintillation light is produced as charged particles travel through the detector. Furthermore, there are difficulties in modeling the Cherenkov light produced by charged hadrons. There are two special light level samples that represent the uncertainties in the scintillation light production and a third MC sample that can be used to determine the error arising from Cherenkov light uncertainties. Furthermore, there are three special MC samples in which the detector response has been altered. These alterations in the detector response represent the uncertainties in the photoelectron energy conversion, giving rise to calibration errors which affect the measurement.

This is the only category of systematics which will not be used as inputs to the multiverse technique. Rather, these exist as separate MC samples, with the relevant changes in the detector response simulated, thus affecting reconstruction. The error bands obtained from these systematic sources will be quadrature-summed with the error bands obtained from doing a multiverse computation.

12.1.2 Sources of largest uncertainties

Among the 97 systematic error sources, some drive the shape of the error bands much more than others. To identify the leading ones, cross sections have been generated with the MC sample with the $+1\sigma$ and -1σ shift applied on the individual systematic. These systematic shifts change the cross-section normalization and/or shape. These systematically shifted cross sections are compared to the nominal cross section and the ones deviating most from the nominal cross section are noted. The deviation of these various cross sections from the nominal comprises the error arising due to systematics. A list of leading systematics is provided below:

1. Uncertainty in the normalization of the neutrino/antineutrino flux in the beam.
2. Uncertainty in the data-driven 2p2h-MEC $|\vec{q}|$ versus q_0 shape distribution.
3. Uncertainty in the normalization and shape of the 2p2h-MEC $E_{\bar{\nu}}$ spectrum.
4. Z-expansion parameters that relate to the CCQE cross-section form factors.
5. The axial-vector and vector mass values used in the dipole form factors of CC baryon resonance events.

12.1.2.1 Flux uncertainty

The uncertainty arising from the beam flux is a leading systematic. As mentioned previously, the flux uncertainty is estimated to be around 10%. Within the NOvA software framework, there are two ways to estimate the uncertainty due to flux systematics. One way is to use the individual flux systematics in conjunction with the PPFX systematics [39]; PPFX is a software package

that corrects the hadron production in the beam based on external data. There is a central value correction to the flux that the PPFX package yields, and a array of weights that can be used to determine the uncertainty on the PPFX correction. This way to determine the systematics is used by the previous ν_μ-CC analysis [37]. The other method of getting the error coverage is to use the flux Principal Component Analysis. Many of the individual flux systematics are correlated. The PCA is a method for taking the individual flux systematics and calculating a set of decorrelated systematics that can be utilized to obtain the full flux coverage. The uncertainty estimated by either method is nearly identical, however the PPFX-based systematics are everywhere slightly smaller. This is shown in the fractional error plots in Fig. 12.6 where a magnified scale is used to elicit the small difference between the two methods.

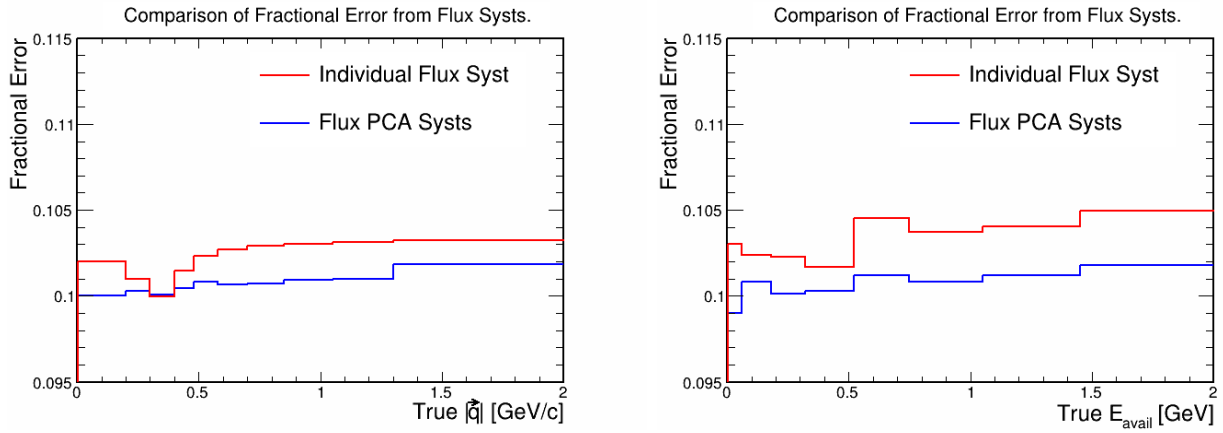

Figure 12.6: Comparison of fractional errors on the $\bar{\nu}_\mu$-CC inclusive cross section due to the flux estimated by the individual flux systematics (red) and with the PCA systematics (blue), shown for the $|\vec{q}|$ (left) and E_{avail} (right) projections.

For this analysis, the flux PCA systematics are used to determine the uncertainty due to flux on the cross section measurements. The effect of the flux uncertainty on the inclusive cross section is shown in Fig. 12.7. It is a relatively constant shift of the cross section up and down with negligible shape effect.

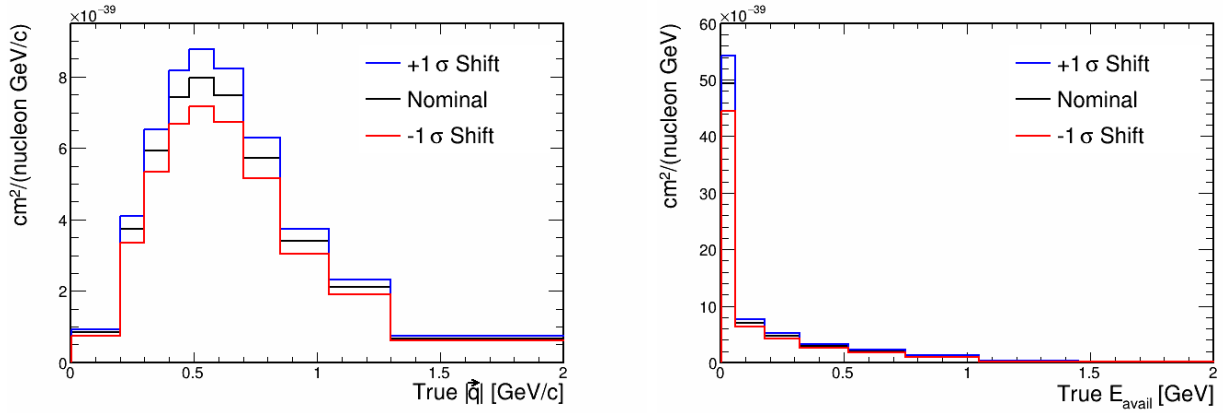

Figure 12.7: Effect of $+1\sigma$/-1σ (blue and red histograms respectively) shifts of the flux systematic on the $\bar{\nu}_\mu$-CC inclusive cross section. The plots on the left and right show the $|\vec{q}|$ and E_{avail} projections of the cross-section respectively. The nominal cross-section projections (shown in black) are superimposed for comparison.

12.1.2.2 Leading cross-section systematics

The 2p2h-MEC systematics introduce the largest uncertainties in the inclusive cross section. This is an uncertainty assigned to the NOvA tune of the Valencia MEC model, which have been developed by ND data-MC fits over q_0-$|\vec{q}|$ variables. The uncertainty on the MEC model is obtained by morphing the shape all non-MEC interactions via systematic shifts, and redoing the fits. Another leading systematic is the uncertainty on the 2p2h-MEC cross section dependence on the neutrino energy. This uncertainty is determined by examining how the cross section of various models differ from the Valencia MEC model as a function of neutrino energy. Details of the assigned MEC uncertainties can be found in Ref. 16.

The next leading cross-section systematics arise from the uncertainties associated with the axial-vector and vector mass values used in cross section predictions of CC baryon resonance events [37].

The final leading GENIE uncertainty is the Z-expansion uncertainty. This is related to the uncertainty associated with the axial form factors associated with CCQE interaction predictions [40]. Again, the details of the uncertainties due to the Z-expansion are explained in Ref. 16.

The effect of the leading GENIE systematics on the inclusive $\bar{\nu}_\mu$-CC cross-section are shown in Figs. 12.8 and 12.9.

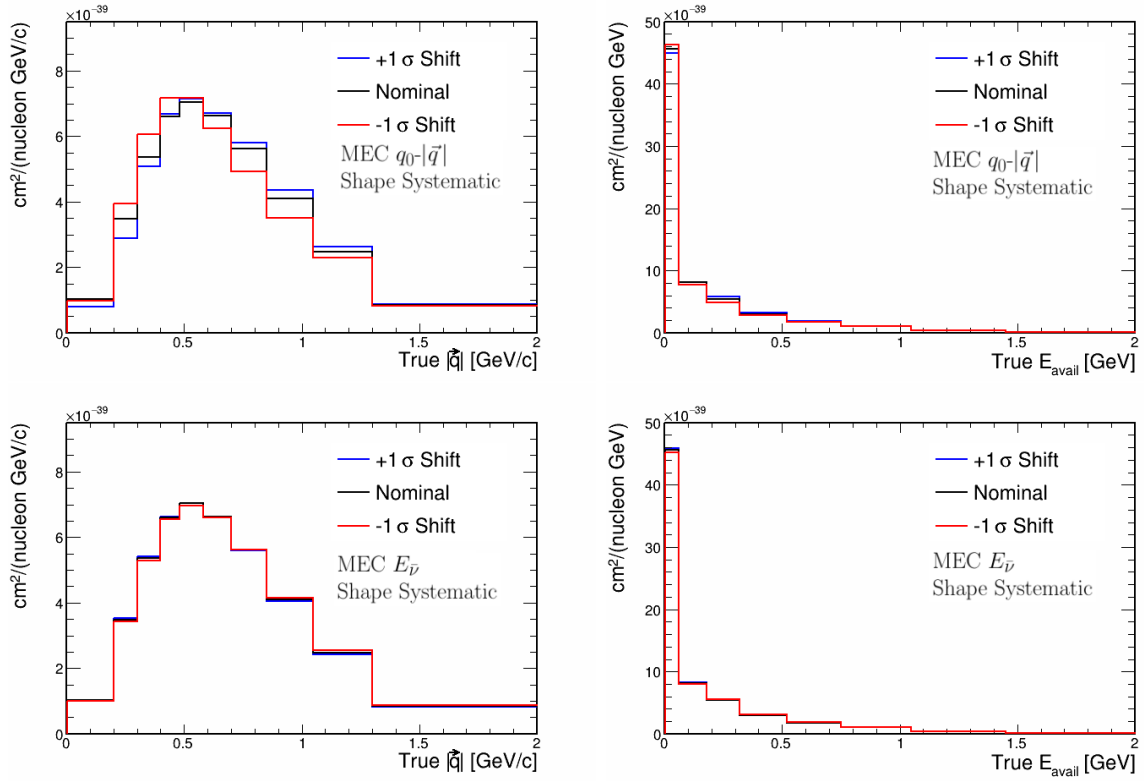

Figure 12.8: Effect of $+1\sigma$/-1σ (blue and red histograms respectively) shifts of the MEC Shape and MEC E_ν shape on the $\bar{\nu}_\mu$-CC inclusive cross section. The plots on the left and right show the $|\vec{q}|$ and E_{avail} respectively. The nominal cross section projections (shown in black) are superimposed for comparison.

A summary of the errors that arise on the cross section from these systematics is provided in Table 12.1.

Systematic from list	Frac. Error on XSec in $(|\vec{q}|, E_{avail})$		
	(0.25, 0.12)	(0.44, 0.25)	(0.78, 0.44)
MEC q_0 vs. $|\vec{q}|$ shape	14.5%	4.7%	7.2%
MEC E_ν dependence	5.9%	2.1%	1.1%
M_A CC RES	1.6%	3.5%	8.7%
M_V CC RES	0.8%	1.2%	5.3%
Z-expansion CCQE	1.7%	0.6%	0.75%

Table 12.1: Errors on the cross section in representative bins of $(|\vec{q}|, E_{avail})$ (in units of GeV/c and GeV respectively) arising from the systematic sources listed in Section 3.2.

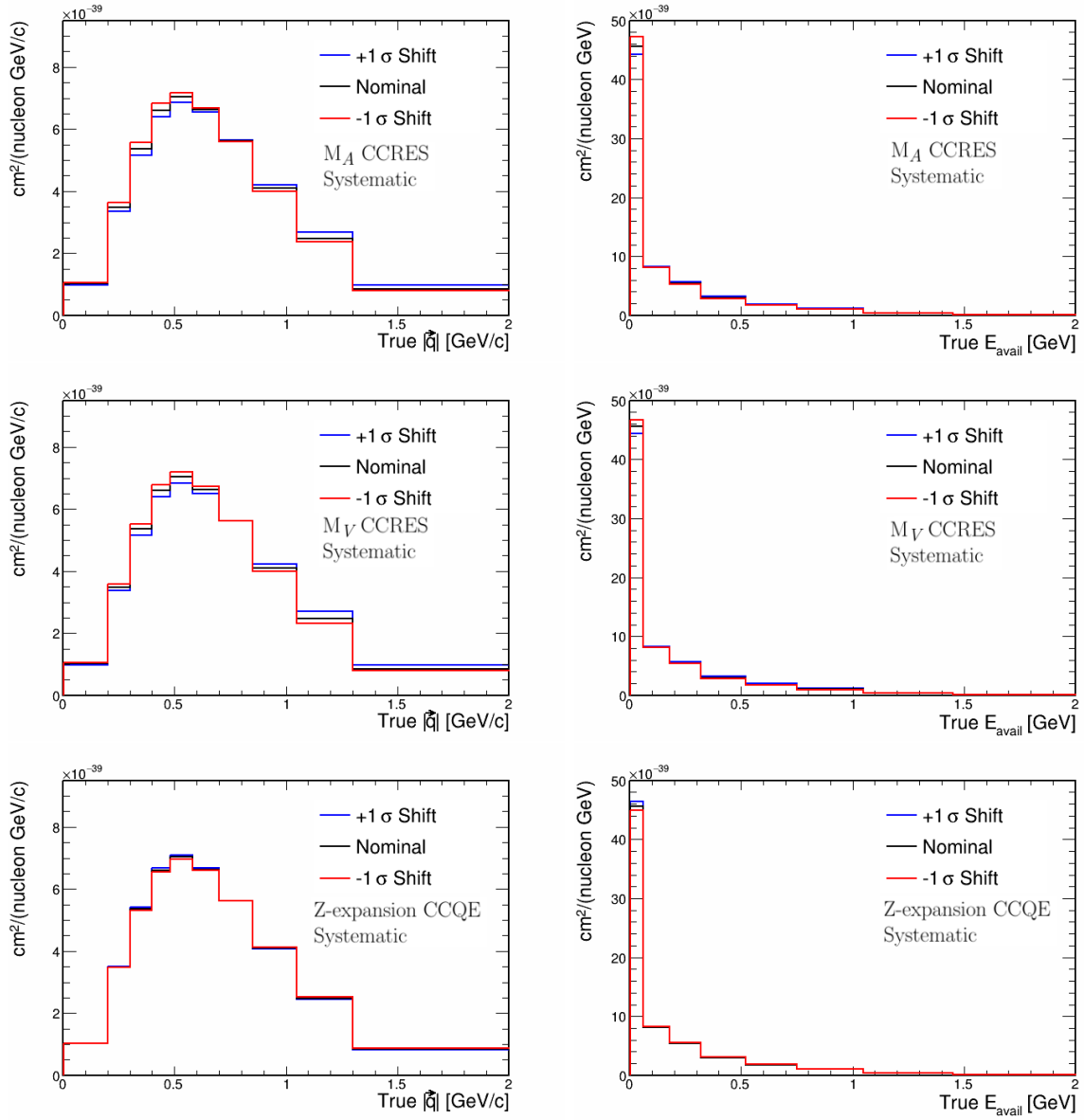

Figure 12.9: Effect of $+1\sigma$/-1σ (blue and red histograms respectively) shifts of the M_A CCRES and M_V CCRES and Z-expansion CCQE systematic parameters on the inclusive cross section. The plots on the left and right show the $|\vec{q}|$ and E_{avail} respectively. The nominal cross section projections (shown in black) are superimposed for comparison.

12.2 Systematic Uncertainties from 2p2h-MEC

As seen in Fig. 1.1, the excess believed to arise from 2p2h events is significant and has a big impact on the final cross-section measurement. Thus it is necessary to estimate the uncertainty arising from lack of knowledge of these reactions. The uncertainties affect the event populations in

different parts of the analysis phase space, thus affecting the cross-section calculations. Work has been previously done to determine this uncertainty [41]. However, these uncertainties have been developed by tuning the 2p2h-MEC model to the Near Detector data, making them unsuitable for a cross-section measurement. An alternative estimate of the 2p2h-MEC systematic uncertainty can be obtained by examining the spread of the event distributions across various models. The available models are the Valencia MEC model [4, 12], the MINERvA tuned MEC models [3, 2] and the Superscaling (SuSA) 2p2h-MEC [5, 13] model. The physics behind these models have been explained in Chapter 2. In the following subsections, these models will be explained further. Projections of the models in each variable are shown in Fig. 12.10. In the $|\vec{q}|$ projection plot (Fig. 12.10, left), the cutoff of the Valencia and MINERvA models at 1.2 GeV of momentum transfer is visible. Only the SuSA model extends further, with the cross section projection going to zero above 2 GeV. It can be seen that most models peak at $|\vec{q}| \simeq 0.5$ GeV. On the q_0 projection, the double-peak structure of the Valencia model and the NOvA-tuned-Valencia model is visible. All the other models have a single peak. In both projections, the SuSA model has a larger cross section than the Valencia model. This is similar to what is observed in the case of neutrino-nucleus scattering. However, the relative strength of the SuSA model is not as large in the case of antineutrino-nucleus scattering [13].

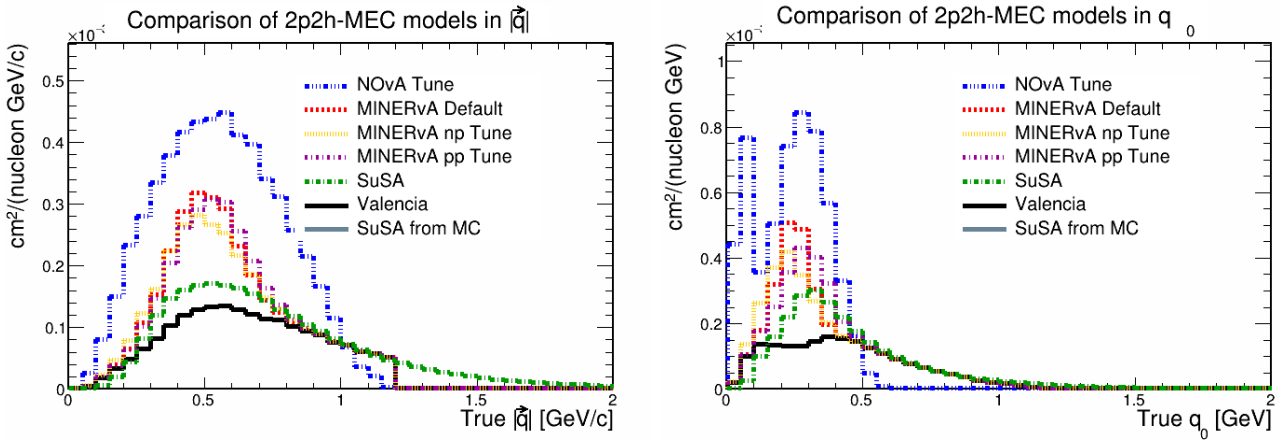

Figure 12.10: Projections of the 2p2h-MEC models in variables $|\vec{q}|$ (left) and q_0 (right).

The cross section of the SuSA model in relation to the Valencia model has been reported in Ref. [9]. The relevant cross sections as functions of neutrino/antineutrino energy are reproduced

in Fig. 12.11. The ratio of the two subsets for each model is given in Table 12.2. The Valencia model and MINERvA default tune have an almost 50% split between the np $\rightarrow$ nn and pp $\rightarrow$ np components. The SuSA model and MINERvA np tune have much smaller proportions of pp $\rightarrow$ np events in comparisons, with the (pp $\rightarrow$ np)/ (np $\rightarrow$ nn) ratio approaching 0.25.

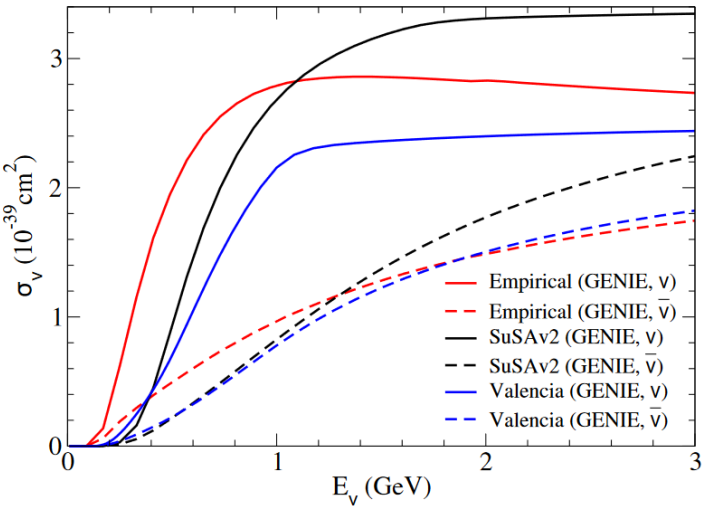

Figure 12.11: Cross sections of the GENIE empirical, SuSA and Valencia 2p2h-MEC models versus neutrino energy (E_ν). The extent to which the cross sections of the SuSA and GENIE empirical models exceed that of the Valencia model, is more pronounced for neutrino interactions (solid-line cross sections) than for antineutrino interactions (dashed-line cross sections).

Model	Ratio of (pp $\rightarrow$ np to np $\rightarrow$ nn evts)
Valencia	0.48
SuSA	0.28
MINERvA Default Tune	0.56
NOvA Tune	0.51
MINERvA np Tune	0.27
MINERvA pp Tune	1.36

Table 12.2: Ratios of the pp $\rightarrow$ np to np $\rightarrow$ nn components of the different MEC models. Note that the SuSA model has a ratio which is nearly half of that predicted by Valencia and by the MINERvA default tune.

12.2.1 Comparison of models by initial state dinucleon pair using true kinematics

It is of interest to examine the $|\vec{q}|$ versus q_0 distributions of the models separated according to the dinucleon pair with which the incoming antineutrino interacts. The comparisons shown in this section are based entirely on MC truth kinematics. Comparisons using reconstructed kinematics are shown in a later section.

Dinucleon transitions: Valencia model

For the Valencia model, the np $\rightarrow$ nn component (Fig. 12.12, left) has the greater strength, and most of the strength is concentrated in two regions. This component contributes events to both peaks of the Valencia model, with many more events being associated with the higher peak in energy transfer. In contrast, the pp $\rightarrow$ np model (Fig. 12.12, right) distributes more uniformly and contributes more evenly to the two Valencia peaks.

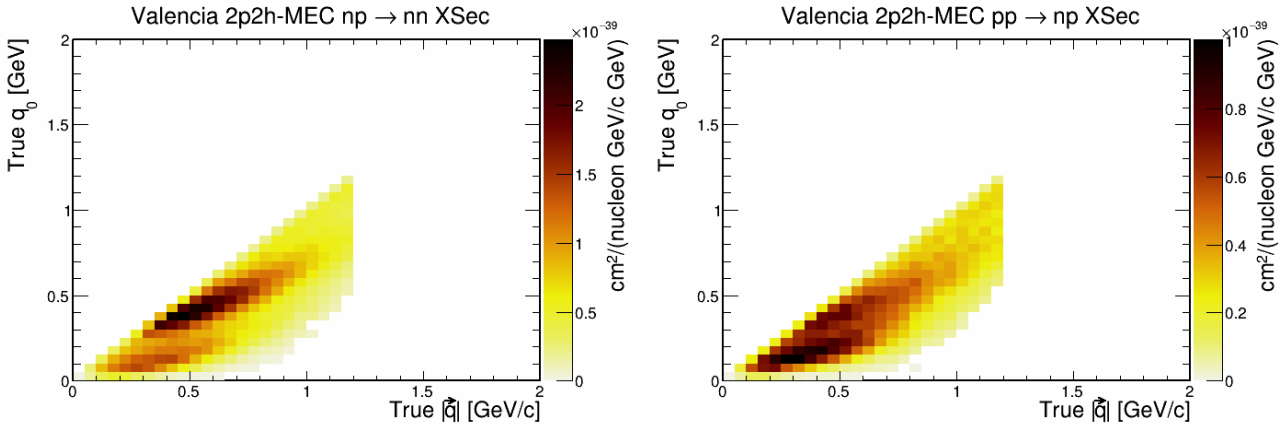

Figure 12.12: Cross section distributions of the np $\rightarrow$ nn (left) and pp $\rightarrow$ np (right) components of the Valencia model in $|\vec{q}|$ and q_0. The latter component contributes much less strength as indicated by the scales on the right hand side of either plot.

Dinucleon transitions: SuSA model

For the SuSA model, the np $\rightarrow$ nn component of the 2p2h-MEC interactions (Fig. 12.13, left) has a distribution similar that of the overall model. There is a narrow strip of events with an average of 0.3 GeV energy transfer. This strip narrows and becomes negligible as the momentum

transfer increases. The pp → np component (Fig. 12.13, right) has a much broader distribution in energy transfer. It can be seen that events in this latter category fall into two regions. One subsample has the same kinematics as the np → nn counterpart. But there also exists another subsample of events that has less energy transfer.

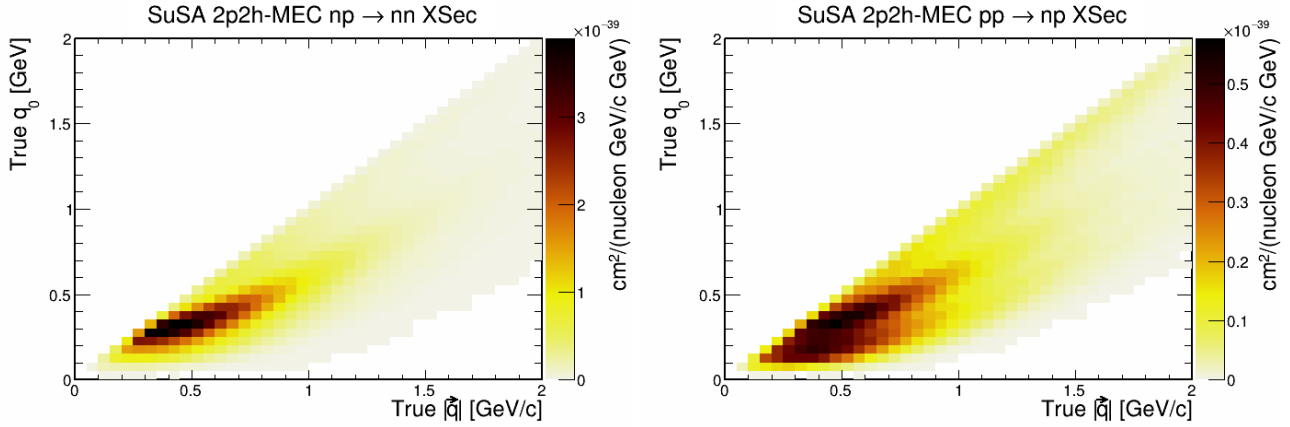

Figure 12.13: Cross-section distributions of the np → nn (left) and pp → np (right) components of the SuSA model in $|\vec{q}|$ and q_0. The latter process has a distantly lower cross section than the former, and that the distribution bifurcates with increasing values of $|\vec{q}|$ and q_0.

Dinucleon transitions: MINERvA default tune

The component processes np → nn and pp → np of the MINERvA default tune are displayed in Fig. 12.14, left and right respectively. Most of the events of each component are concentrated around q_0 of 0.3 GeV with $|\vec{q}|$ ranging between 0.3 and 0.7 GeV.

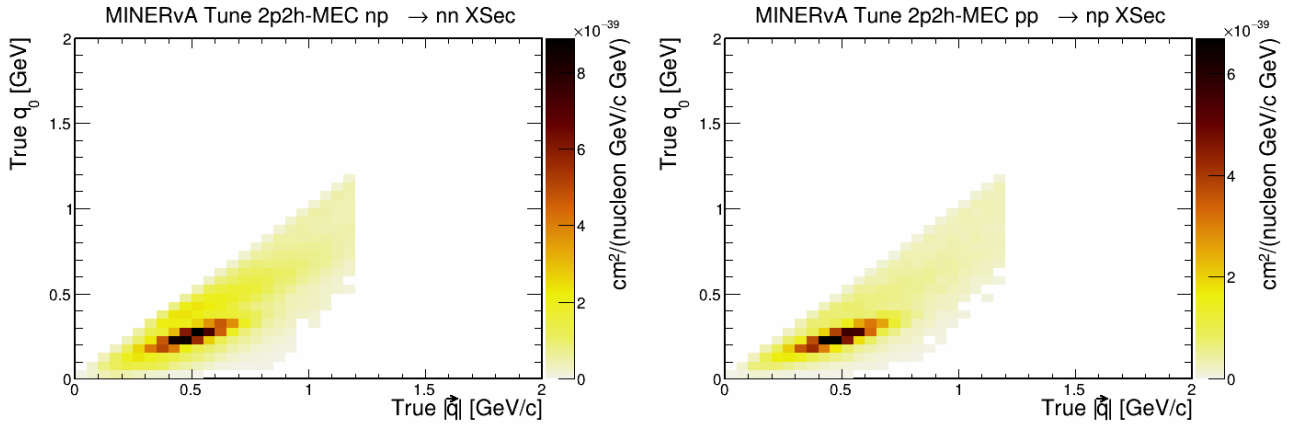

Figure 12.14: Cross section distributions of the np → nn (left) and pp → np (right) components of the MINERvA default tuned 2p2h-MEC in $|\vec{q}|$ and q_0.

Dinucleon transitions: NOvA tune

The two 2p2h-MEC interaction types for the NOvA tuned Valencia model are shown in Fig. 12.15. The np $\rightarrow$ nn interactions have a distinct two-peak structure; there is one peak in the region 300 MeV $< q_0 \leq$ 1 GeV and a slightly greater peak in 0 MeV $< q_0 \leq$ 100 MeV. However, the pp $\rightarrow$ np interactions are mostly concentrated in one region with 300 MeV $< q_0 \leq$ 500 MeV.

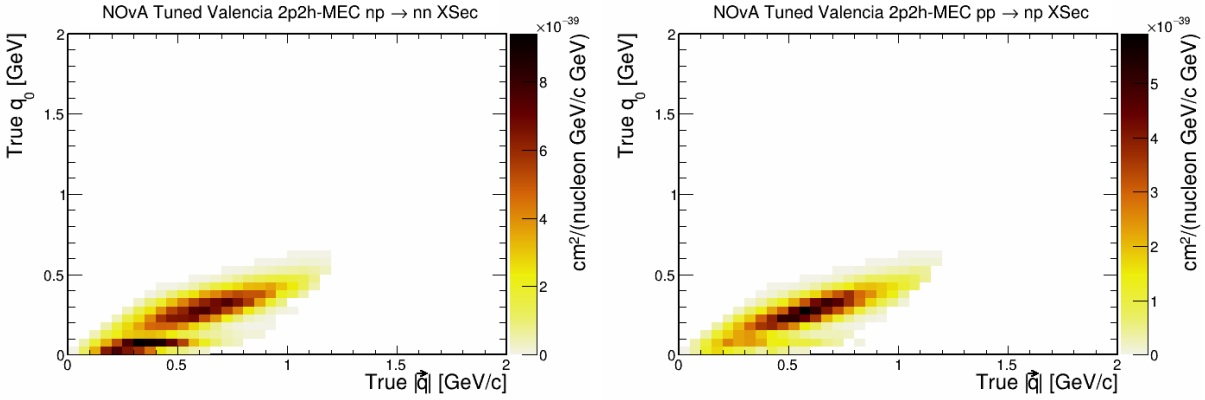

Figure 12.15: Cross section distributions of the np $\rightarrow$ nn (left) and pp $\rightarrow$ np (right) components of the NOvA Tuned 2p2h-MEC in $|\vec{q}|$ and q_0.

Dinucleon transitions: MINERvA np and pp tunes

The MINERvA np tune only affects the cross section of the antineutrino interacting with the np initial state, i.e np $\rightarrow$ nn events. The two-peak distribution of the np $\rightarrow$ nn component of the Valencia model is transformed into a single peak structure. The magnitude of the cross section is also enhanced. The other component of Valencia 2p2h-MEC is left untouched by this tune, which is why the distributions of Figs. 12.12, right and 12.16(b) are identical. Likewise, the MINERvA pp tune involves only modifying the Valencia initial state pp interaction cross section. There is a noticeable difference in the magnitude and the shape of the cross sections of the pp $\rightarrow$ np component of Valencia (Fig. 12.12, right), compared to that of the MINERvA pp tune (Fig. 12.16(d)). For this tune, the np $\rightarrow$ nn component is untouched, which is why the Valencia np $\rightarrow$ pp component (Fig. 12.12, left) is the same as that of the MINERvA pp tune (Fig. 12.16(c)).

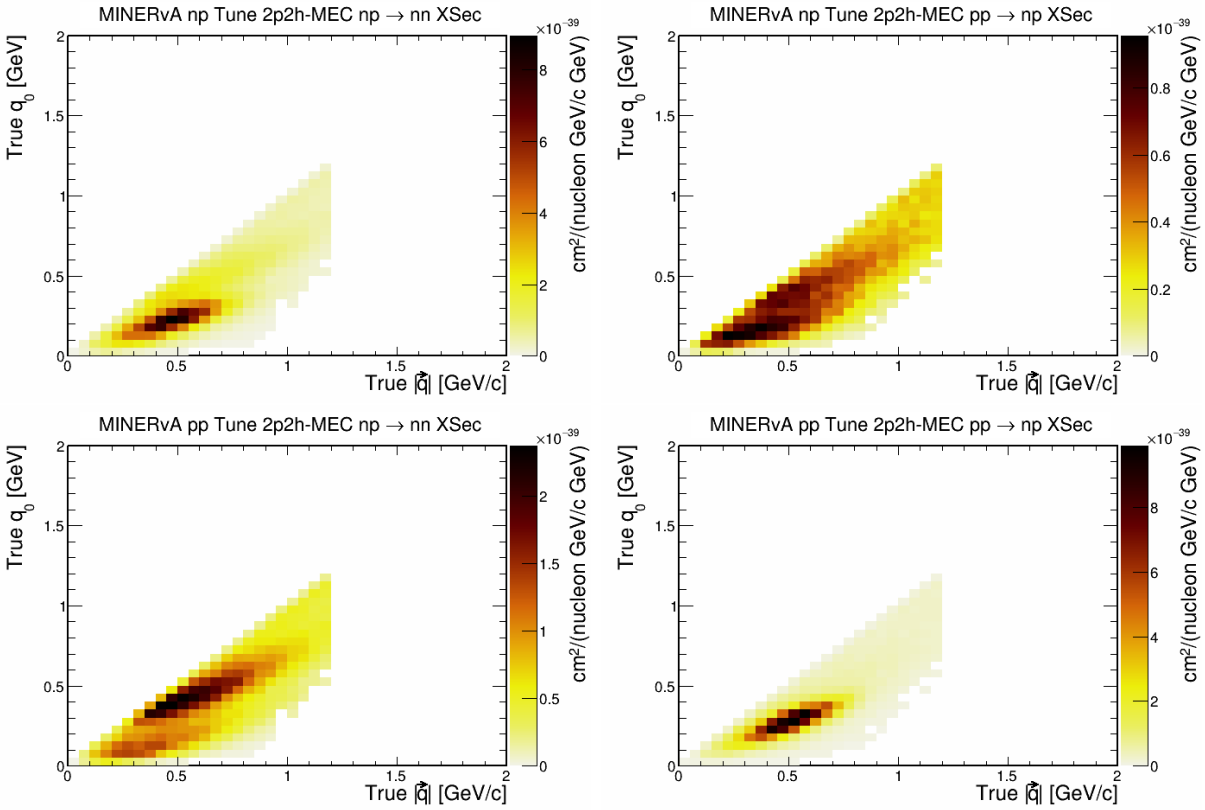

Figure 12.16: Cross-section distributions of the np → nn (left) and pp → np (right) components of the MINERvA np and nn tunes of the Valencia 2p2h-MEC model in $|\vec{q}|$ and q_0.

The projections of the cross sections for each event category in each variable are shown in Figs. 12.17 and 12.19. Most models peak around 0.5 GeV in $|\vec{q}|$ and 0.3 GeV in q_0 the significant enhancements required in the 2p2h-MEC model as reported by the MINERvA and NOvA experiments are apparent in these figures. The NOvA tune has a double peak structure in q_0. This results from the fit of the MEC model to the excess observed at low E_{had}^{vis} in the NOvA data [16].

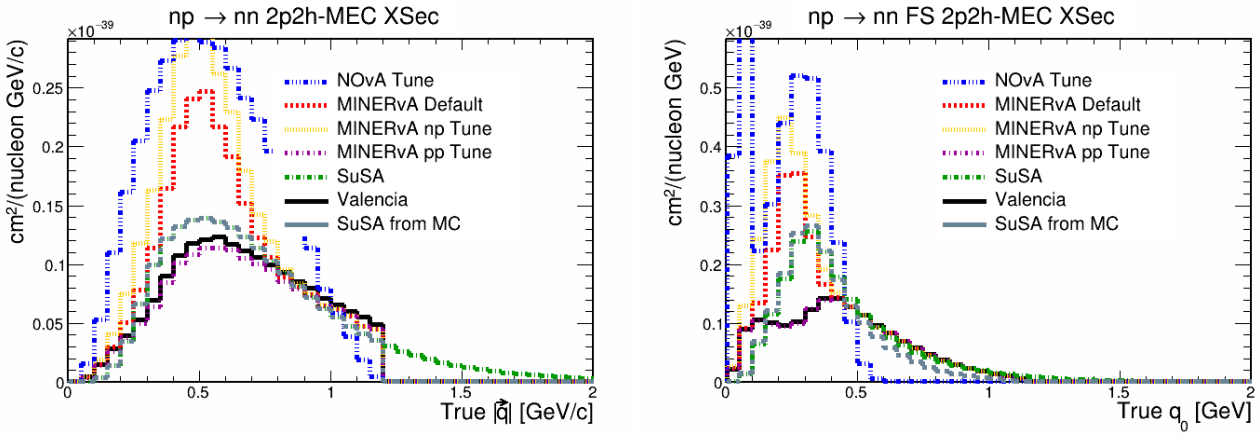

Figure 12.17: Projections of the cross section in $|\vec{q}|$ and q_0 for np $\rightarrow$ nn 2p2h events. The distributions are for the true values of the kinematic variables.

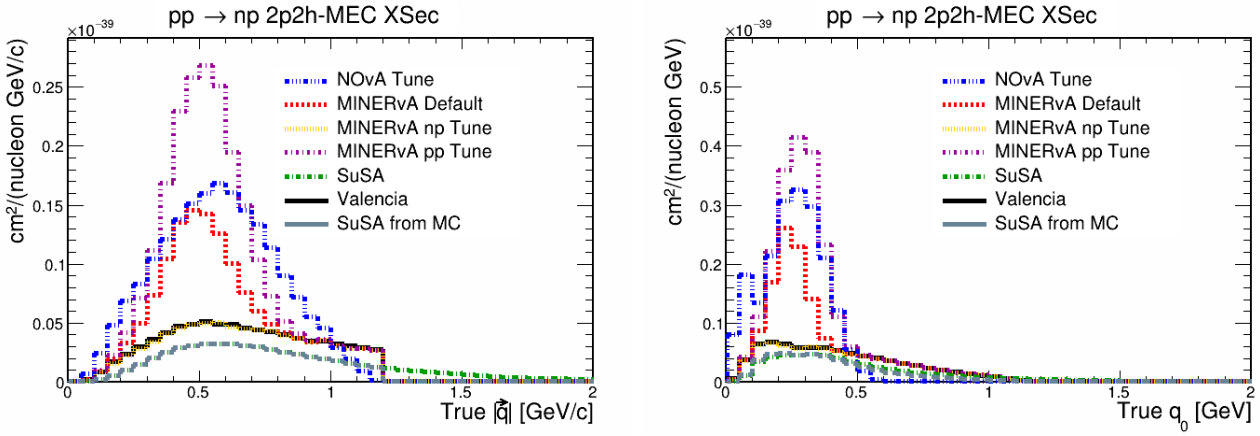

Figure 12.18: Projections of the cross section in true $|\vec{q}|$ and true q_0 for pp $\rightarrow$ np 2p2h-MEC events.

12.2.2 Weighting Valencia events to SuSA events

The SuSA model, as the most recent, fully developed, relativistic model is chosen as one of the models to be used to derive an uncertainty estimate for 2p2h-MEC modeling. In order to reproduce event kinematic predictions of the 2p2h-MEC component in the RHC MC, the base Valenica model is weighted. A weighting scheme for obtaining SuSA has been developed which is similar to one used for the ν_μ-CC inclusive and 2p2h-MEC analysis [29]. The sample of SuSA events as shown in Figs. 2.6 and 12.13 is used to calculate a triple-differential cross section in variables $(|\vec{q}|, q_0, E_\nu)$ for each of the 2p2h components: np $\rightarrow$ nn and pp $\rightarrow$ np. Cross section is calculated similarly for the Valencia model in the same variables. The result of the cross sections

of the SuSA model divided that by that of the Valencia model is used to weight NOvA MC events for SuSA events. To obtain the SuSA model from the Valencia model, the initial dinucleon with which the antineutrino interacts with is examined, and the relative weight is applied. Note that the SuSA model that is obtained by weighting the Valencia model has the characteristic $|\vec{q}|$, | cutoff.

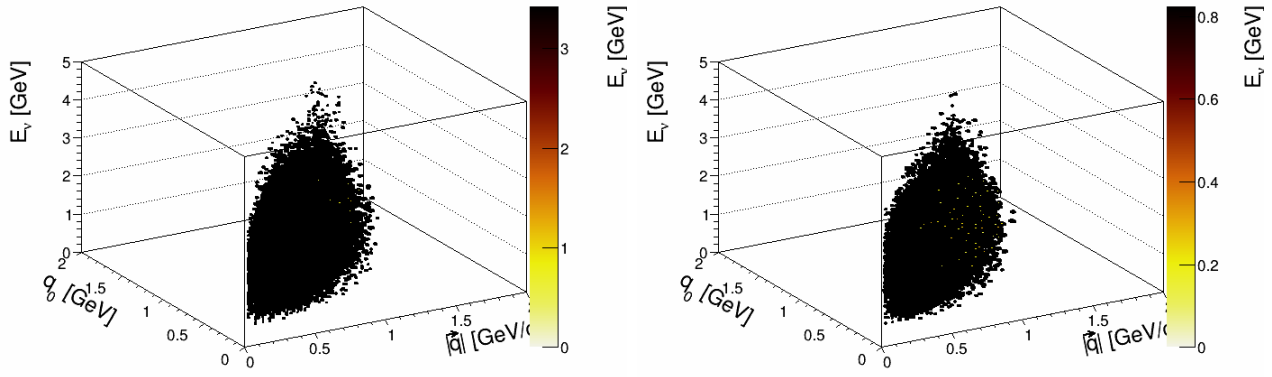

Figure 12.19: Weights for conversion of $\bar{\nu}_\mu$-CC Valencia 2p2h events into SuSA 2p2h events. The plot on the left indicates weights for the subset of events np $\rightarrow$ nn and the one on the right shows weighting for pp $\rightarrow$ np events. The dinucleon upon which the incoming lepton interacts is identified first and the relevant weighting scheme is then used.

12.2.3 Model comparisons in reconstructed $|\vec{q}|$ and E_{avail}

The $\overline{\nu}_\mu$ CC inclusive cross-section measurement will be done in three-momentum transfer and available energy, E_{avail}. It is important to consider what the different 2p2h-MEC models may look like in these variables when reconstructed from observed events. The majority of events are concentrated in the lowest E_{avail} slice. Most 2p2h-MEC events may have an outgoing nn hadron pair, with zero available energy because neutrons do not contribute to this variable. However, some 2p2h-MEC events may have an outgoing np pair, in which case the proton contributes to E_{avail}. Some of the available energy comes from charged hadrons being knocked out due to final-state interactions within the target nuclei.

Valencia 2p2h in reconstructed variables

The event distribution of the Valencia model in $|\vec{q}|$ and E_{avail} is shown in Fig. 12.20. Most of the 2p2h-MEC events are concentrated in the lowest E_{avail} slice, which has an upper limit of 60 MeV. Most of the 2p2h-MEC contribution comes from events in which the outgoing hadron pair is nn, which explains the concentration of events in low E_{avail}. There is also a significant number of events with E_{avail} within 60 - 350 MeV and $|\vec{q}|$ within 300 - 850 MeV/c.

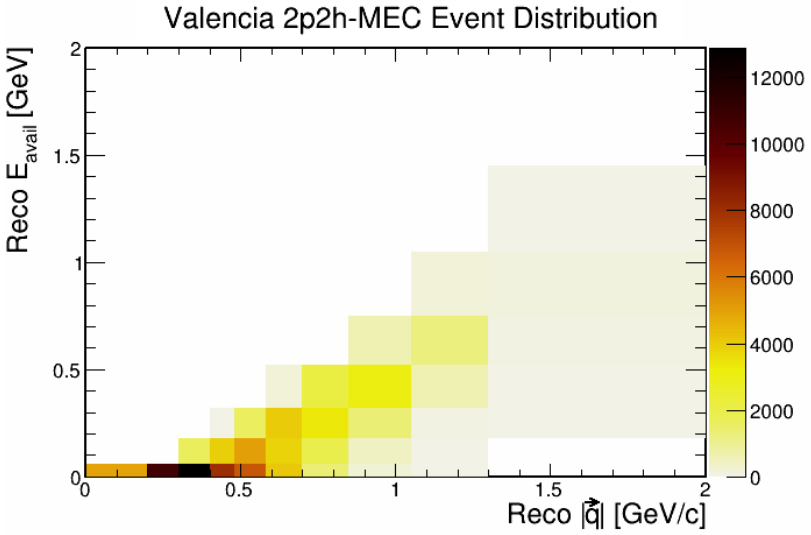

Figure 12.20: Event distribution of $\bar{\nu}_\mu$-CC Valencia 2p2h-MEC event distribution in reconstructed $|\vec{q}|$ and E_{avail}.

SuSA-v2 2p2h in reconstructed variables

The predicted distribution of SuSA 2p2h-MEC events is shown in Fig. 12.21. The pattern of events observed across E_{avail} is very similar to that of the Valencia model. In fact, the E_{avail} distribution of most 2p2h-MEC models is similar. The majority of events populate the lowest E_{avail} slice, with a small portion of events spreading out across higher E_{avail} values. This occurs because 2p2h-MEC interactions favor outgoing nn pairs, which do not have visible signatures. Events that have an outgoing np hadron pair do not produce much available energy as the proton itself often has low energy. The SuSA model predicts a higher event rate than that of the Valencia model. Additionally, the $|\vec{q}|$ spread of the SuSA model in the lowest E_{avail} slice is different from that of the Valencia model. The SuSA model event distribution peaks at a higher $|\vec{q}|$ value.

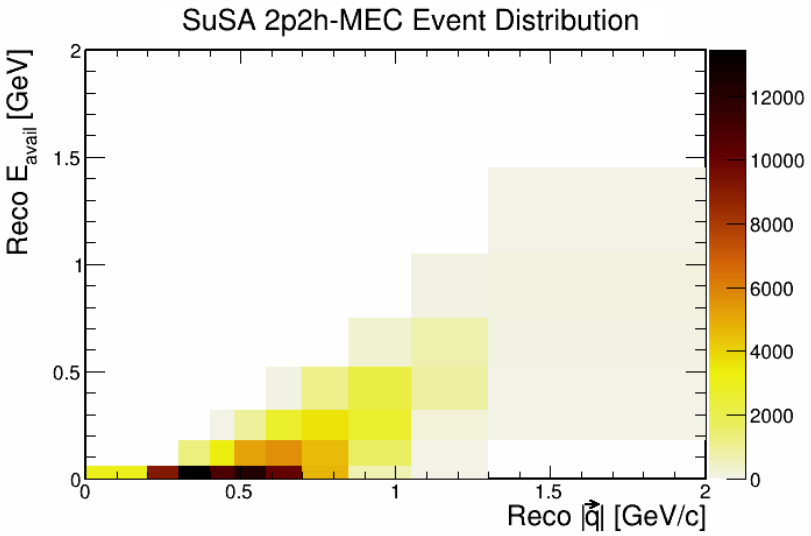

Figure 12.21: Event distribution of $\bar{\nu}_\mu$-CC SuSA 2p2h-MEC event distribution in reconstructed $|\vec{q}|$ and E_{avail}.

MINERvA tune 2p2h in reconstructed variables

The MINERvA 2p2h-MEC models have larger cross sections compared to the Valencia model as can be inferred from the right-side axes in Figs. 12.20 and 12.22. Compared to the Valencia model, the MINERvA tune predicts 60% more events and the shape is more skewed towards the lowest slice of E_{avail}. Each MINERvA tune has a different event distribution because the np→nn and pp→np are tuned differently in each case.

NOvA tune 2p2h in reconstructed variables

The event distributions of the NOvA-tuned Valencia model in $|\vec{q}|$ and E_{avail} is shown in Fig. 12.23. The tuning results in a significant enhancement of events in the lowest E_{avail} slice, in particular, the region $0< |\vec{q}| \leq 600$ MeV/c. It is observed that this is the region where the vast majority of MEC 2p2h-events lie. This enhancement becomes apparent when looking at the z-axis scales of Figs. 12.20 and 12.23.

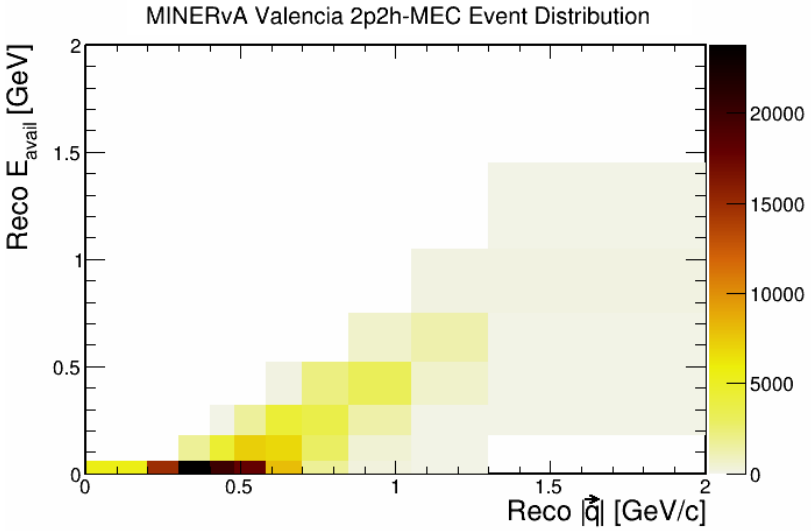

Figure 12.22: Event distribution of the MINERvA tune of the $\bar{\nu}_\mu$-CC Valencia MEC event distribution in reconstructed $|\vec{q}|$ and E_{avail}.

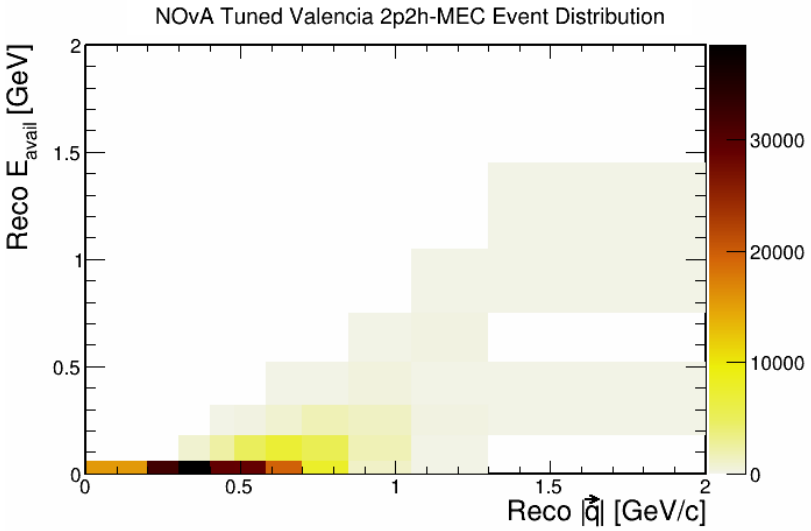

Figure 12.23: Event distribution of the NOvA tuned 2p2h-MEC model in reconstructed $|\vec{q}|$ and E_{avail}.

MINERvA np and pp tunes in reconstructed variables

The MINERvA np tune (Fig. 12.24, left) enhances the strength of the np $\rightarrow$ nn subset, which explains why there is an increase in the cross section in the slice with the lowest E_{avail} values. With the MINERvA pp tune (Fig. 12.24, right), only the strength of the subset of pp $\rightarrow$ np events is enhanced, hence there is an increase in the cross section at higher values of E_{avail} with

the MINERvA pp tune

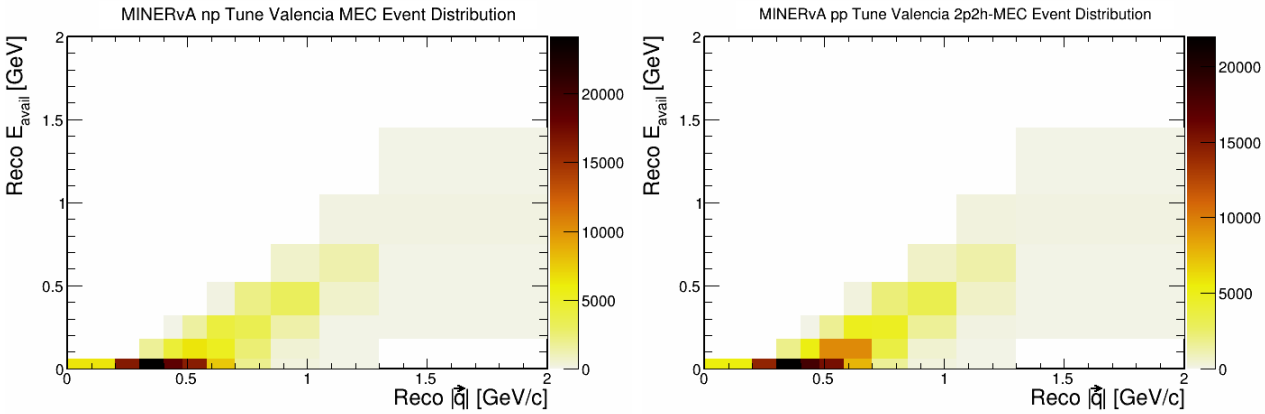

Figure 12.24: Event distribution of MINERvA 2p2h np (left) and nn (right) tunes in variables $|\vec{q}|$ and E_{avail}.

The event distributions of the Valencia model and the MINERvA tunes are similar in $|\vec{q}|$ and E_{avail}, however, there are some differences. For the Valencia model; a smaller portion of the overall population of events is concentrated in the lowest E_{avail} slice.

12.2.4 Projected distributions for reconstructed 2p2h models

Projected distributions for the reconstructed 2p2h models are shown in Fig. 12.25. In each variable, the models show similar trends. From 0 to 0.5 GeV in $|\vec{q}|$, the number of events in each bin increases as $|\vec{q}|$ increases and peaks at 0.5 GeV momentum transfer. Only the NOvA tuning of the Valencia model peaks at a lower momentum transfer. In the E_{avail} projection, most events are in the lowest bin with available energy less than 60 MeV. This subsample of events comes mostly from 2p2h-MEC events with a nn final-state dinucleon and some np events with very low dinucleon kinetic energy. Above 60 MeV, the event population is fed by np dinucleon events and from higher energy nn events which undergo FSI, dislodging charged hadrons from parent nuclei or from rescattering of the outgoing dinucleon.

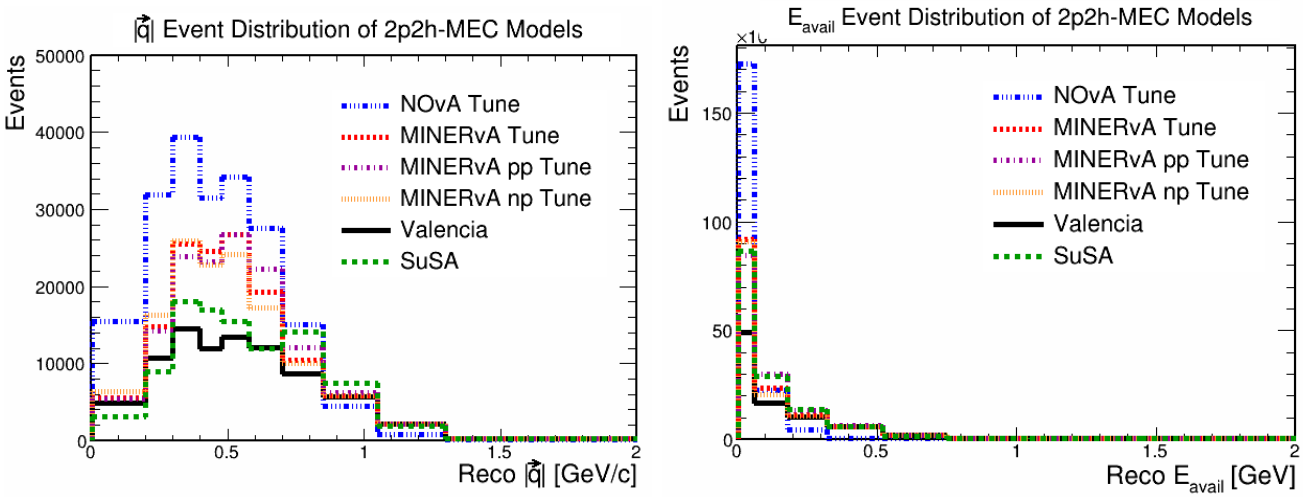

Figure 12.25: Projections of the event distributions in the reconstructed analysis variables $|\vec{q}|$ and E_{avail}.

12.2.5 Model dinucleon transitions in reconstructed variables

Dinucleon transitions: Valencia

A breakdown of the Valencia model in $|\vec{q}|$ and E_{avail} into np $\rightarrow$ nn and pp $\rightarrow$ np components is shown in Fig. 12.26, left and right plots respectively. The two distributions have similar shapes; there is a population of events with $|\vec{q}|$ of 200 to 600 MeV/c and $E_{avail} \leq 60$ MeV. For each component, there also is a second population of events with $|\vec{q}|$ of 400 to 900 MeV and E_{avail} of 60 to 500 MeV.

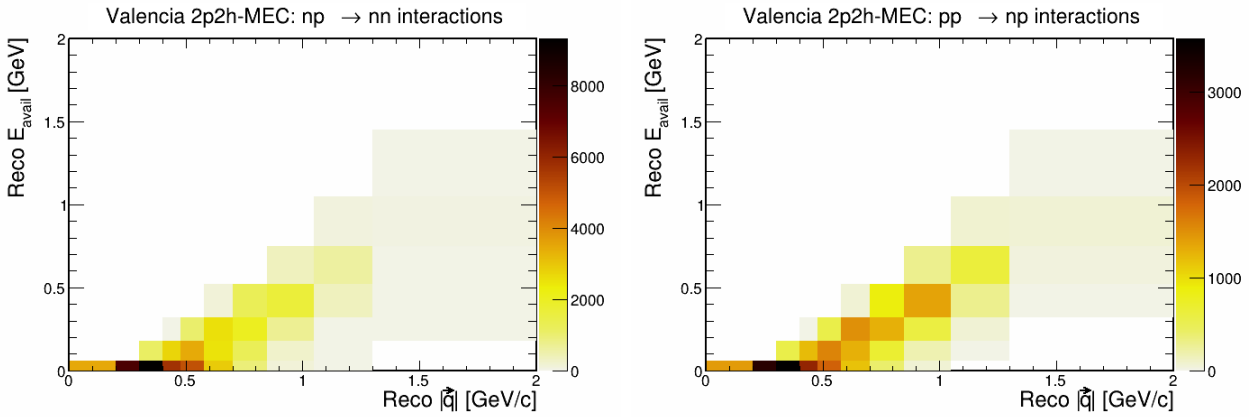

Figure 12.26: Event Distributions of the muon antineutrino CC Valencia 2p2h-MEC event distributions in reco $|\vec{q}|$ and E_{avail} separated into np $\rightarrow$ nn (left) and pp $\rightarrow$ np (right) components.

Dinucleon transitions: SuSA

The event distributions of the SuSA 2p2h-MEC model broken down by dinucleon type is shown in Fig. 12.27. It can be seen that the np $\rightarrow$ nn component has a higher event rate than that of the Valencia model, but a smaller rate for the pp $\rightarrow$ np component. The $|\vec{q}|$ spread of the events in the lowest E_{avail} slice also differs from its Valencia counterparts.

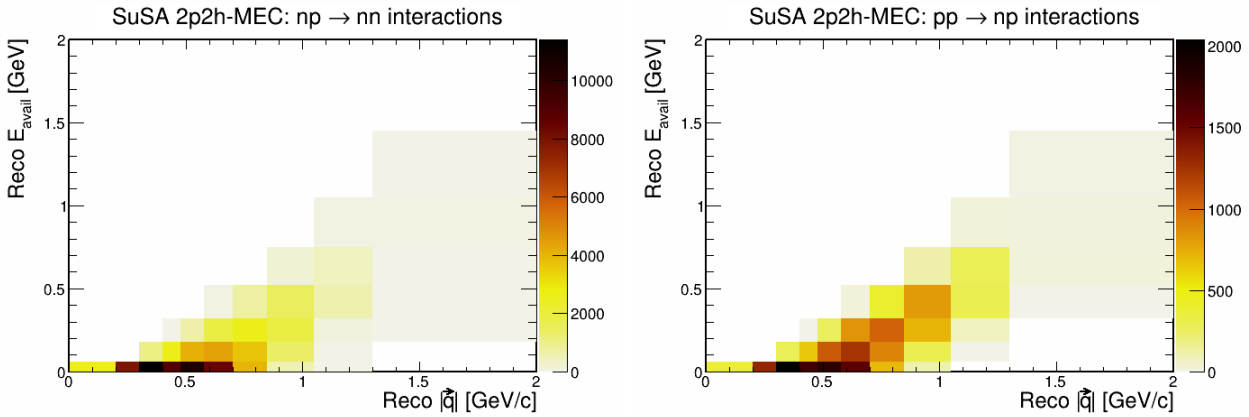

Figure 12.27: Event distributions of the $\overline{\nu}_\mu$ CC SuSA 2p2h model in reco $|\vec{q}|$ and E_{avail} broken down into np $\rightarrow$ nn (left) and pp $\rightarrow$ np (right) components.

Dinucleon transitions: MINERvA default tune

Event distributions of MINERvA default 2p2h-MEC tunes for each component are shown in Fig. 12.28. Like the Valencia model, the np $\rightarrow$ nn and pp $\rightarrow$ np components have similar distributions. For the MINERvA default tune 2p2h-MEC components, the event population is much more concentrated than is the case for the Valencia model. Most MINERvA-tune 2p2h-MEC events have a momentum transfer of 400 to 600 MeV/c.

Dinucleon transitions: NOvA Tune

The components of the NOvA tuned model are shown in Fig. 12.29. It can be seen that each component of this tune has a greater event rate than does the MINERvA tune. Also, the events are much more concentrated in the lowest E_{avail} slice compared to the MINERvA tune.

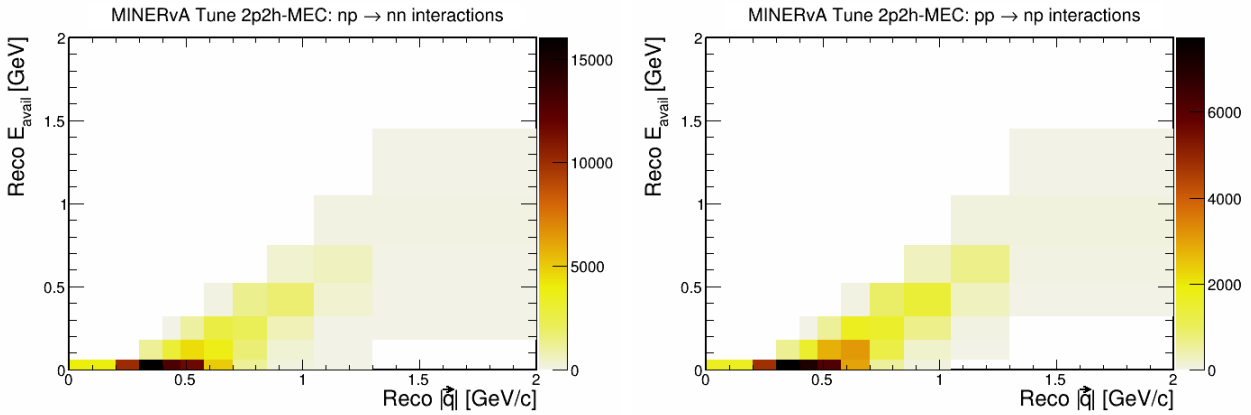

Figure 12.28: Event distributions of the $\overline{\nu}_\mu$ CC MINERvA tuned Valencia 2p2h-MEC model in reco $|\vec{q}|$ and E_{avail} broken down into np $\rightarrow$ nn (left) and pp $\rightarrow$ np (right) components.

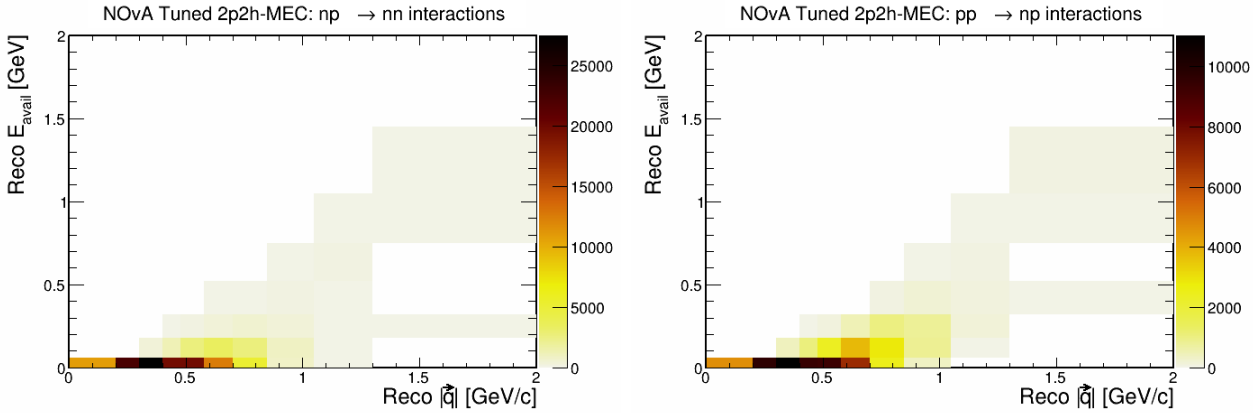

Figure 12.29: 2p2h-MEC event distributions of the $\overline{\nu}_\mu$ CC NOvA tune in reco $|\vec{q}|$ and E_{avail} broken down into np $\rightarrow$ nn (left) and pp $\rightarrow$ np (right) components.

Dinucleon transitions: MINERvA np and pp tunes

Each 2p2h-MEC component of the MINERvA pp and np tune event distributions are shown in Fig. 12.30. For the np tune, the pp $\rightarrow$ np component (Fig. 12.30(b)) of the Valencia model is not affected, hence the distribution of the np tune and Valencia model pp $\rightarrow$ np (Fig. 12.26, left) are identical. The pp $\rightarrow$ np component of Valencia is significantly enhanced. Furthermore, as a result of the transformation, it can be seen that this MINERvA tune favors 2p2h-MEC events having less E_{avail} in general.

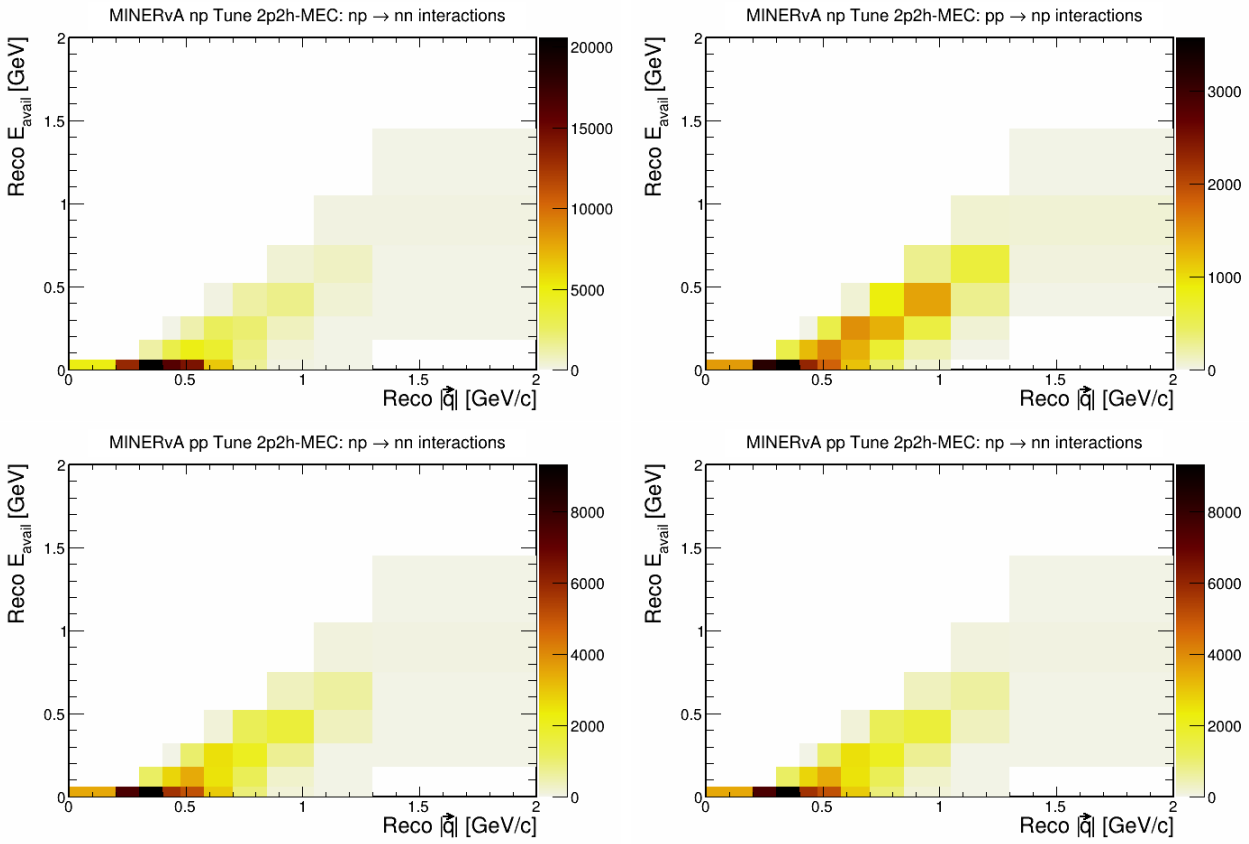

Figure 12.30: MEC event distributions of the $\overline{\nu}_\mu$ CC MINERvA default tune in reco $|\vec{q}|$ and E_{avail} broken down into np $\rightarrow$ nn (left) and pp $\rightarrow$ np (right) components.

For the pp tune, the Valencia pp $\rightarrow$ np component undergoes a similar transformation. The result of this transformation is in Fig. 12.30(d). The overall cross-section magnitude is increased and a larger proportion of events belong in the lowest E_{avail} slice. The np $\rightarrow$ nn component of the Valencia model is untouched by this tune, which is why this component has an identical distribution to that of the Valencia model.

12.2.6 Projected distributions for reconstructed dinucleon transitions

The projections of the np $\rightarrow$ nn and pp $\rightarrow$ np components in each analysis variable are shown in Figs. 12.31 and 12.32 respectively. On the E_{avail} plot (Fig. 12.31, right), most events are concentrated in the lowest bin, which is expected since the outgoing hadronic component of most events are neutrons only. There is subsample of events in which the dinucleon undergoes final state interactions in the medium, knocking out protons and other charged hadrons, which

in turn contribute to the available energy.

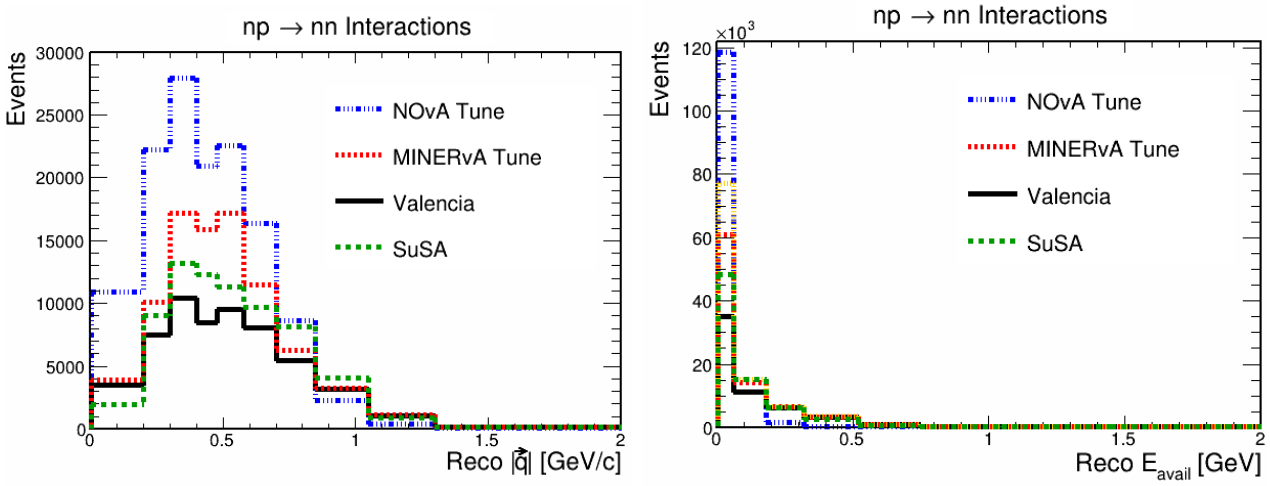

Figure 12.31: Projections of the np → nn component of the 2p2h-MEC models in variables reconstructed $|\vec{q}|$ and E_{avail}. The peak three-momentum transfer for each model is in the vicinity of 0.5 GeV/c.

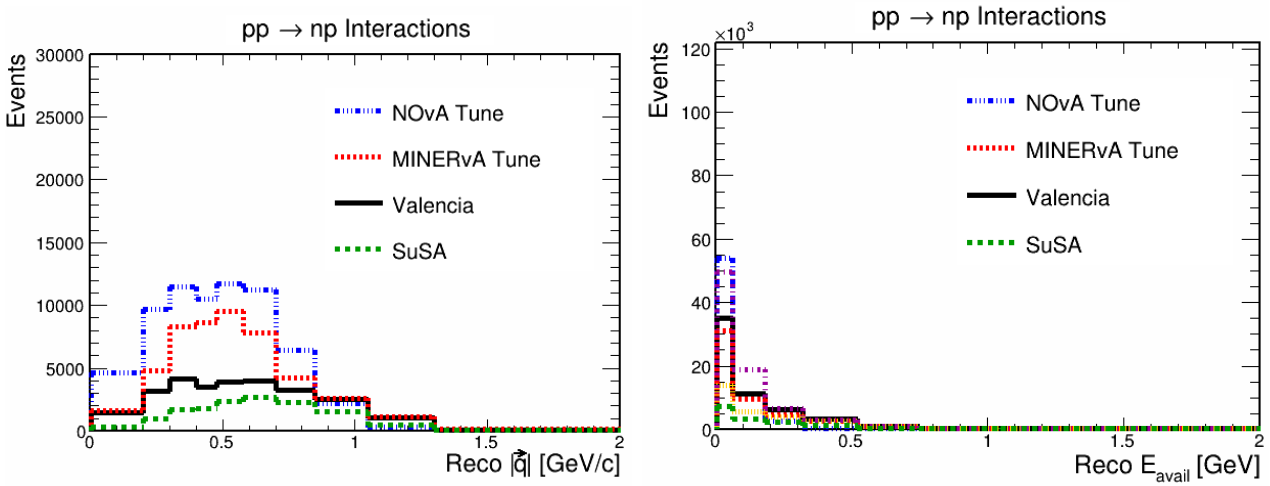

Figure 12.32: Projections of the pp → np component of the 2p2h-MEC models in variables reconstructed $|\vec{q}|$ and E_{avail}. The peak momentum transfer for each model is also at 0.5 GeV/c with the SuSA model peak estimated to be at a slightly higher value.

For the pp → np component (Fig. 12.32), the distributions have a trends have similar to those of the np → nn counterparts. From 0 - 0.5 GeV, the event count increases with increasing $|\vec{q}|$. After the peak, the event count starts to fall off. In E_{avail}, most events are also concentrated in very low E_{avail}, indicating that the outgoing proton in the outgoing dinucleon often does not carry much kinetic energy. As previously, there is a smaller subsample of events that have have higher

available energy. However, the proportion of events with $E_{avail} > 60$ MeV is higher compared to the np $\rightarrow$ nn subset of events.

Each 2p2h-MEC model has a different phase space distribution, with populations of events spread out over in various ways. Not only are the peaks of the event counts different, but the way the activity of 2p2h-MEC events changes while moving away from “hot spots" is also different. This is an important observation. In template fitting methods, a sideband is required to constrain the amount of background in the total data samples. Varying kinematic behaviors across 2p2h-MEC models must be kept in mind when selecting a suitable sideband.

12.2.7 Comparison of fractional uncertainty due to MEC systematics

Traditionally, uncertainty due to 2p2h-MEC processes have been assigned with the systematics developed by the cross-section tuning group [16]. For this analysis, the spread in the 2p2h models is used to determine the error. A comparison of the fractional errors due to 2p2h-MEC processes is shown in Fig. 12.33. The two different methods predict different uncertainties across the phase space. In the $|\vec{q}|$ variable, the nominal 2p2h systematics predict an uncertainty that minimizes between 400 to 600 MeV/c, whereas spread of the 2p2h models indicate that a higher uncertainty at low $|\vec{q}|$ that falls off gradually. In the E_{avail} variable, the 2p2h models indicate a large uncertainty for the lowest E_{avail} slice that falls off with increasing E_{avail}. Above the $|\vec{q}| > 1.2$ GeV/c cutoff, the spread in models cannot be used to assign an uncertainty due to the Valencia cutoff. For this region, the uncertainty from [16] is adopted, this can be seen from the non-zero errors at very high $|\vec{q}|$ and E_{avail}.

12.2.8 Fractional uncertainties of the cross section

The fractional uncertainties on the inclusive cross section from the different sources described in Section 12.2 are shown in Fig. 12.34. The leading source of uncertainty is the beam flux, which has a constant effect across the phase space. The next leading error sources are the 2p2h-MEC systematics followed by uncertainty parameters of the GENIE cross-section models. The

2p2h-MEC

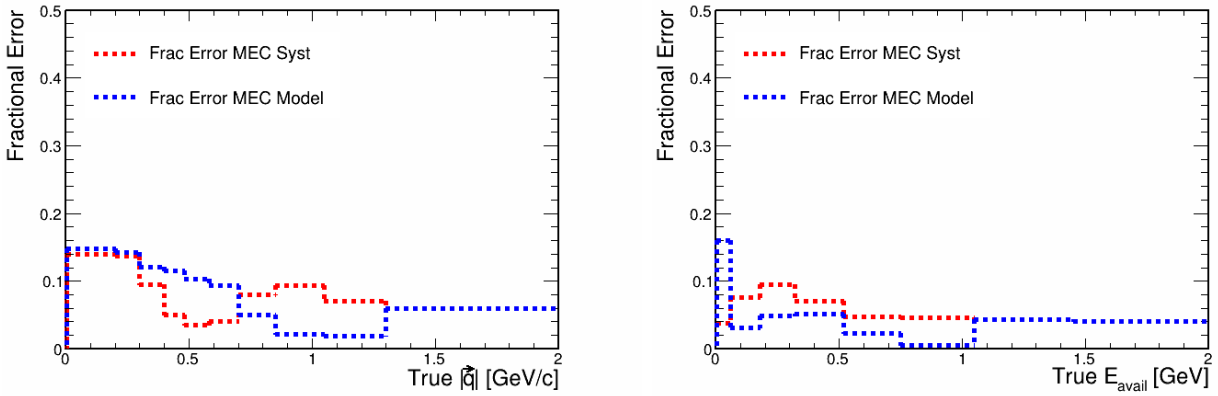

Figure 12.33: Fractional uncertainty in the $\bar{\nu}_\mu$-CC inclusive cross section due to 2p2h-MEC systematics. The red histograms show the fractional 2p2h uncertainty used for oscillation analyses. The blue histograms show the uncertainty derived by examining the spread of the 2p2h models.

systematics have been derived by examining the spread of the cross section with the different 2p2h models discussed in Sec. 12. The uncertainty arising from 2p2h modeling is nearly 15% for $|\vec{q}| \leq 300$ MeV/c; with increasing $|\vec{q}|$ it gradually decreases to $\leq 6\%$. The GENIE cross section errors have a larger effect at higher $|\vec{q}|$ and E_{avail}. The effect of the light, calibration and energy scale errors is negligible. The statistical error is also negligible, except in the highest E_{avail} bin, which has 27 events.

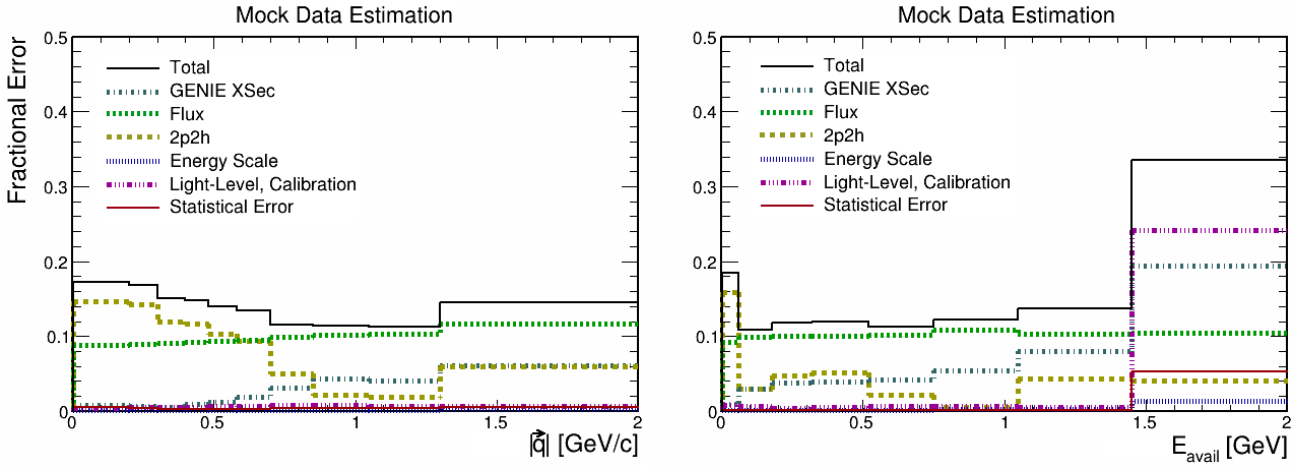

Figure 12.34: Fractional uncertainty in the $\bar{\nu}_\mu$-CC inclusive cross section from the sources listed in Sec. 12.2, estimated using MC mock data. The flux error is nearly constant across the phase space.

The fractional uncertainties due to the different sources across the analysis phase space is shown in Table 12.3. The numbers indicate how much the total, flux-averaged cross section is affected.

Systematic source	Fractional uncertainties on cross section
Cross Section	9.07%
2p2h	9.23%
Flux	10.0%
Energy Scale	0.06%
Light Level, Calibration	0.54%
Statistics	0.11%
Total	17.1%

Table 12.3: Total uncertainty on the $\bar{\nu}_\mu$-CC inclusive integrated cross section from various systematic sources. The total uncertainty is estimated to be 17.1% with the leading source of error being the flux at 10.0%.

Chapter 13

Validation of Unfolding and Background Constraint Code; Mock Data Measurement

With all of the ingredients for a cross section measurement now assembled, we proceed with a measurement of the inclusive cross section of signal events using the NOvA ND data. However, before the methods described in this document are applied to the experimental data, two validation checks must be completed.

There are two coded algorithms that are used in measuring the $\bar{\nu}_\mu$-CC inclusive cross section and the $\bar{\nu}_\mu$-CC 2p2h-MEC interaction cross section: i) RooUnfold, the code used to unfold the reconstructed signal estimate, discussed in Sec. 8, and ii) The data-driven method to constrain backgrounds discussed in Sec. 6.5.2. Mock data tests have been conducted in order to validate these two codes. The first test consists of using RooUnfold to unfold the signal estimate. The second, test uses the background constraint in the muon longitudinal momentum, followed by RooUnfold. For these validation tests, the MC sample has been divided into two. The first half is used as a pseudodata sample from which to extract a cross section, the second half is used as a reference sample to construct the background distribution, unfolding matrix, and efficiency correction. The results from these mock data tests are described below.

13.1 Test of data unfolding with RooUnfold

The first test compares a cross section obtained from the pseudeodata sample with the MC truth cross section. The signal is obtained by subtracting the nominal background estimated by GENIE from the sample of selected pseudo-data events. The unfolding is done via RooUnfold. The extracted cross section is shown in Fig. 13.1 (upper left). A true cross section has been generated with the reference MC sample and is shown in the upper-right of Fig. 13.1. The difference and the ratio between the extracted and the MC truth cross section is shown in the lower plots of the same figure. The latter difference and ratio plots indicate small differences between the two cross sections, with the differences being about three orders of magnitude less than the absolute scale of the measured cross section.

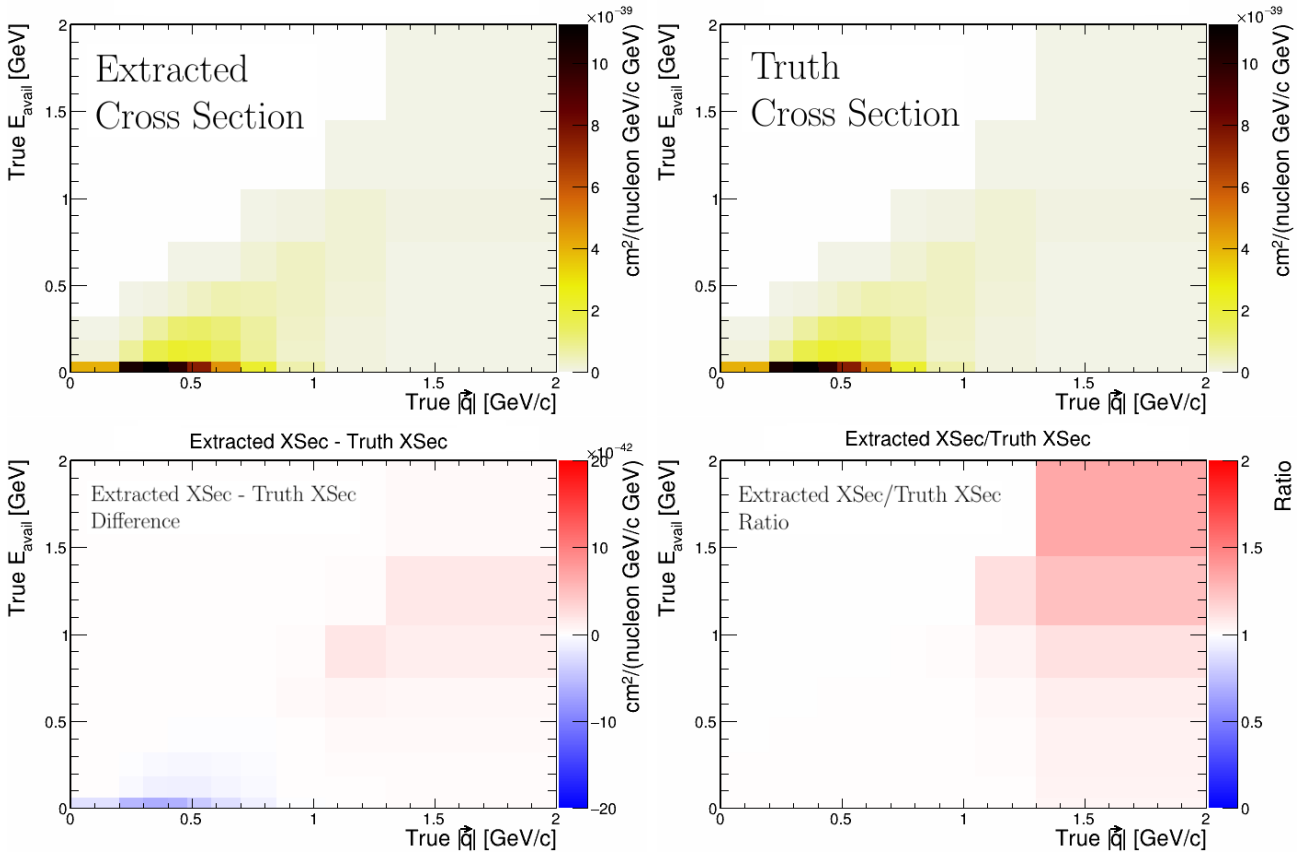

Figure 13.1: Comparison of the $\bar{\nu}_\mu$-CC cross section extracted from a sample of events passing the CC inclusive selection criteria, unfolded with RooUnfold and efficiency-corrected (left), and a cross section calculated from all true signal events simulated in the NOvA near detector (right). The lower plots compare the two cross sections in terms of difference (bottom-left) and ratio (bottom-right).

13.2 Validation of the background constraint code

We now implement the background constraints discussed in Sec. 8 and then proceed with unfolding via RooUnfold. The reference MC is used to construct templates for the NC/Other and wrong-sign backgrounds. The templates are then fitted to pseudodata distributions in longitudinal muon momentum yielding adjusted normalizations for the backgrounds. These adjusted backgrounds are then subtracted from the pseudodata events to obtain the signal estimate. The signal estimate is then unfolded and used to calculate a cross section. The results of the cross section extraction with the background constraint are shown in Fig. 13.2. The differences between the true cross section and the extracted cross section are characterized in terms of difference and ratio in the same figure. It can be seen that the differences are quite small.

Since the proposed cross-section measurement is made more robust by inclusion of constraints on the background, and since our background constraint procedure performs well in this mock data study, the background constraint procedure is included as part of our measurement procedure carried out with real data.

13.3 Mock data measurement of the $\bar{\nu}_\mu$-CC inclusive cross section

The previous test, which incorporated the background constraints into a sample that is unfolded, comprises a mock data measurement. The cross section measured from the pseudo-data is reproduced in Fig. 13.3. A bin-by-bin display of the actual cross section values from this mock data study is provided in Table 13.2. The cross section peaks at $|\vec{q}|$ of 350 MeV/c and E_{avail} of 30 MeV with a value of 1.04×10^{-38} cm^2/(nucleon GeV/c GeV).

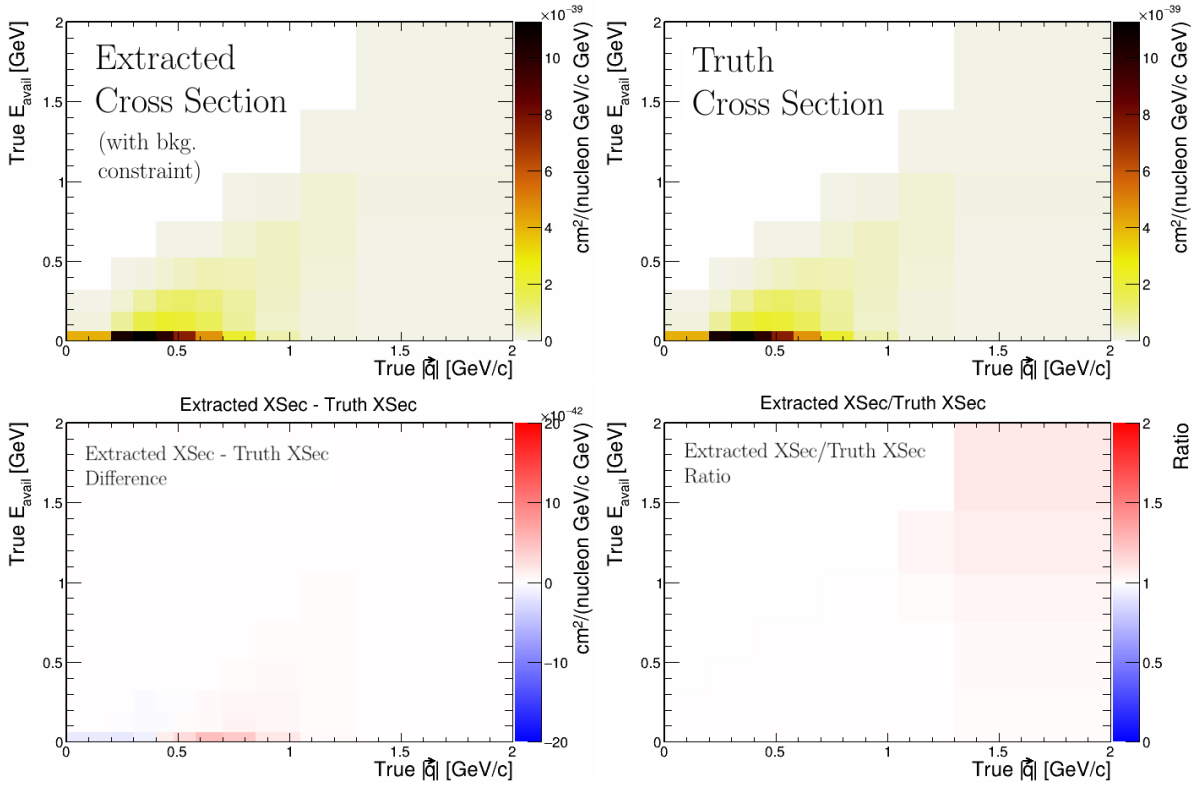

Figure 13.2: Comparison of the $\bar{\nu}_\mu$-CC cross section extracted from a sample of events passing the CC inclusive selection criteria with a background constraint. The signal has been extracted after the background is constrained and then unfolded with RooUnfold (left), and a true cross section calculated from all true signal events simulated in the NOvA near detector (right). The bottom plots compare the two upper plots in terms of difference (bottom-left) and ratio (bottom-right).

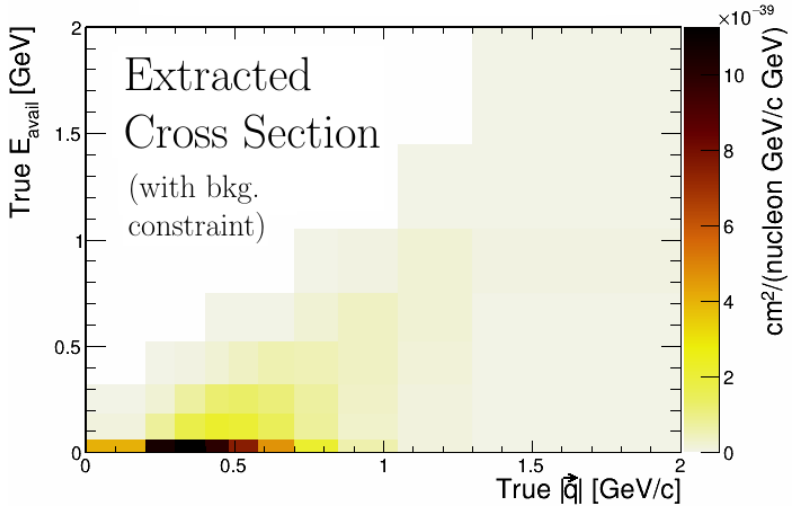

Figure 13.3: Distribution of the $\bar{\nu}_\mu$-CC inclusive cross section obtained with pseudo-data events. Most of the cross section lies within the lowest E_{avail} slice with 0 - 700 MeV/c of $|\vec{q}|$. The cross section falls off by several orders of magnitude in regions of higher $|\vec{q}|$ and E_{avail}.

$\|\vec{q}\|/E_{avail}$	0.00–0.06	0.06–0.18	0.18–0.32	0.32–0.52	0.52–0.75	0.75–1.05	1.05–1.45	1.45–2.00
0.00–0.20	7.40E-39	3.57E-40	3.27E-42	0	0	0	0	0
0.20–0.30	9.68E-39	1.39E-39	4.04E-40	0	0	0	0	0
0.30–0.40	1.04E-38	3.27E-39	1.57E-39	1.45E-40	0	0	0	0
0.40–0.48	7.37E-39	3.31E-39	2.17E-39	5.25E-40	0	0	0	0
0.48–0.58	7.02E-39	3.92E-39	2.85E-39	1.18E-39	1.91E-42	0	0	0
0.58–0.70	5.23E-39	3.30E-39	2.95E-39	2.08E-39	8.88E-41	0	0	0
0.70–0.85	3.02E-39	2.14E-39	2.38E-39	2.48E-39	1.12E-39	9.16E-42	0	0
0.85–1.05	1.05E-39	1.02E-39	1.42E-39	2.28E-39	2.66E-39	7.07E-40	0	0
1.05–1.30	2.11E-40	4.06E-40	6.91E-40	1.30E-39	1.91E-39	2.38E-39	8.98E-41	0
1.30–2.00	6.18E-41	1.47E-40	2.96E-40	5.69E-40	8.85E-40	1.48E-39	1.30E-39	2.60E-40

Table 13.1: Bin-by-bin display of the values of the double differential cross section (MC) shown in Fig. 13.2 (upper left). The units of the cross section are $cm^2/$(nucleon GeV/c GeV). Note that the vertical axis identifies $|\vec{q}|$ bins and the horizontal axis identifies E_{avail} bins.

The fractional error per bin on the extracted cross section of Fig. 13.3, is presented in Fig. 13.4 and the corresponding tabular display of values is given in Table 17. The fractional error is highest on the boundary region where the numerical values of $|\vec{q}|$ are roughly equal to the numerical values of E_{avail}. This is also the region where the event statistics are quite poor. The error in the bins that are populated with events have an error of 20%. This is consistent with the total errors shown in Fig. 12.34.

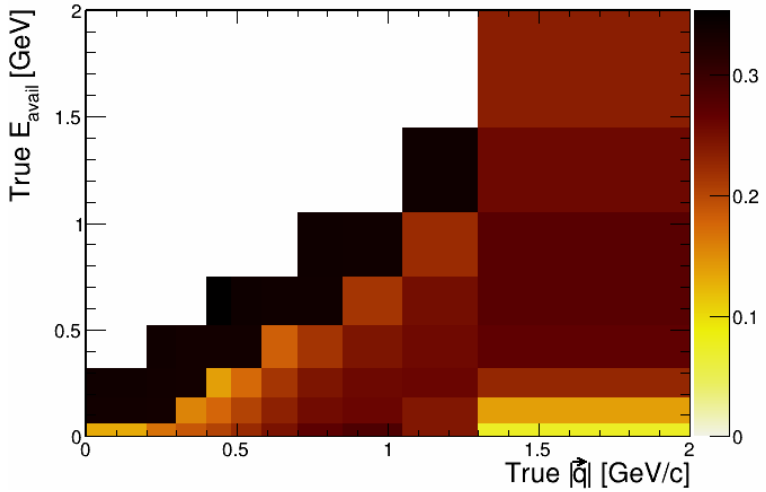

Figure 13.4: Bin-by-bin fractional errors for the inclusive cross section shown in Fig. 13.3.

$\|\vec{q}\|/E_{avail}$	0.00–0.06	0.06–0.18	0.18–0.32	0.32–0.52	0.52–0.75	0.75–1.05	1.05–1.45	1.45–2.00
0.00–0.20	0.13	0.34	0.34	0.00	0.00	0.00	0.00	0.00
0.20–0.30	0.17	0.34	0.34	0.34	0.00	0.00	0.00	0.00
0.30–0.40	0.19	0.16	0.34	0.34	0.00	0.00	0.00	0.00
0.40–0.48	0.20	0.18	0.14	0.34	0.35	0.00	0.00	0.00
0.48–0.58	0.22	0.20	0.17	0.34	0.34	0.00	0.00	0.00
0.58-0.70	0.25	0.23	0.21	0.18	0.34	0.00	0.00	0.00
0.70–0.85	0.27	0.25	0.24	0.22	0.34	0.34	0.00	0.00
0.85–1.05	0.28	0.26	0.26	0.25	0.21	0.34	0.00	0.00
1.05–1.30	0.24	0.24	0.26	0.26	0.25	0.22	0.34	0.00
1.30–2.00	0.07	0.14	0.23	0.27	0.27	0.27	0.26	0.24

Table 13.2: Bin-by-bin display of fractional errors in the double differential cross section calculated with MC events. The errors are typically 15 - 25% in populated bins and 34% along the kinematic boundary.

The single differential cross sections $d\sigma/d|\vec{q}|$ and $d\sigma/dE_{avail}$ are shown in Figs. 13.5 and 13.6. In both figures, the pseudo-data cross section coincides with the NOvA-tuned predicted cross section. This is expected, as both cross sections are calculated with statistically independent samples that are generated the same way. The uncertainty on the pseudo-data measurement covers the full uncertainty due to systematic and statistical errors. The cross section obtained from the pseudo-data, and the prediction of the NOvA tune simulation as well, exceed the cross sections predicted by the MINERvA tune, and by the theory-based SuSA-v2 and Valencia models.

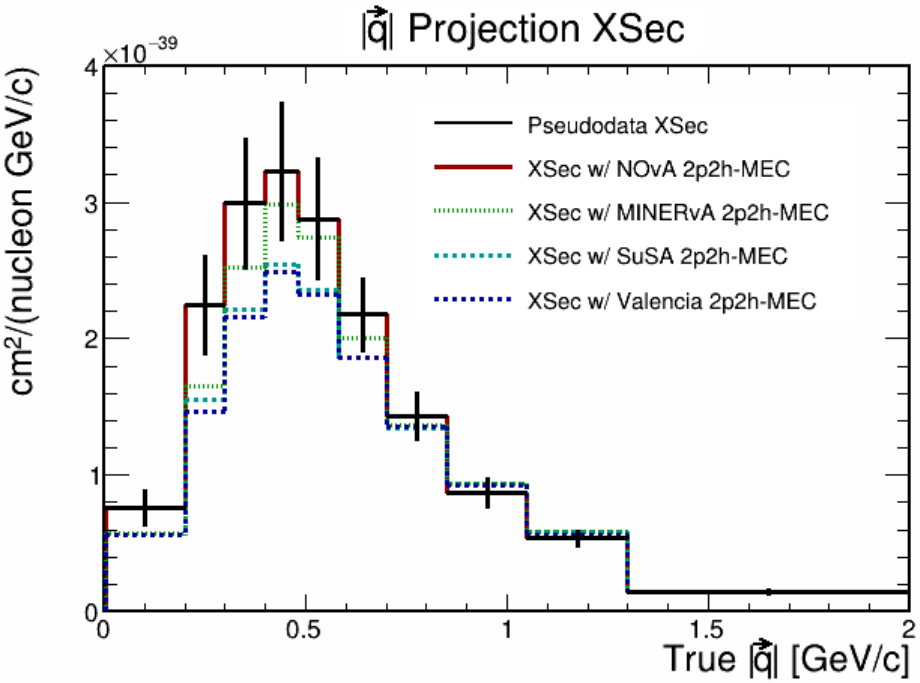

Figure 13.5: The single differential cross section $d\sigma/d|\vec{q}|$. The error bars on the measured cross section depict the total systematic plus statistical uncertainty.

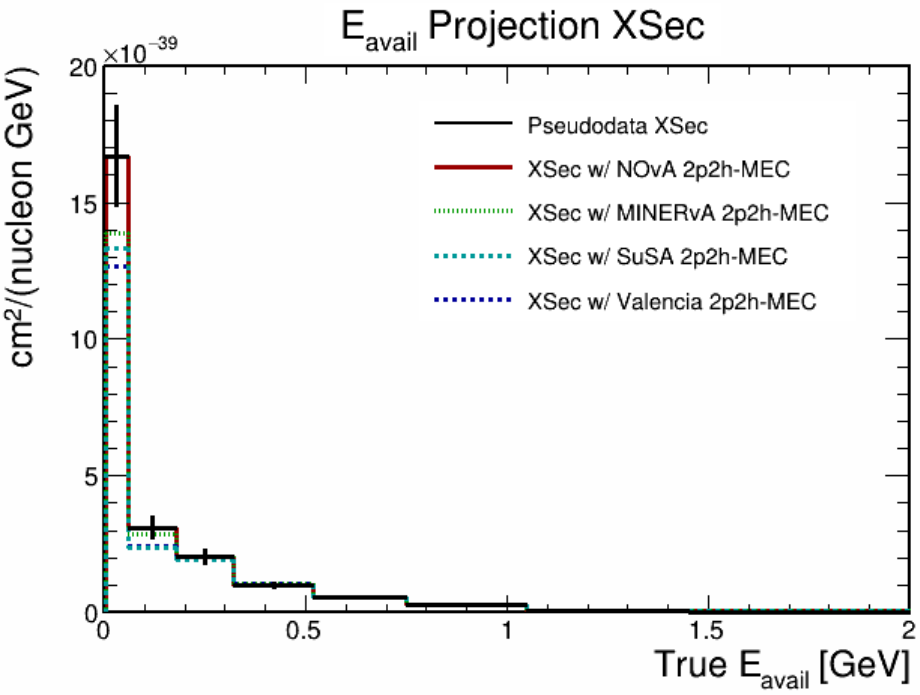

Figure 13.6: The single differential cross section $d\sigma/dE_{avail}$. In the first two bins, the measured cross section lies between the cross-sections predictions based on the NOvA and MINERvA tunes, with the theory models both underpredicitng the data. In the bins $E_{avail} > 180$ MeV the predictions and the data nearly coincide.

Chapter 14

Measurement of the CC Inclusive Cross Section using Near Detector Data

We now apply the methods developed in the previous sections to measure the cross section of inclusive $\bar{\nu}_\mu$-CC inclusive events using the NOvA ND data. The entire ND dataset is passed through the selection criteria defined in Sec. 4.2. The exposure of the NOvA ND RHC data is 12.5×10^{20} POT. The number of selected signal events in the data is 837,038 events. The resulting event distribution is shown in Fig. 14.1.

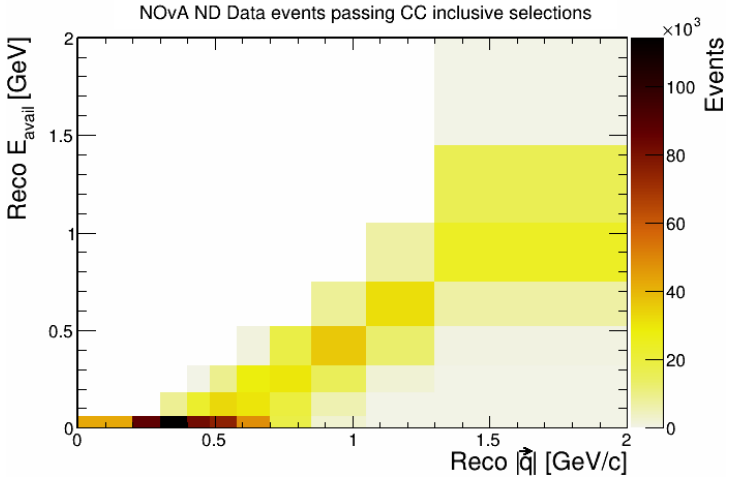

Figure 14.1: Distribution of ND Data events that pass the CC-inclusive selection criteria.

A sample of MC events, passed through the same criteria, are used to determine the background. The POT of the MC sample is 38.2×10^{20}, roughly three times that of the actual data exposure.

The background estimate is constrained to the data using the method described in Sec. 6.5.2. The results from the four stages of the fit (two stages per iteration) are shown in Figs. 14.2 - 14.5. Here, the data distributions (solid line histograms) are being fitted with template distributions constructed from the reference MC. It can be seen that MINUIT performs the fit correctly. In each case, the total prediction better matches the data distributions in $P_{||}$ after the fit. The fits determine what the overall event rate of different background interactions in the data sample should be. A summary of the normalization adjustments relative to the reference predictions by the fit are shown in Table 14.1.

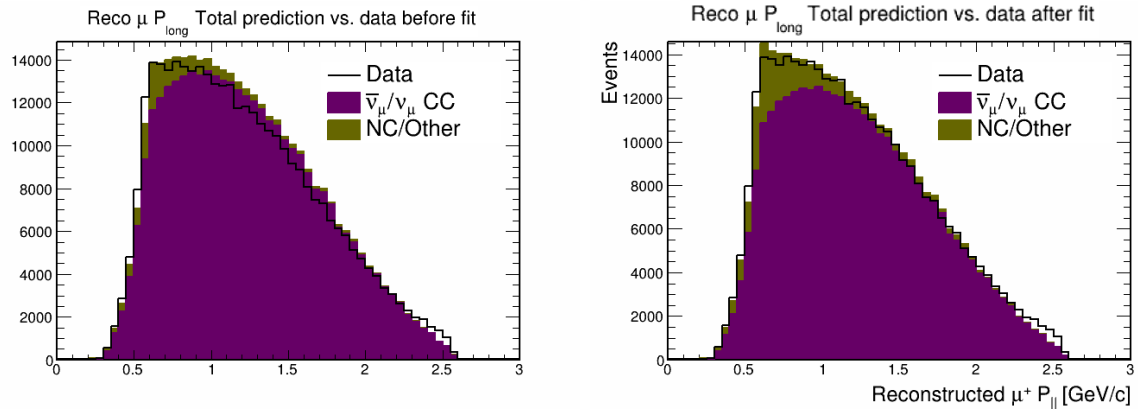

Figure 14.2: First stage of the first iteration of the fitting procedure for background estimate for the inclusive $\bar{\nu}_\mu$-CC cross section measurement.

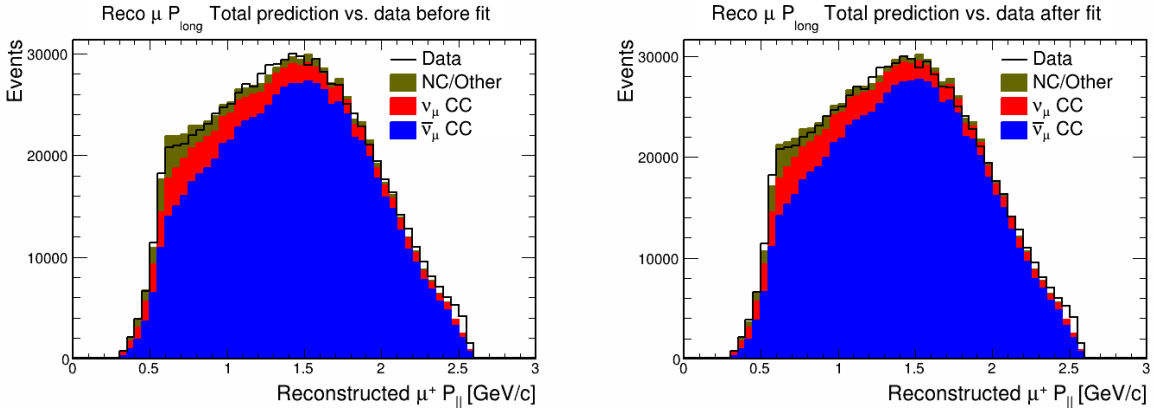

Figure 14.3: Second stage of the first iteration of the fitting procedure for background estimate for the inclusive $\bar{\nu}_\mu$-CC cross section measurement. The wrong-sign template (red) is held constant during the fit.

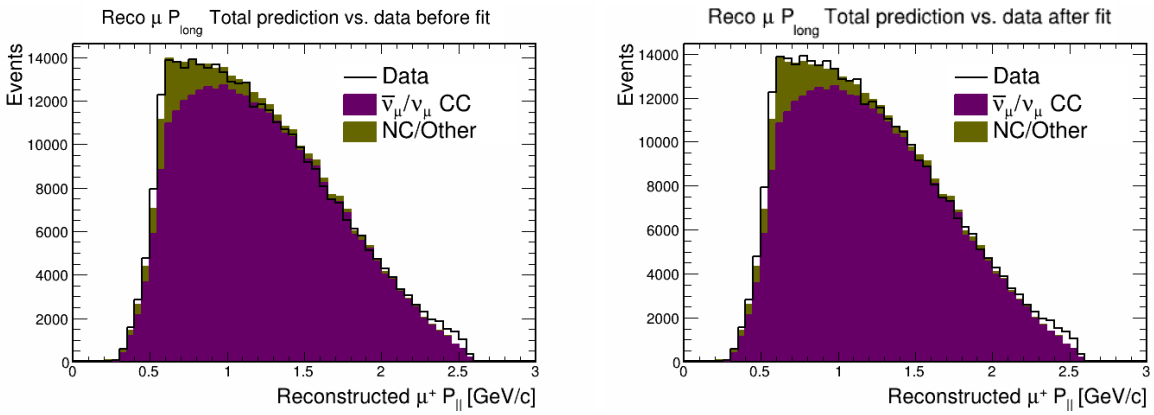

Figure 14.4: First stage of the second iteration of the fitting procedure. As seen from the plots before (left) and after (right) the fit, it can be seen that the fit slightly readjusts the normalizations of the templates. The total prediction at the peak is slightly below that of the data after the fit.

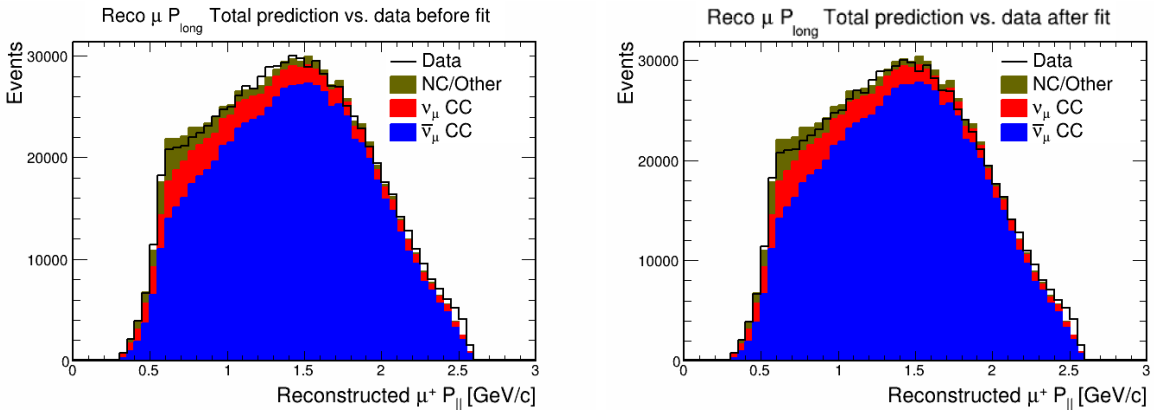

Figure 14.5: Second stage of the second iteration of the fitting procedure. This result is very similar to that of the second stage of the first iteration, indicating that a stable solution is obtained.

Template	Normalization applied
ν_μ-CC wrong sign	0.92
NC/Other	1.41
$\bar{\nu}_\mu$-CC	0.95

Table 14.1: Normalization adjustments assigned to the different interaction types resulting from the fits carried out in the four stages shown in Figs. 14.2 - 14.5.

The constrained background is subtracted from the selected data sample to obtain the data event distribution. The distribution of the background is shown in Fig. 14.6. The nominal background predicted by GENIE (no constraints) is shown in the same figure (right). After normalization,

the shape of the background constraint changes. The peak of the event distribution in the lowest E_{avail} slice is reduced as a result of the reduced CC event rate normalizations. Towards higher $|\vec{q}$ and E_{avail}, the background event count is higher as a result of the increased estimated NC/Other interaction rate.

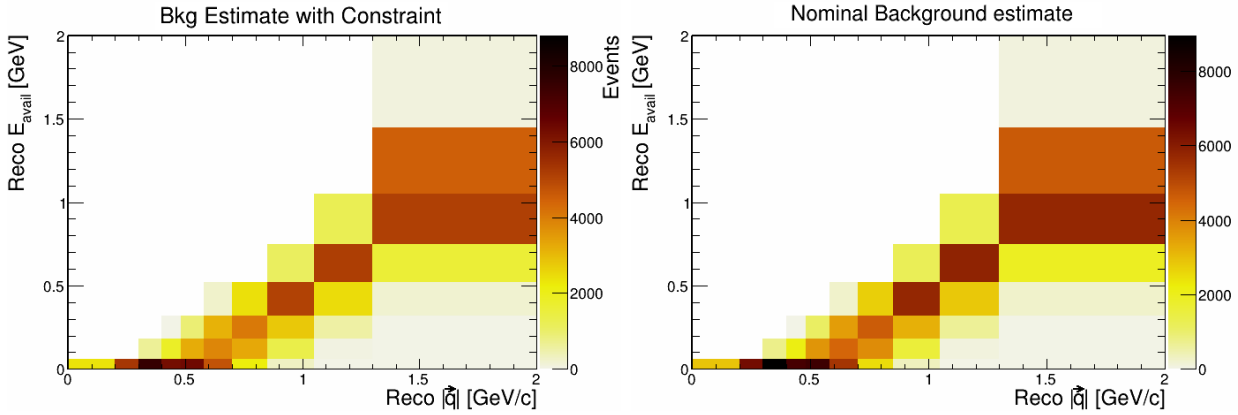

Figure 14.6: Background estimate obtained by fitting templates to data in the distribution of the muon longitudinal momentum (left). The nominal background predicted by NOvA tuned GENIE is shown on the right for comparison.

The signal estimate obtained by subtracting the constrained background from the data is shown in Fig. 14.7. The signal obtained from the data exhibits many of the features observed with $\bar{\nu}_\mu$-CC inclusive event distributions made with the MC. The majority of events populate the lowest bin of E_{avail}. However, there are events with higher E_{avail}, arising from RES, DIS, wrong-sign and other charged current muon-antineutrino interactions.

The signal estimate is unfolded with RooUnfold to yield a signal estimate in $|\vec{q}|$ and E_{avail}. The unfolding matrix is made from the MC and is shown in Fig. 9.1. Two unfolding iterations yields our best estimate of the signal event distribution, as was concluded in Sec. 8.3; the result is shown in Fig. 14.7.

With an unfolded signal estimate in hand, the cross section can be obtained. The distribution of Fig. 134 (right) is then divided by the flux (Fig. 10.5), the efficiency distribution (Fig. 8.1) and the target count (the total number of nucleons in the fiducial volume of the Near Detector) to yield the cross section shown in Fig. 14.8.

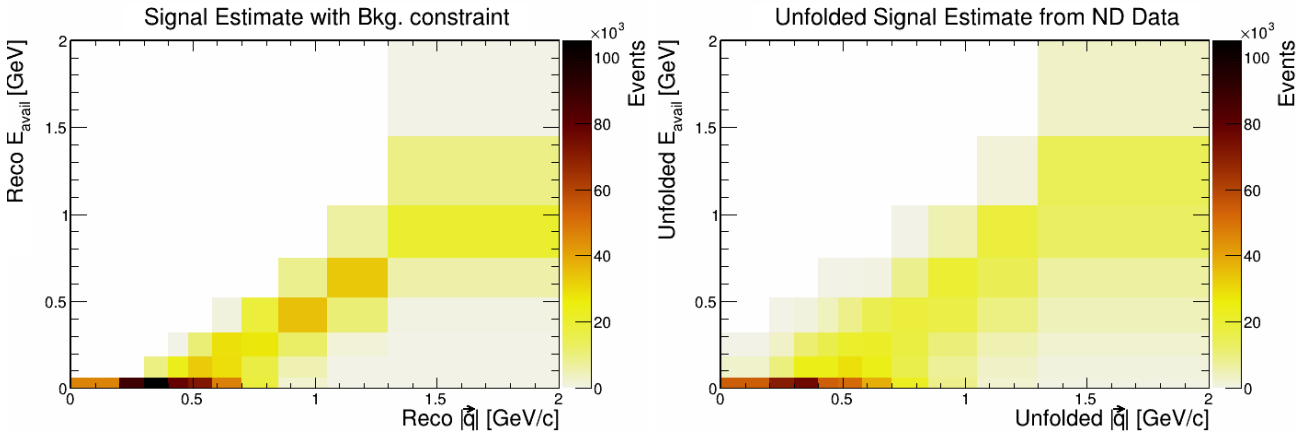

Figure 14.7: Signal estimate obtained after subtraction of background with constraint. The right plot shows the signal estimate after unfolding.

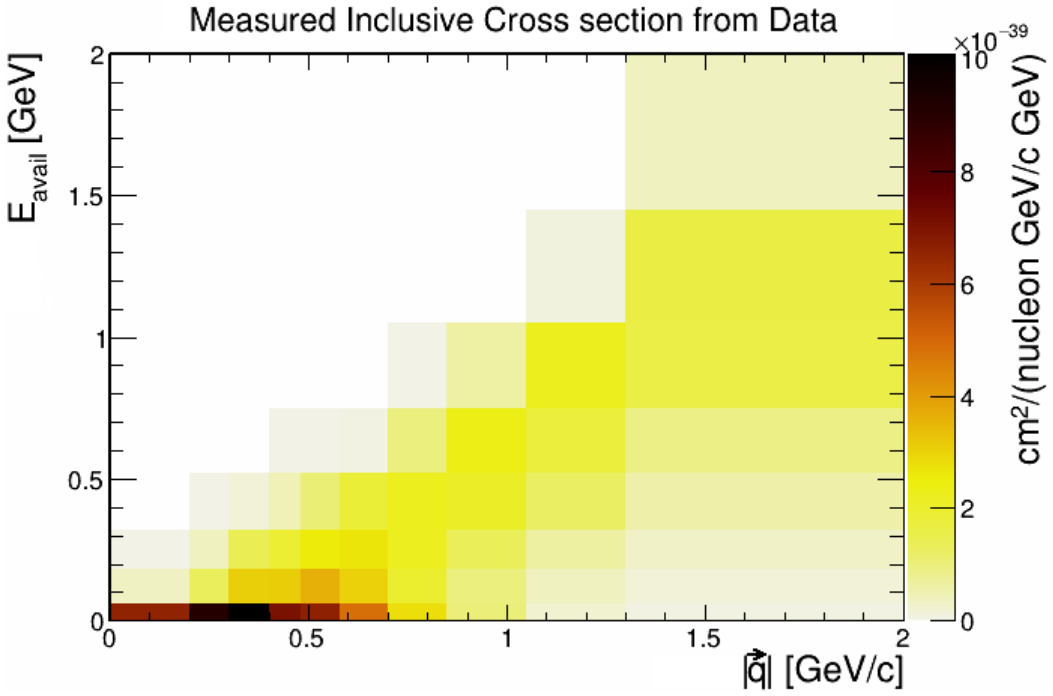

Figure 14.8: Measurement of the double-differential cross section in variables $|\vec{q}|$ and E_{avail} obtained from selected data events in the ND.

The cross section is highest in the lowest E_{avail} slice with $|\vec{q}|$ between 0 and 700 MeV/c, peaking at 350 MeV/c. This region is expected to be rich in $\bar{\nu}_\mu$-CC QE and 2p2h-MEC interactions. With higher E_{avail}, the cross section decreases. Muon antineutrino events that produce a charged particle, such as a pion (for example from RES and DIS interactions), or a proton (through 2p2h-MEC interactions or neutron rescattering) contribute to the cross section at higher E_{avail}

values. The data shows that in the $E_{\bar{\nu}}$ range of the NOvA experiment (roughly 2.0 to 4.0 GeV), high E_{avail} events have relatively small cross sections.

The bin-by-bin values of the measured cross section shown in Fig. 14.8 are given in Table 14.2. The peak of the measured cross section occurs are 0.35 GeV/c $|\vec{q}|$ and 30 MeV E_{avail} with a value of 1.01×10^{-38} cm^2/(nucleon GeV/c GeV).

$\|\vec{q}\|/E_{avail}$	0.00–0.06	0.06–0.18	0.18–0.32	0.32–0.52	0.52–0.75	0.75–1.05	1.05–1.45	1.45–2.00
0.00 – 0.20	6.58E-39	3.25E-40	2.95E-42	0	0	0	0	0
0.20 – 0.30	9.09E-39	1.30E-39	3.68E-40	5.88E-44	0	0	0	0
0.30 – 0.40	1.01E-38	3.08E-39	1.41E-39	1.28E-40	0	0	0	0
0.40 – 0.48	7.08E-39	3.10E-39	1.93E-39	4.65E-40	2.77E-43	0	0	0
0.48 – 0.58	6.65E-39	3.64E-39	2.56E-39	1.04E-39	1.73E-42	0	0	0
0.58 – 0.70	4.86E-39	3.04E-39	2.66E-39	1.83E-39	7.71E-41	0	0	0
0.70 – 0.85	2.79E-39	1.97E-39	2.18E-39	2.24E-39	9.73E-40	7.82E-42	0	0
0.85 – 1.05	1.02E-39	9.56E-40	1.32E-39	2.09E-39	2.35E-39	6.20E-40	0	0
1.05 – 1.30	2.11E-40	3.88E-40	6.52E-40	1.20E-39	1.75E-39	2.23E-39	9.53E-41	0
1.30 – 2.00	6.44E-41	1.49E-40	3.02E-40	5.76E-40	8.89E-40	1.59E-39	1.62E-39	3.56E-40

Table 14.2: Bin-by-bin display of the values of the measured double differential cross section shown in Fig. 14.8. The units of the cross section are cm^2/(nucleon GeV/c GeV). The vertical axis in the table labels the $|\vec{q}|$ bins and the horizontal axis labels the E_{avail} bins.

The single differential cross sections $d\sigma/d|\vec{q}|$ and $d\sigma/dE_{avail}$ are displayed in Figs. 14.9 and 14.10. The cross sections predicted by the MC with different 2p2h-MEC models are superimposed on the same figures. It can be seen from these projections that the NOvA tuned MC overpredicts the measured cross section.

The measured cross section in general lies close to the predictions of the $\bar{\nu}_\mu$-CC inclusive cross sections with NOvA tuned 2p2h and MINERvA tuned 2p2h models. The inclusive prediction with NOvA tuned 2p2h-MEC consistently overpredicts the data. This is the model that is closest to the measurement distribution. From 0 GeV/c to the peak in $|\vec{q}|$, the predicted cross sections with MINERvA tune, SuSA model and Valencia model fall below the measured cross section. Beyond the peak region, the $\bar{\nu}_\mu$-CC cross section with MINERvA tuned 2p2h is higher than the measurement. From 700 MeV/c onwards, all models have a prediction that exceeds the measurement. In the $d\sigma/dE_{avail}$ single differential cross section, the measured cross section lies between the predicted cross sections with NOvA tuned and MINERvA tuned 2p2h models. The

predicted cross sections from the SuSA and Valencia theoretical models fall short of the measured cross section.

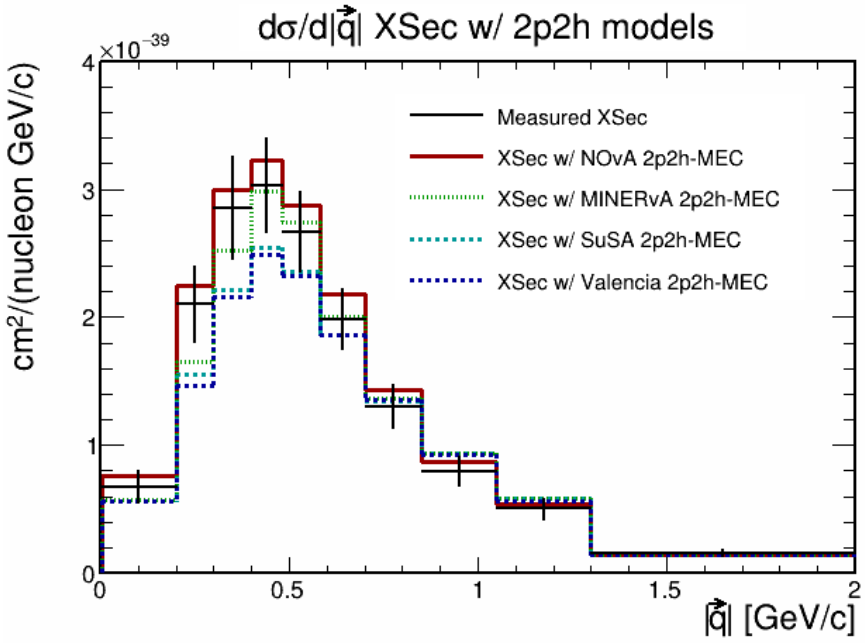

Figure 14.9: The single differential cross section $d\sigma/d|\vec{q}|$. The error bars on the measured cross section depict the total systematic plus statistical uncertainty.

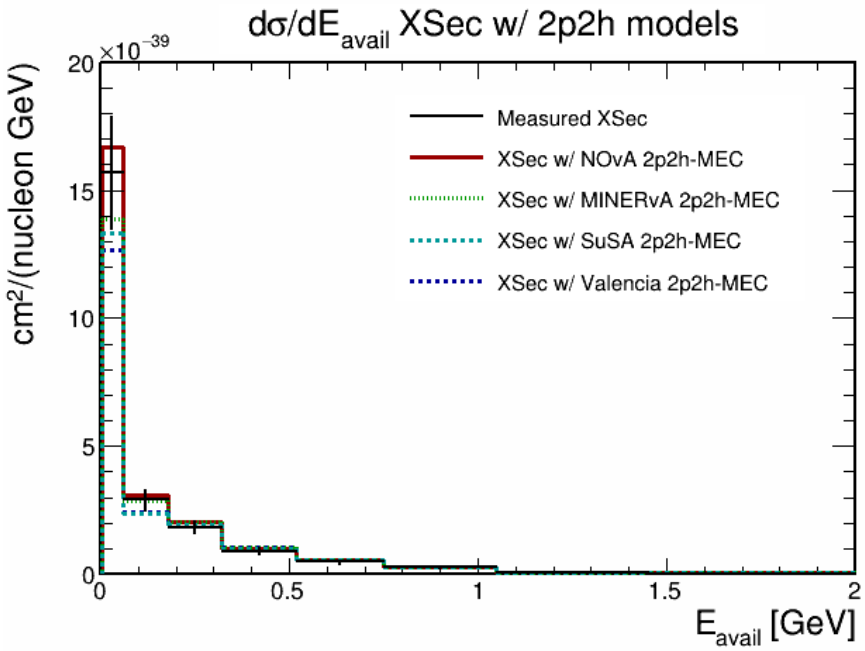

Figure 14.10: The single differential cross section $d\sigma/dE_{avail}$. Most of the cross section is to be found in the lowest E_{avail} bin because the many events have hadronic systemas that are solely comprised of neutrons.

The χ^2 (chi-squared) and χ^2/DoF (chi-squared per degrees of freedom) can be used to quantify the differences between the measured cross section and the predictions obtained with alternative treatments of 2p2h. The χ^2 is calculated as follows:

$$\chi^2 = \sum_{i,j} (Data - Prediction)_i \, V_{ij}^{-1} (Data - Prediction)_j. \tag{14.1}$$

Here, $(Data - Prediction)_i$ is the difference between the measured and model cross sections in the i-th bin and V is the covariance matrix calculated using the measured cross section and cross sections calculated in systematically shifted universes. The covariance matrix is defined as follows [43]:

$$V_{ij} = \frac{\sum_{n=1}^{N} (s_{n,i} - \mu_i)(s_{n,j} - \mu_j)}{N - 1}. \tag{14.2}$$

The indices i and j denote bins of cross-section values, where each bin is defined by an interval in $|\vec{q}|$ and an interval in E_{avail}. The value of the i-th bin of a systematically shifted cross section is $s_{n,i}$ and the value of the measured cross section in the same bin is μ_i. The index n is summed over the universes $n = 1, N$. The degrees of freedom (DoF) are the number of bins over which the χ^2 values are calculated. Only bins that have a significant cross section value are counted. The DoF used for these calculations are the bins that have cross section equal to or greater than the value $1{\times}10^{-41}$cm^2/(nucleon GeV/c GeV). Table 20 shows of the χ^2 and χ^2/DoF values. For the χ^2 calculations, the off-diagonal elements of the covariance matrix are ignored. This means that the bin-to-bin correlations in the cross section are ignored.

The cross section with the NOvA tune has the lowest χ^2, indicating quantitatively that this prediction best matches the data. This is followed by the cross section with the MINERvA tuned 2p2h-MEC. The predictions with Valencia and SuSA theories fall further below the measured cross section, which explains the comparatively higher χ^2 values.

Model	χ^2 (chi-squared)	DoF	χ^2/DoF
NOvA Tune	31.9	41	0.78
MINERvA Tune	32.6	41	0.80
Valencia Model	36.6	41	0.89
SuSA Model	36.5	41	0.89

Table 14.3: Table of χ^2 and χ^2/DoF values of the measured cross section with respect to MC inclusive cross section with different 2p2h models. The cross sections obtained with the NOvA and MINERvA tunes have the lowest and second lowest values respectively, followed by the cross sections obtained with the SuSA and Valencia models.

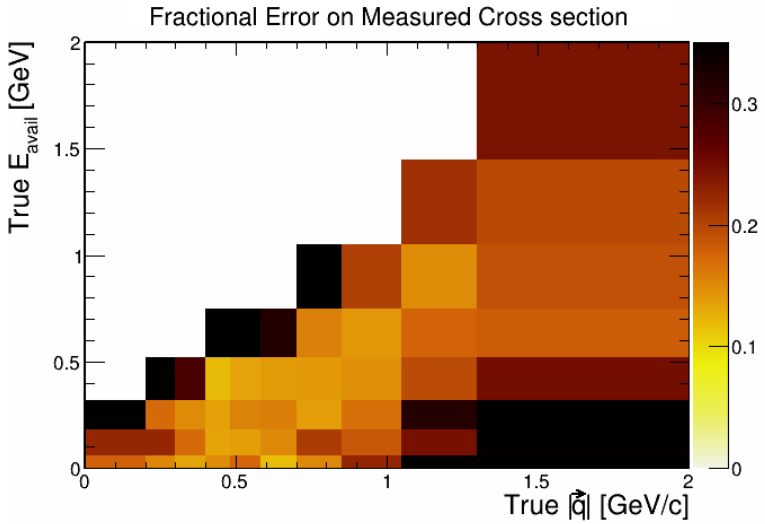

Figure 14.11: Fractional error on the inclusive cross section measured from the sample of data events passing the CC inclusive selections.

The bin-by-bin fractional uncertainties on the measured cross section are plotted in Fig. 14.11, and the corresponding fractional error values are given in Table 20. The fractional error on the measured cross section is highest on the boundary region and at high $|\vec{q}|$ with low E_{avail}. The fractional error values are comparable to the those obtained with the pseudo-data cross section. The uncertainty is calculated with the multi-universe method, which was discussed in Sec. 11.1. For each universe, the background is estimated with fits in the muon longitudinal momentum before the signal is estimated and the cross section is calculated. These fits can affect the systematic uncertainties of the measured cross section. The bin-by-bin fractional error values are given in Table 14.4.

$\|\vec{q}\|/E_{avail}$	0.00–0.06	0.06–0.18	0.18–0.32	0.32–0.52	0.52–0.75	0.75–1.05	1.05–1.45	1.45–2.00
0.00–0.20	0.18	0.22	2.20	0.00	0.00	0.00	0.00	0.00
0.20–0.30	0.15	0.22	0.17	2.00	0.00	0.00	0.00	0.00
0.30–0.40	0.13	0.17	0.15	0.28	0.00	0.00	0.00	0.00
0.40–0.48	0.15	0.13	0.14	0.12	3.56	0.00	0.00	0.00
0.48–0.58	0.18	0.14	0.16	0.13	1.56	0.00	0.00	0.00
0.58–0.70	0.12	0.15	0.16	0.14	0.32	0.00	0.00	0.00
0.70–0.85	0.15	0.21	0.14	0.14	0.16	0.67	0.00	0.00
0.85–1.05	0.23	0.18	0.17	0.15	0.14	0.20	0.00	0.00
1.05–1.30	0.36	0.25	0.31	0.19	0.18	0.15	0.22	0.00
1.30–2.00	0.57	0.35	0.37	0.25	0.18	0.19	0.20	0.24

Table 14.4: Bin-by-bin display of the values of the fractional error in the double differential cross section calculated with ND data events. The vertical axis in the table refers to $|\vec{q}|$ bins and the horizontal axis refers to E_{avail} bins.

A summary of the sources of uncertainty and their contribution to the fractional systematic error of the measured cross section is given in Figs. 14.12 and 14.13. The flux uncertainty is relatively flat around 10% across the analysis phase space space. The effects of the light level and calibration systematics on the uncertainty of the cross section are negligible, which is consistent with what was observed in the MC.

The 2p2h-MEC systematic uncertainty is largest at low $|\vec{q}|$. The cross section uncertainties, which cover the uncertainties due to non 2p2h interactions such as NC/Other, RES and DIS interactions, are larger at higher $|\vec{q}|$.

In the E_{avail} variable, the energy scale uncertainty has a larger role, however it is still below the 2% level in any given bin. The cross-section modeling uncertainty grows with increasing E_{avail}, reaching nearly 25% in the highest E_{avail} bin. The 2p2h modeling uncertainty decreases with increasing E_{avail}, which is sensible as little rate into 2p2h-MEC is expected at very high E_{avail}.

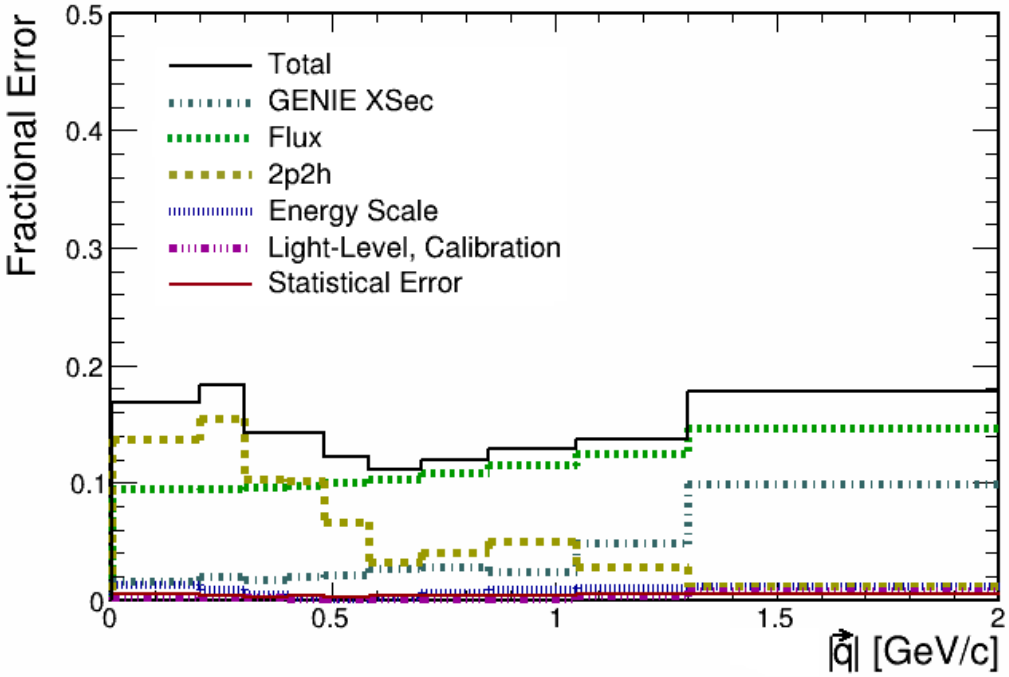

Figure 14.12: Fractional uncertainty on the measured $\bar{\nu}_\mu$-CC inclusive $d\sigma/d|\vec{q}|$ cross section measurement due to different systematic error sources.

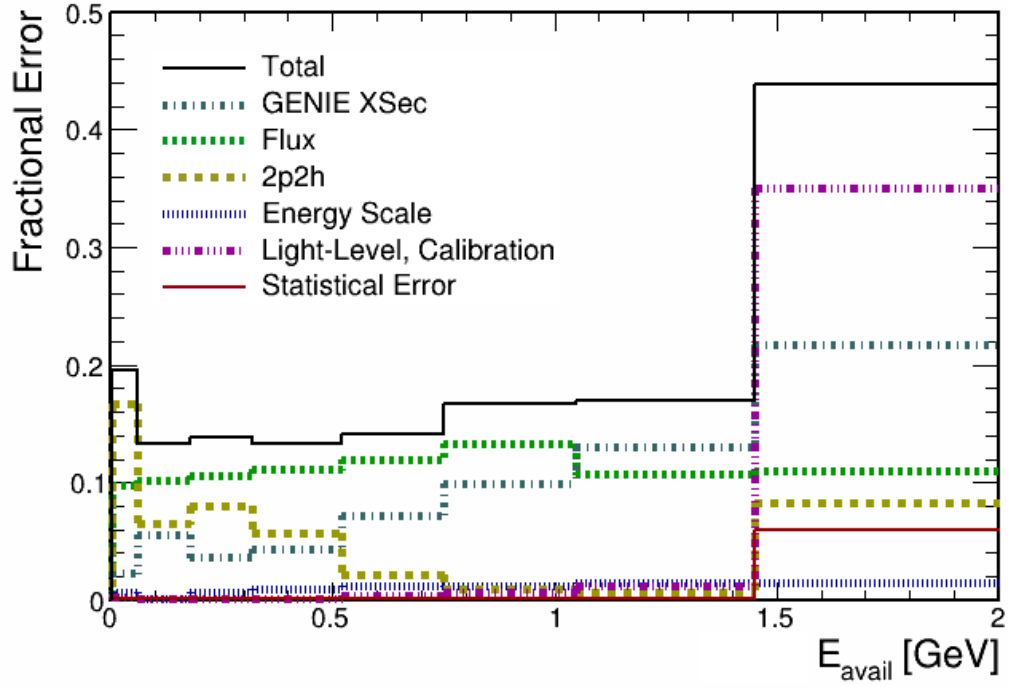

Figure 14.13: Fractional uncertainty on projections of the measured $\bar{\nu}_\mu$-CC inclusive $d\sigma/dE_{avail}$ cross section measurement due to different systematic error sources.

Chapter 15

Estimation of 2p2h in the $\bar{\nu}_\mu$-CC sample

With the inclusive cross-section measurement in hand, we proceed to the extraction of the excess in the data that is not predicted by conventional models of antineutrino-single nucleon scattering. The procedure for this extraction is described in subsection 'Renormalizing the total background' of the Appendix. In brief, a control sample is defined that is devoid of CCQE and 2p2h events; this is the region $E_{avail} > 400$ MeV. Distributions of events of different interaction categories are shown in Fig. 15.1. The CCQE and 2p2h interactions have a substantial presence at $E_{avail} < 400$ MeV/c, however they are estimated by the reference MC to be nearly absent at $E_{avail} > 400$ MeV/c. The latter region is populated by non-CCQE/2p2h events. The $\bar{\nu}_\mu$-CC RES interactions peak at 500 MeV of E_{avail} and then fall off gradually. Neutrino wrong-sign events have a relatively flat distribution, falling off slowly with increasing E_{avail}. DIS events have a relatively flat distribution, but a smaller event rate at low E_{avail}.

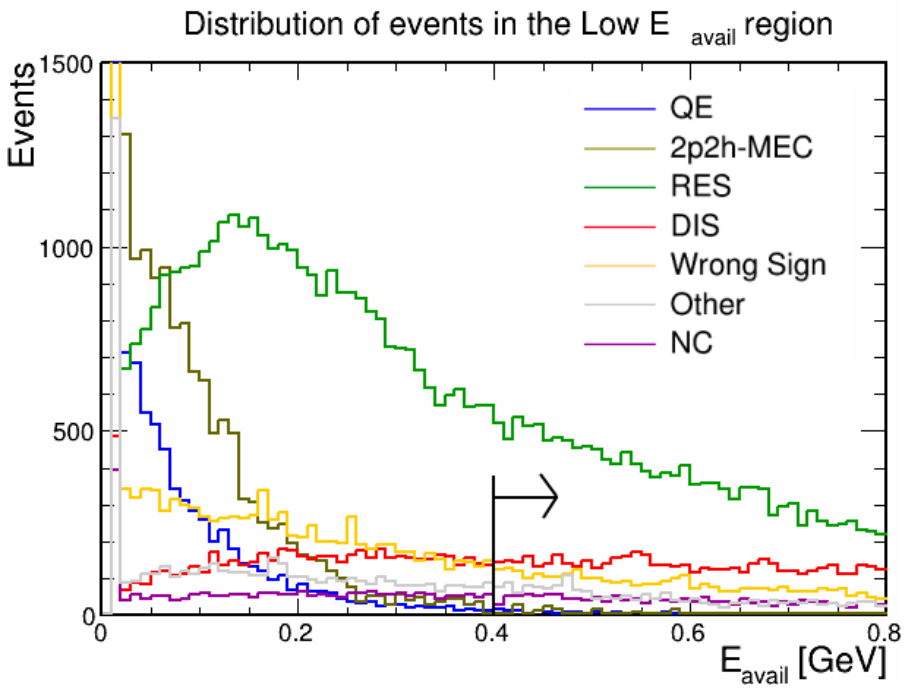

Figure 15.1: Distributions in E_{avail} of antineutrino interaction categories. Events residing in the designated region $E_{avail} > 400$ MeV comprise the control sample used to adjust the background normalization (see text). This region is nearly devoid of the CCQE+2p2h signal while being populated by all other reaction types.

Distributions of selected data and MC events in the control region are shown in Fig. 15.2. Overall, the MC overpredicts the event rate in the sample. In the lowest three bins of E_{avail}, the MC rate prediction exceeds that of the data, while the data is slightly underpredicted in the bin with E_{avail} between 1.05 and 1.45 GeV. The data event count is 80.2% of that of the predicted event count, consequently a scale factor of 0.802 is assigned to normalize the total predicted backgrounds to the CCQE plus 2p2h reactions.

The contributions from reactions which are neither CCQE or 2p2h (i.e. the 'background') is subtracted from the event sample that passes the CC inclusive cuts (shown in Fig. 14.1). This yields an estimate of the $\bar{\nu}_\mu$ CCQE + 2p2h events selected by the analysis. Their distribution is shown in Fig. 15.3.

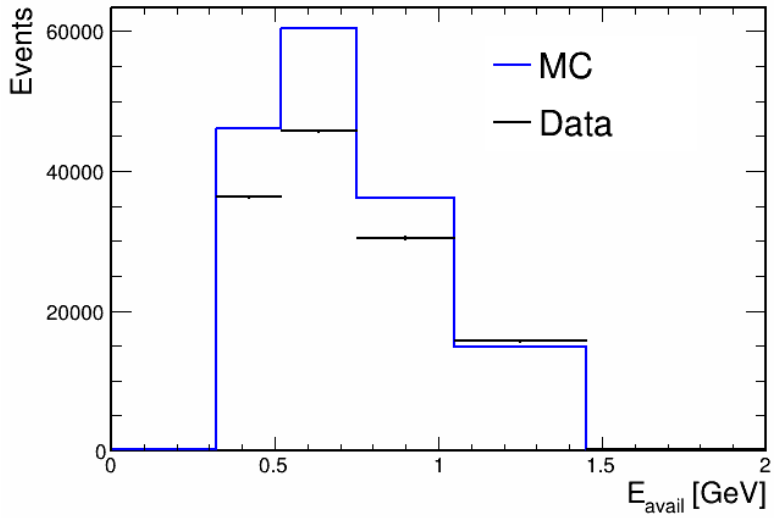

Figure 15.2: Distribution in bins of E_{avail} of events of the selected data sample (black histogram) that populate the control sample, compared to the distribution estimated by the reference MC. The ratio of data to MC events defines a renormalization weight applied to the MC prediction of the non-CCQE + 2p2h event rate at lower E_{avail} values.

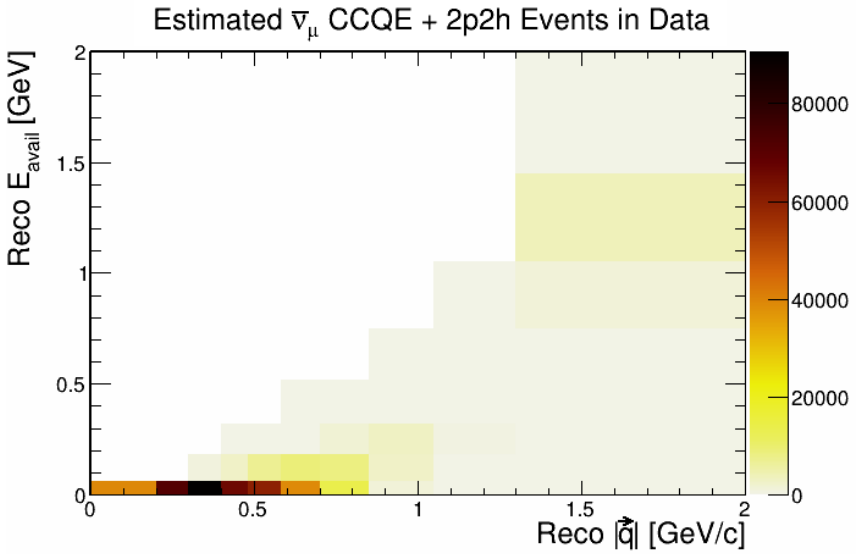

Figure 15.3: Distribution over the plane of $|\vec{q}|$ versus E_{avail} of data events after subtraction of GENIE-based MC templates that describe all selected event categories other than CCQE and 2p2h-MEC. The distribution represents the event rate of CCQE plus 2p2h reactions estimated by this analysis.

To estimate the $\bar{\nu}_\mu$ CC 2p2h event rate by itself, the GENIE estimate of CCQE, which is calculated using the Llwellyn-Smith formalism with a Fermi-gas nuclear model and with an RPA correction applied, is subtracted from the distribution of Fig. 15.3. The GENIE estimate of

CCQE events is shown in Fig. 15.4. Subtraction of Fig. 15.4 from Fig. 15.3 yields the estimate of $\bar{\nu}_\mu$ CC 2p2h events shown in Fig. 15.5. The total estimated 2p2h and CCQE event count is 422,561 events. After subtraction of CCQE events, the number of 2p2h-MEC events in the sample is 170,848 events. Thus the 2p2h-MEC contribution is estimated to be nearly 68% of the strength of CCQE.

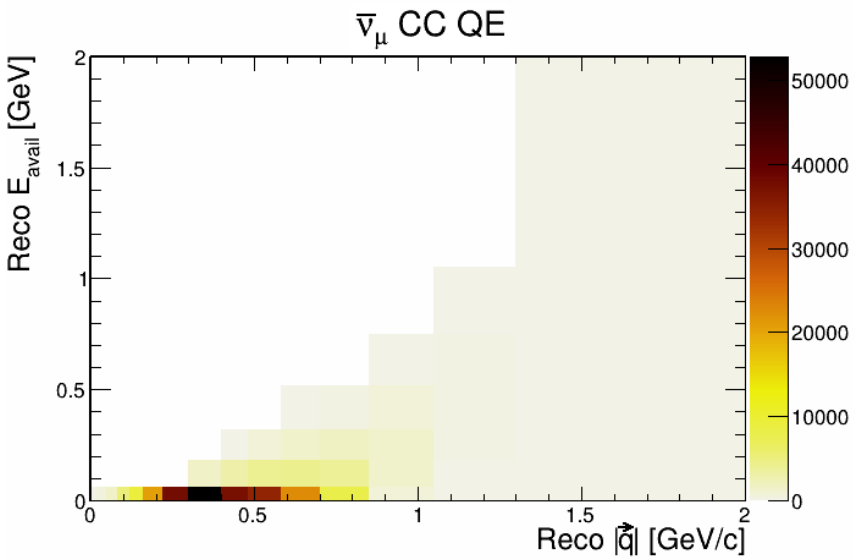

Figure 15.4: Event distribution of true $\bar{\nu}_\mu$ CCQE interactions according to the reference MC. The CCQE cross section is derived from the Llwellyn-Smith formalism for $\bar{\nu}_\mu$p interactions within nuclei modeled as a relativistic Fermi gas.

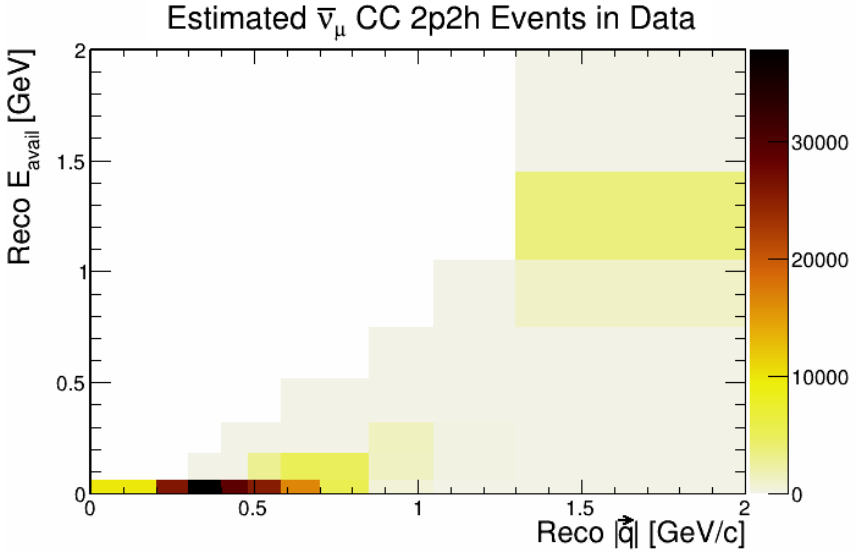

Figure 15.5: Estimated event distribution of $\bar{\nu}_\mu$ 2p2h events in NOvA ND data. The event rate predicted here is 40.4% of the total CCQE + 2p2h estimate.

15.1 Measurement of the 2p2h cross section

This signal displayed in Fig. 15.5 is unfolded with RooUnfold using two iterations. The result is efficiency-corrected, and then divided by the integral of the flux and of the target count to yield the flux-integrated cross section shown in Fig. 15.6. Most of the measured cross section lies in the lowest E_{avail} slice, which goes from 0 - 60 MeV, with $|\vec{q}|$ between 0 - 700 MeV/c. There is a small amount of cross section in the two higher E_{avail} slices, above which the cross section is negligible. This trend confirms the expectation from 2p2h models that nearly all of the hadronic system energy goes into final-state neutrons. Neutron energy is not included in E_{avail}.

The bin-by-bin cross section values are presented in Table 15.1. The peak of the cross section is in the lowest E_{avail} column, in the $|\vec{q}|$ bin ranging from 300 - 400 MeV/c. The value of the measured cross section in this bin is 5.08×10^{-39} cm^2/(nucleon GeV/c GeV).

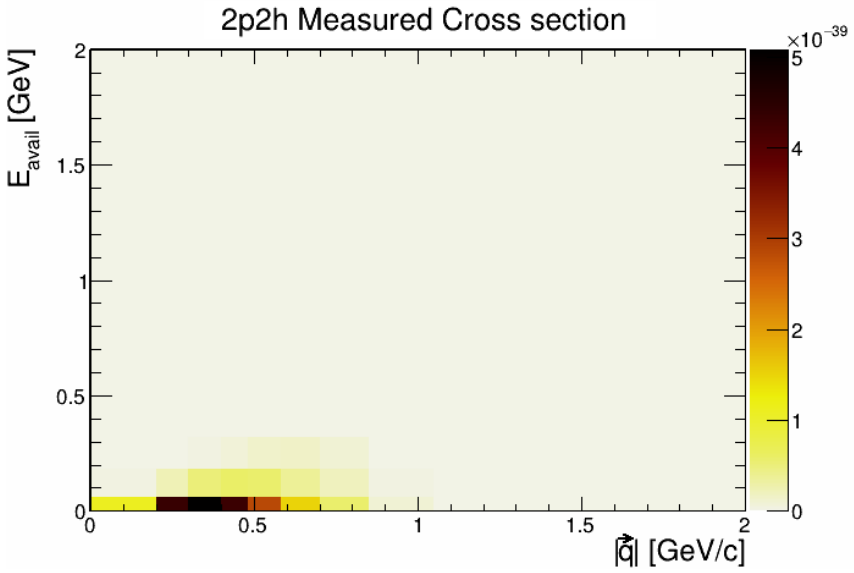

Figure 15.6: Measurement of the $\bar{\nu}_\mu$-CC 2p2h cross section from the NOvA ND data. Most of the cross section is in the lowest E_{avail} slice, with $0 < |\vec{q}| \leq 700$ MeV/c.

$\|\vec{q}\|/E_{avail}$	0.00–0.06	0.06–0.18	0.18–0.32	0.32–0.52	0.52–0.75	0.75–1.05	1.05–1.45	1.45–2.00
0.00–0.20	1.09E-39	2.17E-41	2.89E-43	0	0	0	0	0
0.20–0.30	4.34E-39	2.32E-40	5.48E-42	0	0	0	0	0
0.30–0.40	5.08E-39	5.00E-40	2.74E-41	2.71E-43	0	0	0	0
0.40–0.48	4.30E-39	5.83E-40	6.51E-41	2.71E-43	0	0	0	0
0.48–0.58	2.86E-39	5.63E-40	1.26E-40	1.24E-42	1.23E-43	0	0	0
0.58–0.70	1.50E-39	3.74E-40	1.53E-40	1.67E-42	2.43E-43	0	0	0
0.70–0.85	5.65E-40	1.88E-40	9.14E-41	5.51E-42	3.23E-43	9.94E-44	0	0
0.85–1.05	8.13E-41	2.17E-41	1.72E-41	3.00E-42	2.43E-43	6.94E-43	0	0
1.05-1.30	2.43E-43	9.94E-43	9.70E-43	4.05E-43	1.13E-43	2.94E-43	0	0
1.30–2.00	1.43E-43	5.94E-43	4.70E-43	2.05E-43	1.95E-43	1.24E-43	9.70E-44	4.05E-44

Table 15.1: Bin-by-Bin values of the 2p2h cross section. The cross section peak occurs in the lowest E_{avail} slice in the $|\vec{q}|$ bin 300 - 400 MeV/c, at a value of 5.08×10^{-39} cm^2/(nucleon GeV/c GeV).

The two distributions shown in Figs. 15.7 and 15.8 represent the single differential cross sections $d\sigma/d|\vec{q}|$ and $d\sigma/dE_{avail}$. For $d\sigma/d|\vec{q}|$, the measured cross section lies above all predictions from the data tunes and models. The prediction that best represents the data is that of the NOvA tune, which is to be expected since the tune was constructed by fitting the MC sample to NOvA data. Both the NOvA tune cross section and the measured cross section peak around 350 MeV/c of $|\vec{q}|$. The MINERvA tune has a cross section that falls below the data and the NOvA tune, and it peaks at a higher $|\vec{q}|$ value of 450 MeV/c. The two theory-based models of SuSA and Valencia notably underpredict the data measurement, peaking at 0.15 cm^2/(nucleon per GeV/c) of cross section whereas the measured cross section peaks at 0.55 cm^2/(nucleon GeV/c).

Figure 15.8 shows that most of the measured cross section lies in the lowest bin of E_{avail}. As with $d\sigma/d|\vec{q}|$, the measured $d\sigma/dE_{avail}$ exceeds the predictions of all of the data tunes and models. The NOvA tune gives the closest prediction, followed by the MINERvA tune, and then by the SuSA and Valencia models. In higher bins of E_{avail}, the data and all of the predictions come into agreement. Unfortunately the bin-to-bin uncertainties for both of the measured cross sections are large, with the error bars often encompassing most of the predicted cross sections. Table 15.2 shows the chi-square values comparing the data to the four 2p2h predictions.

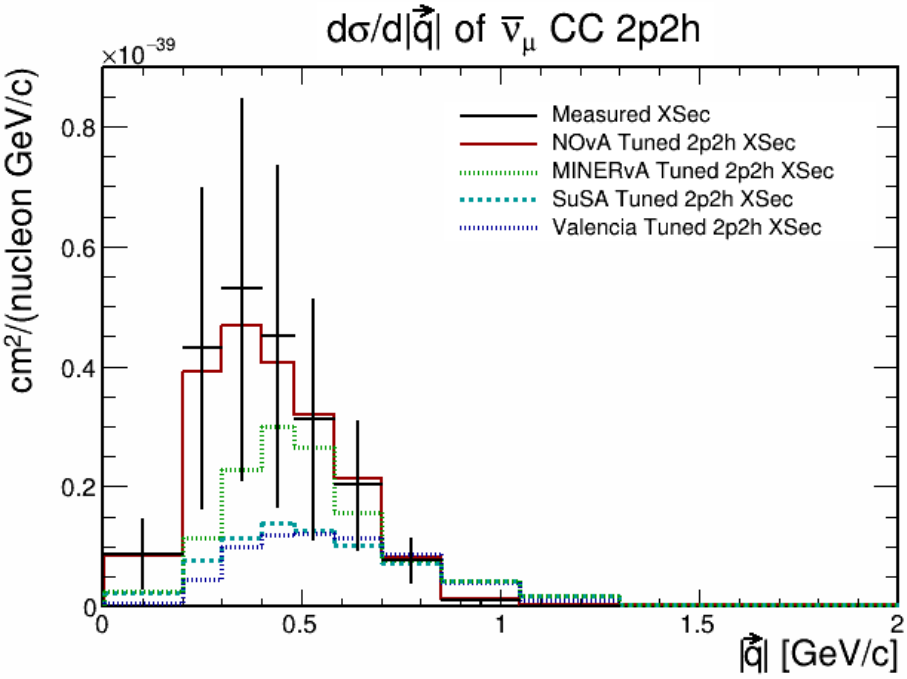

Figure 15.7: The single differential cross section $d\sigma/d|\vec{q}|$ for 2p2h-MEC reactions (black crosses), shown together with predictions based upon two data-tunes and two theory-based models (histograms). The error bars depict the total uncertainty arising from systematics plus statistics.

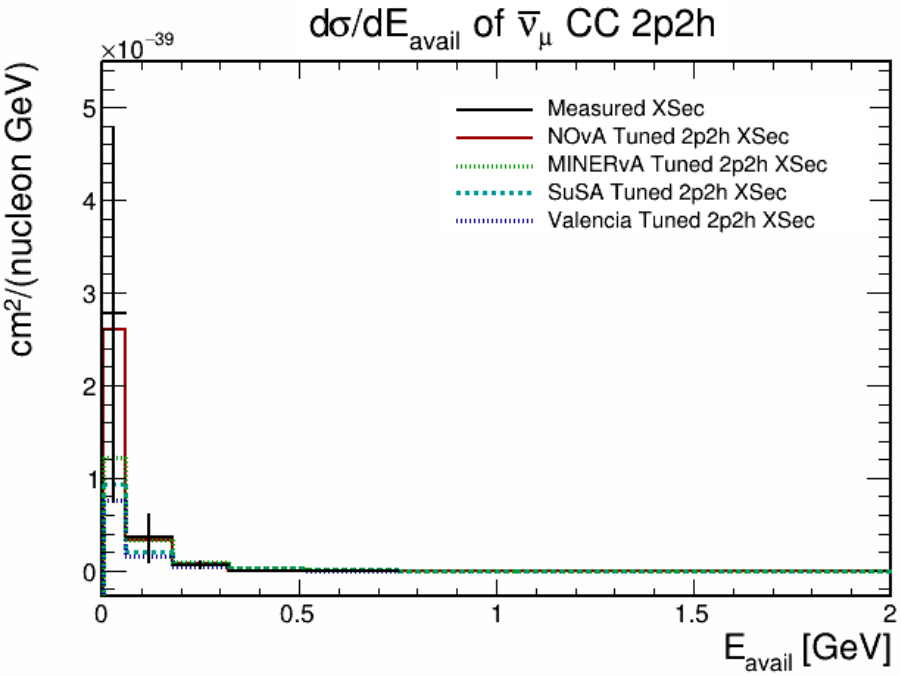

Figure 15.8: The single differential cross section $d\sigma/dE_{avail}$ for 2p2h-MEC, compared to data-tune and model predictions. The limited range of E_{avail} excitation reflects the predominance of neutrons in antineutrino-induced 2p2h final states.

Model	χ^2 (chi-squared)	DoF	χ^2/DoF
NOvA Tune	7.60	22	0.35
MINERvA Tune	13.20	22	0.60
SuSA Model	20.24	22	0.92
Valencia Model	24.19	22	1.10

Table 15.2: Table of χ^2 and χ^2/DoF values of the measured cross section with respect to cross sections of different 2p2h models.

The fractional uncertainty per bin due to systematic and statistical errors is plotted in Fig. 15.9. The bin-by-bin values are provided in Table 15.3. The fractional error across most of the phase space is typically 50 - 60%. The values are higher at bins with $E_{avail} < 300$ MeV and $|\vec{q}| < 400$ MeV, which is where most 2p2h events reside. In regions along the diagonal kinematic boundary, especially at low E_{avail}, the fractional error is much higher as a result of the limited statistics in those regions.

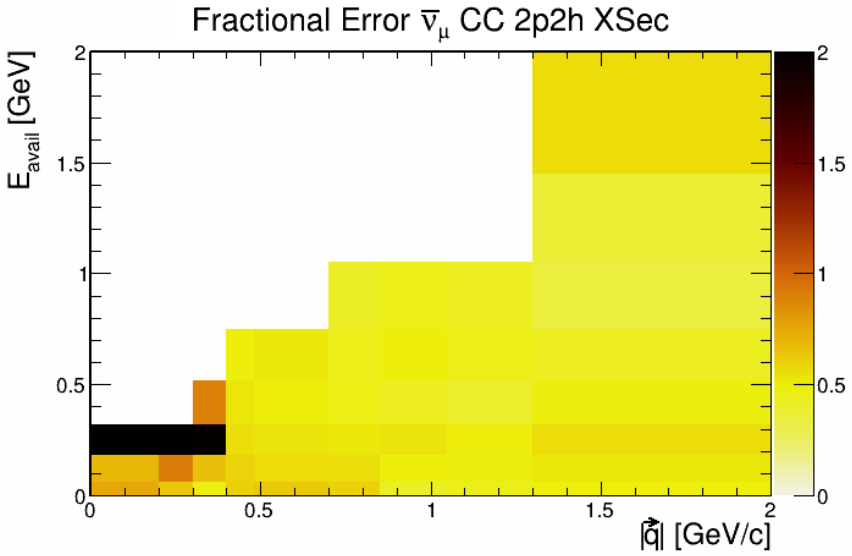

Figure 15.9: Fractional errors for the measured $\bar{\nu}_\mu$ 2p2h-MEC cross section shown in Fig. 15.6.

$\|\vec{q}\|/E_{avail}$	0.00–0.06	0.06–0.18	0.18–0.32	0.32–0.52	0.52–0.75	0.75–1.05	1.05–1.45	1.45–2.00
0.00–0.20	0.75	0.70	2.16	0.00	0.00	0.00	0.00	0.00
0.20–0.30	0.63	0.92	0.96	0.00	0.00	0.00	0.00	0.00
0.30-0.40	0.48	0.67	0.87	0.90	0.00	0.00	0.00	0.00
0.40–0.48	0.60	0.59	0.55	0.53	0.00	0.00	0.00	0.00
0.48–0.58	0.63	0.56	0.54	0.50	0.52	0.00	0.00	0.00
0.58–0.70	0.65	0.60	0.57	0.54	0.59	0.00	0.00	0.00
0.70–0.85	0.60	0.55	0.52	0.47	0.45	0.42	0.00	0.00
0.85–1.05	0.42	0.50	0.54	0.43	0.52	0.46	0.00	0.00
1.05–1.30	0.45	0.48	0.50	0.40	0.46	0.43	0.00	0.00
1.30–2.00	0.50	0.52	0.56	0.49	0.43	0.34	0.37	0.56

Table 15.3: Bin-by-bin fractional errors for the 2p2h cross section.

The fractional uncertainties are shown for bins of the single differential cross sections $d\sigma/d|\vec{q}|$ and $d\sigma/dE_{avail}$ in Figs. 15.10 and 15.11 respectively. A breakdown of the total uncertainty into different systematic sources is also provided in the same figures. At low $|\vec{q}|$, extending from 0.0 to 700 MeV/c, the dominant uncertainty arises from variations allowed by 2p2h models. The uncertainty due to 2p2h-MEC models falls off from $|\vec{q}|$, which is sensible as there are not as many 2p2h events with high $|\vec{q}|$. The next leading systematic uncertainty is due to GENIE cross section uncertaintites, which peaks at around 0.5 GeV/c of $|\vec{q}|$. The third largest systematic is due to the flux, which is relatively constant around at 10%, being slightly higher at high 3-momentum transfer. The energy scale, light level, and calibration systematics, and the statistical uncertainties as well, are small throughout the $|\vec{q}|$ range.

For E_{avail}, uncertainties are largest in the lowest three bins, being driven by uncertainties in 2p2h models and by cross-section modeling uncertainties. The 2p2h error is especially large in the first three bins, above which it falls off and becomes constant at ∼10% in the region 0.5 - 1.5 GeV of E_{avail}. The GENIE cross-section error is relatively constant at 30%. The antineutrino flux uncertainty is also sizable, being approximately 10-15% across the range of E_{avail} bins.

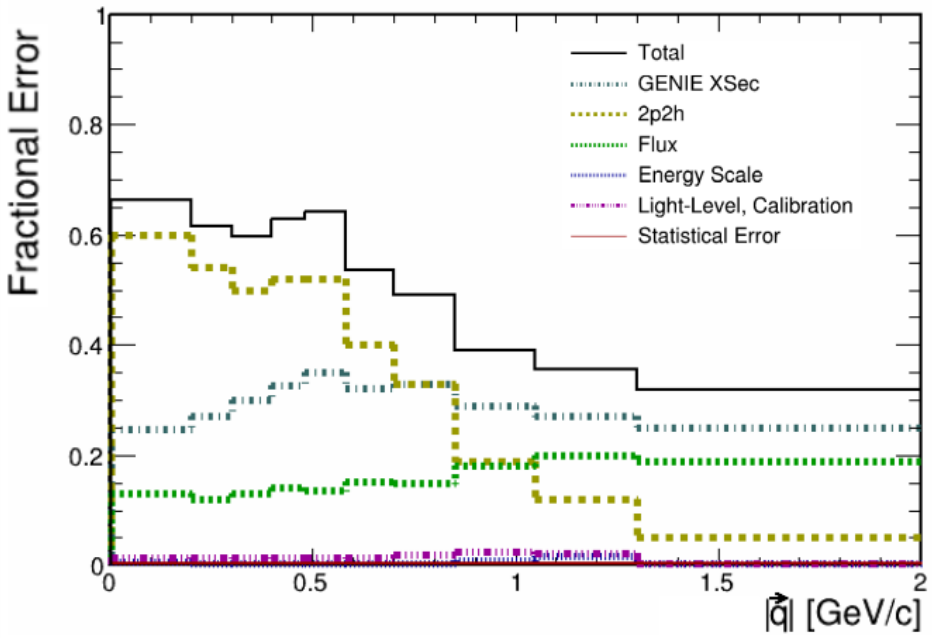

Figure 15.10: Fractional errors for 2p2h $\mathrm{d}\sigma/\mathrm{d}|\vec{q}|$ from different systematic sources. At low $|\vec{q}|$, the uncertainties are led by 2p2h modeling and GENIE cross-section variations.

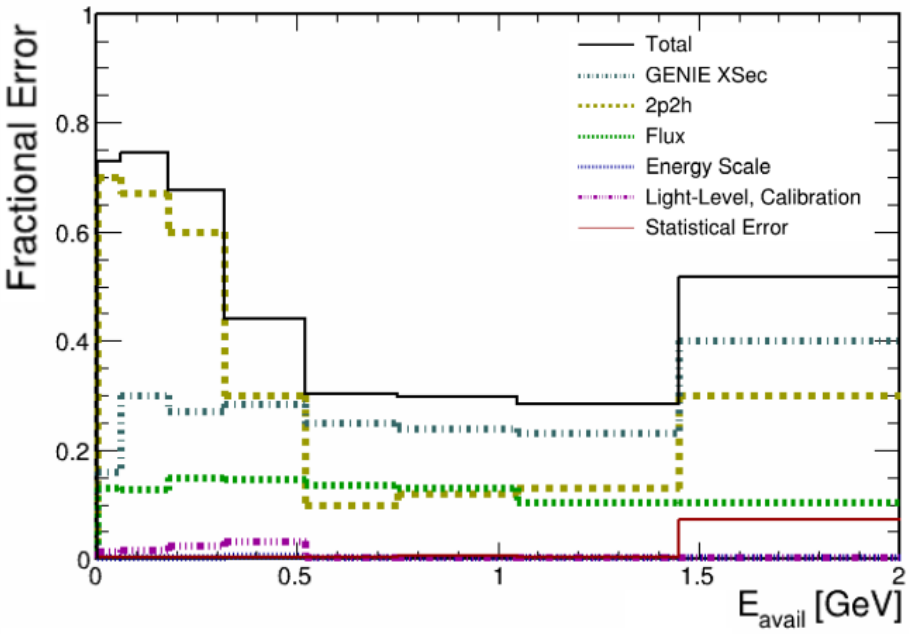

Figure 15.11: Fractional errors for 2p2h $\mathrm{d}\sigma/\mathrm{d}E_{avail}$ from different systematic sources. The fractional error is prominent for $E_{avail} \leq 350$ MeV, which is the region that is expected to be populated by 2p2h events.

Figure 15.12 compares the differential cross section $d\sigma/d|\vec{q}|$ obtained by this work for antineutrino-nucleus scattering to the corresponding differential cross section for neutrino-nucleus scatter-

ing [44].

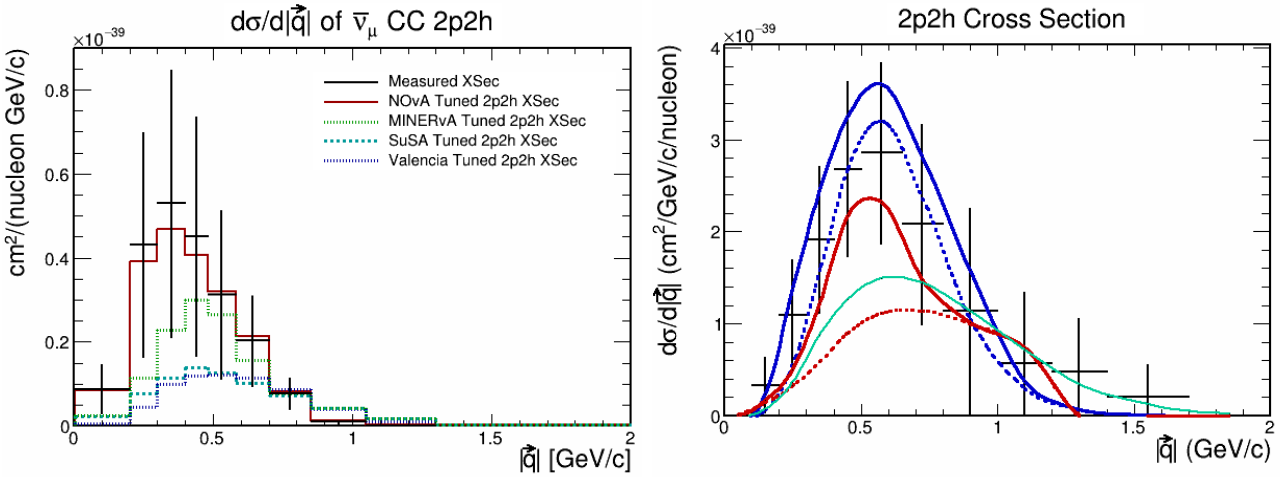

Figure 15.12: Comparison of $d\sigma/d|\vec{q}|$ for $\bar{\nu}_\mu$ 2p2h (left) versus ν_μ (right) measured by NOvA. The 2p2h cross section for antineutrinos is noticably smaller than its neutrino interaction counterpart, e.g. the value of the cross-section peak for $\bar{\nu}_\mu$ 2p2h is $\sim$ 20% of that determined for ν_μ 2p2h using a comparable exposure of the ND detector.

The antineutrino cross section peaks at a lower value than does the neutrino cross section. Notably, the cross section for antineutrinos is distinctly smaller; at its peak value it is approximately 1/5 of the peak cross section measured using neutrinos. Additionally, the antineutrino cross section does not extend as high in $|\vec{q}|$, compared to the neutrino cross section.

Chapter 16

Conclusions

This analysis presents new measurements of CC antineutrino scattering on the predominantly hydrocarbon nuclear medium of the NOvA Near Detector exposed to a narrow-band $\bar{\nu}_\mu$ beam with spectral peak at 2 GeV. The first major measurement is the double differential cross section for inclusive CC $\bar{\nu}_\mu$-nucleus scattering, $d^2\sigma/d|\vec{q}|dE_{avail}$. The number of signal events analyzed is 837,038 events. The analysis uses three-momentum transfer, $|\vec{q}|$, and available hadronic energy, E_{avail}, kinematic variables that are closely related to those used in theoretical treatments of 2p2h-MEC interactions. The measurement is carried out with data-driven constraints applied to NC and to ν_μ-CC reactions. The latter processes comprise the leading backgrounds to the $\bar{\nu}_\mu$-CC signal, representing 3.7% an 9.8% respectively of the selected sample. The bin-by-bin measured values for $d^2\sigma/|d|\vec{q}|dE_{avail}$ are displayed in Fig. 14.8 and Table 14.2. The single differential cross sections $d\sigma/d|\vec{q}|$ and $d\sigma/dE_{avail}$ are also reported, and these are compared to predictions based upon the GENIE v3.0.6 neutrino generator augmented with four different implementation of 2p2h.

The second major measurement is the double differential cross section for the 2p2h-MEC component in the CC inclusive sample. The contribution of 2p2h events is estimated using GENIE-based background templates normalized to a control sample, together with estimation of the CCQE contribution based on conventional charged-current phenomenology and nuclear modeling. The 2p2h signal is estimated to be 170,848 events. The double-differential cross section in $|\vec{q}|$ and

E_{avail} for the extracted 2p2h contribution is presented in Fig. 15.6 and Table 15.1. Single differential cross sections $d\sigma/d|\vec{q}|$ and $d\sigma/dE_{avail}$ are reported in Figs. 15.7 and 15.8. The latter cross sections are compared to the predictions of two different data-tune models and two current theory-based models, namely the SuSA-v2 model and the Valencia model. The analysis indicates shortfalls with the theory-based models: Neither model predicts the strength of the 2p2h-MEC contribution indicated by the NOvA data, and there is tension in predicted versus observed shapes of the cross sections as well. The observations obtained by this analysis enable comparisons to be made with an analysis of 2p2h-MEC that has been carried out using neutrino-nucleus scattering data recorded by the NOvA Near Detector. The measurements of this Thesis shed now light on CC $\bar{\nu}_\mu$-nucleus interactions and 2p2h reactions as occur in the 1.0 - 3.5 GeV region of incident $E_{\bar{\nu}}$. They will facilitate the development of more precise $\bar{\nu}_\mu$ interaction models for this region, an endeavor of great important to present and future neutrino oscillation experiments.

Chapter 17

Appendix

17.1 Methods to determine the $\bar{\nu}_\mu$-CCQE + 2p2h-MEC contribution

This Appendix examines how the CCQE (charged current quasielastic) and 2p2h-MEC events in the NOvA Near Detector RHC sample can be separated from other event types and characterized using the variables three momentum transfer, $|\vec{q}|$, and available hadronic energy, E_{avail}. The MC sample used in the study is a preselected ND RHC event sample that satisfies the quality and containment cuts. The exposure of the MC sample utilized is 5.06×10^{21} POT.

17.1.1 The ND $\bar{\nu}_\mu$ CC inclusive sample

The distribution of the events that satisfy the antineutrino charged current selections is shown in Fig. 17.1 (left). The majority of events have nearly zero available energy. This accumulation of events stands in contrast to what is observed in the FHC sample (Fig. 17.1 (right)) [44], in which events are distributed across higher E_{avail} values. The differences in the distributions indicate that a different approach is required to obtain an estimate of the $\bar{\nu}_\mu$-CCQE and 2p2h-MEC contribution in the RHC sample.

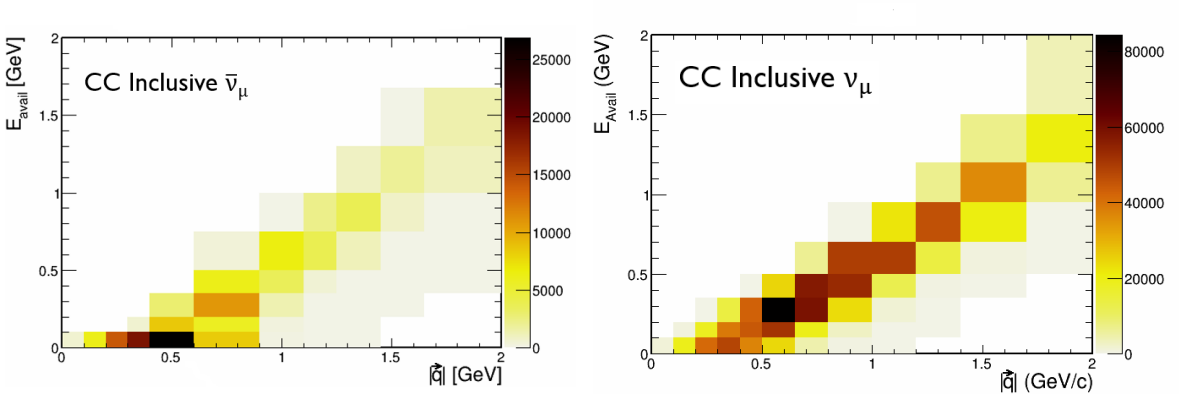

Figure 17.1: Distributions of ND RHC (left) and FHC (right) events that pass CC inclusive selections. Many of the events in the antineutrino sample have zero available energy, as shown by the accumulation of events in the $E_{avail} \leq 100$ MeV slice. The latter contribution comes mostly from CCQE and 2p2h-MEC events.

17.1.1.1 Isolation of CCQE + 2p2h events

The accumulation of events in the lowest E_{avail} slice is due to the large rates of CCQE and 2p2h-MEC interactions; their distributions are shown in Fig. 17.2. Comparing the two distributions, one sees that they mostly occupy the same kinematic phase space. Reaction equations and Feynman diagrams of the two reaction types are displayed in Fig. 17.3. Both reactions involve an incoming leptonic current exchanging a $W^{\pm}$ boson with a proton, turning it into a neutron. However, in 2p2h interactions an additional nucleon is bound to the interacting proton via meson exchange (or perhaps by gluon exchange). Because the CCQE and 2p2h interactions share similar topologies and the same region of the analysis variable phase space, their contributions can be estimated together.

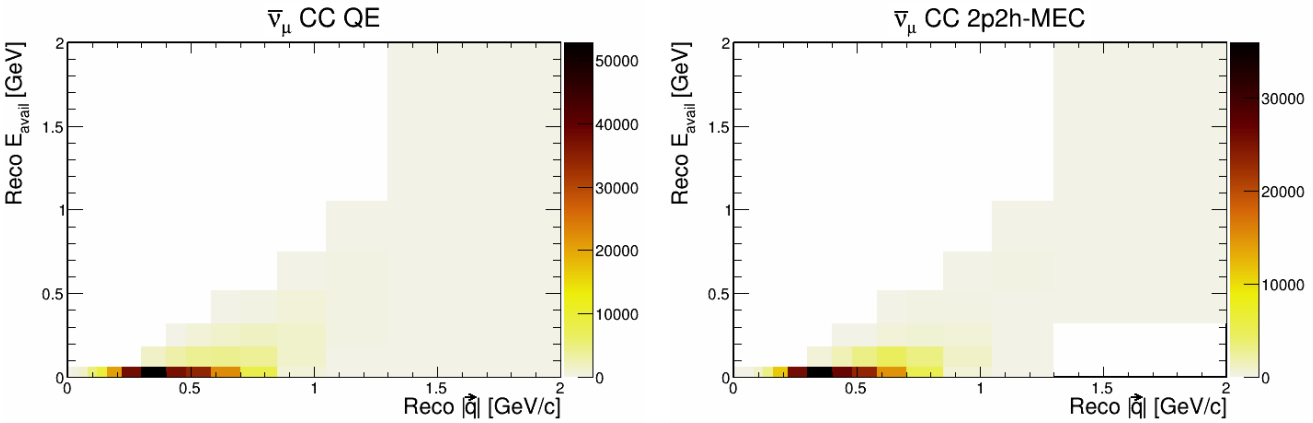

Figure 17.2: Distribution of $\overline{\nu}_\mu$ CCQE (left) and CC 2p2h-MEC (right) in variables $|\vec{q}|$ and E_{avail}. Both reactions tend to have very low available energy in the final state and populate the $E_{avail} <$ 100 MeV region.

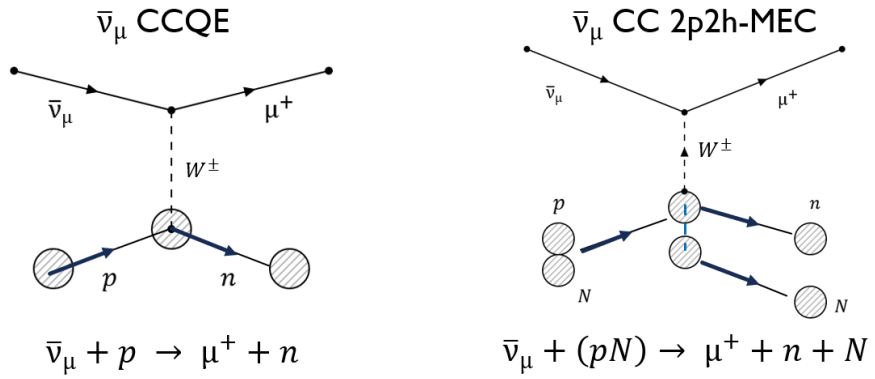

Figure 17.3: Feynman diagrams of the signal reactions $\overline{\nu}_\mu$ CCQE (left) and 2p2h (right). Both interactions produce a μ^+ and a neutron. In the 2p2h reaction, an extra nucleon participates in the interaction.

17.1.2 Background reactions

In addition to the "signal" CCQE and 2p2h-MEC events, there are events from other processes that pass the CC inclusive selection criteria. The total rate of background processes in the sample must be estimated in order to obtain the event rate distribution of signal interactions. The estimation of the background interactions types is based on templates generated by GENIE v3.0. The distributions of the background interaction types are shown are Figs. 17.4 and 17.5. The leading background interaction is baryon resonance production (RES), which typically has available energy values from 0 - 750 MeV in the final state. This is followed by Deep Inelastic

Scattering(DIS) reactions, which have a much smaller rate overall but a broader spread over $|\vec{q}|$ and E_{avail}. The third background is the subsample of neutrino events that occur in the RHC sample, also referred to as Wrong-Sign. Most of the Wrong-Sign events share the same region as the signal interactions, with modest rates occurring at E_{avail} >100 MeV.

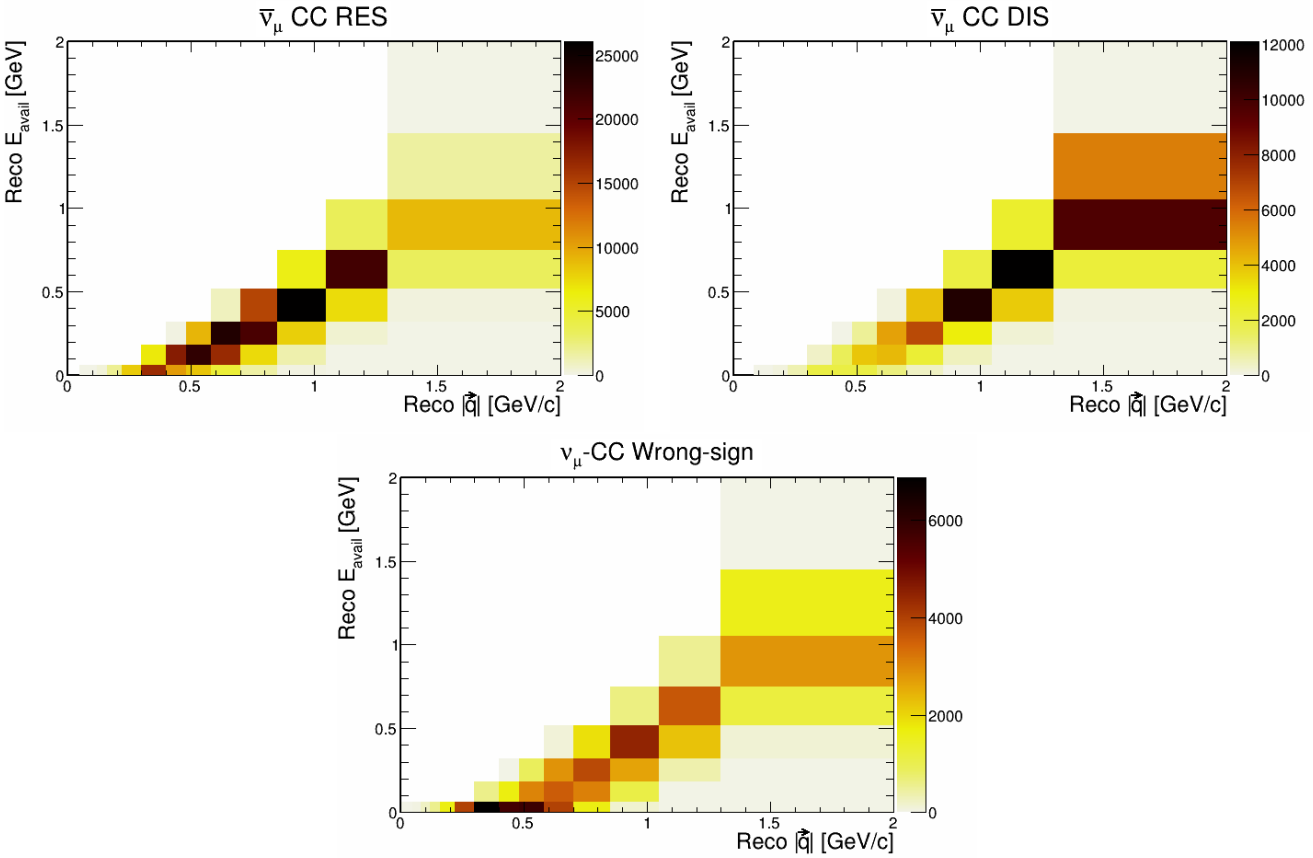

Figure 17.4: Distributions of leading background interactions: RES (upper left), DIS (upper right) and wrong-sign events (lower) in variables $|\vec{q}|$ and E_{avail}. The RES and DIS interactions populate regions with $E_{avail} \geq 200$ MeV whereas most of the wrong-sign events oocupy the bins in the lowest E_{avail} slice.

Other smaller backgrounds in the sample are electron neutrino events, true nonfiducial events, CC coherent interactions ($\overline{\nu}_\mu A \rightarrow \mu^+ \pi^- A$) and neutral current (NC) interactions. Concerning the latter, pions and protons of NC interactions can be mistagged as muons, and this misidentification results in their erroneous selection. These smaller backgrounds, including CC electron neutrino interactions, nonfiducial reactions and antineutrino CC coherent scattering events are accumulated under the umbrella term "Other". The event populations of the Other sample lie away from the signal region.

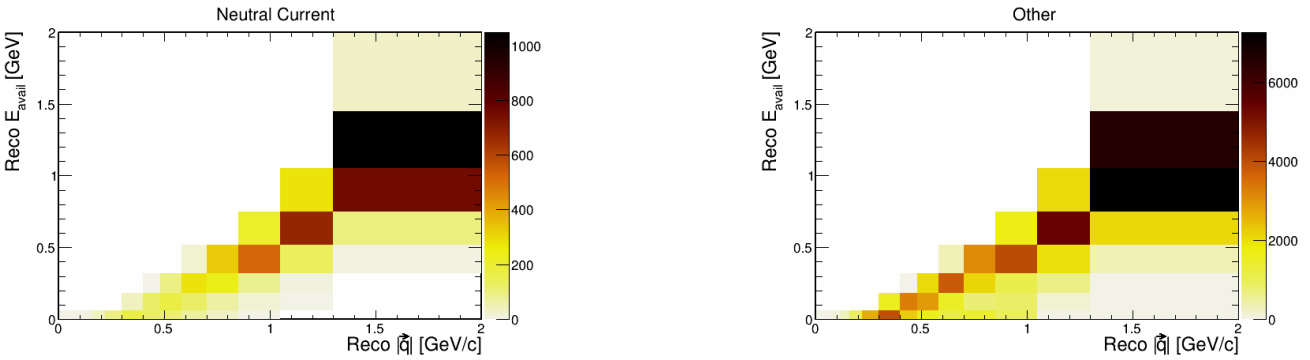

Figure 17.5: Distributions of neutral current (left) and Other (right) events in variables $|\vec{q}|$ and E_{avail}. Both reaction categories have small event rates and are concentrated away from the signal region.

17.1.3 GENIE based background estimation

One way to estimate the CCQE + 2p2h-MEC contributions in a selected sample is to take the total CC selected sample and subtract off the background components as generated by GENIE. This procedure is demonstrated below. A distribution of a 'pseudo'-data sample, which is a sample of RHC events that have passed the CC inclusive selections, in a systematically shifted universe, is shown in Fig. 17.6. As expected, most of the observed events occur in the lowest E_{avail} slice, which is enriched in CCQE and 2p2h-MEC events. Above that is the "remote region", which is mostly occupied by background interactions together with a small proportion of signal events.

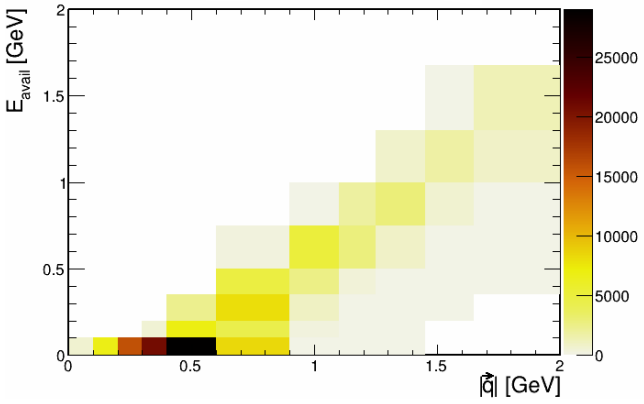

Figure 17.6: A pseudo-data sample of CC inclusive selected events from the ND RHC sample. Most of the events have very low available energy. A relatively sparse population of events are spread over higher E_{avail} values.

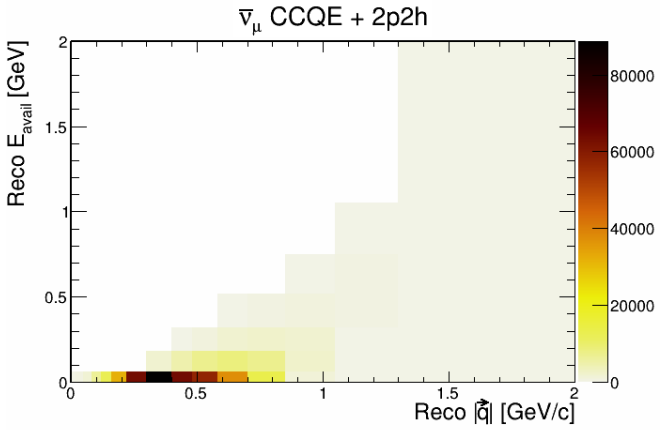

Figure 17.7: Distribution of CCQE and MEC events within the sample shown in Fig. 17.6.

The CCQE + 2p2h contribution according to MC truth is shown in Fig. 17.7. The total background in the sample is estimated to first order by taking all the background distributions shown in Figs. 17.4 and 17.5 and summing them, yielding the distribution shown in Fig. 17.8. Then, the CCQE + 2p2h content in the sample shown in Fig. 17.6 is estimated from the selected sample of Fig. 17.6 by simply subtracting the total background shown in Fig. 17.7 from the selected sample of Fig. 17.6. This yields the distribution in Fig. 17.9, which is to be compared to the MC truth distribution of Fig. 17.7.

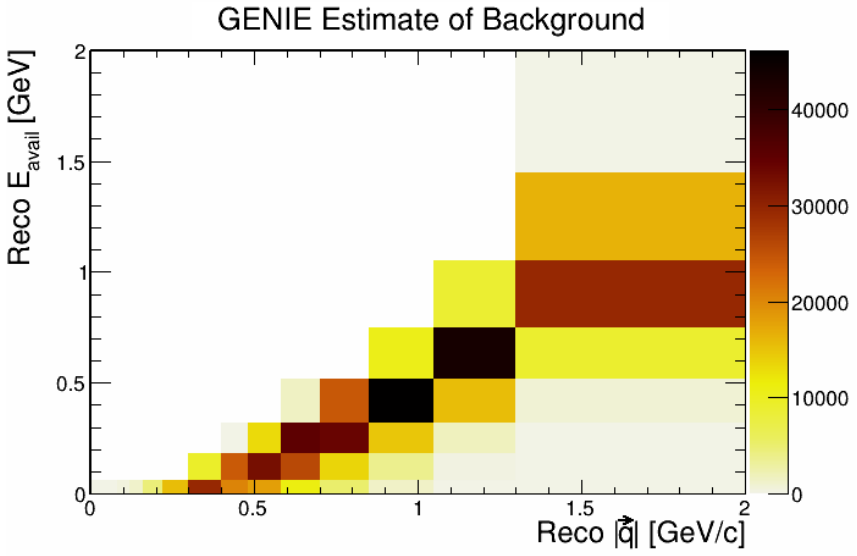

Figure 17.8: Total background event rate in variables $|\vec{q}|$ and E_{avail} of a CC inclusive selected ND RHC sample as estimated by GENIE.

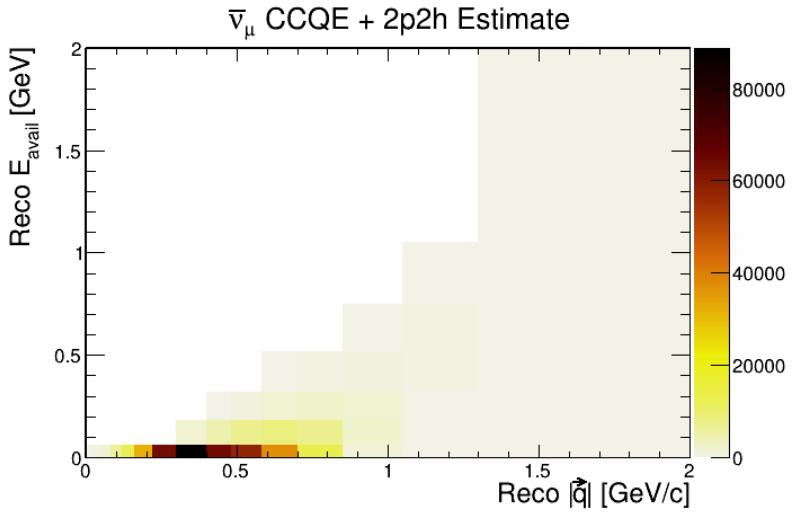

Figure 17.9: A first-order estimate of the signal obtained from the sample shown in Fig. 17.6 via simple background subtraction.

At first glance, the distribution of Fig. 17.9 appears to be very similar to the actual signal. Most of the estimated signal events lie in the lowest available energy slice, and there is a sparser event count at $E_{avail} \geq 100$ MeV.

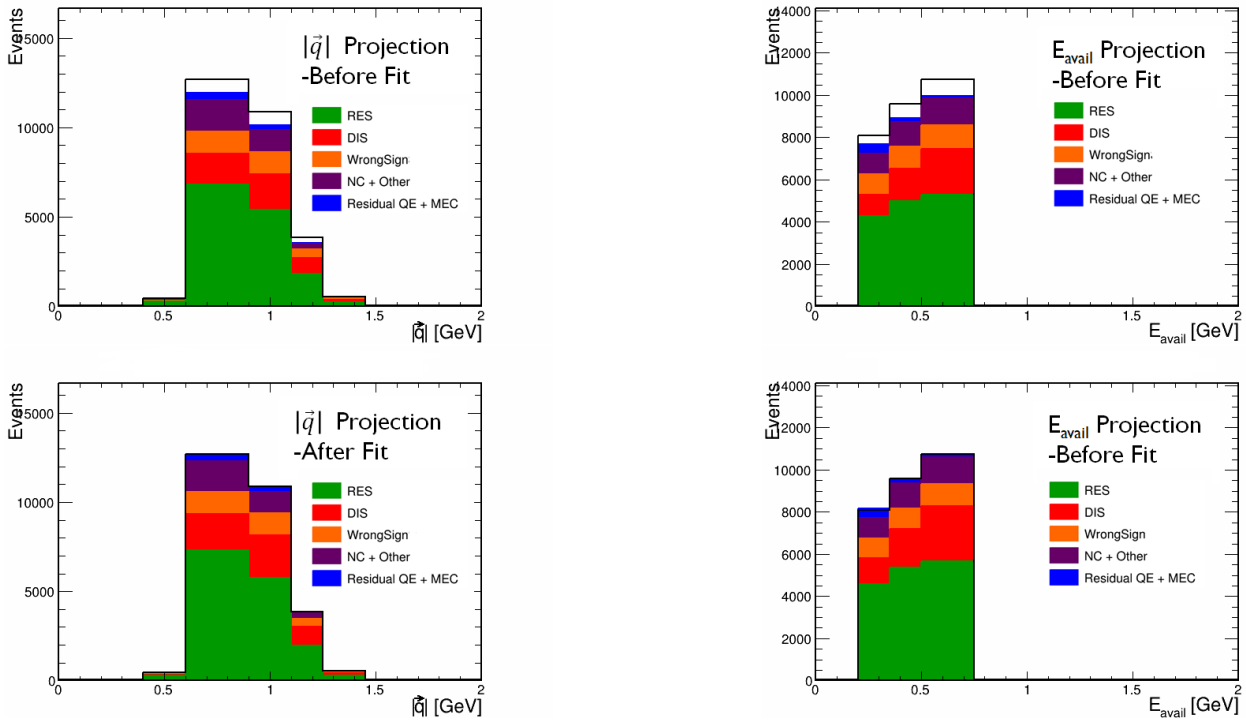

Figure 17.10: Example of a fit done in the base control region for one universe. The plots on the left and right are $|\vec{q}|$ and E_{avail} projections of the base control sample respectively. The plots at the top and bottom are the distributions for MC and data before and after the fit.

An example of a fit to the base control region is shown in Fig. 17.10. In this case, it is observed that the predicted background in the sample falls short of the actual event rate in the control sample. These MC and data distributions are fit with the software package MINUIT [28]. Over many iterations, MINUIT adjusts the normalizations of the individual background templates in order to best match the total prediction to the data event distribution. To obtain reasonable solutions, the templates normalizations are not allowed to change by more than 30% from the nominal. For the base control region fit, only the RES, DIS and Wrong-Sign components are allowed to float. In this case, it is observed that the DIS and Wrong-sign components are enhanced and RES is slightly reduced. The changes in the normalization in this fit are applied to the nominal templates. Changing the normalizations of the individual templates changes the estimated background rate and shape, which in turn affects the estimate of the signal obtained from a CC inclusive selected sample. Figure 17.11 illustrates how the estimate of signal changes when a base control sample fit is done. In this case, it is observed that a fit improves the estimate of the signal.

Unfortunately, when the procedure is repeated across a number of universes, a single base control region fit often fails to return a good estimate of the signal. This is observed with χ^2/DoF values of the estimated signal with respect to the actual signal before and after the fit. A good fit should always return lower χ^2/DoF value. Unfortunately, there are a significant number of universes where the metric value is increased by the fit, indicating a poor estimate.

However, if the comparisons are limited to $E_{avail} \geq 100$ MeV, it is observed that the fit does improve the signal estimation in this region. This is clear when one observes the ratio of estimated signal event count to truth for $E_{avail} \geq 100$ MeV before and after the fit, as shown in Fig. 17.12. Thus a single base-region fit to the control sample provides a better estimate the background for $E_{avail} \geq 100$ MeV. The signal for $E_{avail} < 100$ MeV is yet to be estimated. This issue is addressed in the next section.

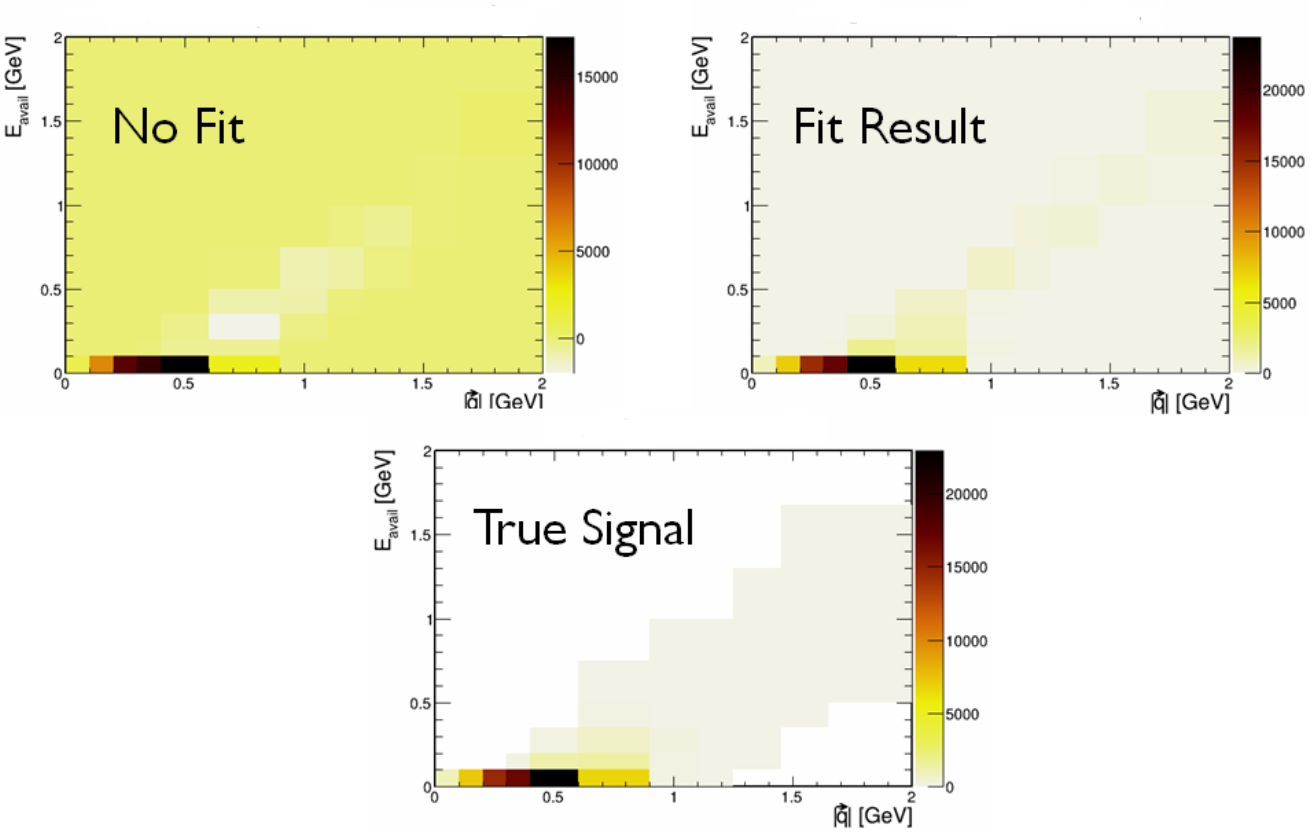

Figure 17.11: Example of signal estimate before the fit (upper left) and after the fit (upper right). The true signal event distribution is shown at the bottom for comparison.

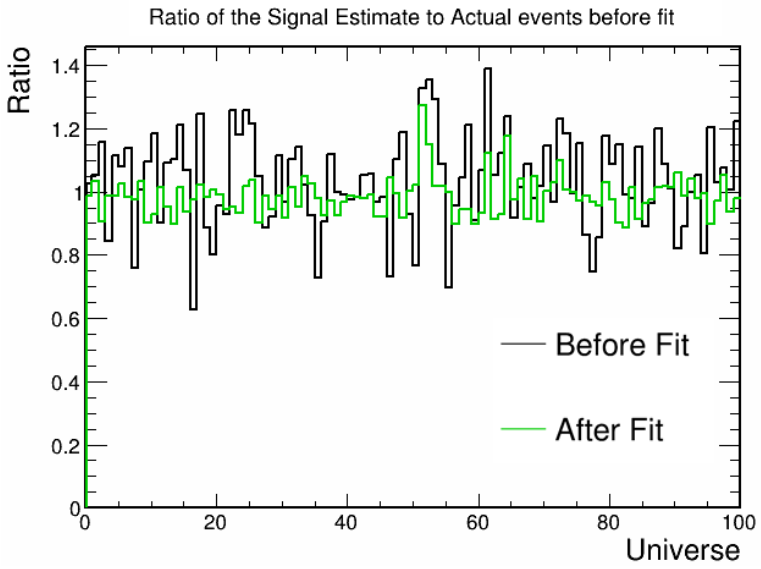

Figure 17.12: Ratio of the estimated signal to the actual before the fit (black) and after the fit (green) for 100 universes.

17.1.4 Estimation of CCQE + 2p2h at low E_{avail}

Examination of the distributions of different interaction types in E_{avail} as shown in Fig. 17.13 reveals that most events in the CC inclusive sample have close to zero available energy. The same E_{avail} distribution has been plotted in three different scales in Fig. 17.13 to show the magnitude of the signal event rates at very low E_{avail} and also the distributions of the different reactions at higher E_{avail}. Most (79%) $\overline{\nu}_\mu$ CCQE and 2p2h-MEC events have less than 10 MeV of E_{avail}. Very few (4.5%) signal events have greater than 400 MeV of available energy. The background interactions are dwarfed by these dominant signal processes at low available energies. However, at $E_{avail} > 10$ MeV the background processes start to dominate. RES is the leading background interaction, with the event rates peaking at around 150 MeV before falling off with increasing E_{avail}. Most wrong-sign events are located in the $E_{avail} \leq$10 MeV bin. But there are also muon neutrino CC events with $E_{avail} > 10$ MeV, with the event rates falling off with increasing available energy. The DIS interaction and NC/Other distributions are almost flat across the range of E_{avail} values.

In order to estimate the background contribution in the region $E_{avail} < 100$ MeV, the event distributions in the variables muon kinetic energy and muon production angle were examined. A distribution of the events in the lowest E_{avail} bin in the aforementioned variables is displayed in Fig. 17.14.

Examination of the individual signal and background event types, as shown on Figs. 17.15 and 17.16 reveals that, most signal events have muon production angles less than 20 degrees, with muon kinetic energies between 1 - 2.5 GeV. The background interactions also have a presence in the region. However, a significant portion of the background events also lie outside this region. To estimate the background for the lowest E_{avail} slice, the RES, DIS, Wrong-sign and NC/Other background templates are fitted outside the $1 <$ muon K.E. ≤ 5 GeV and $0 < \theta \leq 20^o$ region. This region will be referred to as the "muon out-region".

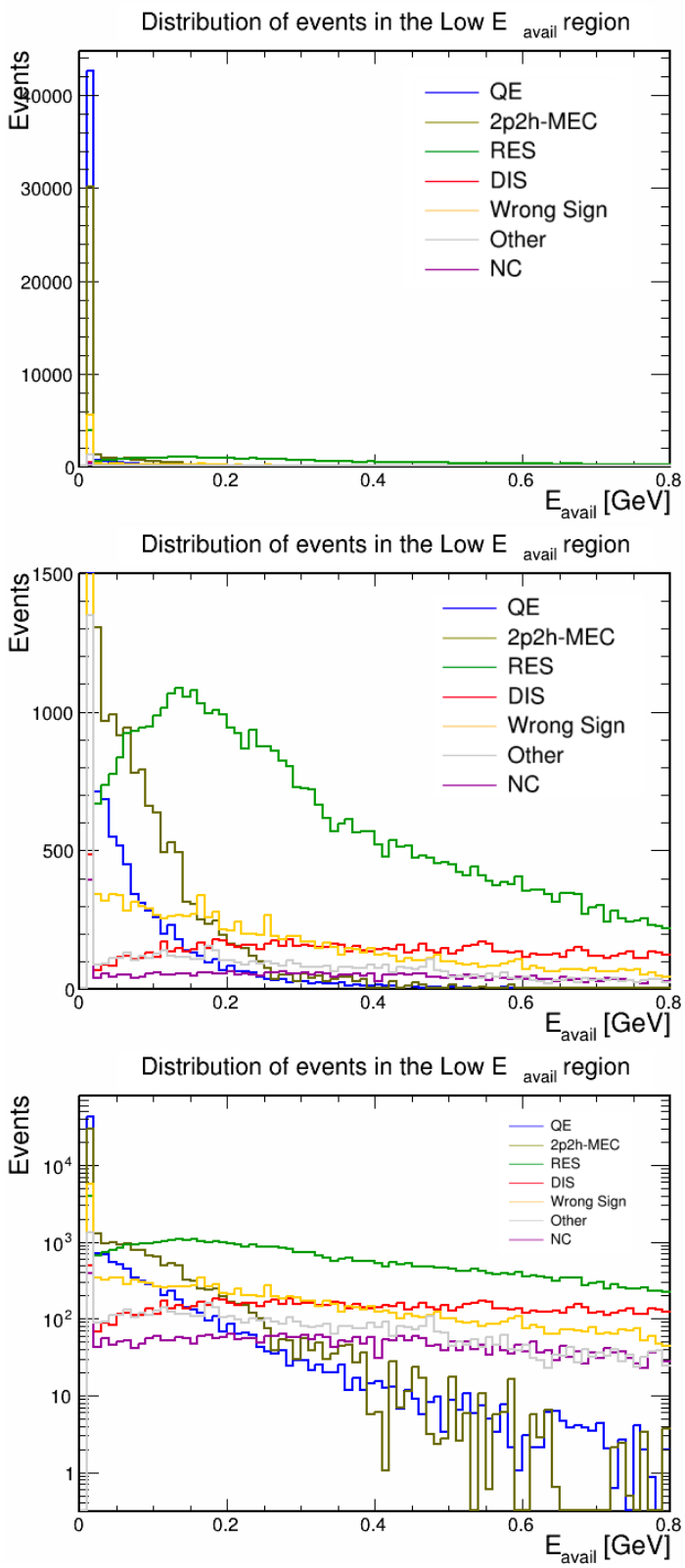

Figure 17.13: Top: Available energy distributions; most QE and 2p2h events have zero E_{avail}. Middle: Distribution with truncated y-axis. Bottom: Distribution displayed on log scale.

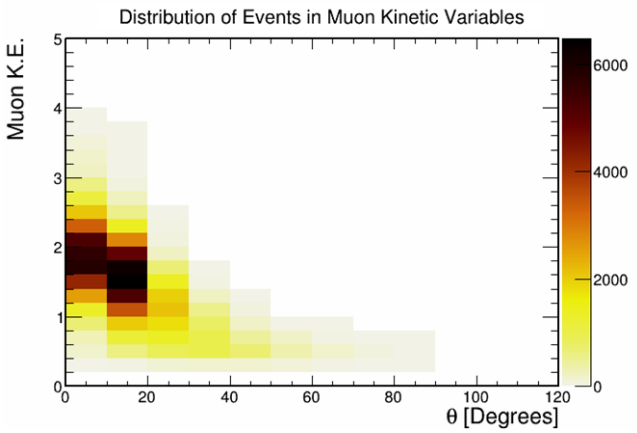

Figure 17.14: Distributions of CC inclusive selected events from the ND RHC sample with $E_{avail} < 100$ MeV in the variables muon production angle and muon kinetic energy.

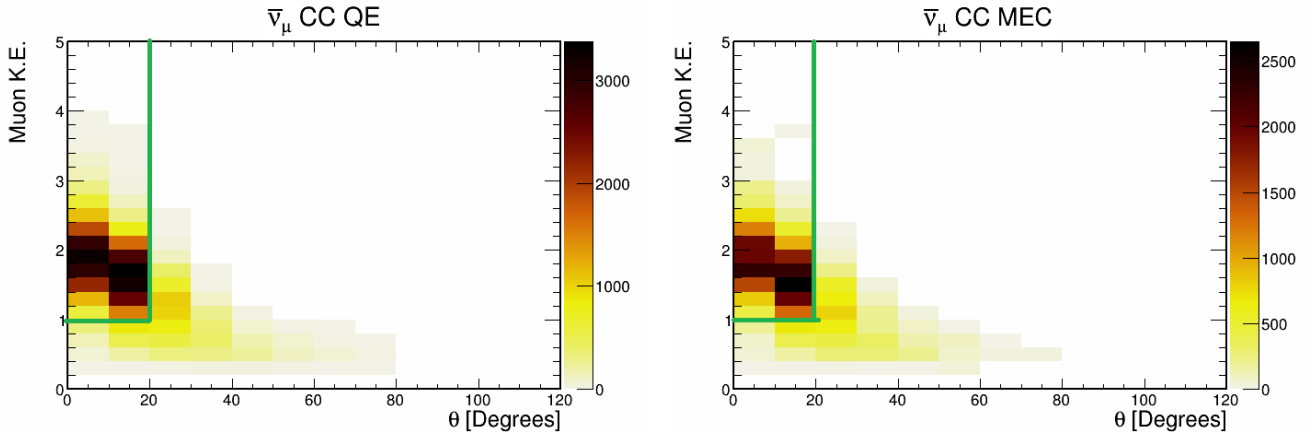

Figure 17.15: Distribution of signal event interactions with $E_{avail} < 100$ MeV in the variables muon production angle and muon kinetic energy.

An example of the fit in the muon out-region is shown in Fig. 17.17. These new normalizations change the total background distribution in the lowest available energy slice, and therefore, changes the estimate of the signal. It is observed that the fit yields a more accurate description of the actual signal compared to a first-order estimate.

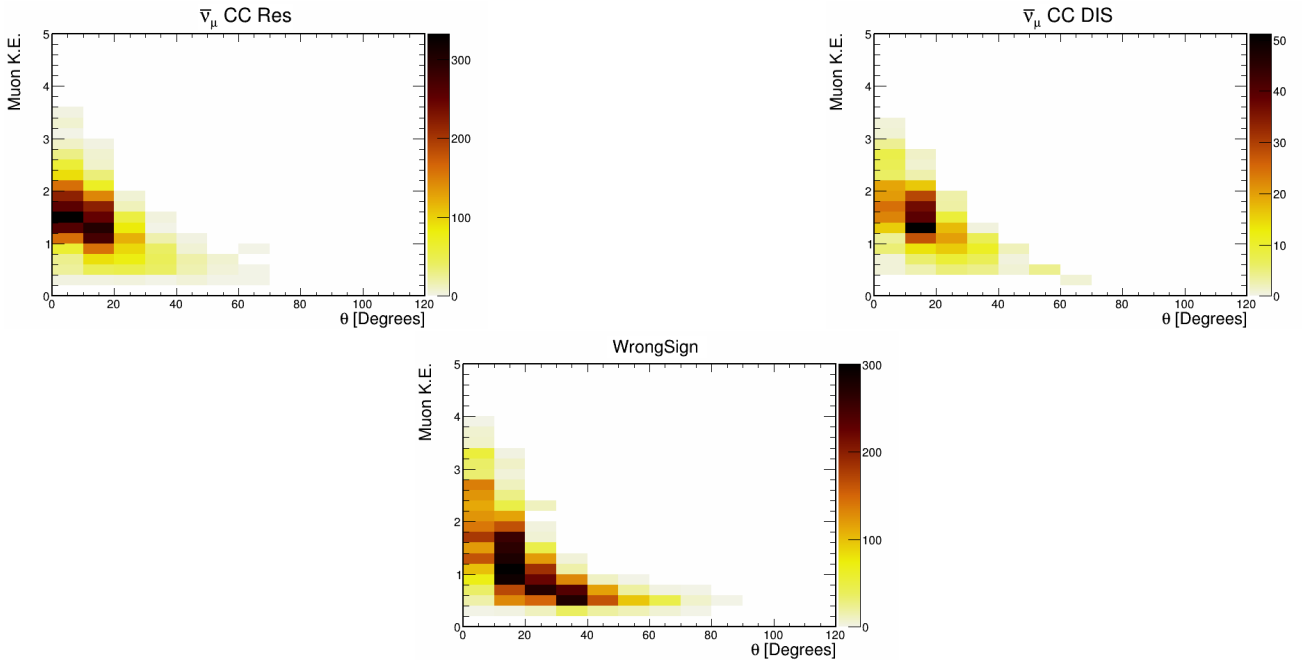

Figure 17.16: Distribution of background event categories with $E_{avail} < 100$ MeV in muon production angle and muon kinetic energy. The superimposed border defines the 'muon out-region' (see text).

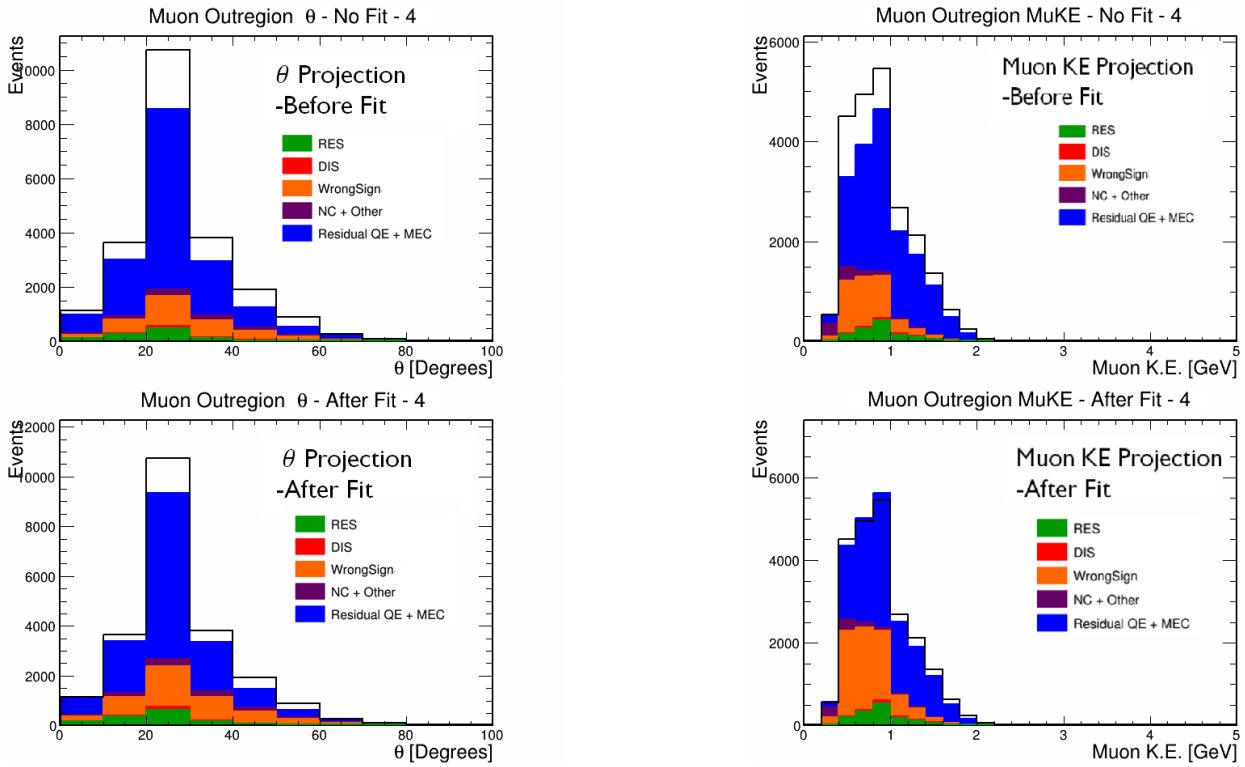

Figure 17.17: Distribution of background event categories with $E_{avail} < 100$ MeV, in variables muon production angle (left) and muon kinetic energy (right).

However, an examination of fit performance in this muon out region, repeated across 100 universes and summarized in Fig. 17.18, shows that the fit does not necessarily improve the estimation as compared to a simple first order estimate. In some universes, the fit brings the signal estimate closer to the truth, and in others, it worsens our estimate of the signal for very low available energy.

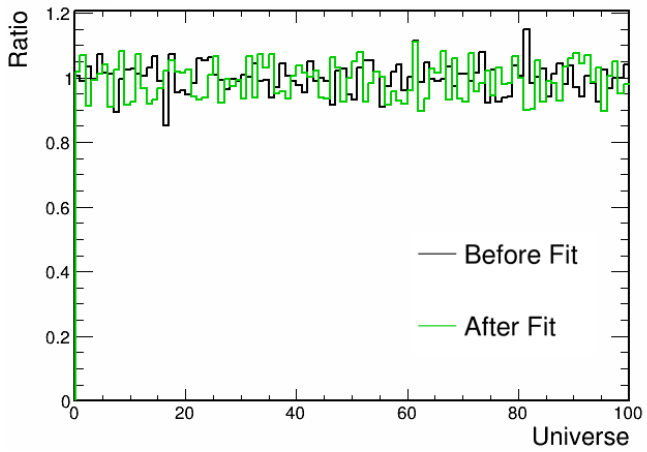

Figure 17.18: Ratio of estimated signal event count to true signal event count for $E_{avail} < 100$ MeV before (black) and after (green) the two-stage fitting is done.

17.1.5 Total signal estimation from the two fits

The estimates of the signal for the two regions $E_{avail} > 100$ MeV and $E_{avail} \leq 100$ MeV, obtained with the methods described in the previous sections can be combined to yield the total estimate of $\bar{\nu}_\mu$ CCQE and 2p2h-MEC contribution in the selected sample. The performance of this two-fit sequence is summarized in Fig. 17.19, which shows the ratio of estimated events to true event count over the entire phase space before and after the fits are carried out.

This plot shows that the estimate of the signal improves as a result of renormalizing the background templates with suitable control samples. In general, the ratios are closer and more stable at a value of 1.0, as opposed to a first-order estimate which often underestimates or overestimates the signal in sample.

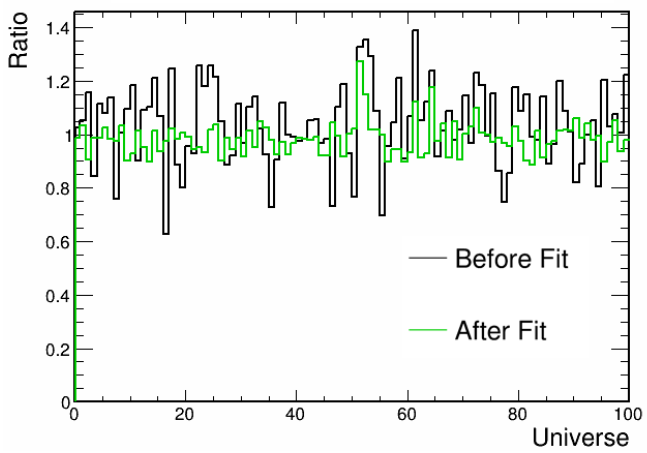

Figure 17.19: Ratio of estimated signal event count to true signal event count over the entire analysis phase space before (black) and after (green) a two-stage fit is done.

The normalizations assigned to the background templates by the two fitting procedures is displayed in Fig. 17.20. The plots show that MINUIT often pushes the normalizations of the templates to the limits allowed, indicating that the fitter is trying to take maximum advantage of what is allowed to get a good fit in the control sample. In general, this is an undesirable situation. One would prefer that the fitter finds the minima lie within the limits.

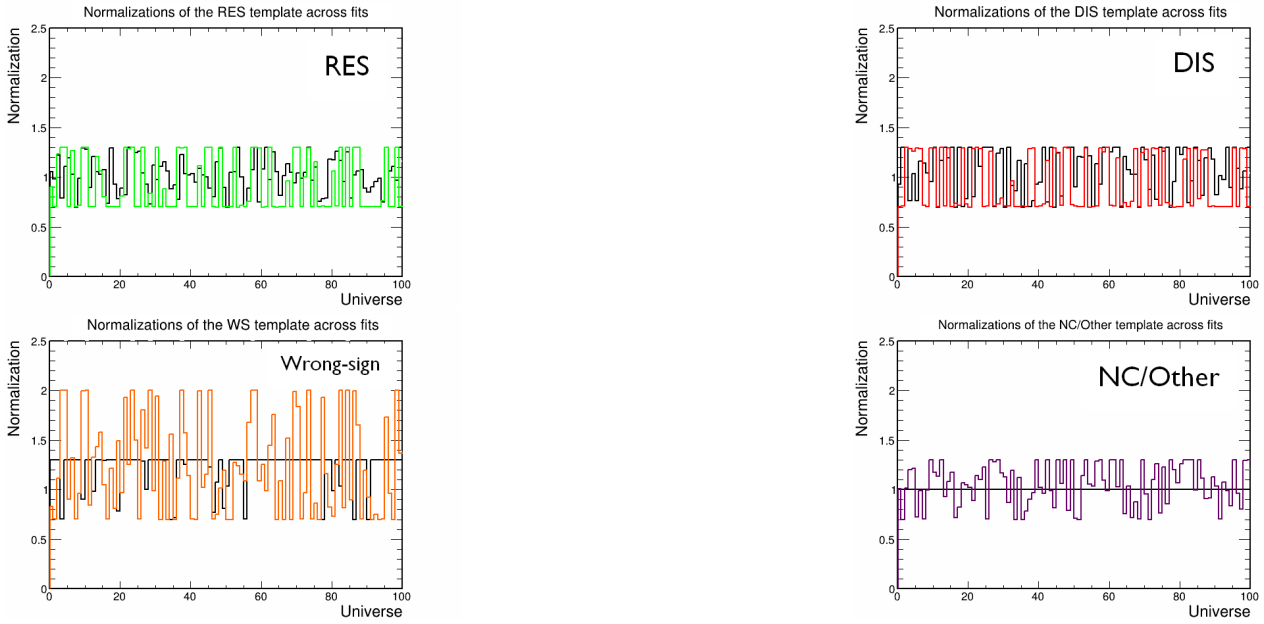

Figure 17.20: Normalizations assigned to the various background templates across the two fits for 100 universes. The solid black lines and the colored lines are the normalizations obtained from the base control region fit and muon out-region fit respectively.

17.1.6 Three-stage fit

The two-stage fitting procedure improves the estimate of the signal. However, there are a number of universes where this fitting procedure does not yield the correct estimate, as evinced by the large ratio values for some of the universes after fit in Fig. 17.19. Upon examining the control sample fits, it is often observed that letting multiple templates float in the same fit may not be the best way to determine the normalizations of the individual templates. The templates can often compensate for each other during the fitting procedure, resulting in normalizations that are not realistic. To better estimate the individual template normalizations, each template should be fit in distinct control sample.

Upon reexamination of the individual background templates in Figs. 17.4 and 17.5, it can be seen that RES is dominant in the original base control sample: 400 MeV $< |\vec{q}| \leq$ 1250 MeV and 100 MeV $< E_{avail} \leq$ 750 MeV. At $E_{avail} >$ 750 MeV and $|\vec{q}| >$ 1250 MeV, RES and DIS interactions have comparable event rates. But NC/Other and wrong-sign have little presence at high available energy. To better understand how to obtain an estimate of the NC/Other and wrong-sign event rates, the muon transverse momentum ($|\vec{P}_t|$) and muon longitudinal momentum ($|\vec{P}_{long}|$) distributions for events with $E_{avail} <$ 100 MeV are examined. The distributions of the different interactions in $|\vec{P}_t|$ and $|\vec{P}_{long}|$ is shown in Fig. 17.21.

In general, the interactions have similar $|\vec{P}_t|$ values, ranging from 150 - 400 MeV. However, the wrong-sign and NC/other templates have much lower $|\vec{P}_{long}|$ than the other templates. The CCQE, RES, 2p2h and DIS templates have $|\vec{P}_{long}|$ values ranging from 0.6 - 2.5 GeV. The wrong-sign events mostly have $|\vec{P}_{long}|$ from 300 - 1500 MeV and the NC/Other templates have $|\vec{P}_{long}|$ values below 500 MeV. This separation of NC/Other and wrong-sign from the other interactions in the $|\vec{P}_{long}|$ variable can be used to constrain the wrong-sign and NC/other normalizations. A boundary drawn at $|\vec{P}_t| = 600$ MeV and $|\vec{P}_{long}| = 400$ MeV roughly separates the NC/Other and wrong-sign contributions from the other interactions.

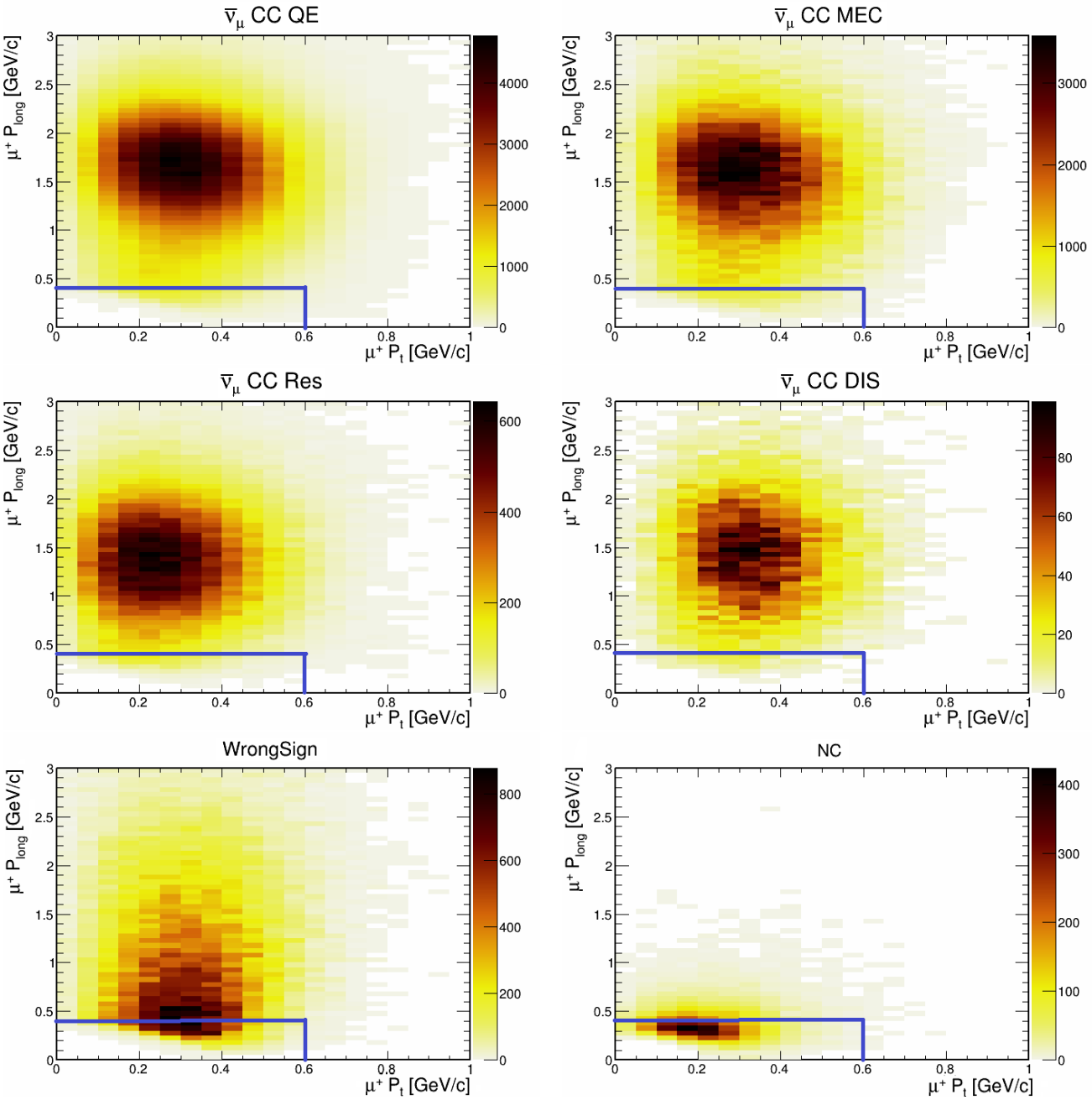

Figure 17.21: Muon transverse momentum vs. muon longitudinal momentum of different interactions with $E_{avail} \leq 100$ MeV. The $\bar{\nu_\mu}$ CC QE, MEC, Res and DIS interactions have higher P_{long} than the wrong-sign and NC/Other interactions.

17.1.6.1 Control samples for the different background templates

With the observations described in this section, the following control samples are defined:

1. $E_{avail} > 100$ MeV and $|\vec{q}| > 1250$ MeV, OR $E_{avail} > 750$ MeV. This will be referred to as the 'outer region'. Only the DIS template will be fit in this control sample.

2. The original base control region: 400 MeV $< |\vec{q}| \leq$ 1250 MeV and 100 MeV $< E_{avail} \leq$ 750 MeV. Only the RES template will be fit in this region.

3. $|\vec{P}_{long}| \leq$ 400 MeV and $|\vec{P}_t| \leq$ 600 MeV with $E_{avail} \leq$ 100 MeV. The NC/Other and wrong-sign templates will be floated in this region.

17.1.6.2 Steps of the fitting procedure

The fit takes place in the order enumerated in Sec. 15.6.1. The first fit takes place in the outer region, in which DIS is floated. An example of this fit is shown in Fig. 17.22. MINUIT varies the normalization of the DIS and retains the scale at which the total MC best matches the data in this control region. This normalization of DIS is retained for the next two fits, and is also used to renormalize the DIS component in the total background estimate. The DIS template is not allowed to vary more than 10% in this fit.

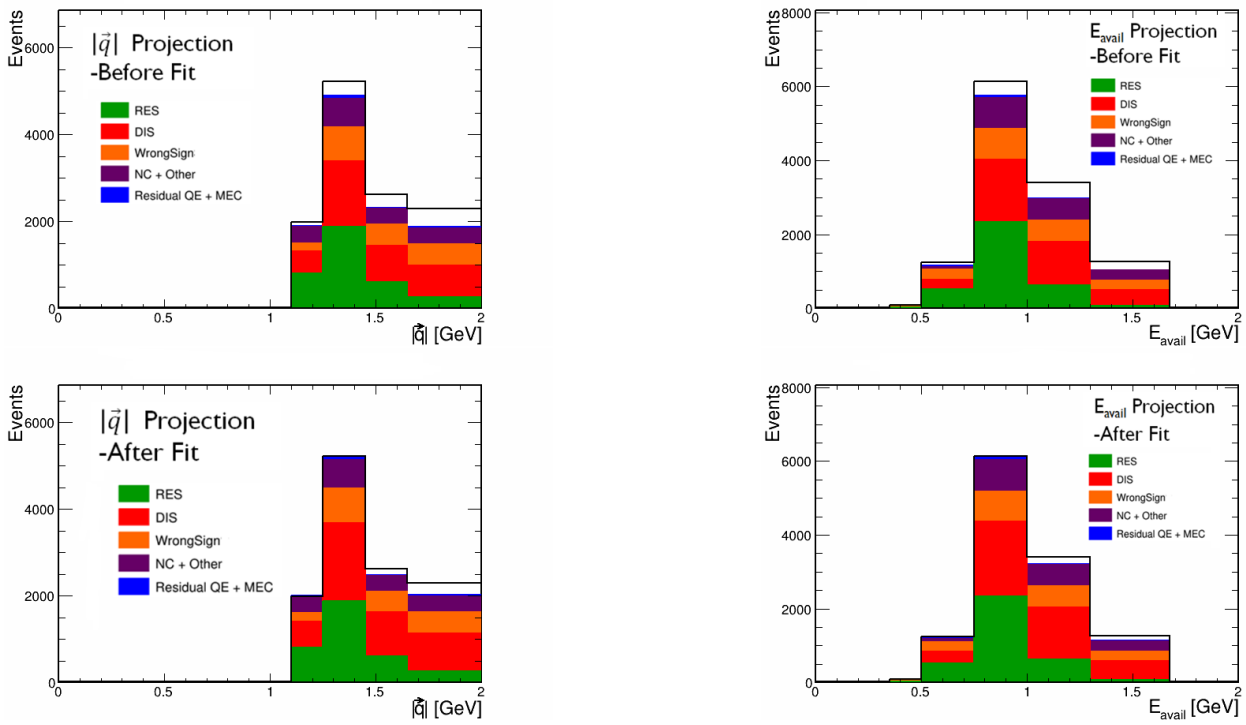

Figure 17.22: Example of a fit done in the outer region. In this fit, only the DIS template is allowed to change normalization.

Now, the second fit performed is in the base control region, shown in Fig. 17.23. The DIS

normalization obtained from step 1 is applied first, and then the RES template is floated by MINUIT over a number of iterations until a good fit between the predicted MC and the data is obtained. This RES normalization is retained for the next fit, and is used to adjust the normalization of the RES component in the total background estimate. The RES template in this fit is only allowed to vary by 20% from the nominal.

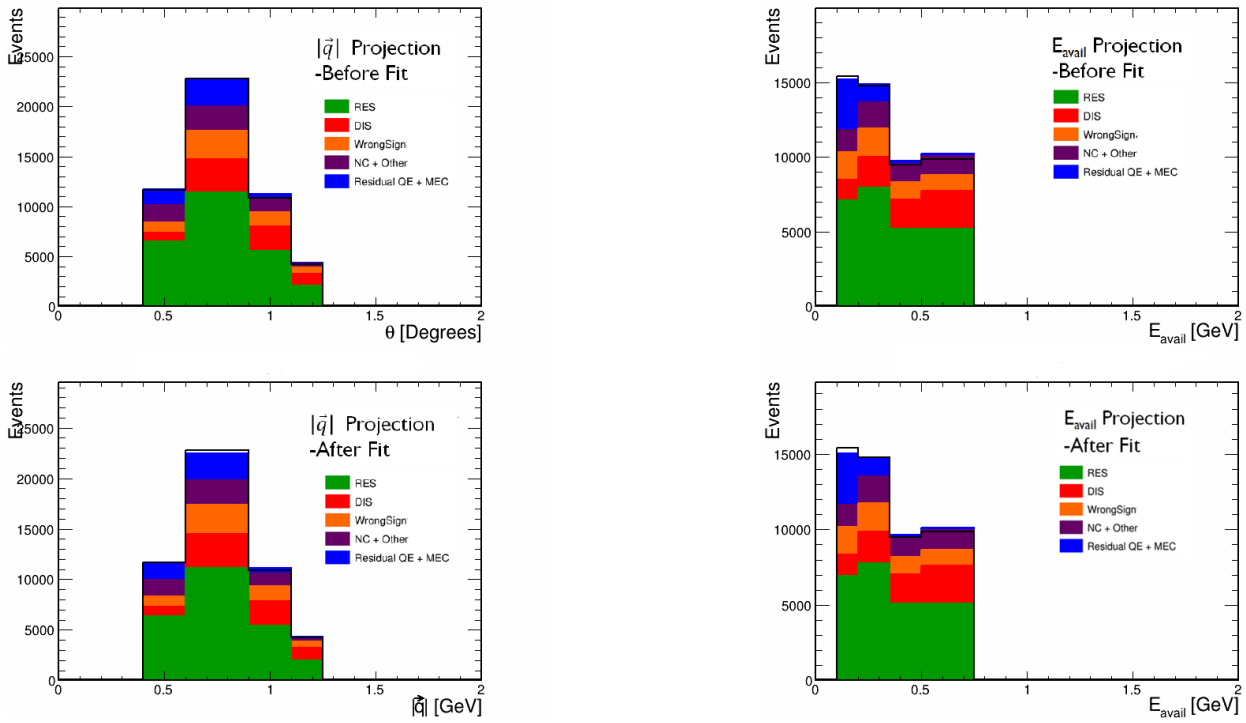

Figure 17.23: Example of a fit done in the base control region. In this fit, only the RES template is allowed to change normalization.

In the final stage of the fit, the data and the predicted MC in the region $P_{long} \leq 400$ MeV and $P_t \leq 600$ MeV in the lowest E_{avail} slice are observed. This is shown in Fig. 17.24. In this stage, the residual signal template, NC/Other and wrong-sign templates are allowed to float. Now, unlike the previous fits, the residual signal must be allowed to float because it has a big contribution in this sample, and any shape and rate change can also arise from systematic shifts in the QE and MEC contributions. No template is allowed to float by more than 20% from the nominal. The NC/Other and wrong-sign normalizations are applied to their respective counterparts in the total background estimation.

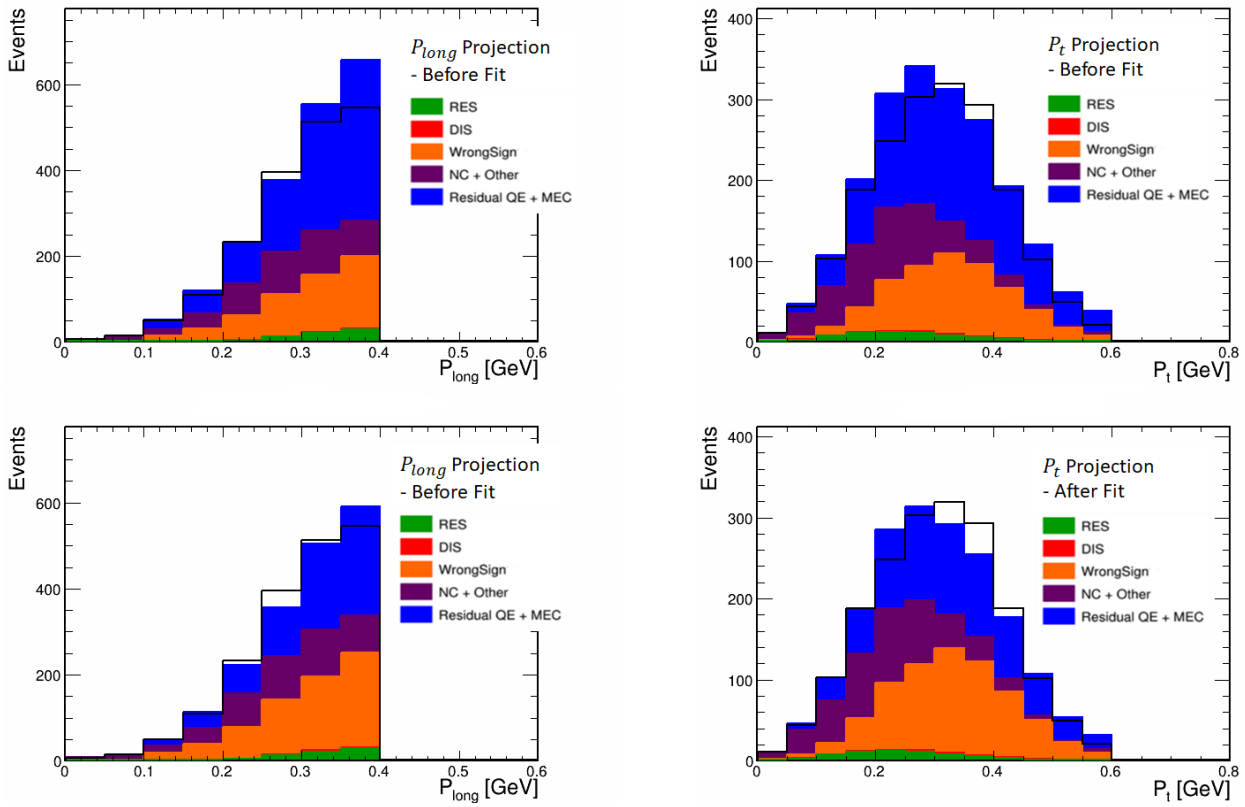

Figure 17.24: Example of a fit done in the muon out-region. In this fit, only the RES template is allowed to change normalization.

Once the four background interactions RES, DIS, NC/Other and wrong-sign have been appropriately normalized with this three-stage fit, the templates are added back together to yield the total background in the CC inclusive sample.

17.1.6.3 Performance metrics for the fitting procedure

Renormalizing the background templates results in the change in the distribution of the total background in the CC inclusive sample. To determine how well this fitting procedure works in yielding the correct signal estimate, the fitting procedure is applied to 100 different CC inclusive samples in which the systematics parameters have been randomly shifted. The ratio of the estimated total signal event count with respect to truth before after a fit is examined for each universe and the values are plotted in the distributions shown in Fig. 17.25. In general, the ratio values after a fit is performed are closer to 1 than before a fit is done. In the distribution of

ratio values, it can be seen that the ratio values of estimated signal to truth after the fit are more nearly peaked at 1.0 when compared to the same distribution of values before the fit. This suggests that the fit improves the estimate of the signal, however the distribution of ratio values is not centered at 1.0.

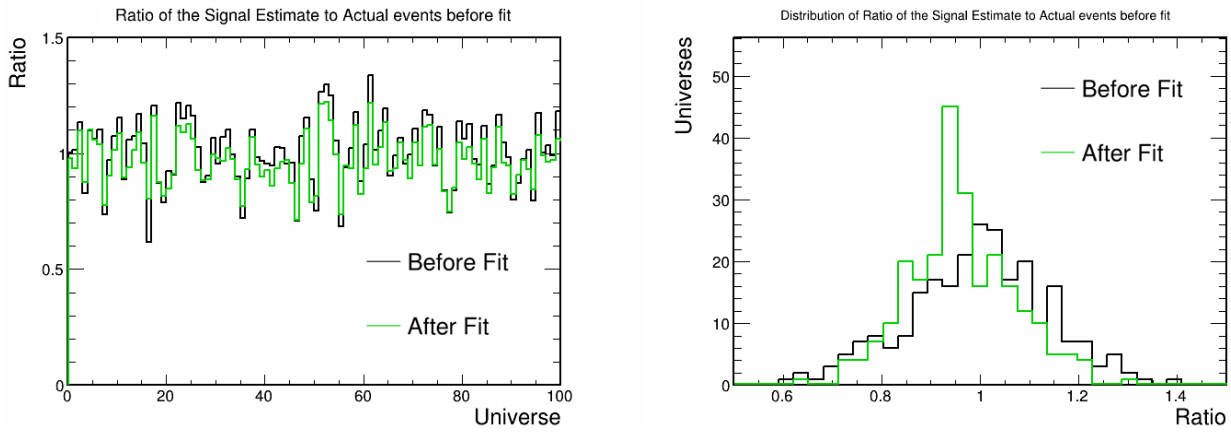

Figure 17.25: Ratio of the estimated total signal event count before (black) and after (green) the three-stage fit. The left plot shows the ratio values for each universe and the right plot shows the distributions of the ratio values before and after a fit.

Another performance metric for this three-stage fitting procedure is the fractional residual of the estimated total signal event count before and after the fit with respect to the truth. Receall that the fractional residual is defined as $(Total\ Signal\ Events - Total\ True\ Events)/(Total\ True\ Events)$. This metric is shown in Fig. 17.26. Ideally, a great fit should reduce this fractional residual value to be close to 0. As observed from the distributions on Fig. 17.26, the fit in general brings the fractional residual value closer to the ideal. The distribution of fractional residual values also peaks more sharply compared to the distribution of values before the fit. The ratio and the fractional residual values indicate that the three-stage fitting procedure improves the estimate of the signal, however a degree if biasing is oberved in the residual distribution.

The normalizations of the different templates assigned by the fit are shown in Fig. 17.27. These plots show that the RES and NC/Other templates take on varied normalization values. On the other hand, the fitting procedure often pushes the normalizations of the DIS and wrong-sign templates to their maximum allowed values.

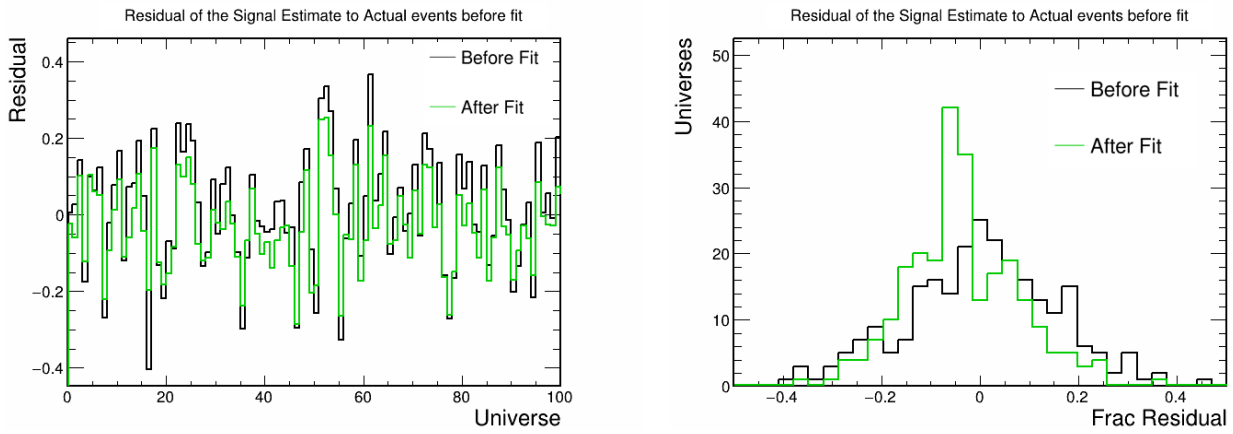

Figure 17.26: Fractional residual of the estimated total signal event count before (black) and after (green) the three-stage fit. The left plot shows the residual values for each universe and the right plot shows the distributions of the values before and after a fit.

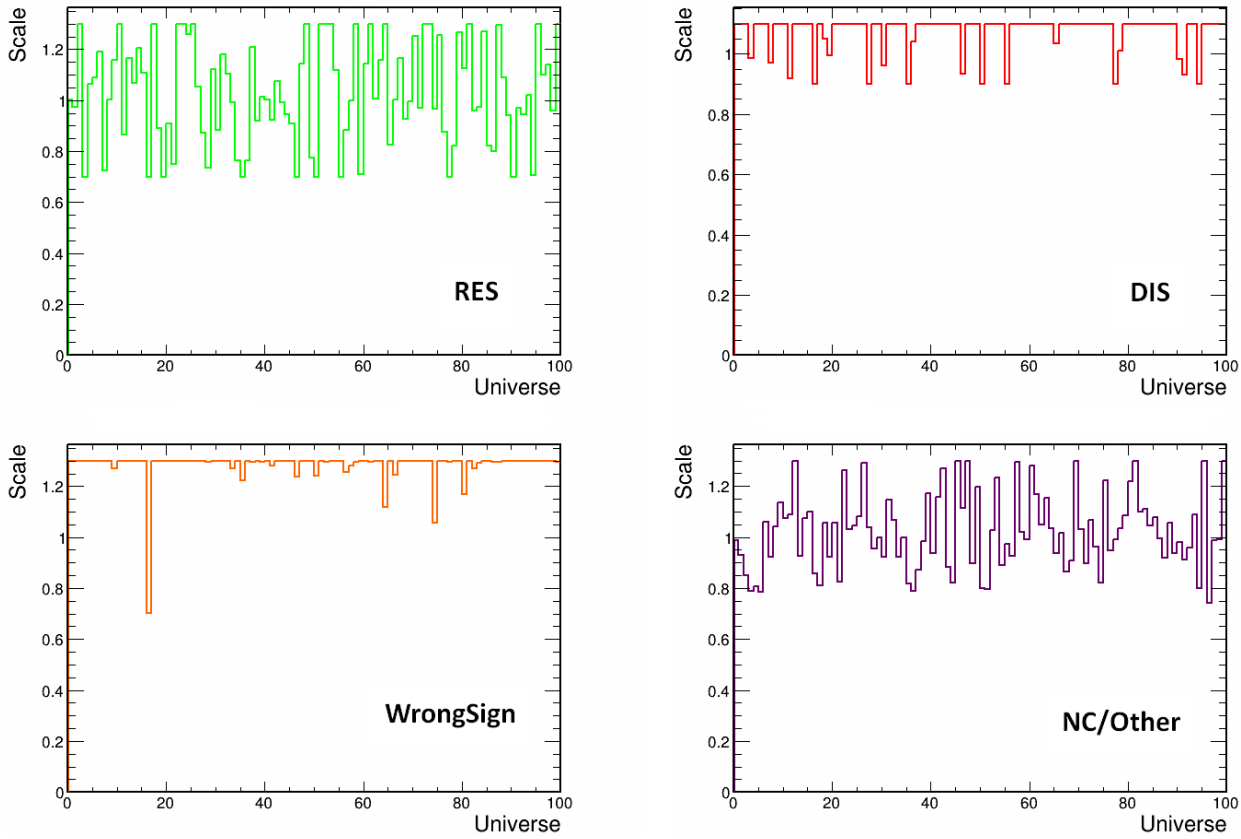

Figure 17.27: Normalizations of the RES (upper left), DIS (upper right), wrong-sign (lower left) and NC/Other (lower right) templates assigned by the three stage fitting procedure for 100 universes.

17.1.7 Renormalizing the total background

Another approach to determining the background involves renormalizing the total background to the total data in a suitable control sample, rather than fitting the individual templates. Looking at the E_{avail} distributions of the different templates in Fig. 17.13, it can be seen that CCQE and 2p2h-MEC interactions are practically nonexistent at E_{avail} greater than 400 MeV, whereas the background interactions still have some presence in this region. Therefore, a control sample can be defined $E_{avail} \geq 400$ MeV. Now, unlike the previous fitting methods described in this Appendix, MINUIT is not used to do a template fit. Rather, the scaling factor which is applied to the background is determined by simply looking at the total predicted MC events and the total data events in the sideband. A scaling factor is then determined:

$$Scaling\ Factor = \frac{Total\ data\ events\ in\ Control\ Sample}{Total\ MC\ events\ in\ Control\ Sample} \tag{17.1}$$

The scaling factor is to be applied to total background estimate spanning the entire phase space. All individual background distributions are scaled by the same value. An example of such scaling in the control sample is shown in Fig. 17.28.

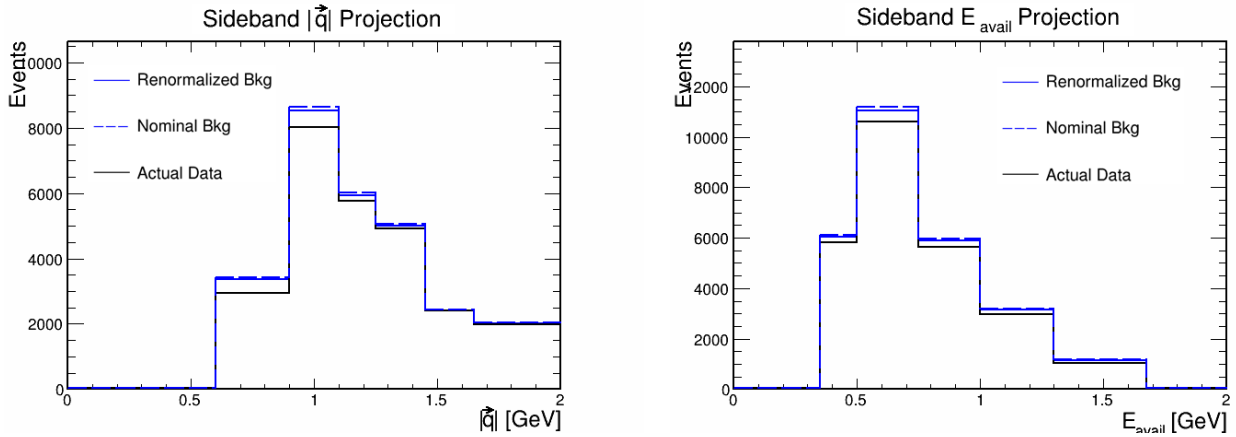

Figure 17.28: Demonstration of background renormalization. The nominal predicted background (dashed blue line) is renormalized to a new prediction (solid blue line) based on the data distribution (solid black histogram) event count according to Eq. (1).

Once the total background has been estimated, the signal distribution can then be obtained by subtracting the CC inclusive sample distribution with this new background. To determine the

performance of this procedure, the ratio of the estimated signal with respect to the truth before and after a fit is done for 100 universes is shown in Fig. 17.29. It can be seen that the ratio of the estimated signal with respect to truth is brought very close to 1.0 after the fit is done. Furthermore, there is little spread in the ratio values that are obtained out of the fit. In fact, this metric demonstrates that this is the most promising way of achieving a good signal estimate out of all the methods discussed in this Appendix. The fractional residuals of the estimated signal with respect to the truth for 100 universes before and after a fit are shown in Fig. 17.30. Ideally, a good estimate should reduce the fractional residual to as close to 0 as possible. This is what is observed consistently across the fits in each of the 100 universes.

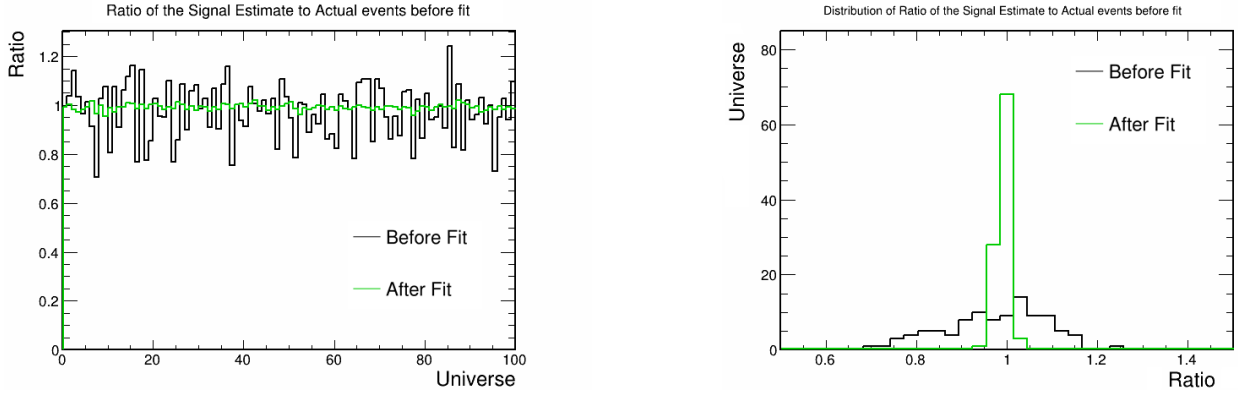

Figure 17.29: Ratio of the estimated signal w.r.t to truth before (black) and after (green) the background normalization in the $E_{avail} > 400$ MeV control sample is carried out. The plots indicate that this method does a very good job of estimating the signal. As indicated by how the after fit ratio values settle around 1.0, with small variance.

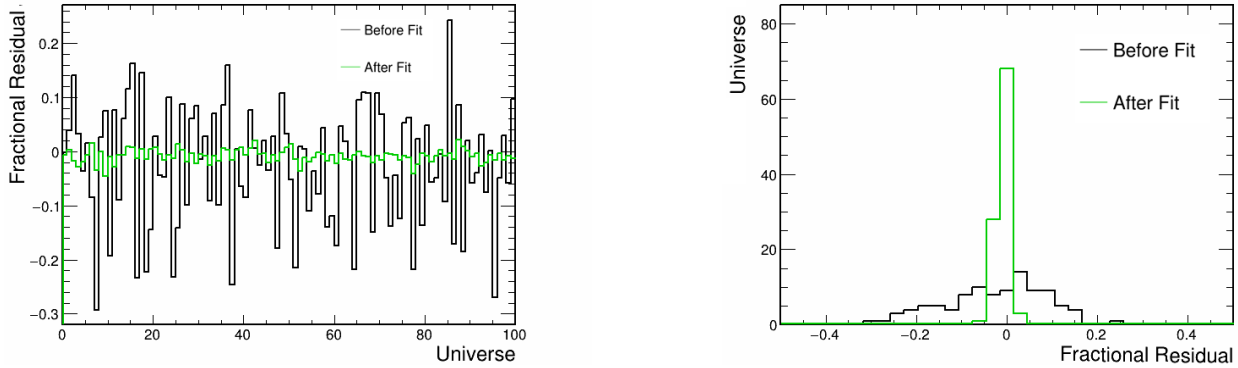

Figure 17.30: Fractional residual of the estimated total signal event count before (black) and after (green) the three-stage fit. The left plot shows the residual values for each universe and the right plot shows the distributions of the values before and after a fit.

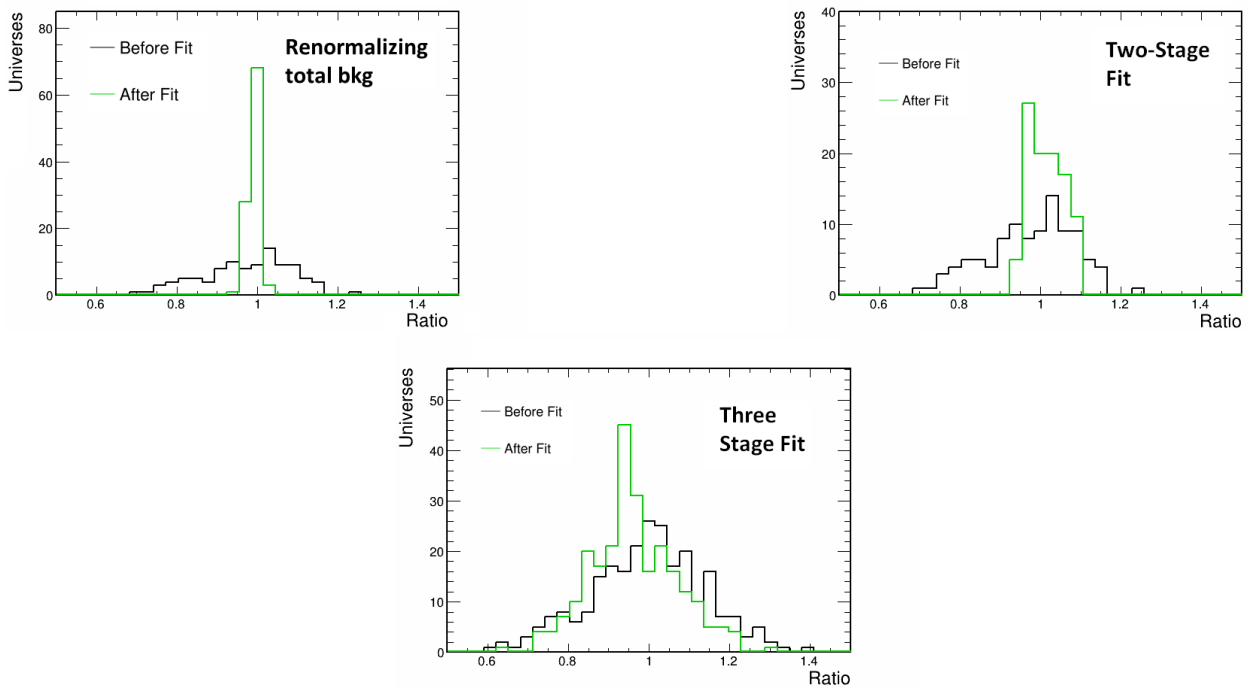

Figure 17.31: Distributions of ratios for estimated signals from the three different fitting procedures, before and after the fits have been performed. The upper left plot shows the ratio values obtained via normalizing the background. The upper right plot shows the values for the two-stage procedure and the bottom plot shows those obtained from the three-stage procedure.

To show how well normalizing the total background in the control sample to the data works, the distributions of ratios before each fitting method are shown side by side in Fig. 17.31. All three distributions after the fit are centered near to 1.0. However, the widths of the distributions of ratio values obtained after the fit are different. The distribution of ratios arising from the two-stage fit has the biggest full-width half maximum (FWHM), 0.15. The FWHM of the distribution of ratios arising from a three-stage fit is 0.06, indicating that floating individual templates in distinct control samples improves the estimate of the signal. The distribution of ratios arising from normalizing the total background to the data at $E_{avail} > 400$ MeV is the sharpest, with a FWHM of about 0.03. These comparisons show that normalization of the total background is the best method for estimating the $\overline{\nu}_\mu$ CCQE plus 2p2h-MEC contribution from the CC inclusive sample.

17.1.8 Final choice of method

In summary, a two-stage fitting procedure involving floating multiple templates in suitable sidebands improves the estimate of the background and the signal in the $\bar{\nu}_\mu$-CC inclusive sample, especially in the region $E_{avail} > 100$ MeV. However, getting a good estimate of the signal for very low E_{avail} often proves difficult, particularly due to the unavailability of a good way to constraint the wrong-sign contribution in the background. A more elaborate method, discussed in Sec. 8, fails to yield a good background estimate. However, simply normalizing the total background to the total data in a control sample $E_{avail} > 400$ MeV yields a good signal estimate, as observed from the metric plots in Figs. 17.29 and 17.30. This method is the one utilized for the estimation of $\bar{\nu}_\mu$ CCQE and CC 2p2h-MEC contributions in the inclusive sample.

References

1. A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), *First measurement of the muon neutrino charged current quasielastic double differential cross section*, Phys. Rev. D **81**, 092005 (2010).

2. P.A. Rodrigues *et al.* (MINERvA Collaboration), *Identification of nuclear effects in neutrino-carbon interactions at low three-momentum transfer*, Phys. Rev. Lett. **116**, 071802 (2016).

3. R. Gran *et al.* (MINERvA Collaboration), *Antineutrino charged-current reactions on hydrocarbon with low momentum transfer*, Phys. Rev. Lett. **120**, 221805 (2018).

4. J. Nieves, I. Ruiz Simo and M.J. Vicente Vacas, *Inclusive charged-current neutrino nucleus interactions*, Phys. Rev. C **83**, 045501 (2011).

5. G.D. Megias, J.E. Amaro, M.B. Barbaro, J.A Caballero, T.W. Donelly and I. Ruiz Simo, *Charged-current neutrino-nucleus reactions within the superscaling meson-exchange current approach*, Phys. Rev. D. **94**, 093004 (2016).

6. M. A. Acero *et al.* (NOvA Collaboration), *Adjusting neutrino models and evaluating uncertainties using NOvA Near detector data*, Eur. Phys. J. C **80**, 1119 (2020); arXiv: 2006.08727.

7. K. Abe *et al.* (T2K Collaboration), *First combined measurement of the muon neutrino and antineutrino charged-current cross section without pions in the final state at T2K*, Phys. Rev. D **101**, 112001 (2020).

8. M. Kakenov, V.I. Kukulin, V.N. Pomerantsev, O. Bayakhmetov, *Properties of dibaryons in nuclear medium*, Eur. Phys. J. A. **56**, 266 (2020).

9. I. Ruiz Simo, J.E. Amaro, M.B. Barbaro, A. De Pace, J. A. Caballero, G.D. Megias, T.W Donelly, *Emission of neutron-proton and proton-proton paris in electron scattering induced by meson-exchange currents*, Phys. Rev. C **94**, 054610 (2016).

10. I. Ruiz Simo, J.E. Amaro, M.B. Barbaro, A. De Pace, J. A. Caballero, T.W. Donelly, *Relativistic model of 2p-2h meson exchange currents in (anti)neutrino scattering*, J. Phys. G **44** 6, 065015 (2017).

11. R. Patrick *et al.* (MINERvA Collaboration), *Measurement of the muon anti-neutrino double-differential cross section for quasi-elastic scattering on hydrocarbon at* $E_\nu \sim 3.5\ GeV$, Phys. Rev. D **97**, 052002 (2018);

12. J. Schwehr, D. Cherdack, R. Gran, *GENIE implementation of IFIC Valencia model for QE-like 2p2h neutrino-nucleus cross section*, arXiv:1601.02038 (2016).

13. S. Dolan, G. D. Megias and S. Bolognesi, *Implementation of the SuSAv2-MEC 1p1h and 2p2h models in GENIE and analysis of nuclear effects in T2K measurements*, Phys. Rev. D. **101**, 033003 (2020).

14. W. Anthony Mann, *Graphs from another world: Neutrino QE, CC(π), 2p2h and RPA a la Valencia*, NOvA docdb-43658 (February 2020).

15. J.E. Amaro, M. B. Barbaro, J.A. Caballero, T. W. Donelly, A. Molinari, and I. Sick, *Using electron scattering to predict charge-current neutrino cross sections in nuclei*, Phys. Rev. C **71** 015501 (2005).

16. K. Bays *et al.*, *Cross-section Tuning 2020 Tech Note*, NOvA Document 43962-v3.

17. P. Adamson *et al.*, *The NuMI Neutrino Beam*, Nucl. Instrum. Meth. **A806**, 279 (2016).

18. S. Mufson, B. Baugh, C. Bower, T.E Coan, J. Looper, L. Corwin, J. A. Karty, P. Mason, M. D. Messier, A. Pla-Dalmau and M. Proudfoot, *Liquid Scintillator production for the NOvA experiment*, Nucl. Instrum. Meth. **799**, November 2015.

19. D. S. Ayers, G. R. Drake, M. C. Goodman, J. J. Grudzinski, V. J. Guarino *et al.*, *The NOvA Technical Design Report*, (2007).

20. Donald H. Perkins. *Introduction to High Energy Physics, Fourth Edition*, 51 - 55. Cambridge: Cambridge University Press (2000).

21. R. Gran, J. Nieves, F. Sanchez and M. J. Vicente Vacas, *Neutrino-nucleus QE and 2p2h interactions up to 10 GeV*, Phs. Rev. D **88**, 113007 (2013).

22. C. Andreopoulos *et al.* (GENIE collaboration), *The GENIE neutrino Monte Carlo generator*. Nucl. Instrum. Methods Phys. Res., Sec. A **614**, 1, 87, (2010).

23. Mark Thomson. *Modern Particle Physics*, 69 - 70. Delhi: Cambridge University Press (2013).

24. B. Behra, *Deep Learning based measurement of the inclusive ν_μ charged-current cross sections*, NOvA docdb 24430-v1, October 2017.

25. N. Raddatz (for the NOvA Collaboration), *Track reconstruction in the NOvA experiment*, Proceedings of the DPF-2011 Conference, 2011.

26. N. Buchanan, L. Cremonesi, C. Johnson, M. Judah, J. M. Paley, P. Singh and L. A. Soplin. *Technical note for Measurement of the Muon-Antineutrino Charged-Current Inclusive Cross Section in the NOvA Near Detector*, NOvA docdb 53925, Feb 2022.

27. S. S. Bashar, *Identification and treatment of background reactions of the $\bar{\nu}_\mu$ CC double differential cross section measurement*, NOvA docdb 46478-v2, July 2020.

28. F. James, *MINUIT Function Minimization and Error Analysis Reference Manual*, CERN Program Library Long Writeup D506.

29. T. Olson and A. Mann, *Plan for measurement of $d\sigma/d|\vec{q}|dE_{avail}$ and estimation of 2p2h contribution using CC ν_μ interactions in the Near Detector*, NOvA docdb 39508, September 2021.

30. Stefan Schmitt, *Data Unfolding Methods in High Energy Physics*, EPJ Web Conf. **137**, 11008 (2017).

31. G. D'Agostini, *A Multidimensional Unfolding Method Based on Bayes' Theorem*, Nucl. Instrum. Methods Phys. Res., Sect. A **362**, 2, 487 (1995).

32. H. B. Prosper and L. Lyons, editors, *Proceedings, PHYSTAT 2011 Workship on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding.*

33. S. S. Bashar, *Cross Section Closure Test for $\bar{\nu}_\mu$ Charged Current Dytman MEC Events on Carbon in the NOvA RHC Sample*, NOvA docdb 37530-v1, May 2019.

34. Shih-Kai Lin, *Note of the Multi-Universe approach for GENIE systematics*, NOvA Document 21635-v2.

35. T. Olson, *Determination of systematic uncertainties for $d^2\sigma/d|\vec{q}|dE_{avail}$ measurement using ν_μ interactions in the Near Detector*, NOvA Document 38676-v2.

36. Kun Il Park, *Fundamentals of Probability and Stochastic Processes with Applications to Communications.* Springer, ISBN 978-3-319-68074-3.

37. M.A. Acero *et. al.* (NOvA collaboration), *Measurement of the Double-Differential Muon-neutrino Charged-Current Inclusive Cross Section in the NOvA Near Detector*, arXiv:2109.12220 [hep-ex].

38. Yu-Ji Shi, W. Wang, Zhen-Xing Zhao, Ulf-G. Meibner, *Towards a heavy diquark effective theory for weak decays of doubly heavy baryons*, The European Physical Journal C **80**, no. 398 (2020).

39. L A. Soplin, *PPFX and NuMu*, NOvA Document 16955-v1.

40. A. S. Meyer, M. Betancourt, R. Gran and R. J. Hill, *Deuterium target data for precision neutrino-nucleus cross sections*, Phys. Rev. D **93**, no. 11, 113015 (2016).

41. M. Martinez-Casales, *MEC Systematic Uncertainties*, NOvA Document 43392-v1.

42. A. Mislivec, *GENIE Skew Reweight and the FSI Tune*, NOvA Document 53973-v1.

43. Kun Il Park, *Fundamentals of Probability and Stochastic Processes with Applications to Communications*, Springer, ISBN 978-3-319-68074-3.

44. T. Olson and T. Mann, *Box Opening and Update for Double-Differential Cross Section and 2p2h Contribution using ν_μ-CC Interactions in the NOvA Near Detector*, NOvA docdb-54939-v2, May 2022.