Measurement of $d^2\sigma/d|\vec{q}|dE_{avail}$ and 2p2h contribution using charged current ν_μ interactions in the NOvA Near Detector

A dissertation submitted by

Travis Olson

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physics

Tufts University

June 2021

Adviser: Prof. W. Anthony Mann

Abstract

This Thesis presents the analysis methods and measurement of the $d^2\sigma/d|\vec{q}|dE_{avail}$ ν_μ-CC inclusive cross section that describes neutrino scattering in a predominantly hydrocarbon medium. Studies that are prerequisite for the analysis including, variable determinations, event selections, efficiency, purity determinations, and systematic uncertainty estimations, are summarized. Tests of the analysis are also detailed, and shown to reproduce the input distributions. The double-differential cross-section measurement provides the foundation for determination of the inclusive rate for the excess induced by 2p2h processes together with nuclear medium effects that are not described by the Fermi gas model. The methodology for the extraction of this rate, defined as the excess observed relative to the rate estimated for known single-nucleon interactions, is described in detail. The following new measurements are reported in this Thesis: (1) The ν_μ-CC inclusive double-differential cross section as a function of three-momentum transfer and available hadronic energy, for an average E_ν of 1.8 GeV, is obtained. The double-differential cross section value that obtains when the

bin widths are taken to be the dimensions of the entire analysis domain, e.g. $0.2 \leq |\vec{q}| \leq 2.0$ GeV/c and $0.0 \leq E_{avail} \leq 2.0$ GeV, is inclusive cross section is $(6.05 \pm 0.75) \times 10^{-39}$ cm^2/GeV/GeV/c/nucleon. An event excess attributed to 2p2h-MEC processes is observed; most of the rate occurs in a contiguous phase space region $0.3 \leq |\vec{q}| \leq 1.0$ GeV/c $0.0 \leq E_{avail} \leq 0.35$ GeV. This rate represents $12.0 \pm 6.5\%$ of the observed CC inclusive cross section; its cross section ratio relative to CCQE scattering is estimated to be $44.3 \pm 23.9\%$.

Acknowledgments

Most of the work presented here comes as the result of the support of many people who have encouraged me over the years. Firstly, I would like to thank my parents and my brother for always supporting and encouraging me during my studies and research.

I would like to thank Professor Anthony Mann who was instrumental in mentoring me through my entire time in graduate school. I would also like to thank Professor Hugh Gallagher for all of his very useful input into many aspects of this analysis. The rest of the Tufts neutrino group, especially Jeremy Wolcott, Joao Coelho, and Nate Mayer were extremely helpful in my research, with providing guidance on mastering the Root framework and other analysis tools.

I would like to thank my collaborators of the NOvA experiment. Everyone involved worked hard to make the experiment and the analysis presented in this Thesis a reality. I would like to specifically thank the members of the Near Detector analysis working group, especially Jon Paley, Mat Muether, and Linda Cremonesi for all of the helpful comments over the years. Greg Pawloski, Jeff Nelson, and Aaron Mislivec provided invaluable feedback to this work as my analysis committee.

Finally, I want to thank all of my friends in the Tufts Physics department: Many thanks to Ozgur and Jon for our numerous discussions about physics during lunch breaks. Many thanks also to Katie, Josh, Saqib, and Alec for all of the good times we spent together over the years.

Contents

List of Tables

List of Figures

Chapter 1

Rationale for this Thesis

Detailed knowledge of neutrino cross sections is essential to precision neutrino oscillation experiments because it enables accurate predictions for neutrino interaction rates. In ν_μ-induced charged-current events at the NOvA Near Detector an excess event rate, relative to a prediction based on known exclusive channels, is observed as shown in Fig 1.1. Previously a similar excess was seen in the scattering of electrons on carbon 12 [2]. This extra rate in electron scattering is attributed to 2-particle 2-hole (2p2h) nuclear medium effects. In 2p2h processes two nucleons are ejected from a struck nucleus, leaving behind two unoccupied nucleon states. In neutrino scattering, the 2p2h enhancement mainly increases the rate of quasielastic-like events; the extra rate into other processes is predicted to be small [3, 4]. Theoretically, 2p2h receives contributions from virtual meson exchange (MEC) diagrams and so the phenomenon is also referred to as 2p2h-MEC. Currently MINERvA has obtained the only direct measurement of the 2p2h rate enhancement from neutrino scattering on carbon [5]. The measurement performed in this Thesis expands the range of neutrino energies and kinematic phase space for which there is an experimentally measured cross section. The new cross section information obtained by this Thesis will provide guidance for the development of more refined models of 2p2h-MEC neutrino scattering.

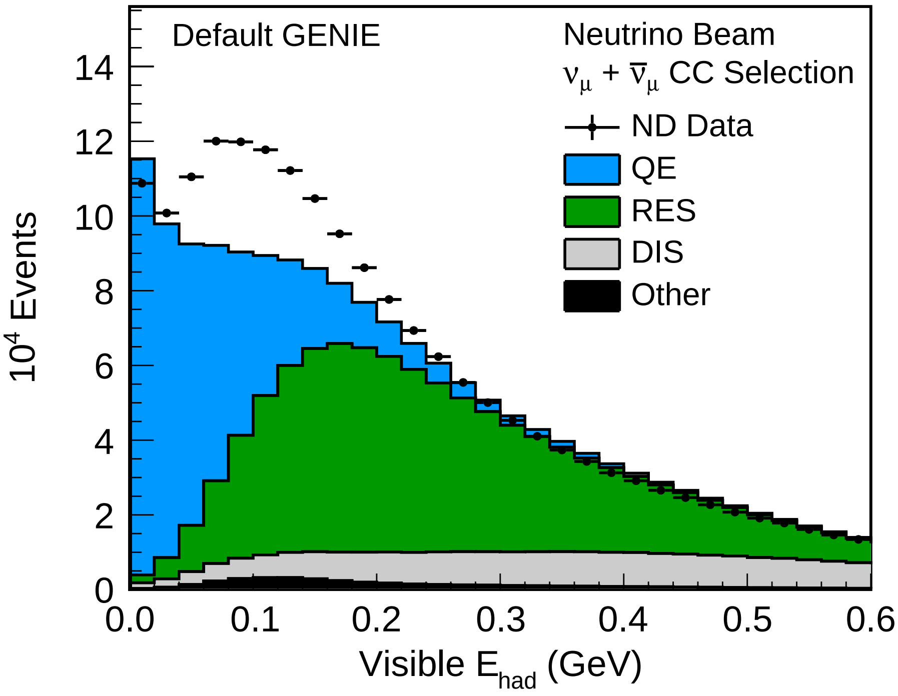


Figure 1.1: Distribution of visible hadronic energy, for data (black crosses) and for simulation that does not include 2p2h processes (solid colors). The data exhibits a clear excess over the simulation indicating the presence of a 2p2h nuclear medium cross section enhancement. (Ref. [6])

Chapter 2

Neutrino interaction with multi-nucleon systems: the 2p2h process

Two-particle two-hole (2p2h) interactions that, theoretically, likely involve meson-exchange-current (MEC), were first proposed as an explanation for the apparent excess event rate observed "dip region" in electron-nucleus scattering. The 2p2h process for electron-nucleus scattering proceeds as follows:

$$e^- + \mathscr{A}_A \rightarrow e^- + N + N' + \mathscr{A}_{A-2}, \tag{2.1}$$

where N and N' are nucleons, either protons or neutrons, that are ejected from the struck $\mathscr{A}$ nucleus of nucleon number A. Figure 2.1 gives a schematic characterization of the inclusive $e^- - \mathscr{A}$ scattering cross section versus the electron energy transfer [7]. The dip region lies between the quasi-elastic peak (solid line) and the Δ-resonance peak (dashed dotted line).

The quasielastic (QE) and the $\Delta(1232)$ peaks are well-described by most models, however many models falter when trying to predict the dip region. By including a 2p2h-MEC process some models significantly improve their agreement with the data [8], an example of which is shown in Fig. 2.2.

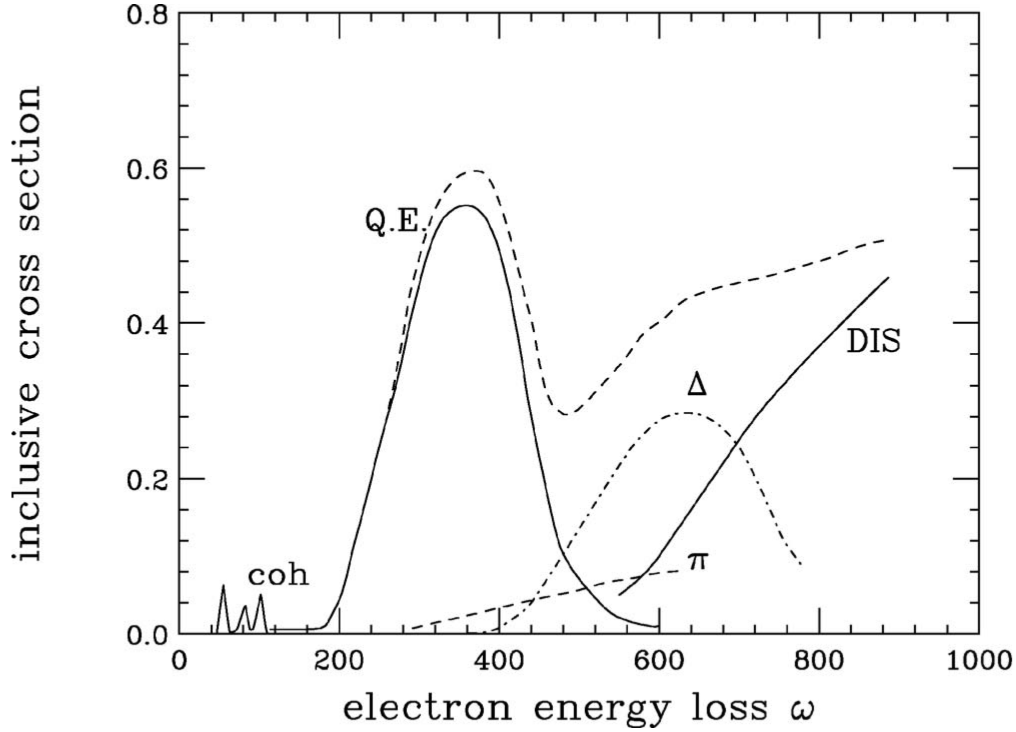


Figure 2.1: Electron scattering cross section verses electron energy loss. The dip region of interest can be seen between the QE peak (black line) and the Δ resonance (Ref. [7]).

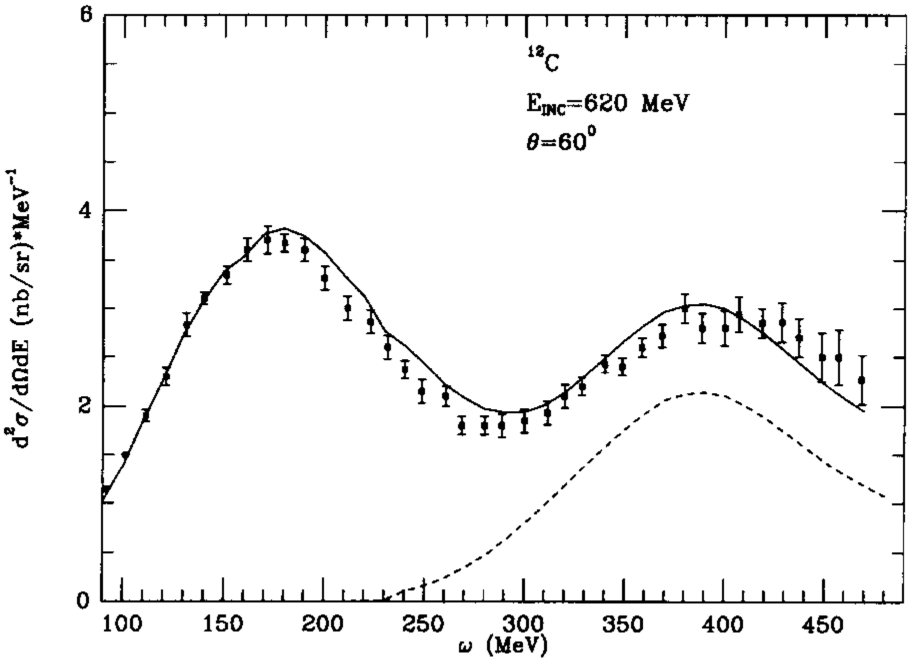


Figure 2.2: Electron scattering model with 2p2h-MEC compared to data. The dip region is more filled in than is the case for predictions without 2p2h-MEC (Ref. [8]).

While it was known for some time that this process was necessary to account for electron-nucleus scattering, 2p2h-MEC in neutrino-nucleus scattering was ignored until proposed by Martini *et al.* of the Lyon group. [9] in response to an excess of quasi-elastic events seen by the MiniBooNE experiment [10]. The MiniBooNE cross-section excess is shown in the upper plot of Fig. 2.3. The measured cross section exceeds the prediction based on neutrino-nucleon scattering in a Fermi gas nucleus, $\nu_\mu + n \to \mu^- + p$, with the axial vector mass of the axial dipole form factor set to $M_A = 1.03$ GeV (red dashed line). The measured data exceed the prediction by at least one standard deviation in all bins. The cross section prediction based on M_A of 1.35 GeV (blue solid line) aligns better with the data at higher E_ν but an excess still remains at lower E_ν. Figure 2.4 shows a comparison of the Lyon model to the MiniBooNE data, the inclusion of a np-nh model, where n=2 gives the dominant contribution, greatly improves agreement over the conventional description of quasielastic scattering on free nucleons.

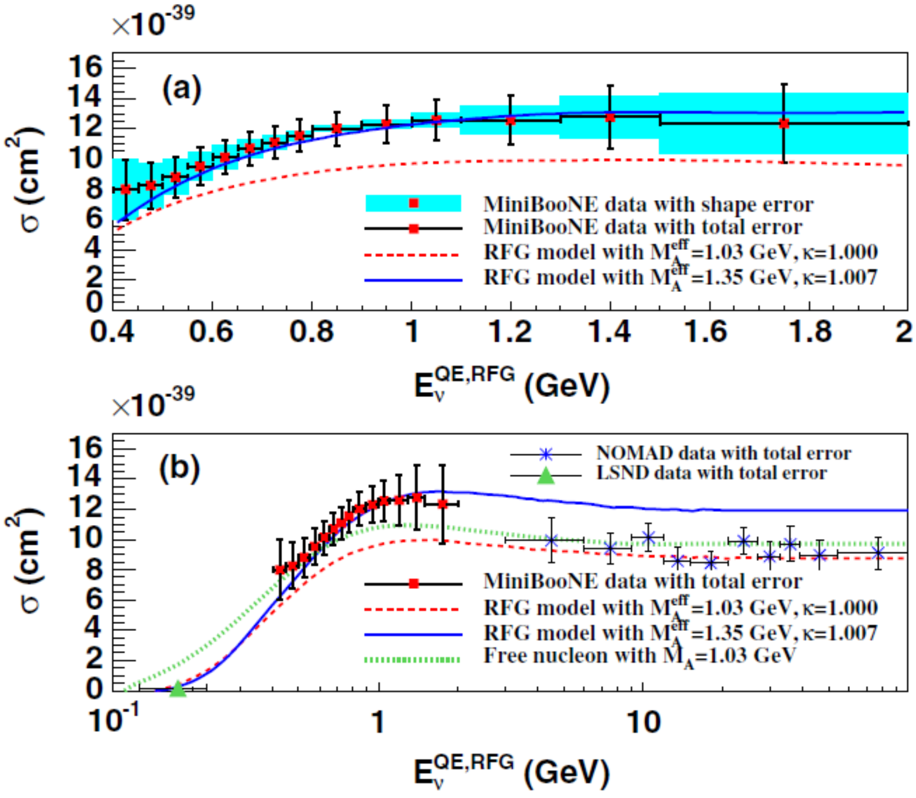


Figure 2.3: The quasielastic cross section versus neutrino energy measured by MiniBooNE (top); comparison with other experiments shown in an expanded neutrino energy range (bottom). There is an excess over the cross section predicted with $M_A = 1.03$ GeV (red) and even over a prediction with an axial mass value of 1.35 GeV (Ref. [10]).

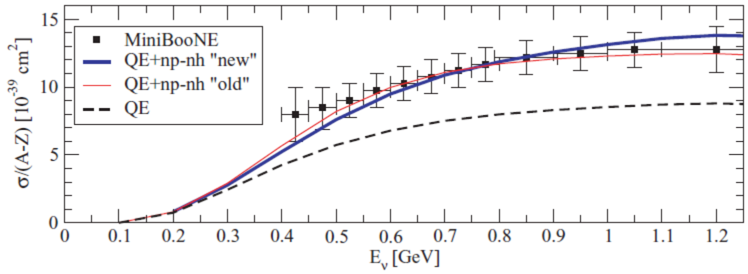


Figure 2.4: The QE + MEC cross section prediction by Martini *et al.* compared to MiniBooNE data. When np-nh cross sections are included the mode aligns much better with the MiniBooNE data. (Ref. [9]).

Subsequently other models that included 2p2h-MEC processes also obtained improved agreement with the MiniBooNE QE results [10] as well as with a later MiniBooNE neutral current elastic result [11, 12, 13, 14, 15, 16].

2.1 Models of 2p2h-MEC interactions in neutrino scattering

Currently there are several models of neutrino-induced 2p2h-MEC that are available to the NOvA neutrino experiment. Neutrino-induced 2p2h-MEC is defined as follow:

$$\nu_\mu + \mathscr{A}_A \rightarrow \mu^- + p + N + \mathscr{A}_{A-2}, \tag{2.2}$$

where N is either a proton or a neutron. These include but are not limited to the Valencia model [4, 13], version 2 of the SuSa model that incorporates a phenomenological prescription called superscaling [17, 18], plus the empirical models available in the GENIE [19] and GiBUU [20] neutrino event generators.

2.1.1 Valencia model

The Valencia 2p2h-model is a microscopic, semi-relativistic model developed by J. Nieves *et al.* The model uses a many-body expansion of the gauge boson self-energy tensor using QFT formalism [4]. The range of the MEC process in $|\vec{q}|$ of this model is limited to below 1.2 GeV/c. A comparison of the Valencia 2p2h-model with the Lyon calculation and the MiniBooNE results is shown in Fig. 2.5 (left). The Valencia QE plus 2p2h model (solid green line) undershoots that MiniBooNE data at low E_ν and overshoots the MiniBooNE data at high E_ν. Figure 2.5 (right) shows the Valencia prediction for all 2p2h processes, including but not limited to those from MEC scattering, versus $|\vec{q}|$ and q_0 [21]. The predicted cross section peaks between 0.5 GeV/c and 0.7 GeV/c in $|\vec{q}|$ and 0.35 and 0.45 GeV in q_0. There is also a secondary peak around 0.4 GeV/c of $|\vec{q}|$ and 0.2 GeV of q_0. The prediction has rate above 1.2 GeV/c in $|\vec{q}|$ due to the inclusion of non-MEC 2p2h processes such as Δ absorption in the nucleus resulting in two nucleons being ejected.

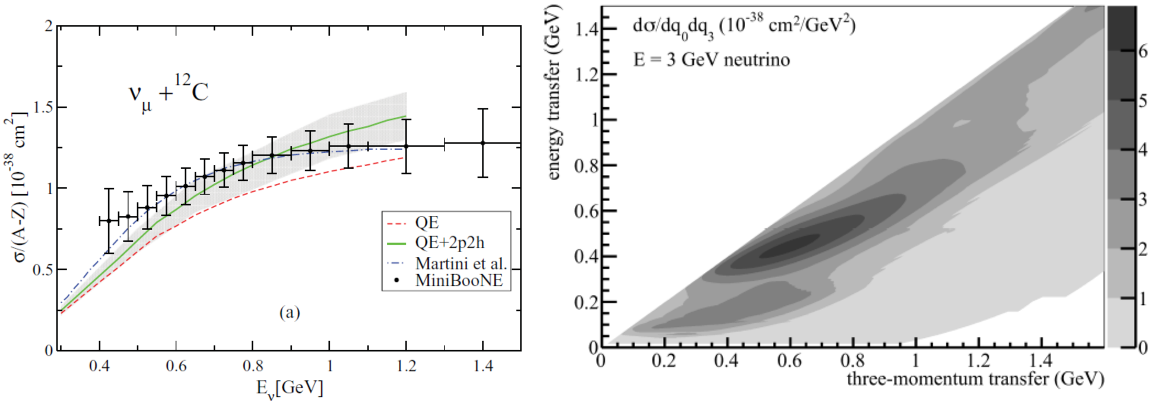


Figure 2.5: Left: Valencia cross section compared to Lyon model and to MiniBooNE results versus E_ν. The Valencia model tends to undershoot the MiniBooNE data at very low E_ν and overshoots slightly at E_ν above 1.0 GeV (Ref. [22]). Right: Valencia prediction for the 2p2h cross section versus $|\vec{q}|$ and q_0 for E_ν of 3 GeV. The cross section peaks between 0.5 GeV/c and 0.7 GeV/c in $|\vec{q}|$ and 0.35 and 0.45 GeV in q_0 (Ref. [21]).

2.1.2 The Super-Scaling version two model

The Super-Scaling (SuSa) model is based on a phenomenological scaling algorithm first established in electron-nucleus scattering data [23]. The model has been modified to describe neutrino-nucleus scattering and further refined to include MEC processes [24, 18]. Figure 2.6 shows the SuSa prediction for the 2p2h-MEC cross section versus neutrino energy (left) and versus three momentum and energy transfer (right). The SuSa cross section, compared to the GENIE-v3

Empirical 2p2h-MEC and the Valencia model, is larger than both overall, has a steeper rising edge than the Valencia model, and reaches a higher constant cross section than does the GENIE empirical model.

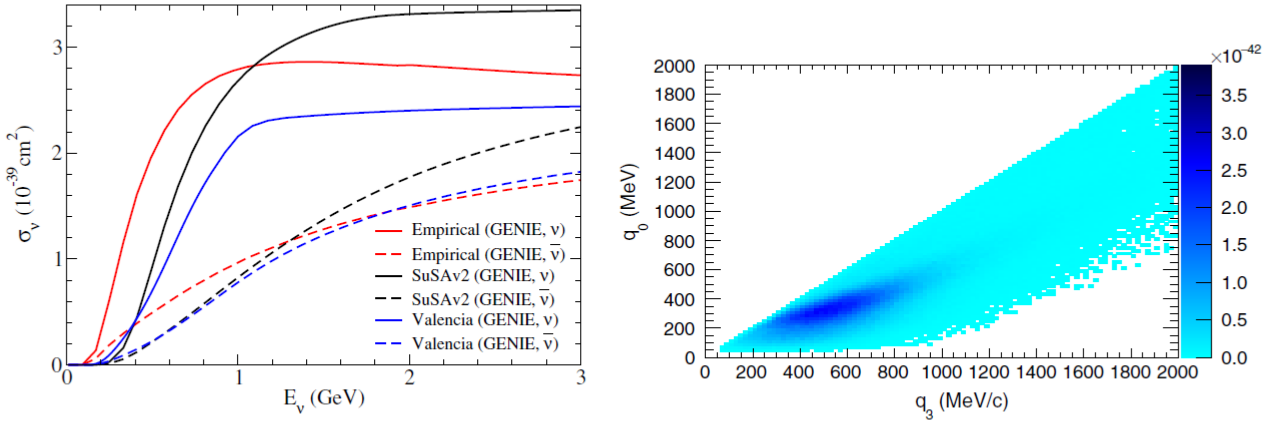


Figure 2.6: Left: SuSa 2p2h-MEC cross section compared to the Valencia 2p2h-MEC model and GENIE 3 empirical 2p2h; Right: the SuSa 2p2h-MEC cross section in $|\vec{q}|$ and q_0. The SuSa model peaks at a higher E_ν than the other models and also plateaus to a larger cross section. The bulk of the cross section lies between $200 \leq |\vec{q}| \leq 800$ MeV/c and $200 \leq q_0 \leq 500$ MeV. (Ref [18]).

2.1.3 GiBUU

The GiBUU generator uses quantum kinetic transport theory. For the GiBUU treatment of 2p2h-MEC, a structure function obtained from electron scattering data is adapted for neutrino scattering [20].

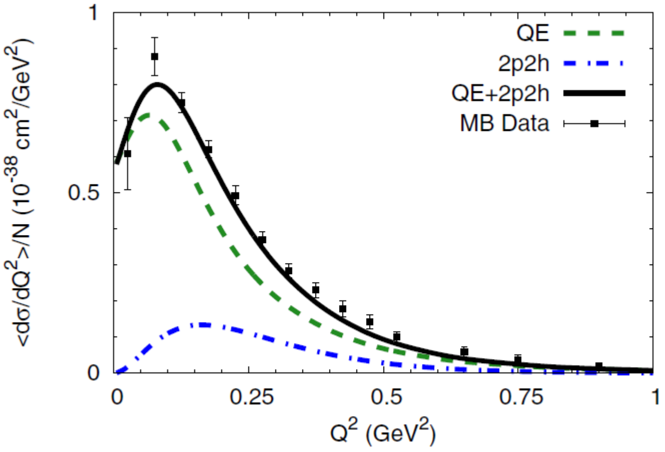


Figure 2.7: GiBUU 2p2h-MEC cross section compared to MiniBooNE data in Q^2. The GiBUU prediction under predicts the data at the peak but otherwise shows good agreement. (Ref. [20]).

Figure. 2.7 compares the GiBUU 2p2h-MEC prediction to the MiniBooNE QE-like differential cross section in four-momentum transfer squared, Q^2. The data Q^2 distribution is under-predicted in the peak at 0.1 GeV^2, and there is also a mild under-prediction between 0.35 and 0.45 GeV^2.

2.2 Experimental measurements of 2p2h

Currently the premier measurements of 2p2h-MEC processes are those carried out by the MINERvA collaboration for both ν_μ-nucleus and $\overline{\nu}_\mu$-nucleus scattering [25, 26]. The MINERvA measurements were carried out using a mostly carbon detector with an event sample with an average E_ν of 3.9 GeV. Figure 2.8 compares the MINERvA $\nu_\mu - CH$ measurement to the predicted neutrino cross section from the Valencia model. The MINERvA results show a clear excess over the Valencia prediction which is especially prominent between $|\vec{q}|$ of 0.4 and 0.6 GeV/c.

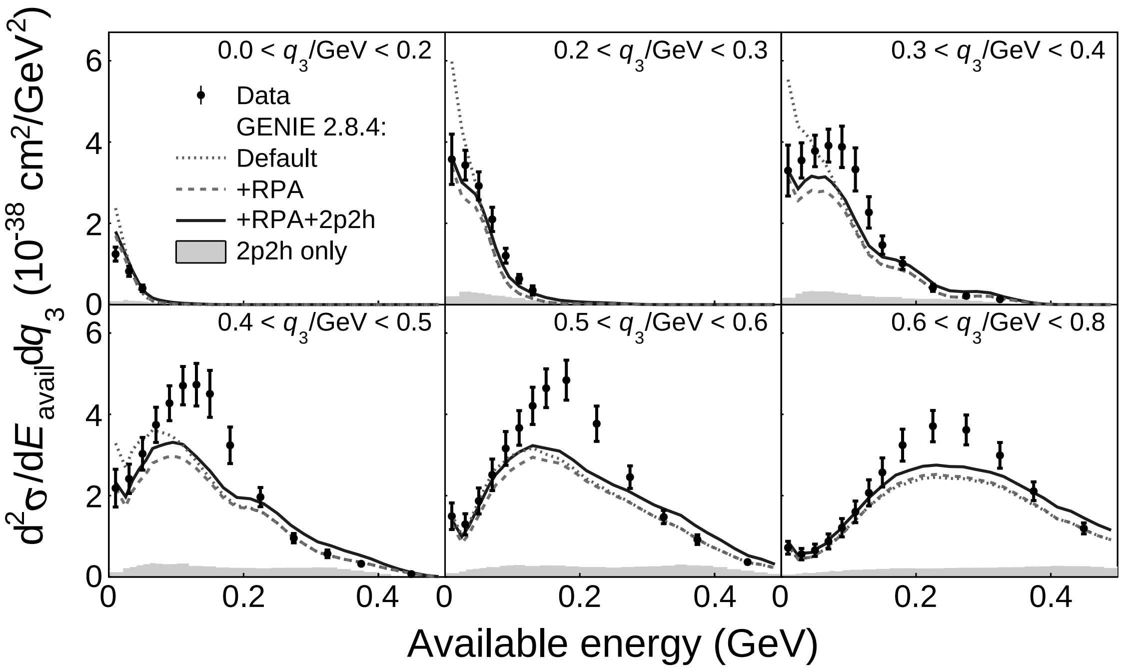


Figure 2.8: Cross section versus available energy in bins of |q|, reported by the MINERvA experiment. (Ref. [25]).

The T2K collaboration has also made measurements of nuclear effects that are driven by 2p2h-MEC processes [27].

2.3 The measurement of this thesis

Our measurement of the inclusive $\nu_\mu + A$ cross section compliments that obtained by the MINERvA at a higher mean neutrino energy 3.9 GeV. The MINERvA measurement is based on 74,949 selected signal events [5], while this measurement uses 394,101 selected signal events. Our inclusive cross section will enable new comparisons with the predictions of current models for the CC inclusive cross section. Such measurements are important for improving model descriptions that pave the way for higher precision in neutrino oscillation measurements. Our measurement covers the region in E_ν from 0.8 to 3.2 GeV, the region most relevant to the NOvA neutrinos oscillation measurements. This region lies above the sub-GeV region of E_ν analyzed by T2K [27], while being moderately below the 2 to 6 GeV examined by MINERvA. It covers the lower half of the high-flux plateau in the ν_μ energy spectrum planned for the DUNE experiment [28].

Chapter 3

The NOvA experiment

The NuMI Off-axis ν_e Appearance (NOvA) experiment is a long baseline neutrino oscillation experiment designed to measure the $\nu_e(\overline{\nu}_e)$ appearance rate from a mostly $\nu_\mu(\overline{\nu}_{mu})$ beam. The NOvA experiment uses the Neutrino Main Injector (NuMI) beam at Fermi National Accelerator Laboratory (Fermilab) as a source of neutrinos for the experiment and two detectors that are constructed to operate in a similar fashion. The detectors are made up of PVC cells that are filled with liquid scintillator, which are segmented, to detect the passage of charged particles. The Near Detector (ND) has an active scintillator volume of mass 193 metric tons and a downstream "muon catcher" which brings the total detector mass to approximately 300 tons. It is located 1 km from the start of the NuMI beam. The Far Detector (FD) is 14 kilotons is 810 km from the start of the NuMI beam. Together the detectors measure the neutrino energy and neutrino flavor composition of the NuMI beam before and after neutrino oscillations have occurred. Both the Near Detector and the Far Detector are located 14 mrad off axis of the primary direction of the NuMI beam line. The NuMI beam uses two magnetic horns to create a focused beam of neutrinos. The NuMI beam can operate in either forward horn current mode (FHC) in which the beam consists primarily of neutrinos, or in reverse horn current mode (RHC) which gives a primarily antineutrino beam.

3.1 Fermilab accelerator complex

Fermilab's accelerator complex is services several particle beams that are made for use by high energy physics experiments. A diagram of the accelerator complex is shown in Fig. 3.1.

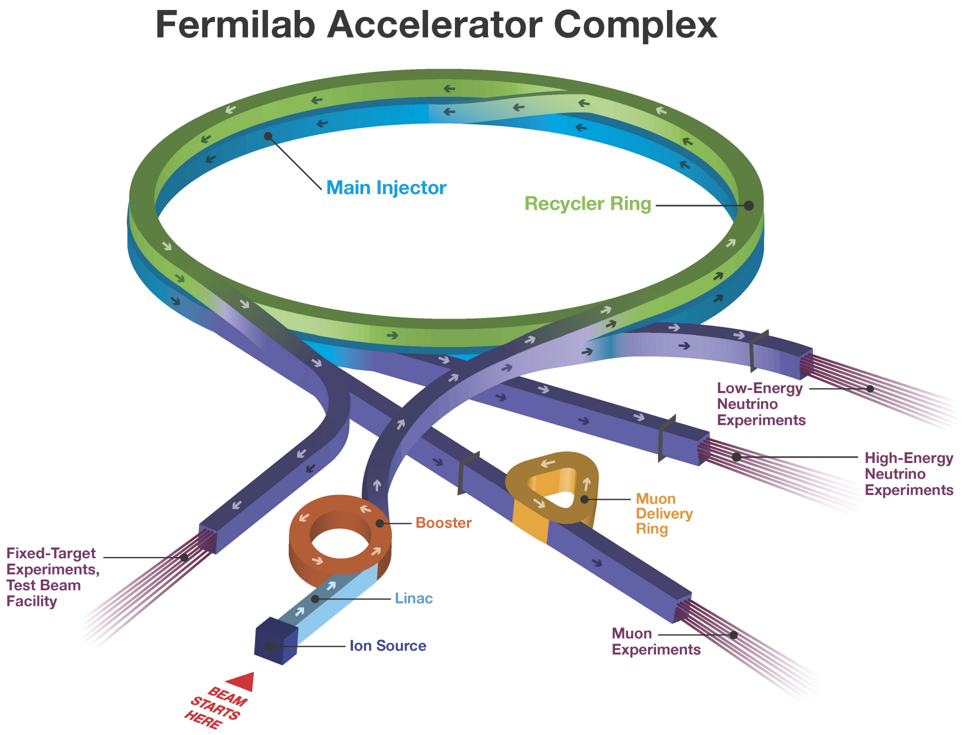


Figure 3.1: The Fermilab accelerator complex. The protons start at the Linac. The Linac feeds into the booster, which then feeds into the recycler. The protons then head into the main injector, pictured under the recycler in light blue. From the main injector the protons are directed into the NuMI beamline labeled "High-Energy Neutrino Experiments." (Ref. [29]).

The accelerator complex starts with hydrogen atoms at the ion source, shown in the bottom middle in Fig. 3.1. The electrons are removed from the atoms, then the remaining protons are accelerated to 0.4 GeV in the Linear Accelerator, which is the straight section immediately downstream of the Ion Source. The accelerated protons then enter the Booster Ring, where the protons are accelerated up to 0.8 GeV. From the booster Ring the protons enter the Recycler, the ring in the center of the diagram, which combines the protons into batches to create a more intense, higher flux, proton beam. The protons enter the Main Injector synchrotron accelerator, under the recycler (light blue ring), which accelerates the protons to 120 GeV. Finally the protons are extracted in batches and are directed onto a graphite target to produce neutrino beams for experiments. The high energy neutrino beamline in the diagram is the one used by NOvA.

3.2 NuMI Neutrino Beam

The source of neutrinos for the NOvA experiment is the NuMI beam [1]. Figure 3.2 shows the various beam elements that comprise the NuMI neutrino beam. The beam contains six major components, the target, the magnetic horns, the decay pipe, the hadron monitor, the absorber and the muon monitors. The 120 GeV protons from the accelerator complex are directed onto a graphite target during a beam spill.

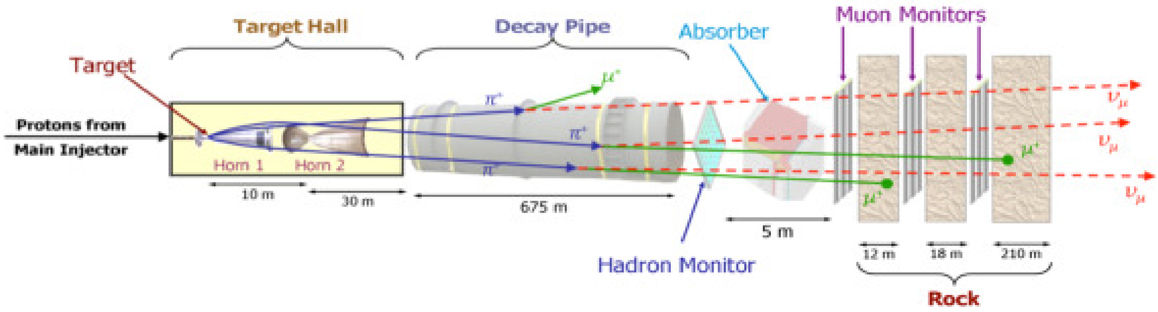


Figure 3.2: The NuMI neutrino beam. Protons from the Main Injector collide with the target to make hadrons. The hadrons are focused in the Target Hall. Then the Hadrons decay to neutrinos used by experiments. (Ref. [1]).

Each beam spill is 10 μs long and there is typically 1.3 seconds between each spill. The spill is then segmented into bunches as displayed in Fig. 3.3. The beam being bunched in time this way allows for easier discrimination between beam neutrinos and cosmic ray background based on timing. There are approximately 4.8×10^{13} protons impinging upon the target per beam spill. The proton interactions on the graphite target result in the production of charged mesons which get focused into a beam by the magnetic horns.

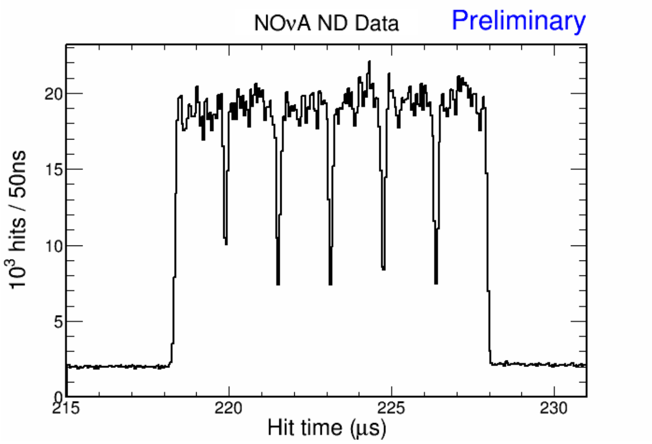


Figure 3.3: Time profile of the proton bunches. The protons are bunched into five 2 μs bunches over a 10 μs spill.

The horns are designed to select either positively or negatively charged $\pi^{\pm}$ and $K^{\pm}$ mesons to focus into a beam while deflecting the mesons of other charge, based on the sign of the current going through them. The current can be reversed to select oppositely charged mesons. This results in neutrino beam that is either primarily made up of neutrinos or antineutrinos depending on the horn current. The focused beam of charged mesons then enters the decay pipe, which is 675 m and filled with helium. In the decay pipe the charged mesons decay into charged leptons and their associated neutrinos:

$$\pi^{+} \to \mu^{+} + \nu_{\mu}, \quad \pi^{-} \to \mu^{-} + \overline{\nu}_{\mu}, \tag{3.1}$$

$$K^{+} \to \mu^{+} + \nu_{\mu}, \quad K^{-} \to \mu^{-} + \overline{\nu}_{\mu}. \tag{3.2}$$

After traversing the decay pipe the beam goes through the hadron monitor, the absorber, the muon monitors, and then 240 meters of rock. The remaining hadrons and muons are absorbed by the absorber and the rock, which leaves the beam to be pure neutrinos.

The neutrino flux at the NOvA near detector for both FHC (left) and RHC (right) beam modes is shown in Fig. 3.4. The FHC flux is made up primarily of muon neutrinos with the second largest component being anti-muon neutrinos, while the RHC flux is mostly anti-muon neutrinos followed by muon neutrinos. The electron and anti-electron neutrino components are small for both horn current modes.

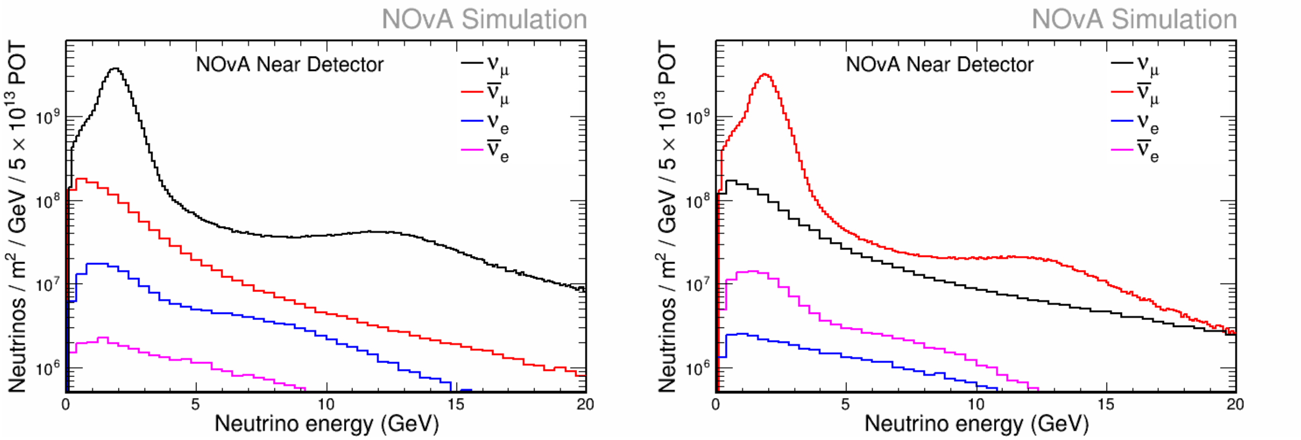


Figure 3.4: The FHC flux (left) and the RHC flux (right). Muon neutrinos are the largest component of the FHC beam, while anti-muon neutrinos are the largest for the RHC beam. Electron neutrinos and antielectron neutrinos are a small contribution to both samples.

3.2.1 Off-axis neutrinos

The NOvA detectors are located 14 mrad off-axis from the primary direction of the NuMI beam. This approach is designed to use two-body decay kinematics of the pions that are the primary source of neutrinos in the beam to create a neutrino beam that is more sharply peaked around 1.8 GeV. This comes at the cost of a lower overall flux of neutrinos. For small angles, the energy and flux of neutrinos from the decay of pions is given by the follow equations:

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

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

where E_ν is the neutrino energy, E_π is the energy of the decaying pion, $\gamma = E_\pi / m_\pi$, θ is the angle between the pion direction and the neutrino direction, Φ_ν is the neutrino flux, A is the cross sectional area of the detector, and L is the distance from the source of the beam to the detector. Equations 3.3 and 3.4 are plotted in Fig. 3.5. The neutrino flux and neutrino energy for detectors that are 4 mrad off-axis are shown by the solid red line. At 14 mrad off-axis neither the flux (left) nor the energy (right) depend strongly on the incident pion energy; they are relatively flat at neutrinos of energy 2 GeV.

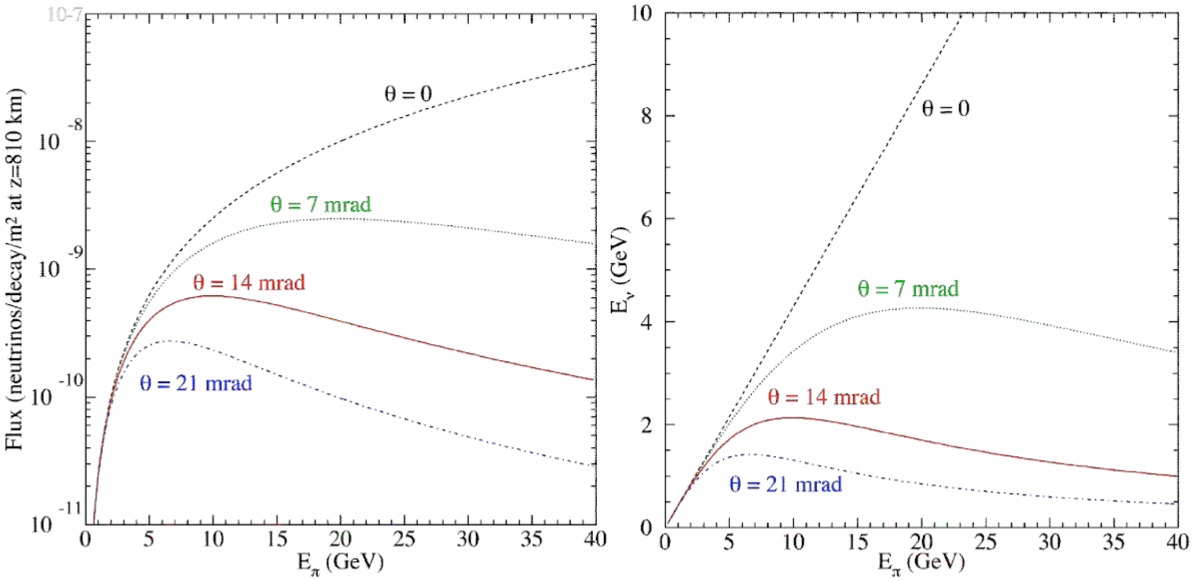


Figure 3.5: The off-axis neutrino flux (left) and the off-axis neutrino energy (right). (Ref. [30]).

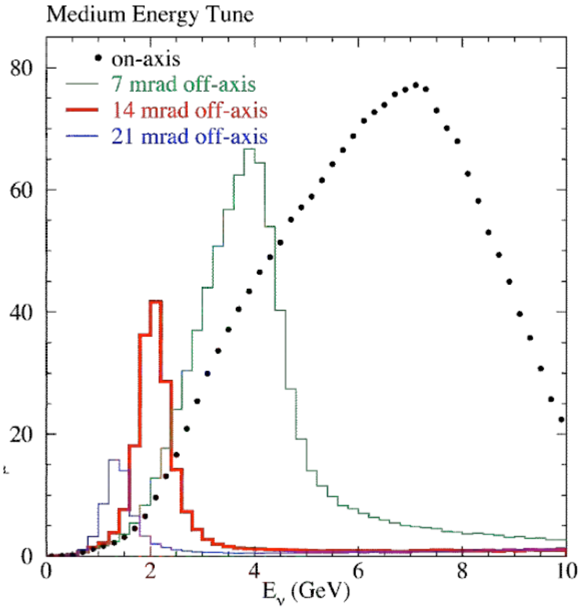


Figure 3.6: The neutrino flux at different angles off axis for a detector placed 810 km away from the beam. (Ref. [30]).

Figure 3.6 shows the neutrino flux at a detector located 810 km away from the origin of the beam at different angles off-axis. The flux at 14 mrad off-axis is significantly lower than the on-axis flux but has a much narrower energy range. However at 2 GeV of neutrino energy the predicted flux of neutrinos 14 mrad off-axis is about 5 times the flux from an on-axis beam. The first oscillation maximum is near 2 GeV for a baseline near 810 km which results in higher statistics measurements of relevant neutrino oscillation parameters.

3.3 The NOvA Detectors

The NOvA experiment has two detectors, the Near Detector located 1 km from the origin of the NuMI beam, and the Far Detector, located 810 km from the origin of the beam. The two detectors are placed 14.6 mrad off-axis of the center or the NuMI beam and are functionally identical. The similarity in the two detectors design allows for a reduction in the impact of systematic uncertainties during comparisons of the data obtained at each detector. A diagram of the detectors and their relative sizes is shown in Fig. 3.7. The diagram also shows how the planes alternate between vertical and horizontal orientations.

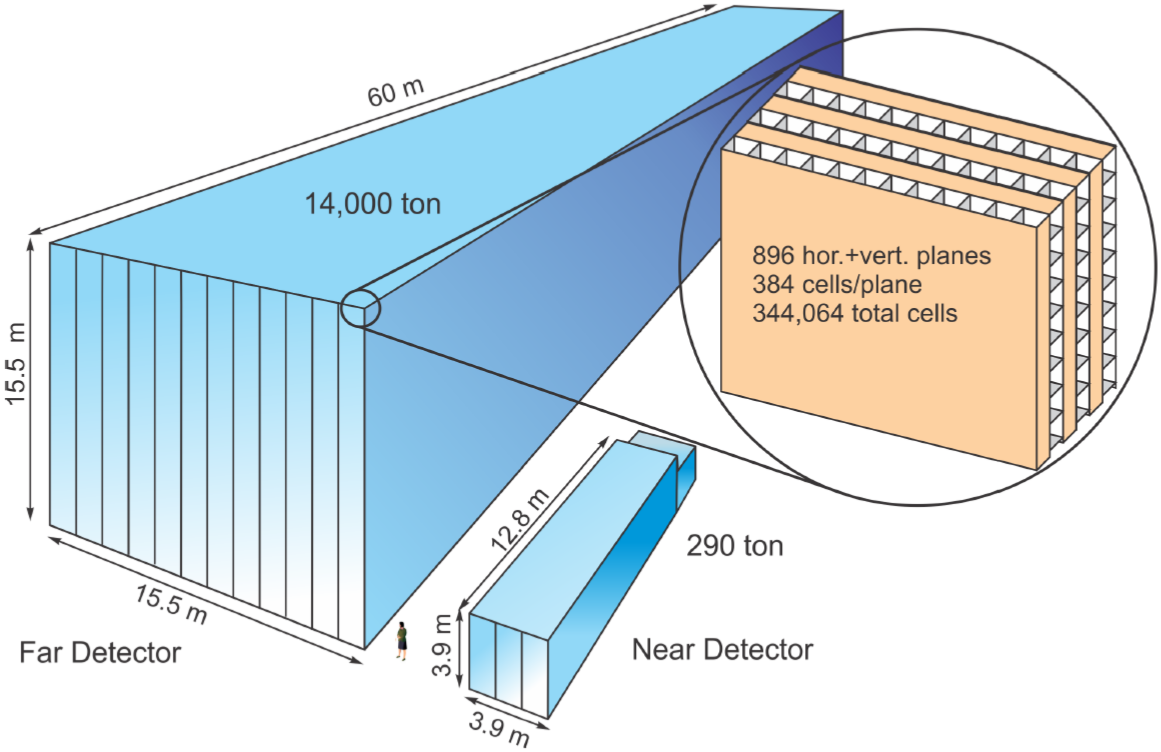


Figure 3.7: Diagram of the Near and Far Detectors compared to the size of a person. The alternating vertical and horizontal orientations of the planes is shown in the top right. (Ref. [31]).

3.3.1 Detector components

3.3.1.1 The extruded PVC cells

The basic building blocks of the NOvA detectors are PVC cells containing liquid scintillator and wavelength shifting fiber, shown in Fig. 3.8 (left). The cells are extrusions of a reflective dioxide rigid PVC. The cells have a 3.8 cm interior width, outer walls that are 2 to 4.5 cm thick, and are 15.5 meters in length.

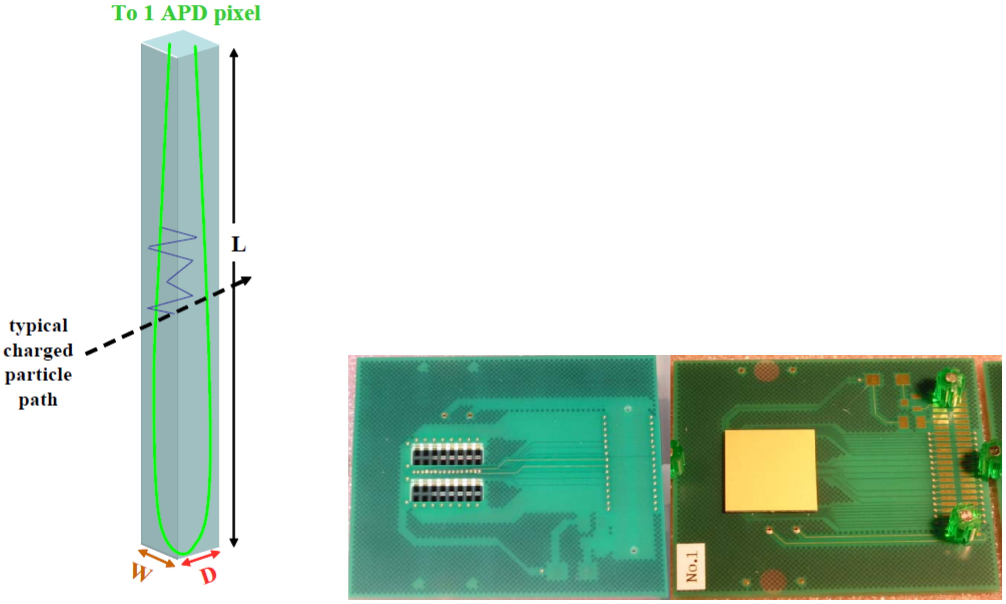


Figure 3.8: A diagram of a NOvA cell (left) (Ref. [30]) and a photograph of an APD board with 32 pixels as input (right) (Ref. [32]). A charged particle traversing a cell causes scintillator light, which is collected by the wave-shifting fiber. The fiber channels the light to an APD pixel to be read out as an interaction in the cell.

3.3.1.2 The liquid scintillator

The NOvA detectors are filled with a liquid scintillator that is a mineral oil solvent doped with 4.1% pseudocumene scintillation which emits light in the range of 360 - 390 nm [30]. The mixture also includes additives to shift the initial light to 400-450 nm which is the fiber optics absorption spectrum. There is also an anti-static agent added to the liquid to prevent the build-up of charge when the oil is being added to the cells.

3.3.1.3 Wavelength-shifting fiber

Each 15.6 m cell contains a loop of 33.5 m of wavelength shifting fiber [30]. The fiber captures blue light of 400-450 nm from the scintillator and shifts the wavelength to green light of 490 - 550 nm to increase the efficiency of the avalanche photodiode. The fiber is polystyrene mixed with 300 parts per million of R27 dye as a wave-shifter.

3.3.1.4 Avalanche Photodiode

The NOvA experiment uses an avalanche photodiode (APD) manufactured by Hamamatsu [30] shown in Fig. 3.8 (right). The APDs are the component that converts the light from the cells into electronic signals. An APD works as follows: electron-hole pairs are excited by photons, which are accelerated by a strong electric field. Due to the impact-ionization process, which is that accelerated electrons collide with other electrons causing them to dislodge from the atoms of the APD, secondary electron-hole pairs are created, which accelerate and strike more electrons, causing the process to repeat. The avalanche of electrons leads to amplification of the current. The APD used by NOvA has a quantum efficiency of 85% for the 520-550 nm light from the wavelength-shifting fiber. In order to minimize the thermal noise, the APDs are cooled to - 15 degrees Celsius. The APDs are operated at a voltage of 425 V which produces a gain in the output current of about 100. The signal is then read out by the Front End Board (FEB) which digitizes the hits above a certain threshold. The information on the hits from the FEB is accumulated by a Data Condenser Module (DCM) that passes the data along to a processing farm for analysis.

3.3.2 Detector Modules

The NOvA detectors are made up of a collection of cells. Sixteen cells are connected together to form an extrusion as shown in Fig. 3.9 (top). Two 16 cell extrusions are put together to form a 32 cell extrusion module as shown in Fig. 3.9 (bottom left). The extrusion modules also contain an end plate, side seal, manifold cover, snout and electronics bow. The end plate covers the end of the module to seal the detector. The manifold cover is the cap on the other end of the module and directs the output of 32 wave-shifting fibers in the cells to the APD pixels. Multiple extrusion modules are put side-by-side to form a plane. The planes are layered in an alternating pattern such that the orientation of the cells of one plane are orthogonal to the orientation of the cells of the planes next to it as shown in Fig. 3.9 (bottom right).

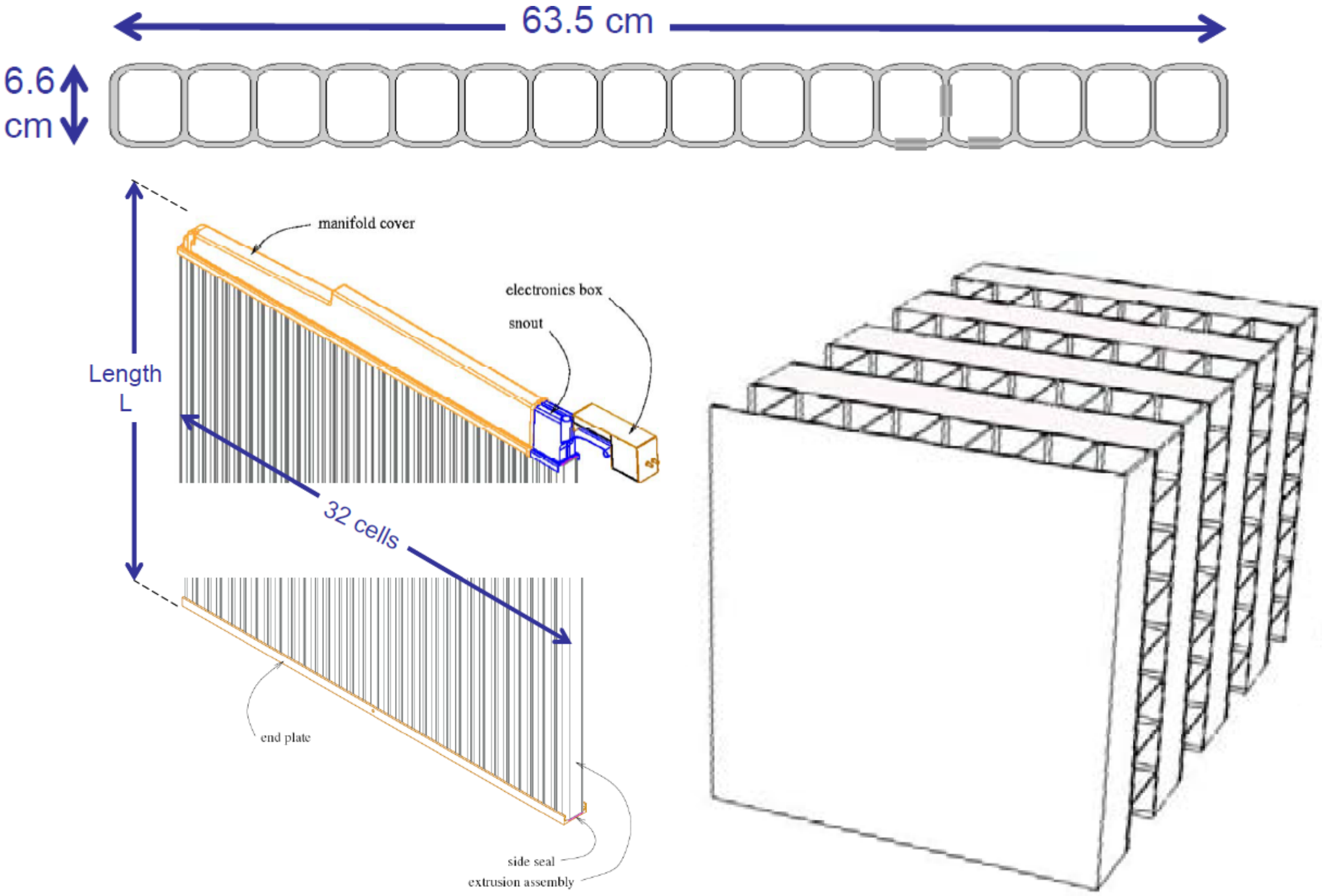


Figure 3.9: Diagram of NOvA cells assembled together into a 16 cell extrusion (top), how the extrusions are assembled into a plane (bottom left), and how planes are assembled into a block (bottom right). (Ref [30]).

The orthogonal assembly allows for three dimensional track reconstruction if a track passes through multiple planes as one orientation measures X-Z coordinates and the other measures Y-Z coordinates. The planes are glued together to form a block, which consists of 24 planes in the Near Detector and 32 planes in the Far Detector.

3.3.3 The Near Detector

The NOvA Near Detector is located 1 km away from the target of the NuMI beam on the Fermilab site. A photograph of the Near Detector is shown in Fig. 3.10. The beam is incident upon the face of the detector shown in the photo. Due to the proximity of the Near Detector to the origin of the NuMI beam it sees a much higher flux of neutrinos than does the Far Detector. It is positioned 105 m underground to reduce the rate of cosmic ray interactions. The Near Detector is made up of 20,192 cells arranged into 214 planes with a total mass of 300 tons when fully filled. The top and sides of the Near Detector is outfitted with the electronics described previously in this Chapter. The electronics can be seen in Fig. 3.10 on top of the Near Detector, the APD and FEB enclosures are the gold colored boxes that connect via cables to the DCMs. The dimensions of the Near Detector are 4.2 meters in width and height. In addition to the Active Region of the Near Detector consisting of PVC planes, a muon catcher, shown in Fig. 3.11, is placed at the end of the Near Detector opposite the NuMI beam to increase the stopping power for muons initiated by neutrino interactions in the detector.

Figure 3.10: Photo of the Near Detector. The beam is incident onto the near edge of the detector. The top of the Near Detector shows the APD and FEB enclosures (gold colored boxes) wired to the central DCMs.

The muon catcher consists of steel planes in addition to the PVC modules in the active region of the detector. The steel planes are 10 cm thick and are separated by two PVC planes, one in each orthogonal direction. The vertical planes in the muon catcher consist of three modules but the horizontal one only consist of two, thus the muon catcher is as wide as the rest of the Near Detector but only two-thirds as tall. The muon catcher preventing muons from exiting the back of the detector and greatly improves the quality of the muon energy reconstruction.

Figure 3.11: Photograph of the muon catcher attached to the Near Detector. The NuMI beam is directed towards the camera.

3.3.4 The Far Detector

The Far Detector is located in Ash River Minnesota, 810 km away from the start of the neutrino beam and is also 14.5 mrad off of the primary NuMI axis. The Far Detector is located 10 m below ground at an elevation 372 m above sea level. A photograph of the Far Detector is shown in Fig. 3.12. The detector consists of 896 planes and comes to 14 kilotons when completely filled with liquid scintillator. The Far Detector is designed to be functionally identical to the Near Detector's active region where the only difference is that the Far Detector is much larger. Notably the Far Detector is large enough that muons do not typically exit the back of the detector and so the detector does not require a muon catcher. To mitigate cosmic ray interactions the Far

Detector site has an overburden that is 122 cm thick of concrete and 15 cm thick of barite on top of the building. This overburden provides 12 radiation lengths of shielding to help eliminate photons coming from cosmic rays.

Figure 3.12: Photo of the Far Detector. The beam enters the detector from the opposite end than the one shown. The top of the detector is covered by the readout electronics (APDs, FEBs, and DCMs). The lattice of pipes above the electronics is the water cooling system that services the APDs.

3.4 Detector capabilities

Figure 3.13 shows a display of the Near Detector during a typical beam spill. The event display is separated into two two-dimensional images; the top image shows the X and Z coordinates of the Near detector, while the bottom image represents the Y and Z coordinates. The pixels are colored according to the time from the start of the beam spill. The timing resolution for an event is 1 μs which is indicated by the bottom color axis of the event display. Typically an event occurs within a few ns; since this is less than the 1 μs window, all ionizations initiated by the same neutrino are grouped together in time. The spatial resolution of events is the same as the size of a cell, namely 5.8 to 8.3 cm in length. This corresponds to the size of the pixels in the event display. The vertical and horizontal directions are only measured in every other plane due to the alternating orientations of the planes which causes the gaps between each pixel in the display.

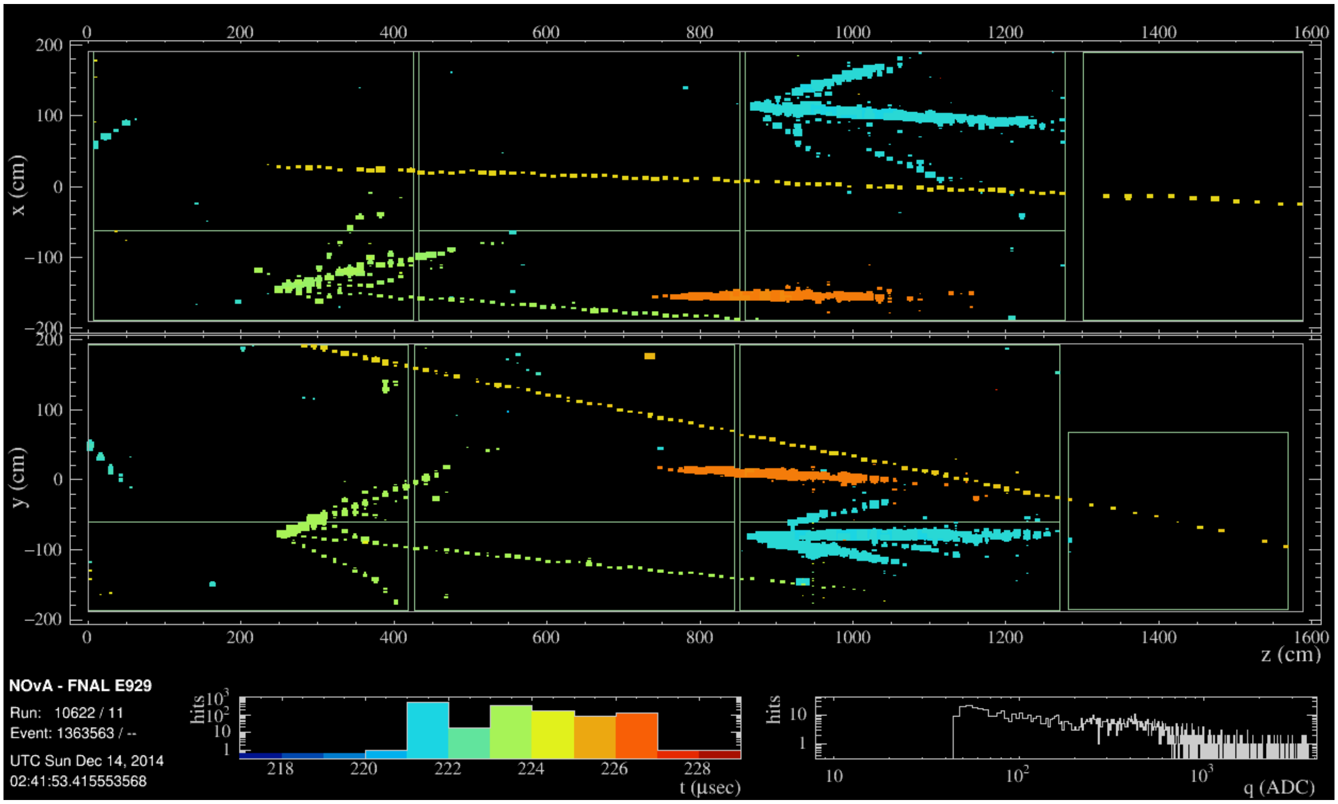


Figure 3.13: Event display of the Near Detector during a typical beam spill. The beam enters from the left. Two events displayed in green and blue start in the middle of the detector and are likely neutrino events. The yellow track enters from outside, traverses the scintillator volume, and exits via the muon catcher.

The construction of the Near Detector leads to very good tracking of muons, and to accurate reconstructions of the muon energy and direction. The amount of light deposited in the each cell can be reconstructed which leads to knowledge of the energy deposition of leptons and hadrons as they travel through the detector. This Thesis makes use both of the reconstruction of muon kinematics and hadronic energy to make a precise measurement of cross-section distributions.

Chapter 4

Data, Monte Carlo, and Analysis Variables

4.0.1 Data and reference Monte Carlo

The analysis is carried out using the NOvA fourth analysis forward-horn-current Near Detector Monte Carlo and data datasets. The data was taken from August 2014 to February 2017 in the Near Detector. The beam power during that period ranged from 250 kW to 700 kW and yielded an exposure of 8.09×10^{20} protons on target (POT).

The analysis reference Monte Carlo is obtained using the simulation software GENIE, version v2_12_2 [19], and corresponds to 3.05×10^{21} of simulated POT.

4.0.2 Monte Carlo central value weights

Weights are applied to the central value Monte Carlo to account for current knowledge of cross section modeling and the NuMI neutrino flux. The central value cross section weight is called in CAFAna by, *kXSecCVWgt2018* [33]. The tune changes the quasielastic axial mass (M_ACC-QE) from 0.99 to 1.04. The random phase approximation (RPA) is applied to quasielastic and baryon resonance production by default. Deep inelastic scattering cross section for events with $\mathrm{W} < 3$ GeV is increased by 10%. The 2p2h-MEC model is weighted such that the two-dimensional distribution of energy and momentum transfer agree with data, which is described in detail in the 2018 analysis weight technical note [33]. The flux weighting is called in CAFAna by *kPPFXFluxCVWgt* and is based on hadron production measurements using the PPFX package.

It is known that 2p2h contributions to $\nu_\mu - CC$ inclusive scattering *(i)* exist, and *(ii)* are present in NOvA data. The analysis described here aims to improve the World's knowledge concerning the absolute rate and kinematic distribution of this contribution. The approach taken is to include a 2p2h-MEC model in the reference MC for the purpose of estimating selection efficiencies, sample purity, influence of non-$\nu_\mu - CC$ backgrounds, and optimal use of unfolding. The effects of introducing alternative 2p2h distributions into these estimations is then evaluated using the multiverse method, wherein bin-by-bin variations to the 2p2h contribution are allowed in accord with the maximal bin-by-bin variation admissible according to current, very different models of 2p2h. The range of variation evaluated in this way defines the measurement systematic uncertainty that arises from lack of knowledge concerning 2p2h-MEC processes. A full description of this approach is given in Sec. 8.1.2.

4.1 Event selections

4.1.0.1 Signal definition

The signal definition used for the analysis is the same as the one used by the NOvA CC-ν_μ inclusive analysis selections [34]. For simulated observables a distinction must be made between the (true) or generated value from the "reco" value reconstructed in the detector. The signal definition as applies to Monte Carlo events is follows:

1. The event is a true CC-ν_μ event.
2. The event has its primary true vertex within the fiducial volume.
3. The event passes a true muon kinematic selection. Selection criteria expressed in the true quantities muon kinetic energy, T_μ, and muon-angle-cosine with respect to the neutrino beam, $cos(\theta_\mu)$, are shown in Eq. (4.1):

$$\begin{aligned}
0.50 < \text{ True } T_\mu < 1.1\,\text{GeV} \text{ and } 0.50 < \text{True } \cos(\theta_\mu), \text{ or}\\
0.50 < \text{ True } T_\mu < 1.2\,\text{GeV} \text{ and } 0.56 < \text{True } \cos(\theta_\mu), \text{ or}\\
0.50 < \text{ True } T_\mu < 1.3\,\text{GeV} \text{ and } 0.62 < \text{True } \cos(\theta_\mu), \text{ or}\\
0.50 < \text{ True } T_\mu < 1.4\,\text{GeV} \text{ and } 0.68 < \text{True } \cos(\theta_\mu), \text{ or}\\
0.50 < \text{ True } T_\mu < 1.8\,\text{GeV} \text{ and } 0.85 < \text{True } \cos(\theta_\mu), \text{ or}\\
0.50 < \text{ True } T_\mu < 1.9\,\text{GeV} \text{ and } 0.88 < \text{True } \cos(\theta_\mu), \text{ or}\\
0.50 < \text{ True } T_\mu < 2.2\,\text{GeV} \text{ and } 0.91 < \text{True } \cos(\theta_\mu), \text{ or}\\
0.50 < \text{ True } T_\mu < 2.5\,\text{GeV} \text{ and } 0.94 < \text{True } \cos(\theta_\mu).
\end{aligned} \tag{4.1}$$

Selected events are binned and unfolded using the variables $|\vec{q}|$ and E_{avail} and no cuts are imposed using these variables. For the final result, bins having very low efficiency are not reported. We define the domain ($|\vec{q}| \leq 2.0\,\text{GeV/c}$, $E_{avail} \leq 2.0$ GeV) to be the region analyzed by this Thesis.

4.1.0.2 Event selections

The event selections are the same as the NOvA CC-ν_μ inclusive analysis selections [34]. They are as follows:

1. The event passes initial quality cuts. That is, the event has enough hits in the Near Detector to enable reconstruction, has a reconstructed track, and has an ionization hit in at least four different planes.

2. The event has a reconstructed muon. A reconstructed muon is a Kalman track [35], a track that is found using the Kalman filtering algorithm, that receives a muon identification score greater than 0.24 from the boosted decision tree developed by the NOvA CC-ν_μ inclusive analysis [34].

3. The event has a vertex, the beginning of the reconstructed muon track, within the fiducial volume: a a right rectangular prism of dimensions 270 cm in X by 270 cm in Y by 900 cm in Z, centered inside the detector.

4. The event is fully contained within the Near Detector, including the muon and all hadronic tracks and showers.
5. The event passes the same muon kinematic selection defined above in Eq. 4.1 but in reconstructed quantities instead of true quantities.

4.2 Analysis variables

The first goal of the 2p2h enhancement cross-section analysis is to deliver a double differential cross section using kinematic variables that describe the muon and hadronic energies. Previously considered variables include muon energy, E_{μ}, muon production angle, ϑ_{μ}, four momentum transfer-squared, Q^2, and invariant hadronic mass, W. These variables either did not contain enough information (E_μ and θ_μ) or did not separate 2p2h from other processes well (Q^2 and W). Three-momentum transfer $|\overrightarrow{q}|$ is the best choice for muon kinematics because it contains information on both the muon energy and muon angle with respect to the beam.

A second, independent variable is needed to describe the hadronic system. Reconstructed hadronic energy is typically found by subtracting the reconstructed muon from reconstructed neutrino energy, but this is too model-dependent for cross-section purposes. Visible hadronic energy, E_{vis}, is the sum of non-leptonic energy deposited in the Near Detector; it is less model dependent, but too detector-specific to be useful to researchers outside of NOvA. Available energy, E_{avail}, is designed to be as close as possible to the energy that can be reliably observed in the detector with minimal model dependence. The variable E_{avail} is a correction to E_{vis} to estimate the amount of visible energy if the detector were perfect. This variable has been previously used by the MINERvA experiment for a similar analysis [5].

The observable quantities that are used to construct the working variables are muon energy, E_μ, muon angle with respect to the incoming neutrino, θ_μ, and calorimetric hadronic energy deposited in the detector, E_{vis}. The reconstruct energy from all non-lepton hits in the detector, E_{had}, is obtained by correcting E_{vis} for unseen energy such as that lost to dead material in the detector or to neutrons. The reconstructed neutrino energy variable used in this analysis is identical to the one used in the ν_μ disappearance analysis [36]. The neutrino energy is constructed by summing the reconstructed muon energy and the reconstructed energy from all non-lepton hits in the detector, E_{had}, as given by Eq. (4.2).

$$E_\nu = E_\mu + E_{had}. \tag{4.2}$$

In order to obtain $|\overrightarrow{q}|$, the four-momentum transfer squared is computed according to Eq. (4.3):

$$Q^2 = 2E_\nu \left(E_\mu - p_\mu \, cos\theta_\mu\right) - m_\mu^2. \tag{4.3}$$

Then the three-momentum transfer is constructed by combining the observables using Eq. (4.4):

$$|\vec{q}| = \sqrt{Q^2 + (E_\nu - E_\mu)^2}. \tag{4.4}$$

The relationship of the above variables to a CC neutrino-nucleus interaction is illustrated by the Feynman diagram shown in Fig. 4.1.

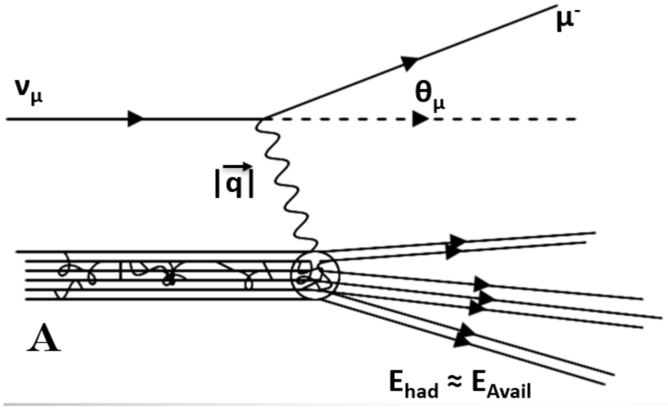


Figure 4.1: Feynman diagram for the 2p2h reaction with time flowing left to right. The diagram shows a muon neutrino interaction with a nucleus. The W gauge boson transfers three momentum and energy to the hadronic system. The $|\vec{q}|$ is reconstructed lepton variables; E_{avail} provides an estimation of energy transfer.

4.2.1 Three-momentum transfer

Three-momentum transfer between the leptonic and hadronic currents is calculated using the muon energy based on track range and the neutrino energy estimated using hadronic energy plus muon energy [36]. These quantities are assembled according to Eqs. (4.2), (4.3), and (4.4) above to make $|\vec{q}|$. Figure 4.2 shows the relationship between reconstructed $|\vec{q}|$ versus the true $|\vec{q}|$ obtained from the reference MC. The relationship is seen to be a nearly linear, with most events clustered around a central ridge. This clustering can be quantified using the absolute resolution for reconstructed $|\vec{q}|$. Absolute resolutions are used to guide the choices of bin widths in this analysis, as described in the next Section. The absolute resolution is defined as the standard deviation of the absolute residual distribution. Also useful is the fractional resolution, defined as the standard deviation of the fractional residual distribution. The absolute residual and the fractional residual are defined: in Eqs. (4.5) and (4.6) respectively.

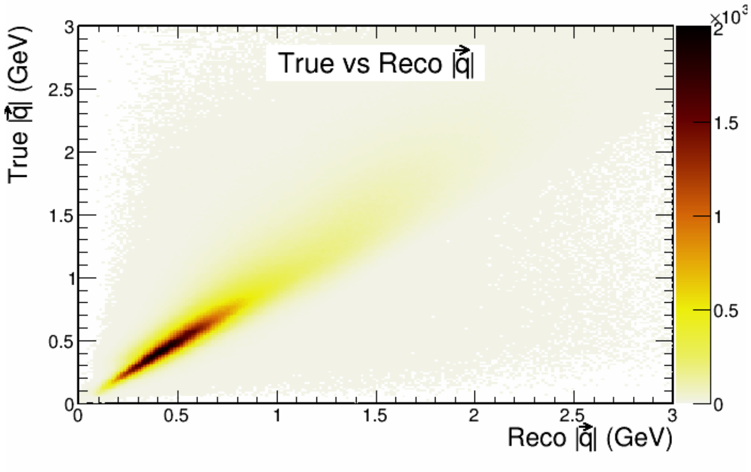


Figure 4.2: Plot of reconstructed $|\vec{q}|$ versus true $|\vec{q}|$ with the distribution height displayed using a linear scale on the color axis.

$$\text{absolute residual} = true\,|\vec{q}| - reco\,|\vec{q}|, \tag{4.5}$$

$$\text{fractional residual} = \frac{(true\,|\vec{q}| - reco\,|\vec{q}|)}{true\,|\vec{q}|}. \tag{4.6}$$

The absolute and fractional resolutions for $|\vec{q}|$ can be inferred from the residuals in Fig. 4.3. The absolute $|\vec{q}|$ resolution, defined as the standard deviation of the absolute residual distribution, is shown in Fig. 4.3 (left) to be 0.14 GeV. The fractional $|\vec{q}|$ resolution is defined as the standard deviation of the fractional $|\vec{q}|$ residual distribution as shown in Fig. 4.3 (right); it is 21%. Figures

4.4 and 4.5 show the absolute and fractional residuals for $|\vec{q}|$ broken out into four bins of reconstructed $|\vec{q}|$. In Fig. 4.4, the absolute residual is seen to broaden with increasing $|\vec{q}|$. In both sets of plots, the $|\vec{q}|$ residuals are centered close to 0.0. Furthermore a mild skewing towards higher values is observed, indicating that the reconstruction has a tendency to underestimate the $|\vec{q}|$ value at high values of reconstructed $|\vec{q}|$.

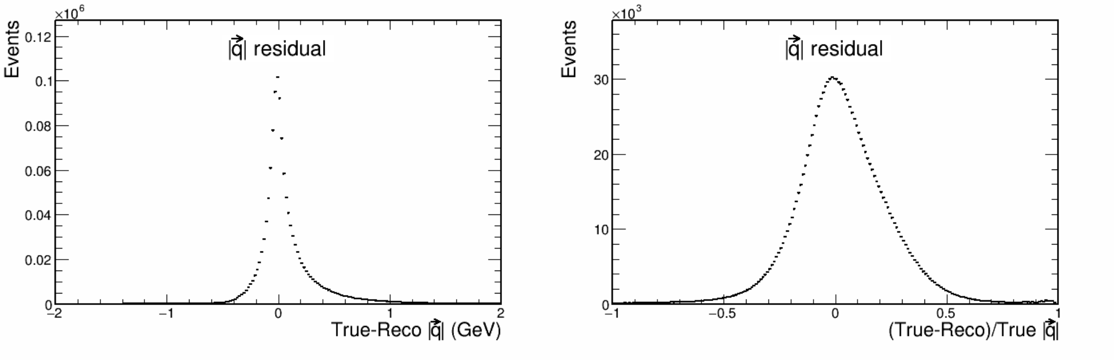


Figure 4.3: Distributions of absolute $|\vec{q}|$ residual (left) and fractional $|\vec{q}|$ residual (right). The mean of the absolute residual is centered at 0.05 GeV, and the distribution has a σ of 0.14 GeV. The fractional residual has a mean of 0.03, and a σ of 0.21. The resolution of $|\vec{q}|$ is 0.14 GeV with a fractional resolution of 21%.

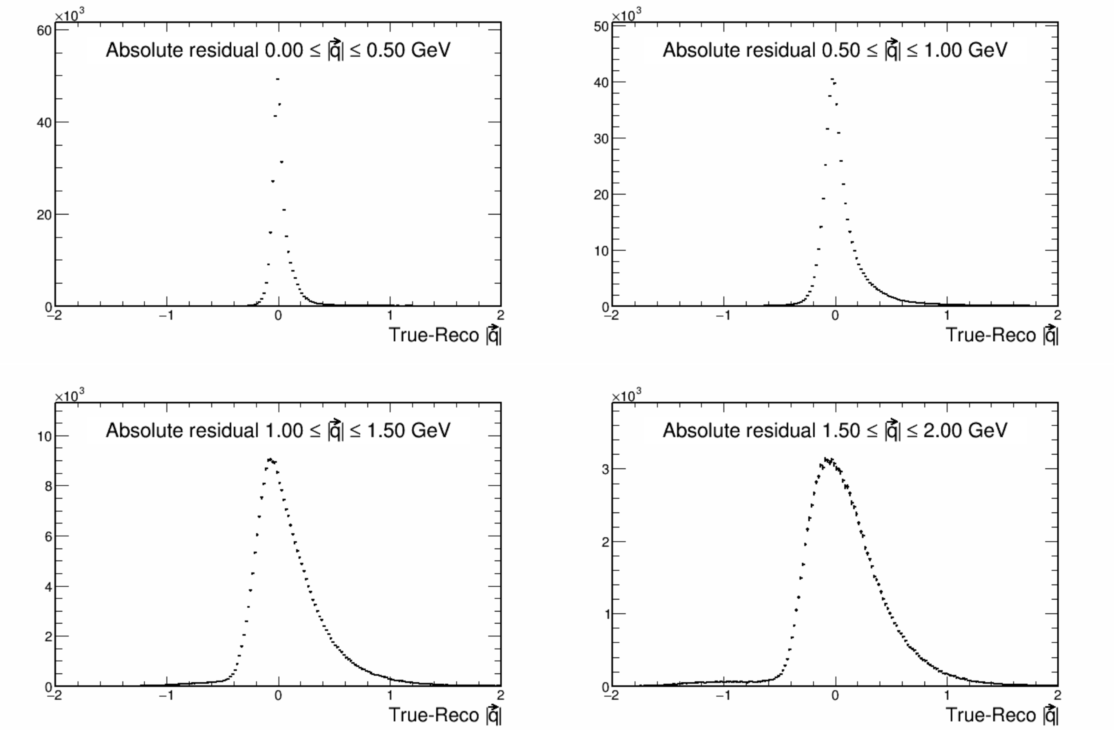


Figure 4.4: Distributions of the absolute $|\vec{q}|$ residual in increasing bins of reconstructed $|\vec{q}|$.

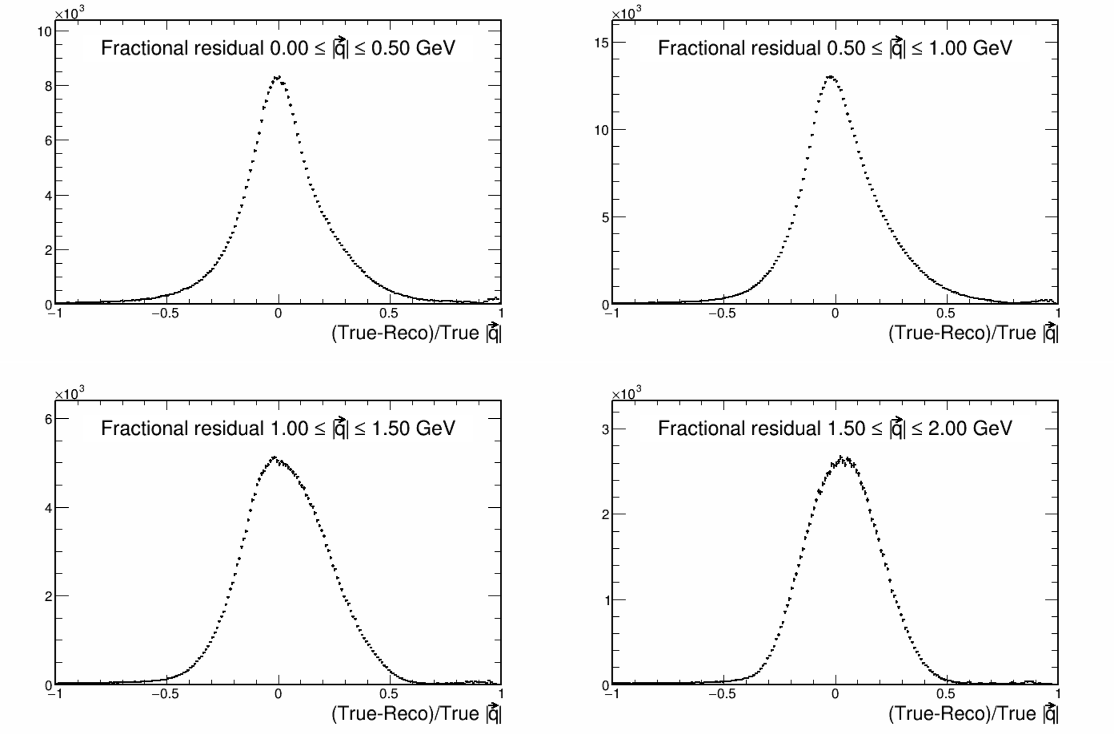


Figure 4.5: Distributions of the fractional $|\vec{q}|$ residual in increasing bins of reconstructed $|\vec{q}|$.

4.2.2 Definition of available energy

As described above, E_{avail} is the sum of all detectable hadronic energies deposited in the detector whose evaluations are very nearly model-independent. Hadronic energies included in the sum are the following:

1. The kinetic energy of protons, and charged pions.
2. The total energy of electrons, photons, neutral pions, and kaons.
3. The total energy minus the nucleon mass for hyperons.
4. The total energy for anti-nucleons.

Available energy does not include energy initiated neutrons, as neutron collisions with nuclei generate very little ionization in the detector.

4.2.3 Estimator for E_{avail}

The variable $\mathrm{E}_{\mathrm{avail}}$ is constructed using the analysis reference Monte Carlo, defined in Section 4.0.1, to create a mapping from event-by-event visible energy, E_{vis}, to true E_{avail}. An estimator for reconstructed E_{avail} is assembled as follows: For each selected CC event a mapping is made from the sum of non-leptonic energy deposited in the detector, namely reconstructed E_{vis}, to true E_{avail} obtained from simulation. Then a profile of the mode value (the peak value) of true E_{avail} is made for each bin of reconstructed E_{vis}. Finally a function is fit to the profile, providing a map that relates reconstructed E_{vis} to reconstructed E_{avail}.

Figure 4.6 shows the reconstructed E_{vis} versus true E_{avail} mapping. The distribution appears to have a single population, indicating that a fit of the mode value of E_{avail} in bins of E_{vis} is feasible. The distribution of true E_{avail} is however not symmetric about the mode of E_{avail} in most bins of E_{vis}. As shown in Fig. 4.7, the mode of true $\mathrm{E}_{\mathrm{avail}}$ in each bin of E_{vis} of Fig. 4.6 is linear at lower E_{vis} but starts to fluctuate at higher $E_{vis} > 1.2$ GeV. However the region above 1.2 GeV contains very few events, as can be seen in Fig. 1.1.

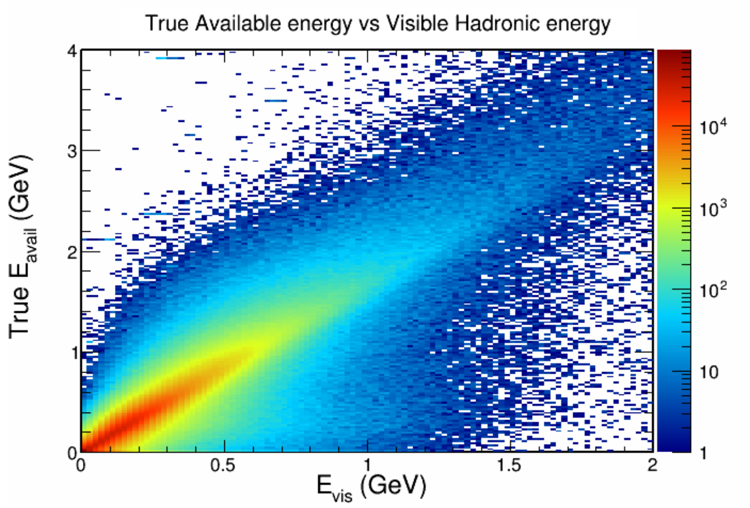


Figure 4.6: Plot of reconstructed visible hadronic energy, E_{vis}, versus true available energy, E_{avail}, using a log scale on the color axis. There appears to be a single population of reconstructed E_{vis} values whose mode values can be mapped onto true E_{avail} in a linear way. For example the mode E_{avail} with $E_{vis} = 0.5$ GeV indicates a mapping to 0.8 GeV.

The fit in Fig. 4.7 is made using the same fitting package as the NOvA CC-ν_μ second analysis neutrino energy estimator [37] with the only modification being the use of different input variables. The fitting function is a second-order quadratic equation:

$$E_{avail} = Y_0 + A * E_{vis} + B * E_{vis}^2. \tag{4.7}$$

Here, intercept Y_0, A, and B are parameters to be fit. The Y_0 value is the E$_{avail}$ intercept, A is the linear slope, and B is the quadratic strength. The result of the fit is shown in Fig. 4.7 and the fit parameters are summarized in Table 4.1.

Table 4.1: Fit results. The Intercept is near zero, and B, the coefficient of the quadratic term is small.

Variable	Value	Error
Y_0	-5.12e-03	9.78e-05
A	1.68e+00	6.71e-04
B	2.35e-02	6.10e-04

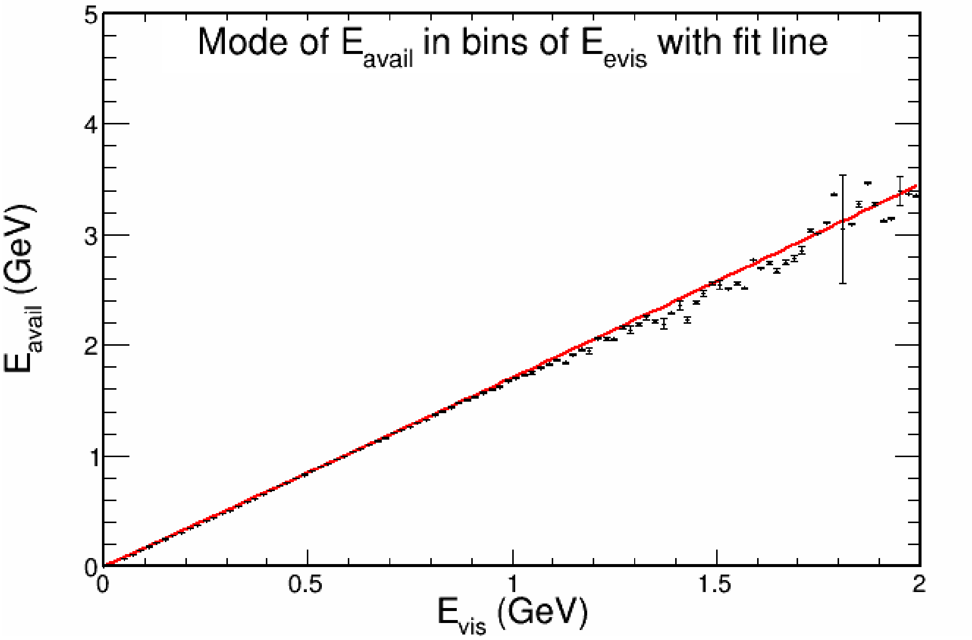


Figure 4.7: Plot of the mode of E$_{\text{avail}}$ for each bin of E$_{vis}$ in Fig. 4.6. The viability of a linear map from reconstructed E$_{vis}$ to E$_{avail}$ is clearly indicated. The fit lies on top of the lower E$_{vis}$ bins, but deviates at E$_{vis}$ greater than 1.2 GeV.

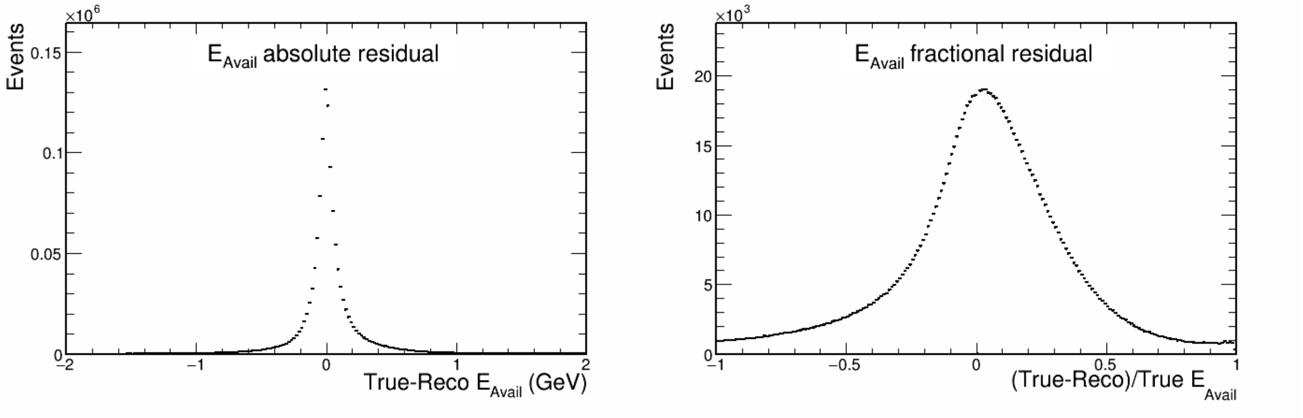


Figure 4.8: Distributions of absolute E_{avail} residual (left) and fractional E_{avail} residual (right). The mean of the absolute residual is centered at 0.02 MeV, and the distribution has a standard deviation of 0.21 GeV. The fractional residual has a mean of 0.03, and a standard deviation of 0.32. The resolution of E_{avail} is 0.13 GeV with a fractional resolution of 32%.

Figure 4.8 shows the absolute and fractional residuals of E_{avail}. The absolute $\mathrm{E}_{\mathrm{avail}}$ resolution, defined as the σ of the absolute residual distribution shown in Fig. 4.8 (left), is 0.21 GeV. This is better than the hadronic energy resolution used in the NOvA CC-ν_μ disappearance analysis, which is 0.23 GeV on average [36]. The fractional $\mathrm{E}_{\mathrm{avail}}$ resolution is similarly defined as the standard deviation of the fractional $\mathrm{E}_{\mathrm{avail}}$ residual distribution, shown in Fig. 4.8 (right); it is 32%. This is better than the MINERvA resolution for the same variable, which varies from 38% to 55% for a range of $|\overrightarrow{q}|$ values [5].

Figures 4.9 and 4.10 show the distributions of absolute and fractional residuals for E_{avail}, broken out into four bins of increasing E_{avail}. As with the $|\vec{q}|$ residuals, the residuals for E_{avail} remain centered around 0.0 as the bin energy is increased. The distributions are roughly Gaussian but exhibit broadened tails for $E_{avail} > 1.0$ GeV.

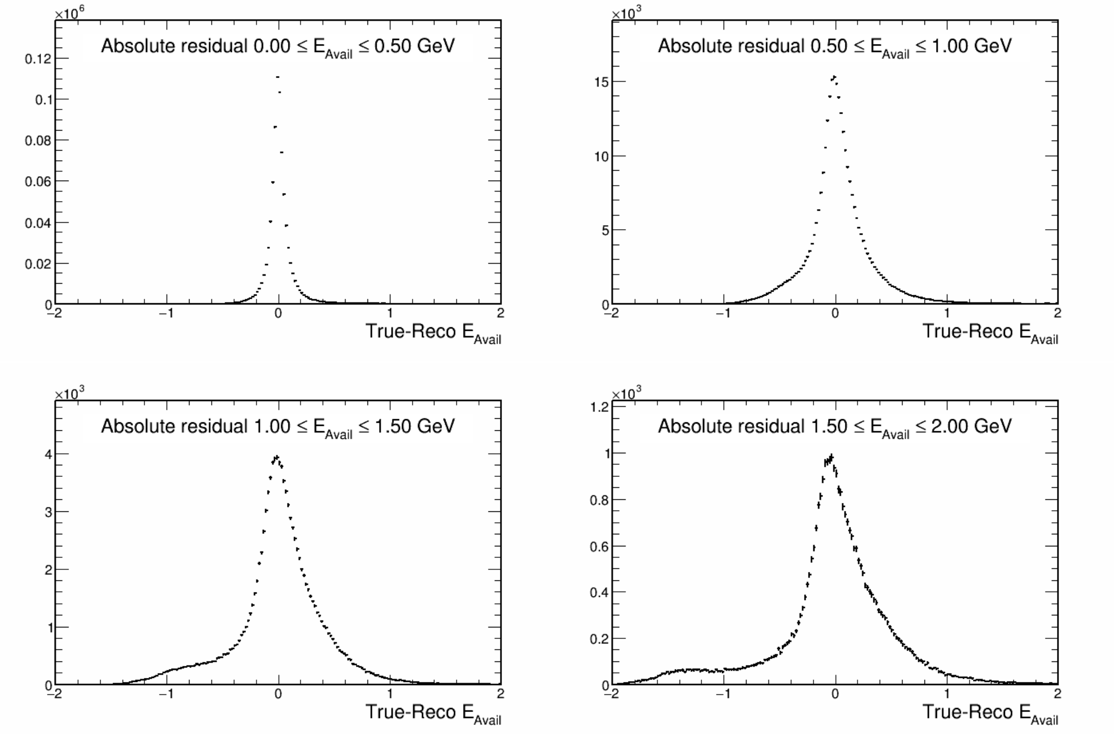


Figure 4.9: Distributions of absolute E_{avail} residual in increasing slices of reconstructed E_{avail}.

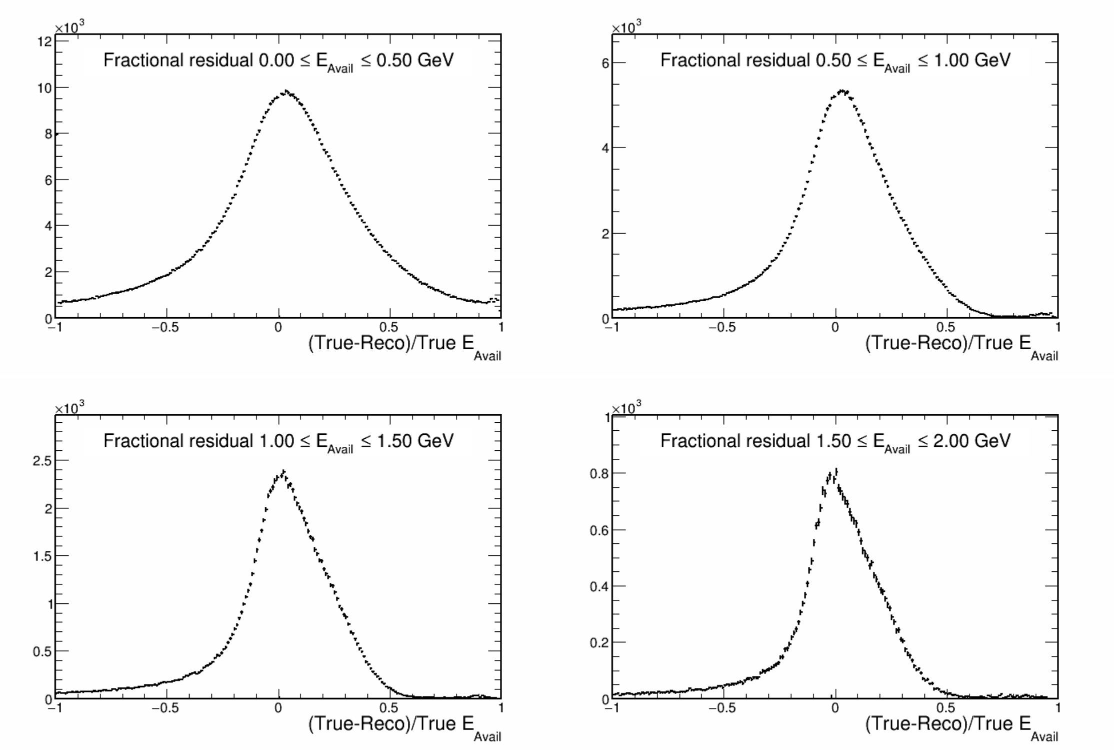


Figure 4.10: Distributions of fractional E_{avail} residual in increasing slices of reconstructed E_{avail}.

4.3 Optimization of Binning

The binning for each reconstructed variable is chosen according the experimental resolution. Figure 4.11 (left) shows the residual versus reconstructed $|\vec{q}|$ with the resolution overlaid; Fig 4.11 (right) shows the resolution versus reconstructed $|\vec{q}|$. The resolution worsens roughly linearly with increasing $|\vec{q}|$ though there is a small inflection point between 1.6 and 2.0 GeV.

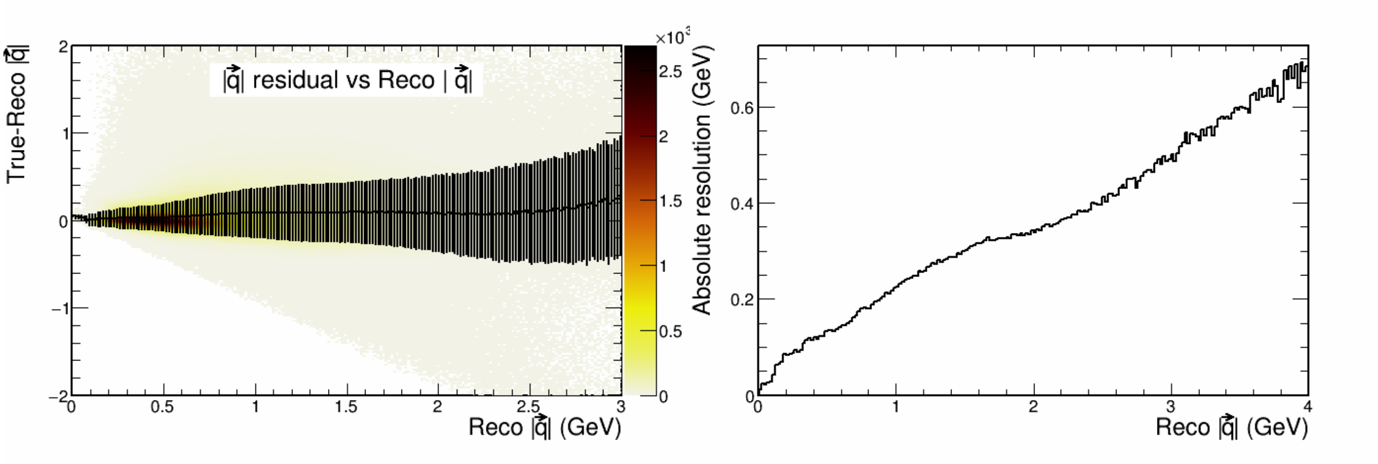


Figure 4.11: Residual $|\vec{q}|$ versus reconstructed $|\vec{q}|$ (left) and $|\vec{q}|$ resolution versus reconstructed $|\vec{q}|$ (right). The resolution worsens approximately linearly as reconstructed $|\vec{q}|$ increases.

The binning is then constructed by taking the resolution at the largest value of $|\vec{q}|$ in the analysis region, namely 3 GeV and using a bin of that size with 3 GeV as the upper edge, and working back towards zero $|\vec{q}|$ from there. Table 4.2 shows the fourteen $|\vec{q}|$ bins, which grow with increasing $|\vec{q}|$, used in the analysis.

Table 4.2: Binning in three-momentum transfer.

$\lvert\vec{q}\rvert$ binning (GeV)	0.0-0.10	0.10-0.20	0.20-0.30	0.30-0.40	0.40-0.50	0.50-0.65	0.65-0.80
$\lvert\vec{q}\rvert$ binning (GeV)	0.80-1.00	1.00-1.20	1.20-1.40	1.40-1.70	1.70-2.00	overflow	

The residual and resolution for E_{avail} is shown in Fig. 4.12. The resolution has a linear relationship with reconstructed E_{avail} in the analysis region of $E_{avail} \leq 2.0$ GeV.

The E_{avail} binning is determined in the same fashion as for the $|\vec{q}|$ binning however the starting (maximum) value for E_{avail} is 2.0 GeV. The resulting 9 bins are given in Table 4.3.

Table 4.3: Binning for available energy.

E_{avail} Bins (GeV)	0.0-0.10	0.10-0.20	0.20-0.35	0.35-0.50	0.50-0.70
E_{avail} Bins (GeV)	0.70-0.95	0.95-1.20	1.20-1.50	1.50-2.00	overflow

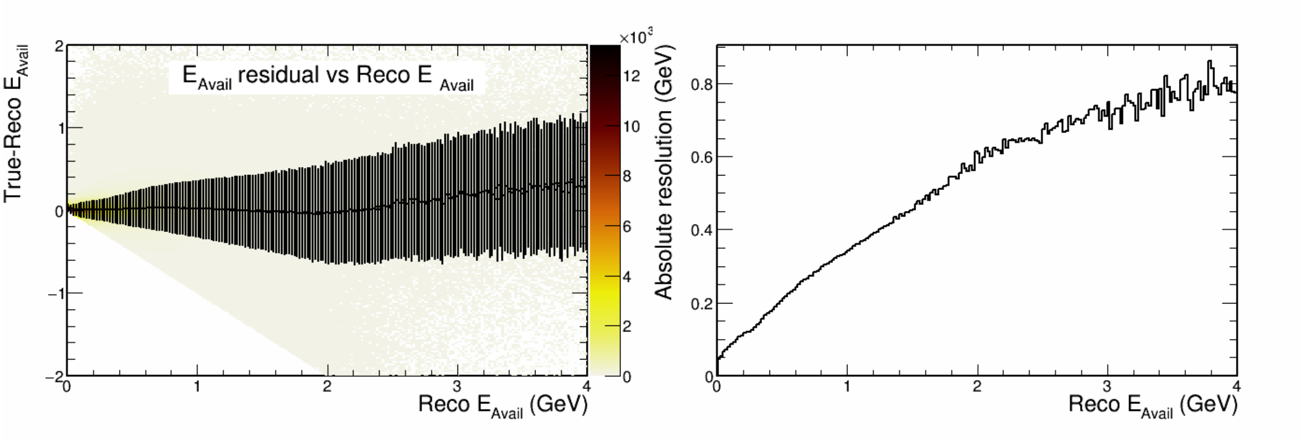


Figure 4.12: Residual E_{avail} versus reconstructed E_{avail} (left) and E_{avail} resolution versus reconstructed E_{avail} (right). The resolution is approximately linear with available energy for $E_{avail} < 2.0$ GeV.

Chapter 5

Efficiency, Purity, and Backgrounds of the Selected Sample

5.1 Selection efficiency

Recall that the signal definition requires an event to be a true CC-ν_μ event whose primary vertex lies within the fiducial volume, and whose muon has T_μ and $cos(\theta_\mu)$ that satisfy the requirements of Eq. 4.1; the analysis domain is taken to be $0.2 \leq |\vec{q}| \leq 2.0$ GeV/c and $0.0 \leq E_{avail} \leq 2.0$ GeV. The restriction of $|\vec{q}|$ to values greater than 200 MeV/c arises from the muon kinematic selections given my Eq. 4.1. Selection efficiency distributions are presented in the variables E_{avail} and $|\vec{q}|$. Selection efficiencies for the variables T_μ and $\cos(\theta_\mu)$ have also been determined - see DocDB: 38860. Sample purity distributions in the same kinematic variables are also shown, and the overall selection efficiency and purity for the $d^2\sigma/d|\vec{q}|dE_{avail}$ analysis are determined.

Selection efficiency and sample purity are defined according to the following ratios of simulated events:

$$Efficiency \equiv \frac{Selected\,true\,signal\,events}{All\,true\,signal\,events}, \tag{5.1}$$

$$Purity \equiv \frac{Selected\,true\,signal\,events}{All\,selected\,events}. \tag{5.2}$$

The kinematic domain of the measurement is $0.2 \leq |\vec{q}| \leq 2.0\,\mathrm{GeV/c}$ and $0 \leq E_{avail} \leq 2.0\,\mathrm{GeV}$ [38]. The selection efficiencies for each of these variables individually are shown in Fig. 5.1, with $|\vec{q}|$ displayed in the left-hand plot and E_{avail} shown in the right-hand plot. The efficiency for $|\vec{q}|$ is highest at low values, and decreases as the value increases. The E_{avail} efficiency is largest at low and high E_{avail} with a dip in the mid-range around 0.9 GeV. Quasielastic scattering and multi-nucleon scattering occupy low $|\vec{q}|$ and E_{avail}, while baryon resonance and deep inelastic scattering occupy higher $|\vec{q}|$ and E_{avail}. This leads to changes in efficiency when $|\vec{q}|$ and E_{avail} lie beyond the region where quasielastic scattering and 2p2h-MEC scattering are maximal and where baryon resonance and deep inelastic scattering become significant portions of the total cross section.

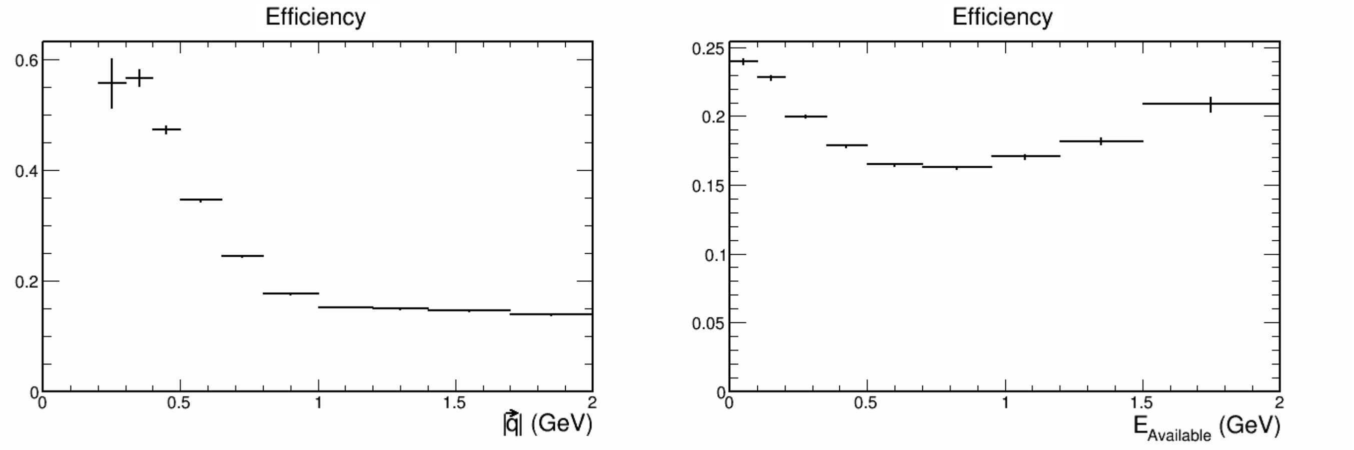


Figure 5.1: Selection efficiency as a function of three-momentum transfer (left) and available energy (right). The selection efficiency in terms of $|\vec{q}|$ is highest when $|\vec{q}|$ is small and diminishes as $|\vec{q}|$ increases; the efficiency levels off when $|\vec{q}| > 1.0\,\mathrm{GeV/c}$. The selection efficiency as a function of E_{avail} is largest at $E_{avail} < 0.2$ GeV and steadily diminishes until ∼1.0 GeV and gradually rises thereafter.

The efficiency for one parameter is correlated with that of the other. The selection efficiency as a function of both variables is displayed in Fig. 5.2. The efficiency is largest when E_{avail} in units of GeV and $|\vec{q}|$ in units of GeV/c are approximately equal in magnitude, and falls off when both become larger than 2.0 or when one becomes numerically very different than the other. The region with $E_{avail} < 0.4\,\mathrm{GeV}$ and $0.6 \leq |\vec{q}| \leq 1.2\,\mathrm{GeV/c}$ has a fairly constant selection efficiency of 0.15.

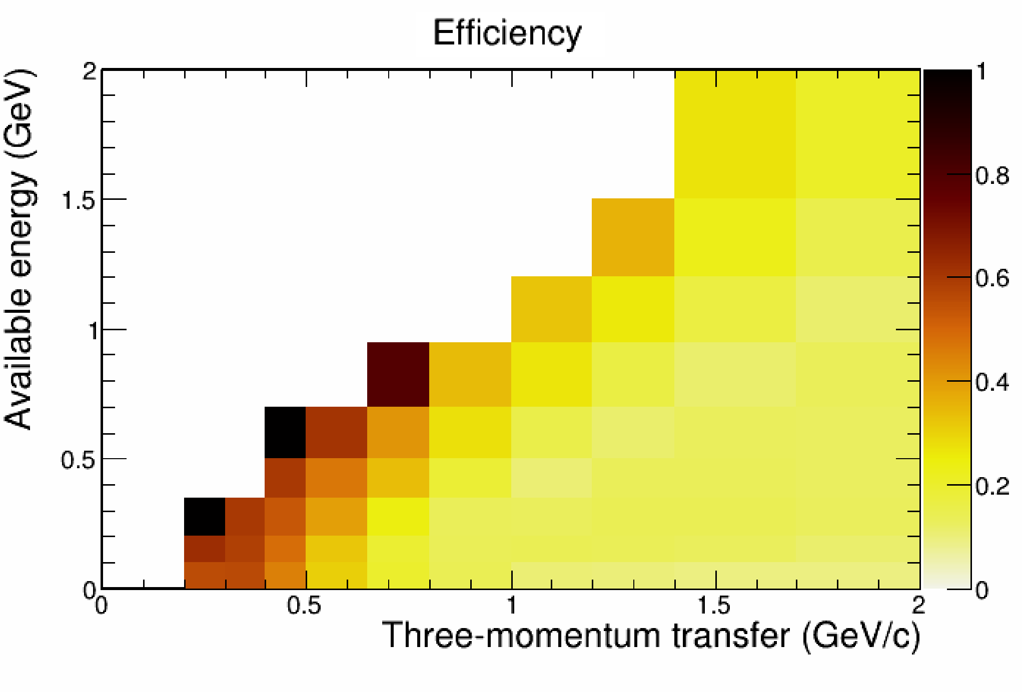


Figure 5.2: Selection efficiency for the analysis region in bins of three-momentum transfer versus available energy with binning determined by the resolution. The efficiency is highest when E_{avail} in GeV approximately equals $|\vec{q}|$ value expressed in GeV/c and falls off when $|\vec{q}|$ is larger than E_{avail}.

The purpose of including selection cuts is to reduce the amount of background in the analysis; this comes at the cost of selection efficiency. The plot sequence (a) through (f) in Fig. 5.3 shows the evolution of efficiency versus selection cuts displayed over the phase space of the analysis.

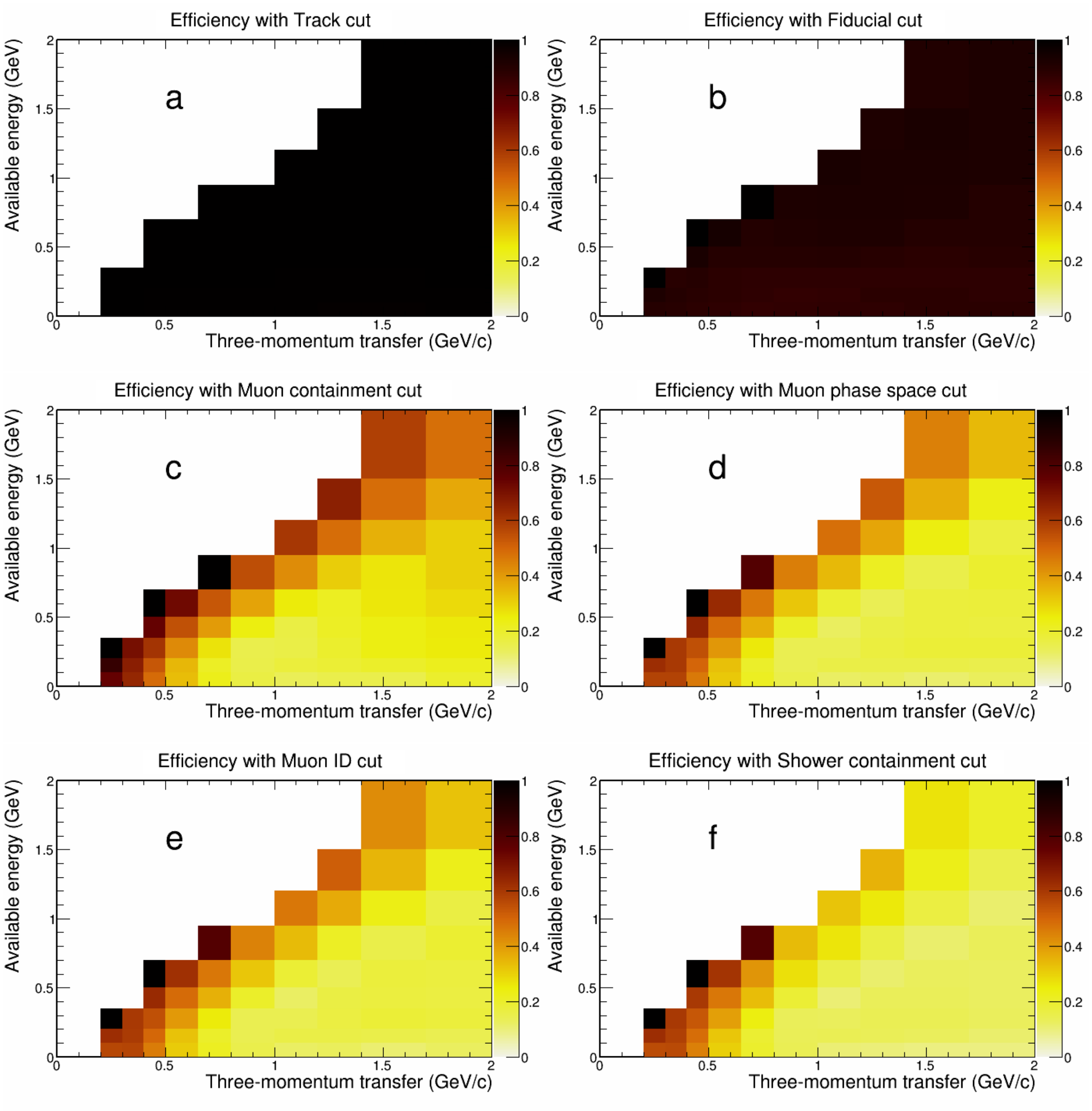


Figure 5.3: The plots (a) through (f) show the evolution of efficiency versus selection cuts displayed over the phase space of the analysis. Figure 16 (f) shows that the final efficiency after all selections approaches 60% over the region populated by 2p2h-MEC and decreases at larger values of $|\vec{q}|$ and E_{avail}.

The initial quality cut and the track reconstruction cut (a) have little impact on the efficiency. The muon identification cut (b) reduces the efficiency to ~85.3%. Further requiring a vertex contained within the fiducial volume (c) brings efficiency down to ~85%. Requiring the muon to be contained (d) has the largest effect on the efficiency, which brings it to ∼25%. The muon phase space (e) reduces the efficiency to ∼22%. Applying the shower containment cut (f) makes the final efficiency for the analysis to be 18.7%. Table 5.1 shows the reduction in efficiency as selection cuts are added sequentially. The selection cuts shown in the left-hand column includes all cuts that appear above them.

Table 5.1: Variation of efficiency with event selections. Selection cuts listed in the left-hand column include all cuts that appear above them. The efficiency falls off as more selection cuts are added. The single largest efficiency drop comes from adding the muon containment cut.

Selection cut	Events	Efficiency
All true signal	1,987,000	100%
Quality	1,985,000	99.9%
Track reconstruction	1,983,000	99.8%
Muon identification	1,667,000	85.3%
Vertex fiducial volume	1,609,000	82.3%
Muon containment	482,600	24.7%
Muon phase space	439,700	22.1%
Shower containment	371,800	18.7%

The effect of applying every cut except the one that is specifically called out per row is shown in Table 5.2. That is, every cut is included in the event selection except the cut specified in the left-hand column. The Table shows that muon containment removes the most signal events from the sample, 776,200 events, followed by the hadronic shower containment cut, 67,900 events, with other cuts making smaller reductions. For example, the vertex fiducial volume cut removes 10,800 signal events whose primary vertices were erroneously reconstructed to be outside the fiducial volume.

Table 5.2: The N-1 cut Table shows the effect of each selection cut on the total number of simulated signal events. The first row shows the number of signal events after all selection cuts have been applied. For all other rows, all selection criteria but the one named on that row have been applied. The muon containment cut removes the most events from the selected sample.

Selection cut	Signal Events	Signal events removed
All selection cuts	371,800	-
Quality + track reconstruction	371,900	100
Muon identification	376,500	4,700
Vertex fiducial volume	381,600	10,800
Muon containment	1,148,000	776,200
Muon KE vs Angle	411,600	39,800
Shower containment	439,700	67,900

5.2 Sample purity

The sample purity as a function of the two analysis variables is shown in Fig. 5.4, with dependence on $|\vec{q}|$ given on the left and dependence on E_{avail} shown on the right. The purity of $|\vec{q}|$ grows from 0.60 at $|\vec{q}| = 0.2$ GeV/c to 0.95 at 0.9 GeV/c , and then steadily diminishes with increasing $|\vec{q}|$ values. The purity is largest at $E_{avail} \leq 0.5$ GeV and then declines gradually as E_{avail} increases.

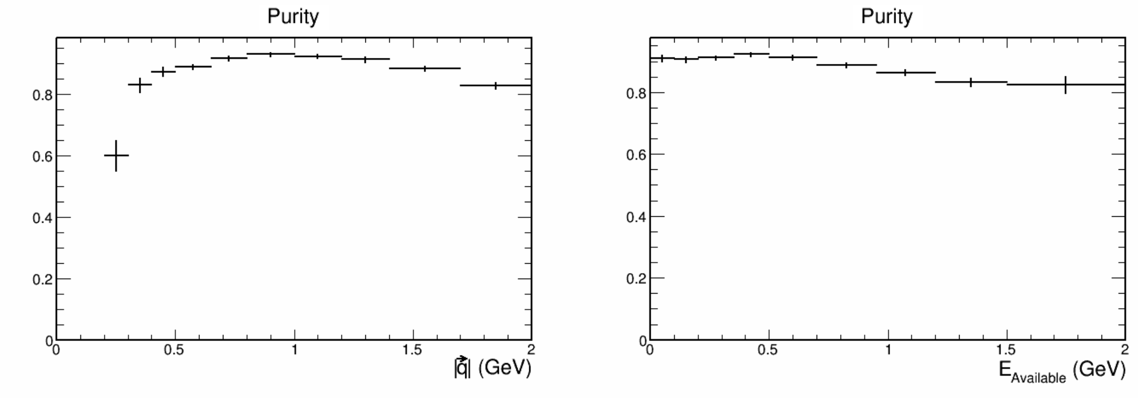


Figure 5.4: Sample purity as a function of three-momentum transfer (left) and as a function of available energy (right). The purity in $|\vec{q}|$ is largest at $|\vec{q}| = 0.9$ GeV/c. The maximum purity in terms of E_{avail} is 0.95 at $E_{avail} = 0.4$ GeV.

As is the case with efficiency, the sample purity in bins of one parameter is correlated with values of the other parameter. Figure 5.4 shows that the purity is largest when $|\vec{q}|$ and E_{avail} are small and have numerical values close to each other. When E_{avail} is much less than $|\vec{q}|$ the efficiency slowly decreases.

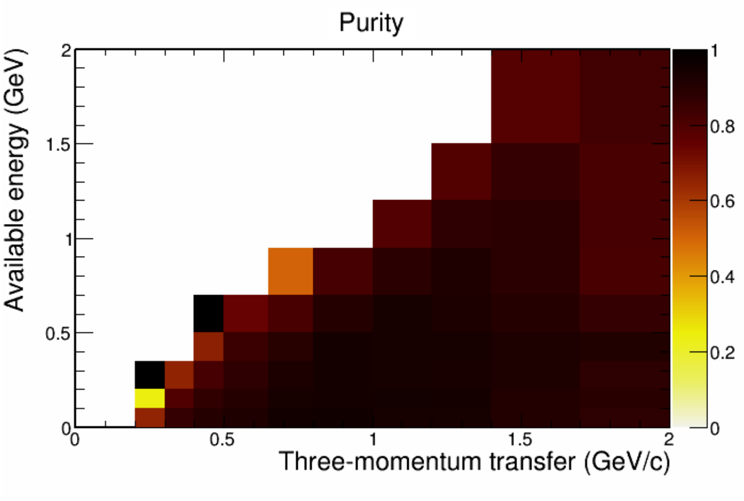
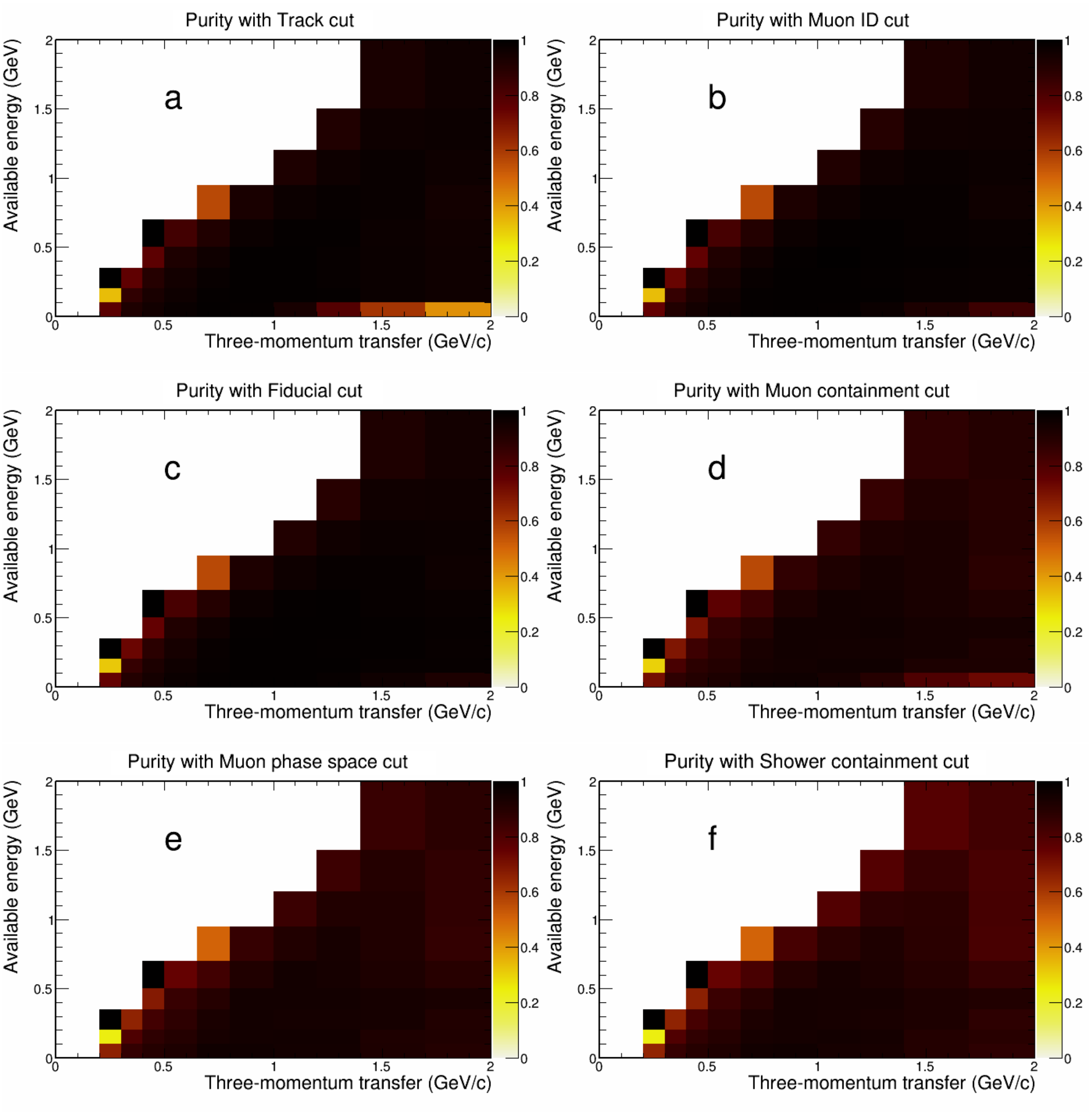


Figure 5.5: Sample purity in bins of of three-momentum transfer versus available energy with binning determined by the resolution.

Figure 5.5 shows sample purity in bins of of three-momentum transfer versus available energy with binning determined by the resolution. The purity is largest when E_{avail} in GeV is roughly equal to the value of $|\vec{q}|$ in GeV/c. The purity is reduced when $|\vec{q}|$ is larger than E_{avail} or when both are large in magnitude. Although purity is lower in regions of $|\vec{q}| > E_{avail}$ the purity stays above 0.6 in the domain of the measurement, namely $E_{avail} \leq 2.0$ GeV and $|\vec{q}| \leq$ 2.0 GeV.


Purity with Track cut

Purity with Muon ID cut

Purity with Fiducial cut

Purity with Muon containment cut

Purity with Muon phase space cut

Purity with Shower containment cut

Figure 5.6: Plots (a) through (f) show the evolution of purity versus selection cuts displayed over the plane of E_{avail} versus $|\vec{q}|$. The shower and muon containment cuts improve the purity dramatically so that the final purity approaches 88% over most of the phase space, as shown in Figure 5.6 (f). The regions along the diagonal kinematic border are populated by small numbers of NC events, giving rise to a few bins of low purity.

The overall purity is improved by making selection cuts, as is shown by the plot sequence in Fig. 5.6. The quality, track reconstruction, muon identification, and vertex fiducial volume selections bring the purity to 39% and the remaining, muon containment, muon phase space, and hadronic shower containment selections raise the purity to nearly 92%. Table 5.3 shows that small improvements accrue with the inclusion of quality and fiducial vertex volume cuts. Upon requiring the presence of a muon KE vs angle phase space cut, the purity increases by 45%, which is the largest purity increase. The overall purity after all cuts are applied is 91.85%.

Table 5.3: Summary of purity versus event selections. Selection cuts listed in the left-hand column include all cuts that appear above. Starting with the quality cut, the purity is 5.51%. Adding the fiducial volume cuts increases the purity to 38.9%. Then the muon containment increases the purity to 46.5%. Finally, including all cuts makes the overall purity 91.9%.

Selection cut	Signal events	Background events	Total events	Purity
Quality	1,985,000	34,686,000	36,671,000	5.41%
Track reconstruction	1,983,000	34,011,000	35,994,000	5.51%
Muon identification	1,695,000	25,733,000	27,428,000	6.18%
Vertex fiducial volume	1,635,000	2,572,000	4,207,000	38.86%
Muon containment	491,000	1,055,000	1,546,000	46.54%
Muon phase space	440,000	39,000	479,000	91.86%
Shower containment	372,000	33,000	405,000	91.85%

5.3 Background processes

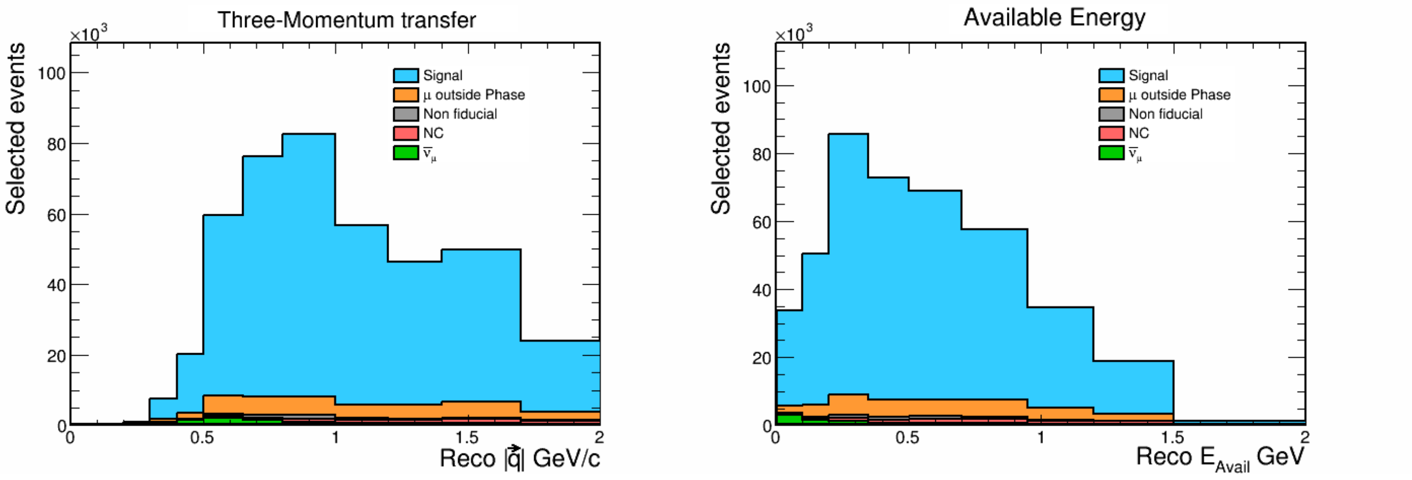


Figure 5.7: Distributions of selected events in $|\vec{q}|$ and E_{avail} (left and right plots respectively) subdivided into signal (blue), and background (red, brown) distributions.

Background reactions for this analysis fall into three main categories. These consist of (i) muons with true kinetic energy and angle outside the signal phase space, (ii) neutral current (NC) interactions reconstructed as ν_μ-CC interactions, and (iii) $\bar{\nu}_\mu$ interactions arising from the wrong sign component of the NuMI beam. The background also receives small contributions from ν_e-CC events and from CC events with vertices originating outside the fiducial volume.

Distributions of the signal and background components in $|\vec{q}|$ and E_{avail} are shown in Fig. 5.7. Figure 5.8 shows the ratios of the signal to selected events (left) and the backgrounds to selected events (right).

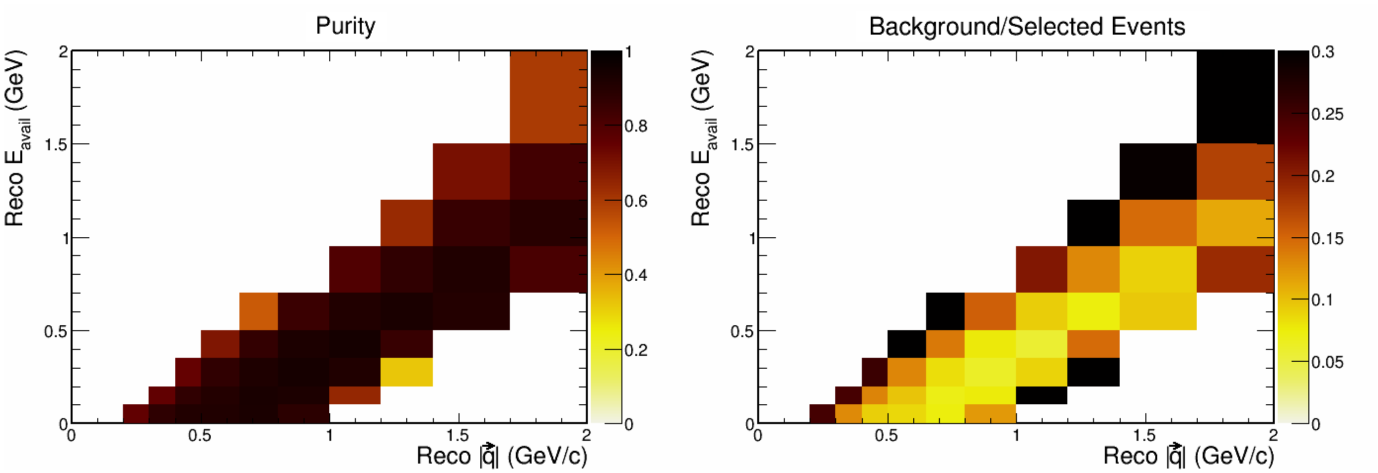


Figure 5.8: The fraction of signal to all selected events (left) and the fraction of background in all selected events (right).

Muons with true kinematics outside the phase space defined in the signal definition are the largest background. They populate phase space regions where the numerical values of $|\vec{q}| \approx E_{avail}$ and $|\vec{q}| \approx E_{avail} + 0.5$ (GeV). Events with vertices outside the fiducial volume distribute evenly across the analysis region and are 45% of the background and 3.70% of the total sample. The $\bar{\nu}_\mu$-CC events account for 18% of the total background and 1.48% of the total selected sample. The $\bar{\nu}_\mu$-CC distribution mostly populates lower values of E_{avail} due to the fact that neutrons contribute little visible energy. Neutral currents comprise 12.7% of the total background and 1.04% of the selected sample.

Figure 5.9 shows the fractions of events with muon kinematics outside the muon phase space, with vertices outside the fiducial volume, and ratios of NC, and $\bar{\nu}_\mu$-CC events to the total of selected events, plotted over the $|\vec{q}|$ versus E_{avail} phase space. The remaining background is made up of ν_e-CC interactions and events that are reconstructed inside the fiducial volume but are truly outside of the fiducial volume. The various background processes are shown in Table 5.4

as fractions of selected events. The leading backgrounds are (1) events where the muon should have failed the kinematic phase space selection but did not (3.7%), and (2) CC events induced by beam anti-neutrinos.

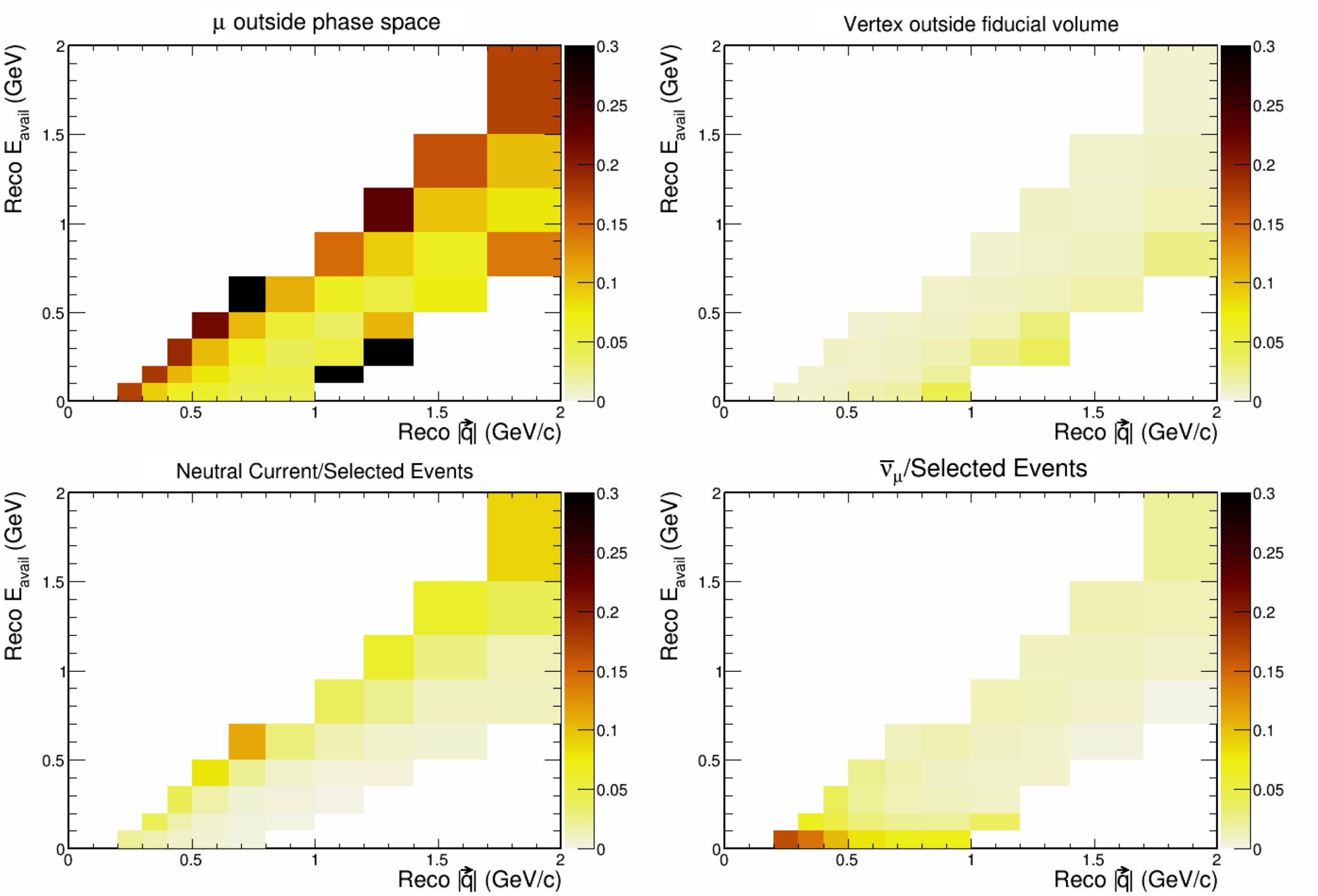


Figure 5.9: The fraction of events that have muon kinematics outside the defined phase space (top right), the fractions of events with a true vertex outside the fiducial volume (top left), the fraction of neutral current background to all selected events (bottom left), and the fraction of anti-neutrino background to all selected events (bottom right).

Table 5.4: Background processes as a fraction of selected events.

Processes	Selected events	Fraction of selected events
Signal	372,000	91.85%
Total Background	33,000	8.15%
Outside muon KE vs angle	15,000	3.70%
CC Anti-neutrino	6,000	1.48%
Non-fiducial	7,600	1.88%
Neutral current	4,200	1.04%
Electron neutrino	160	0.04%

Note that 7,600 signal events whose true primary vertices lie outside the fiducial volume are actually reconstructed inside the fiducial volume and are therefore background. Roughly speaking, one expects a degree of detailed balancing here. That is, there should be a similar amount of true signal events removed by the vertex fiducial volume cut. Table 5.2 shows the latter subsample to consist of 10,800 events. One sees that the numbers are indeed comparable however the number of true signal events removed is larger. This difference can be attributed to the fact that vertex reconstruction is most commonly foiled by the presence of prongs ejected into the backward hemisphere relative to the vertex. The reconstruction is then inclined to assign the primary vertex to the ends of backward-produced prong segments.

Chapter 6

Data Unfolding

6.1 Data unfolding

A cross-section measurement is carried out by measuring event count distributions in bins of variables of interest. Detector effects cause these variables to be smeared away from their true values. The process of correcting for smearing effects due to the finite resolution of the detector is known as unfolding [39]. Unfolding is necessary to present a cross section result in terms that are useful to the neutrino community, and not just the NOvA experiment. In this analysis, D'Agostini [40] unfolding is implemented using the RooUnfold [41] package in ROOT. D'Agostini unfolding is chosen because of the neutrino community's extensive experience using it in cross-section measurements [42, 34, 43], and because it preserves total event count [40]. D'Agostini unfolding is an iterative process that repeats the unfolding over a number of iterations determined by the user.

To determine the optimum number of iterations a metric known as the mean squared error (MSE) given in Eq. (6.1) was used.

$$MSE = \sum_{j=1}^{Bins} \frac{(Unfold_{,j} - True_{j})^2}{True_{,j}} \tag{6.1}$$

Here, $Unfold_j$ is the event count in the jth bin of the unfolded distribution, and $True_j$ is the jth bin in the truth distribution. The underlying truth of the data is unknown, so instead of optimizing the MSE to the data, it was optimized using 250 systematically independent simulated samples to find the best number of iterations on average. The averaged value, $\overline{MSE}$, is calculated according to Eq. (6.2), where M represents the number of systematic universes.

$$\overline{MSE} = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{Bins}\frac{(Unfold_{i,j} - True_{i,j})^2}{True_{i,j}} \tag{6.2}$$

6.1.1 Double differential unfolding

The unfolding procedure uses the reconstructed-to-true-matrix, shown in Fig. 6.1, to determine how the unfolding proceeds. The matrix is a four-dimensional object, with a convolution of reconstructed E_{avail} and reconstructed $|\vec{q}|$ on the X-axis, while the Y-axis has the true quantities. The matrix has a block-like structure, where each large block can be interpreted as a bin in E_{avail} that contains many bins of $|\vec{q}|$ that have that particular reconstructed and true E_{avail}. The reconstructed-to-true matrix is diagonal with most events populating either the diagonal, one bin off of the diagonal, or one block away from the diagonal.

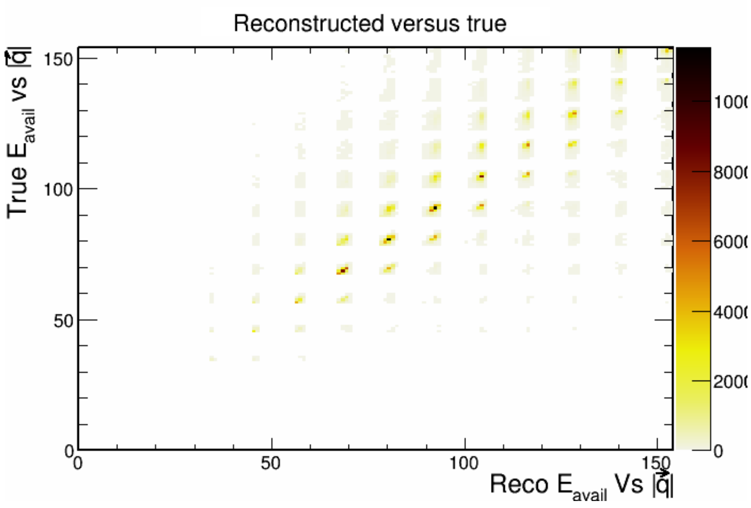


Figure 6.1: Reconstructed-to-true matrix. The matrix is a four-dimensional object in which the outer blocks are bins in E_{avail}, within which are distributions in $|\vec{q}|$. The matrix is diagonal, indicating that the reconstructed quantities fall into the same bin as the true quantities more often than not.

To perform the unfolding optimization, 250 cross section systematic universes were generated using the GENIE multiverse method [44] and a systematic universe was examined for each of the energy scale and light calibration systematic samples as well. Then the MSE was calculated for successive of iterations between 0 (no unfolding) and 15 iterations. The results are shown in Fig. 6.2. The left-hand plot shows the MSE including no unfolding in the furthest left column of the histogram. The MSE improves by an order of magnitude with just one iteration of unfolding. The right-hand plot shows the MSE after at least one iteration of unfolding; the MSE is minimized after two iterations of unfolding.

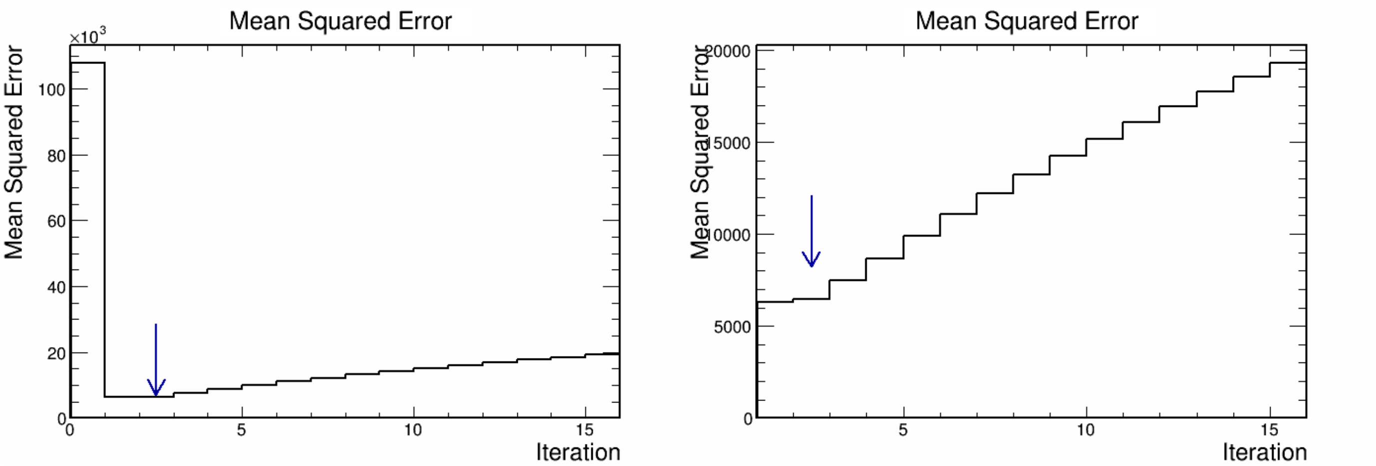


Figure 6.2: Mean square error per iteration. The left-hand plot shows the MSE with no unfolding iterations and one through fifteen iterations, while the right plot only shows the MSE after unfolding at least once. The MSE improves dramatically with one iteration and is minimized with two iterations, after which is begins to climb.

In addition to the MSE, the χ^2 was also used as a metric. Because more iterations of unfolding increases the statistical error, the χ^2 has a problem that, given infinite iterations, it will eventually converge to zero; however if limited to few iterations, it will find a local minima. Figure 6.3 shows the results of testing wth the χ^2. The plot on the left includes no unfolding, while the plot on the right does not. There is a large decrease in χ^2 from no unfolding to one iteration, and the χ^2 is minimized at two iterations.

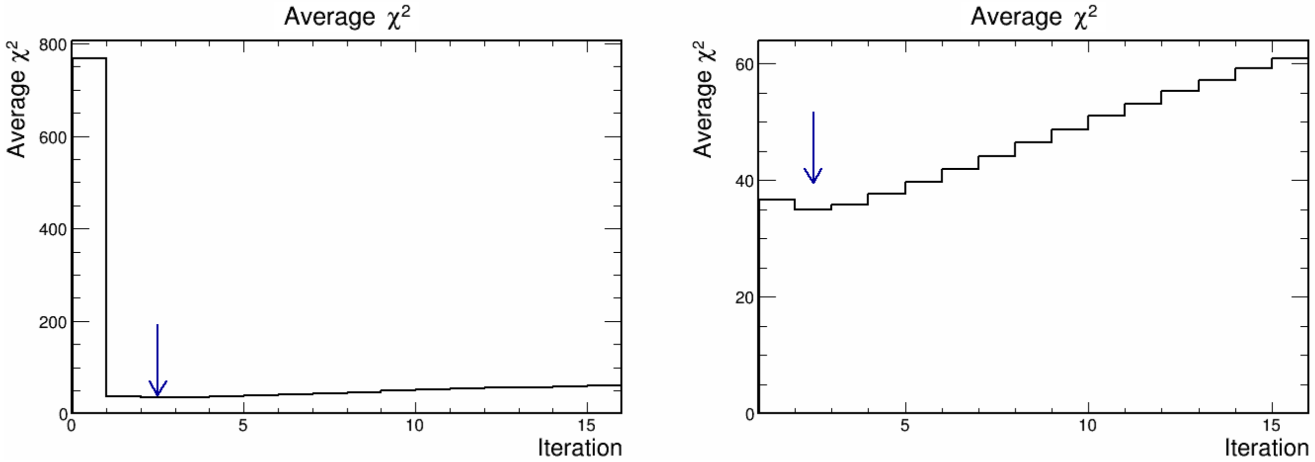


Figure 6.3: The χ^2 per iteration. The left plot shows the average χ^2 with no unfolding and up to fifteen iterations of unfolding, while the right plot omits a bin with no unfolding. The χ^2 improves by an order of magnitude from out-of-the-box reconstructed to one iteration of unfolding. There is then minor improvement from one iteration to two iterations; iterations beyond two increase the χ^2.

Since both the MSE and the χ^2 show optimization at two iterations, two is the number of iterations that is used in the analysis.

6.1.2 Unfolding Results

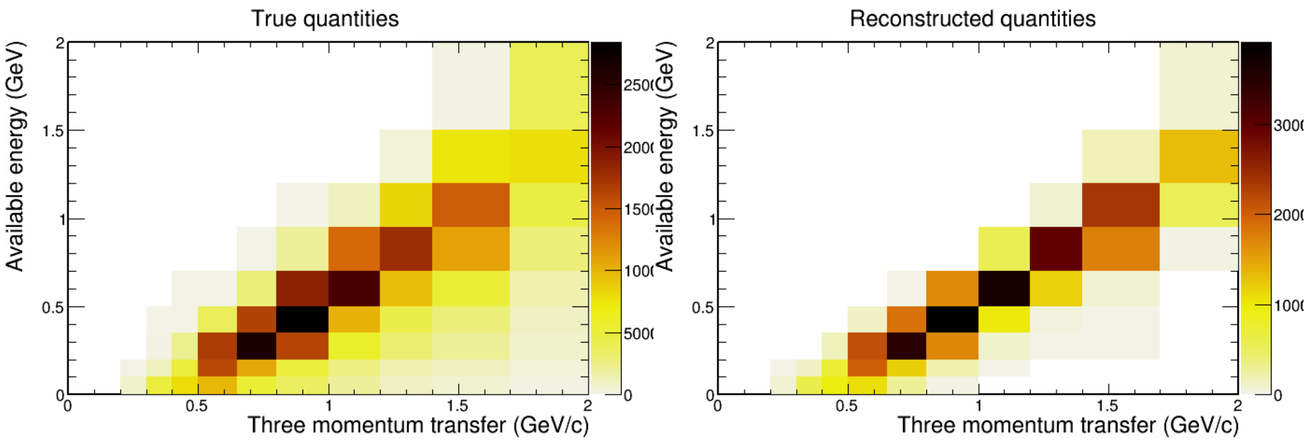


Figure 6.4: True distribution (left) and the reconstructed distribution (right) for a sample that is statistically and systematically independent from that used to create the reconstructed-to-true matrix. Both distributions have most events along $E_{avail} \approx |\vec{q}|$ however the true distribution has more spread, while the reconstructed distribution has most of the events concentrated along the diagonal.

With the unfolding optimized at two iterations, it is important to compare the output distributions. For these comparisons a random systematic universe that is independent from the default Monte Carlo was chosen. Figure 6.4 shows the true distribution on the left and the reconstructed distribution on the right. The two distributions are similar but not the same, the true distribution is wider, while the reconstructed one is concentrated at $|\vec{q}| \approx E_{avail}$.

The distribution after two iterations of unfolding, shown in Fig. 6.5, looks much closer to the truth distribution of Fig. (6.4) (left).

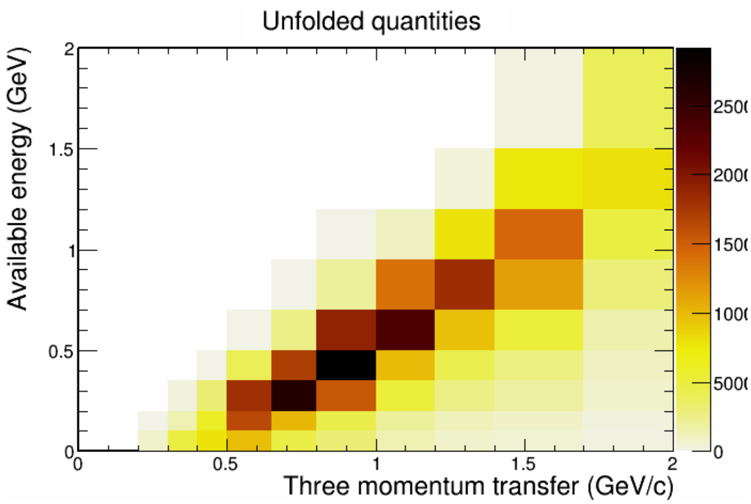


Figure 6.5: $|\vec{q}|$ versus E_{avail} after unfolding. After unfolding the distribution is very similar to the corresponding true distribution.

6.1.3 One dimensional unfolding results

The one-dimensional distributions of $|\vec{q}|$ and E_{avail} are also of interest. These distributions are obtained by taking projections of the two-dimensional distribution. Figure 6.6 shows the results obtained from a systematically shifted universe for which unfolding is carried out using the central value migration matrix. Figure 6.6 (left) shows the event distribution in $|\vec{q}|$ and Fig. 6.6 (right) shows the bin-by-bin ratio to the truth distribution. The $|\vec{q}|$ distributions all follow the same trends. The ratio plots show that the unfolded distributions is closer to the truth in all bins. At high values of $|\vec{q}|$ the reconstructed distribution is especially distorted; fortunately the unfolding remedies the situation.

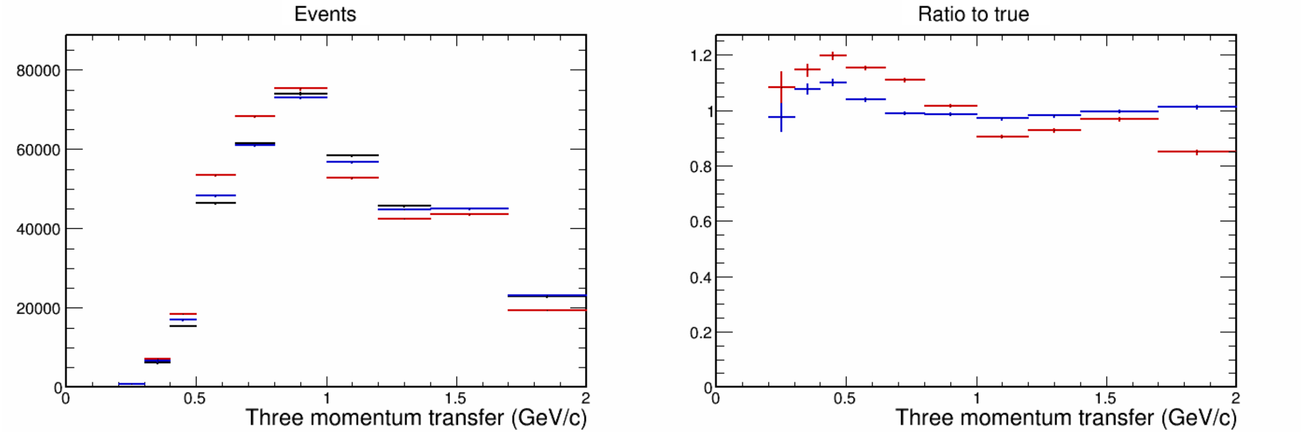


Figure 6.6: Distribution of $|\vec{q}|$ events (left) and ratios to truth (right). The reconstructed distribution (red) is farther away from truth (black) than the unfolded distribution (blue). The unfolded distribution (blue) is seen to be closer (ratio near 1) than the reconstructed distribution (red) in all bins.

The E_{avail} event distribution in the left-hand plot Fig. 6.7 shows that the reconstructed (red), unfolded (blue), and truth (black) distributions follow the same trends, but the ratios to truth in the right-hand plot, show that the unfolded distribution is much closer to the actual truth. The unfolding correction in the highest E_{avail} bin is particularly impressive.

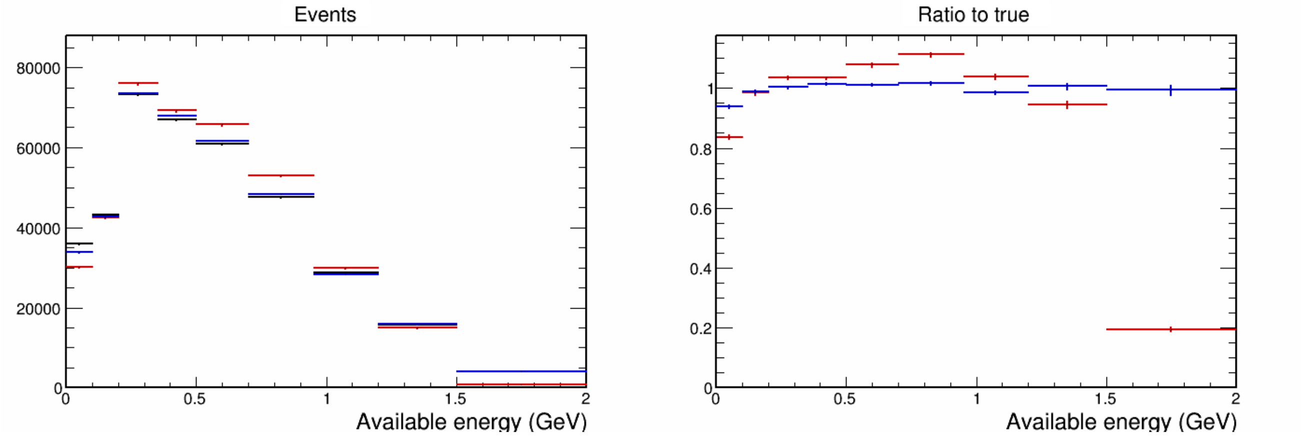


Figure 6.7: E_{avail} event distributions (left) and ratios to truth (right). The reconstructed distribution (red) is further away from the truth distribution (black) than the unfolded distribution (blue), but both follow the same trend, the distribution of ratios (right) shows improvement with unfolding in all bins.

Chapter 7

Cross section determination and validation of analysis framework

7.1 Cross section calculation

The double differential cross section with respect to $|\vec{q}|$ and E_{avail}, $\frac{d\sigma^2}{d\vec{q}dE_{avail}}$, is calculated as follows:

$$\left(\frac{d\sigma^2}{d|\vec{q}|\,dE_{avail}}\right)_{ij} = \frac{\sum_{\alpha\beta} U_{ij,\alpha\beta}\left(N^{Data}_{\alpha\beta} - N^{Bkgd}_{\alpha\beta}\right)}{\epsilon_{ij}\,(\Phi_\nu T_N)\,(\Delta|\vec{q}|)_i(\Delta E_{avail})_j}. \tag{7.1}$$

The quantity $N^{Data}_{\alpha\beta}$ is the number of selected data events, and $N^{Bkgd}_{\alpha\beta}$ is the number of estimated background events that is subtracted from the data to get the estimated signal. The unfolding matrix, $U_{ij,\alpha\beta}$, converts reconstructed bins (α, β) to unfolded bins (i,j); ϵ_{ij} is the efficiency correction in the ($|\vec{q}|$,E_{avail}) bin designated by (i,j), Φ_ν is the integrated neutrino flux, T_N is the number of nucleons in the fiducial volume, $(\Delta|\vec{q}|)_i$ and $(\Delta E_{avail})_j$ are the widths of the 2D bin (i,j).

7.2 Analysis framework

Validation is done by using the analysis package to analyze a set of simulated data with a known cross section. Distributions obtained by the package in available energy and three-momentum transfer are compared to the corresponding distributions obtained using Monte Carlo truth. The output distributions of the analysis package after unfolding and efficiency corrections are observed to replicate the input truth distributions almost exactly. Previously we carried out an In-Out Test of the standard cross section analysis package for the determination of cross section as a function of a single variable, namely $d\sigma/dE_\nu$ [38]. We determined from this simpler test that the package can reproduce $\sigma(E_\nu)$ for the so-called Dytman MEC cross section [45]. This analysis uses a modified version of the standard NOvA cross section analysis package, *crosssectionanalysis.h* [42], that has been changed to return a double differential cross section instead of a single differential one in neutrino energy.

The goal of an In-Out Test is ensure that an analysis package (AP) returns the underlying cross section of the data that it is being used to analyze. The AP processes the simulated data to produce a true signal event distribution in the following sequence:

1. The AP creates a distribution of events that pass the selection cuts from the data sample.
2. The AP then makes a distribution using the reference sample (the central value MC), of events that pass the selection cuts, but are not part of the signal definition.
3. The package then subtracts the distribution from step 2 from the distribution made in step 1. The result is a distribution of true signal events that pass the selection cuts.
4. The package unfolds the distribution obtained in step 3 using one iteration of the iterative unfolding method [40] to map reconstructed variables to truth variables.
5. The AP divides the result of step 4 by the detection efficiency to produce a distribution of true signal events.

The resultant true signal event distribution from the AP, is then compared to a distribution of true signal events made using Monte Carlo truth, in order to determine whether or not the AP produces the correct distribution.

In order to compare the output of the AP to the Monte Carlo truth, the two distributions are divided bin-by-bin, and the result is shown on the left of Fig. 7.1. The ratio is 1.0 everywhere the distribution is populated, seen on the right in Fig. 7.1, which shows that the AP successfully reproduces the underlying true Monte Carlo distribution.

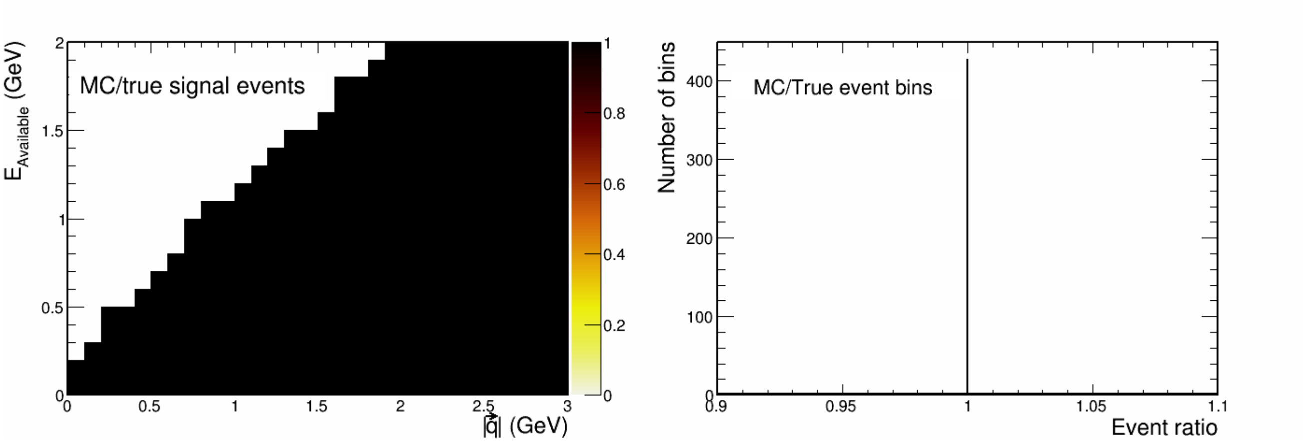


Figure 7.1: Ratio of AP output to MC truth (left), and the value of each bin (right). The ratio of the output of the AP to the MC truth is 1.0 in all bins that are populated. This shows that the AP reproduces the MC truth.

The In-Out test described above utilizes GENIE empirical MEC as its test channel and it is carried out for a double-differential cross section. The test cannot directly reproduce the input GENIE spline which is stored in neutrino energy. Consequently we have also carried out a total cross section In-Out test (also using GENIE empirical MEC). The only difference between this test and the one outlined above is that this test is carried out using E_ν instead of $|\vec{q}|$ and E_{avail}. This latter test allows the AP output to be compared directly to the input cross-section spline used by GENIE. The comparison to the GENIE spline directly checks the AP use of flux estimation and target counting. The results of the total cross section In-Out test are displayed in Fig. 7.2. The AP data points (black crosses) are seen to lie on the GENIE cross-section spline (red curve).

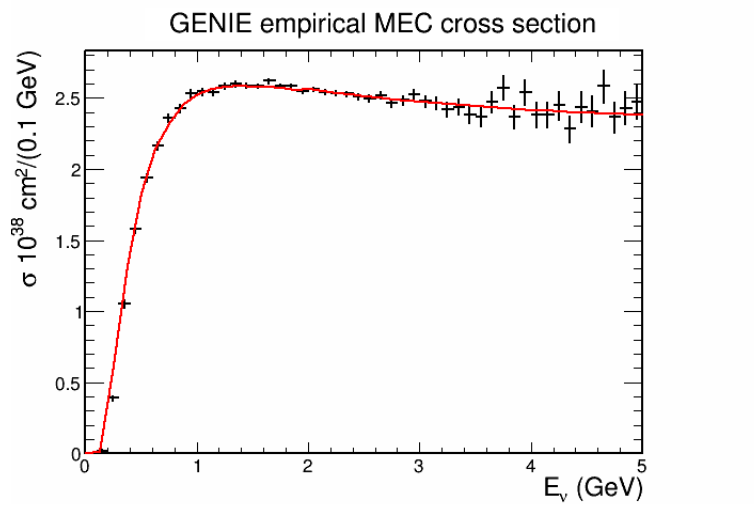


Figure 7.2: Comparison of the AP prediction for the GENIE empirical MEC cross section versus neutrino energy (black crosses) to the input GENIE spline (red curve), as obtained in the total cross section In-Out test.

Since the AP has passed the In-Out-Test, it is ready to use to analyze data.

Chapter 8

Estimation of Systematic Uncertainties

8.1 Systematic uncertainties

8.1.1 The multi-universe method

The multi-universe method for determining the total uncertainty range, has been used previously in NOvA by the ν_μ-CC inclusive cross section analysis [46]. The method involves randomly varying all of the systematic parameters to create a new, reference simulation. In the new MC simulation the background estimate is altered, as are the unfolding matrix, efficiency correction, and flux estimation for the cross section calculation. Consequently the new simulation implies a new cross section for this particular universe. The cross sections, calculated using the alternate MC samples, are then compared to the nominal MC used by the analysis. The error band is constructed taking the root mean square (r.m.s.) of the bin-by-bin upward and downward excursions as they occur in the ensemble of universes, to create an error band, that may, in general, be asymmetric.

The method is illustrated here with a toy model example. Consider a nominal prediction for a parameter whose value near 2.0 (in some units) for each of five independent bins as shown in Fig. 8.1a. The test parameter is then randomly altered by rolling each bin with a Gaussian of mean 1.0 and width 2.0 to produce a value that characterizes a different universe. The outcomes from an ensemble of 25 universes are shown in Fig. 8.1b. The uncertainty band is constructed by taking the r.m.s. of the upward excursions and separately for the downward excursions and using those values to define the excursion of the error band. Figure 8.1c shows the error band obtained using 25 universes. In the large N limit of universes the error bands of this simple example would presumably converge to the value of 1.0. In the analysis of this work there are 140 systematic parameters that must be randomly changed for each universe in order to estimate the total uncertainty associated with the cross section measurement.

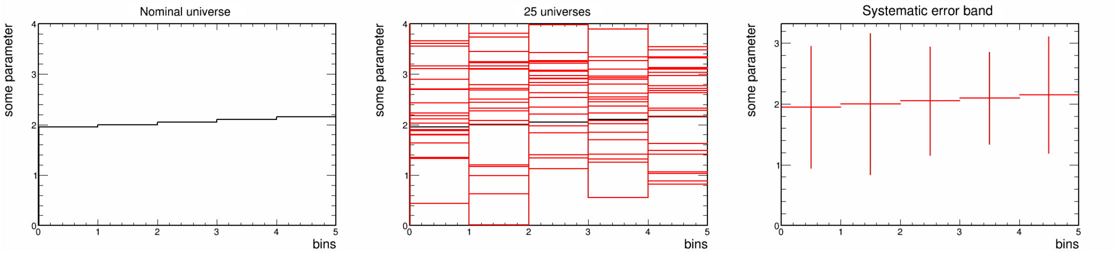


Figure 8.1: Toy model of the multi-universe method: Consider five bins of independent data, where each bin initially has a value near 2.0 shown in the left plot. Another “universe” is generated by randomly altering the content of each bin by drawing from a Gaussian of mean 0 and $\sigma = 1.0$. Values obtained with 25 universes are shown in the middle plots. With generation of a large number of universes the uncertainties for each bin should equal the σ of the Gaussian used to generate the universes.

8.1.2 Non-2p2h-MEC systematic uncertainties

There are 99 sources of systematic uncertainly that are relevant to the double-differential cross section measurement. The parameters associated with these sources can be grouped into four categories, as follows: (1) parameters associated with neutrino-nucleus interaction modeling; (2) parameters associated with the neutrino flux; (3) parameters associated with estimation of energy deposited by particles in the detector; (4) and parameters associated with the calibration constants of the Near Detector.

8.1.2.1 Neutrino nucleus modeling

The reference MC is used to estimate backgrounds, correct for efficiency losses, and construct the unfolding matrix. Consequently uncertainties in the parameters of GENIE modeling propagate to the error band of the measurement. The systematic uncertainty sources of this category are the parameters used by GENIE to model neutrino-nucleus scattering [19]. These include neutrino cross section parameters, neutrino reaction final-state hadronization, and intranuclear hadron transport, of mesons and nucleons.

8.1.2.2 Neutrino flux modeling

Sources of uncertainty associated with the forward horn current neutrino flux include beam focusing, beam transport, secondary modeling, hadron production uncertainties, and PPFX fractional uncertainty. These uncertainties have been aggregated using a principal component analysis, “PCA” [47]. This analysis uses the Near-Detector-only PCA, which does not consider the Far Detector at all.

8.1.2.3 Energy deposition scale

There are six sources of uncertainty in this category. The most significant one for this analysis is the amount of visible energy deposited by neutrons. Visible hadronic energy is used to estimate E_{avail} and changing the amount of visible energy from neutrons affects that estimate. Of the remaining systematics, two correspond to the amount of muon energy deposited in the active region and in muon catcher of the Near Detector [48]. A fourth systematic is the amount of muon energy deposition that is correlated between the two regions. There is also an uncertainty for change in energy deposition resulting from multiple overlapping muon tracks. Finally, there is an uncertainty associated with whether or not it is appropriate to apply Coulomb corrections to the muon energy deposition, which NOvA currently does not.

8.1.2.4 Detector light calibration

There are four sources of systematic uncertainty that arise from the calibration of the detector. These are, positive or negative offsets in the calibration, and variations in the shape of the calibration. There are also uncertainties in the amount of scintillator light expected from particles, including the uncertainty associated with the parameters of Birks’ empirical formula [49]. This uncertainty is accounted for by changing the amount of light in the simulation and then recalibrating the reconstruction to that different amount of light. The light yield from protons was measured with the JINR test stand [50] and so the uncertainty associated with Birks' parameters is small. Finally, there are uncertainties in the amount of Cherenkov light produced by particles moving faster than the speed of light in the liquid scintillator. The systematic uncertainties are estimated by generating new MC simulations. This approach is not compatible with the multi-universe approach, and so detector calibration systematics must be added in quadrature to the results from the multi-universe method.

8.1.2.5 Nucleon target counting

Estimation of the number of target nucleons is documented in Section 8 of DocDB 32688 [34]. The mean target count was found to be $5.689 \times 10^{31} \pm 1.785 \times 10^{26}$ nucleons. The error is seen to be negligible compared to the error sources listed above.

8.1.3 Effect of selections on the neutrino energy spectrum

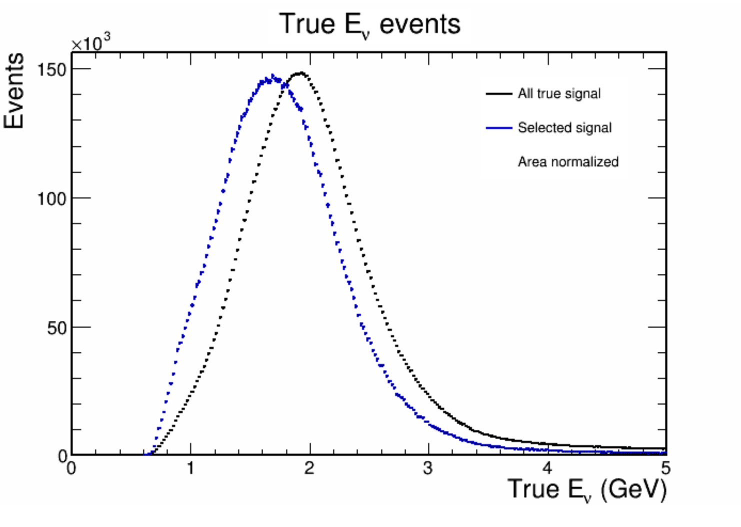


Figure 8.2: True signal events (black), and selected signal events (blue) versus true neutrino energy. The selections cuts, particularly containment, reduce mean value of E_ν for the sample and shift the distribution towards lower values.

The event selection cuts have an effect on the initial, uncut neutrino energy spectrum. Figure 8.2 compares the true neutrino energy distribution of all true signal events (black) to the E_ν distribution of true signal events after event selection (blue). The distributions are shown area-normalized to each other. The event selections lower the neutrino energy peak and increase the low end of the distribution, while decreasing the high-energy tail. These effects arise primarily from the muon containment cut and are to be expected, as lower energy muons are more likely to be contained.

8.1.4 Dominant uncertainty sources

The sources of error have been ranked according to the fractional change on the integrated cross section. Five sources were found to contribute at greater strength than the remaining 94. These sources are as follows:

1. The flux modeling (10.9%).
2. The axial mass of the axial vector form factor for CC baryon resonance events (MaCCRES) (1.1%).
3. The axial mass of the axial vector form factor for CC quasi-elastic events (MaCCQE) (0.7%).
4. Shape of random phase approximation (RPA) enhancements of $|\vec{q}|$, q_0 distributions. (RPA shape: enh2019) (0.7%).
5. The vector mass of the vector form factor for CC baryon resonance events (MvCCRES) (0.6%).

We note that the effects of uncertainty from detector calibration systematics cannot be quantified using the above-mentioned fractional change. Rather, detector calibrations introduce bin-by-bin uncertainties in the differential cross section. These are be accounted for by stating the bin-by-bin uncertainties for the final differential cross section. A summary of the total systematic uncertainty is presented after discussion of the 2p2h-MEC systematic (see Table 8.1).

8.2 Uncertainty associated with 2p2h-MEC

A systematic error associated with 2-particle 2-hole, meson exchange currents (2p2h-MEC) is required for the $d^2\sigma/d|\vec{q}|dE_{avail}$ ν_μ-CC inclusive cross section analysis. The uncertainties affect the event populations in kinematic regions where the efficiency changes rapidly and thus may change the unfolding calculations. The NOvA oscillation analyses include a treatment of these systematics, which factor into the shape of the distributions in energy transfer and three-momentum transfer, the ratio of final state nucleons, and the energy dependence on the total cross section. Unfortunately the momentum and energy transfer shape systematic is based on the Near Detector data, making it unusable for an analysis that is using the same data. Instead of using the oscillation systematics, alternate models of 2p2h-MEC are used to determine the range of possible cross-section distortions in the analysis. The alternative models are, GENIE empirical 2p2h [19], Valencia 2p2h [51, 52], three MINERvA tunes to Valencia 2p2h [5, 26], and the Superscaling (SuSa) 2p2h-MEC model (version 2) [17, 24, 18].

8.2.0.1 GENIE empirical 2p2h

GENIE empirical 2p2h is modeled as a two-dimensional Gaussian over the phase space defined by $|\vec{q}|$ and q_0, with a peak in event rate occurring between the quasielastic and $\Delta(1232)$ scattering peaks. The strength of the 2p2h-MEC process is set to agree with the MiniBooNE and NOMAD data. This model was originally motivated by knowledge of 2p2h enhancements in electron-nucleus inelastic scattering [45].

8.2.0.2 Valencia 2p2h

Valencia 2p2h is a microscopic model developed by J. Nieves and collaborators. The model invokes a many-body expansion of the gauge boson self-energy tensor using QFT formalism [4]. This is a semi-relativistic model whose range in $|\vec{q}|$ is limited to below 1.2 GeV/c.

8.2.0.3 MINERvA tunes of the Valencia model

The MINERvA Collaboration has carried out a fit to their data that applies a two-dimensional Gaussian to the kinematic region where a data excess appears relative to the Valencia model [25, 26]. MINERvA has reported three separate cross section fits. In their reference fit, the shape of 2p2h distribution according to Valencia is fit to the data; in that fit, the final state nucleon fraction, $R_N = (np \to pp)/(nn \to np)$, is set to the value 2.8/1.0. In the other two tunes, the final state nucleon pair is set to either only *pp* or to *np*. The latter fits are intended to quantify the possible spread inherent to uncertainty with R_N which cannot be directly calculated in the Valencia model.

8.2.0.4 Superscaling MEC

Superscaling refers to a phenomenological scaling algorithm first observed in electron nucleus scattering data. The model has been adapted for use with neutrino nucleus scattering and further refined to include MEC processes [24, 18].

8.2.1 Comparison of models

The models vary in their theoretical pedigree, hence some are to be taken more seriously than others. It is useful nevertheless to consider the similarities and differences among all of the models. The analysis developed here is carried out in the phase space defined by three-momentum transfer, $|\vec{q}|$, available energy, E_{avail}, and neutrino energy, E_ν. Each of the models provides a prediction for the distribution of 2p2h-MEC in these three kinematic variables.

To better understand the performance of the different models and how they effect the analysis, it is useful to examine and compare how events are distributed in $|\vec{q}|$ and E_{avail} within each model. All of the models have a similar dependence on E_ν, as will be shown in subsection 11.1.1. Figure 8.3 displays predictions of the GENIE empirical 2p2h model (top left), the Valencia model (top right), and the SuSa model (bottom). The empirical 2p2h prediction is concentrated heavily at 0.5 GeV/c in $|\vec{q}|$ and 0.3 GeV in q_0, and the strength falls quickly away from the peak. The

Valencia model predicts a peak at higher $|\vec{q}|$ and q_0 values at 0.7 GeV/c and 0.5 GeV respectively, and has a lower peak and total cross section. Valencia 2p2h, which is based on a non-relativistic calculation, is limited to $|\vec{q}| < 1.2$ GeV/c and does not make a prediction beyond that cutoff value. The SuSa model predicts a peak roughly in the same place as GENIE empirical 2p2h, but the tail of the distribution extends much further.

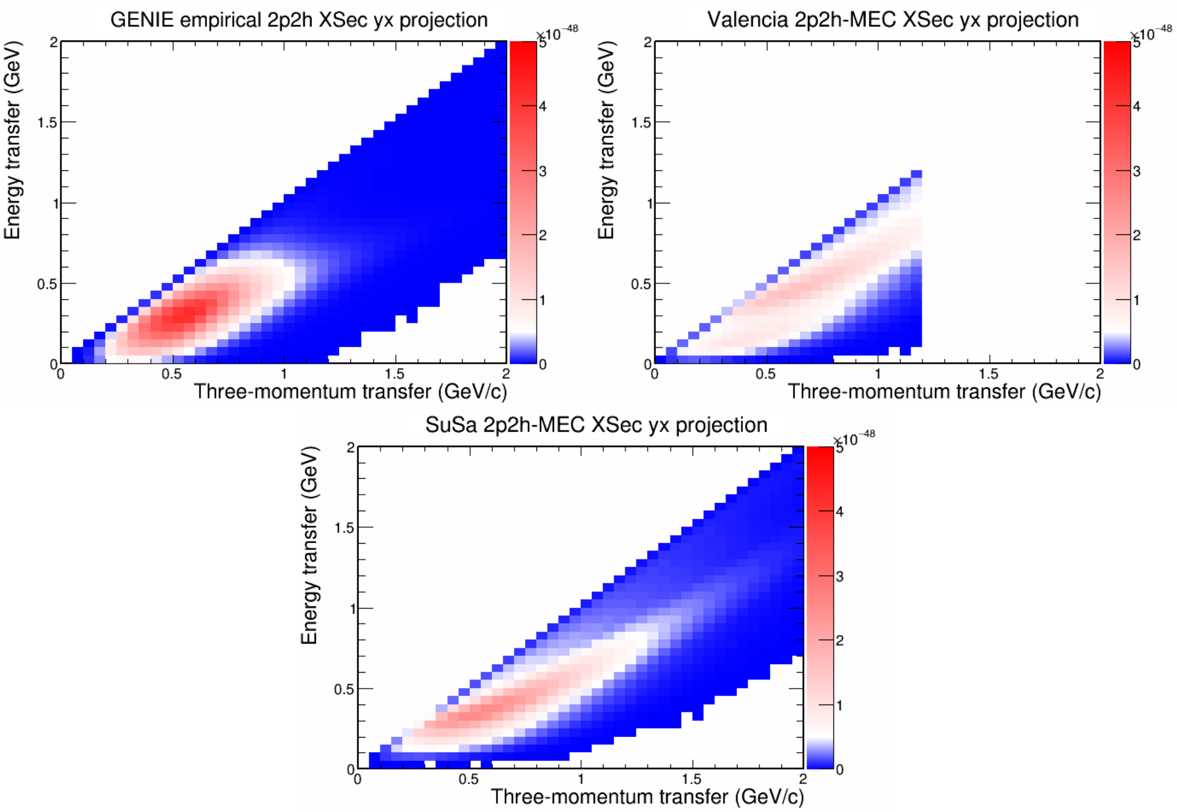


Figure 8.3: Cross section predictions of GENIE empirical 2p2h (top left), the Valencia model (top right), and the SuSa 2p2h-MEC model (bottom). The empirical and SuSa predictions place the peak of the distribution in the same location, 0.5 GeV/c in $|\vec{q}|$ and 0.3 GeV in q_0. SuSa predicts a more extended distribution than do the other models, while empirical 2p2h falls of very quickly. Valencia predicts a peak at higher $|\vec{q}|$ and q_0 and a much lower overall rate.

All of the models predict different ratios of final state nucleons pairs, and some models predict different cross section distributions depending on the nucleon pairs involved in the reaction. In Fig 8.4 the $np \rightarrow pp$ cross section predictions are shown with GENIE empirical on top left, followed by Valencia in the top right, and SuSa on the bottom. Figure 8.5 shows the $nn \rightarrow np$ cross sections in the same order. The empirical model has a value of $R_N = 4.0$, otherwise the pp and np distributions are identical. Valencia 2p2h has $R_N = 2.8$, which is the lowest of the three models. The pp final state of the Valencia model exhibit a double-peak structure with a higher peak at higher $|\vec{q}|$ and q_0 , while the np final state also has two peaks, but the peaks have similar magnitude. The SuSa model prediction of $R_N = 7.4$ for the final states is much higher than for the other models. In the np final state, SuSa also predicts a double peak, however the pp final state gives only a single peak.

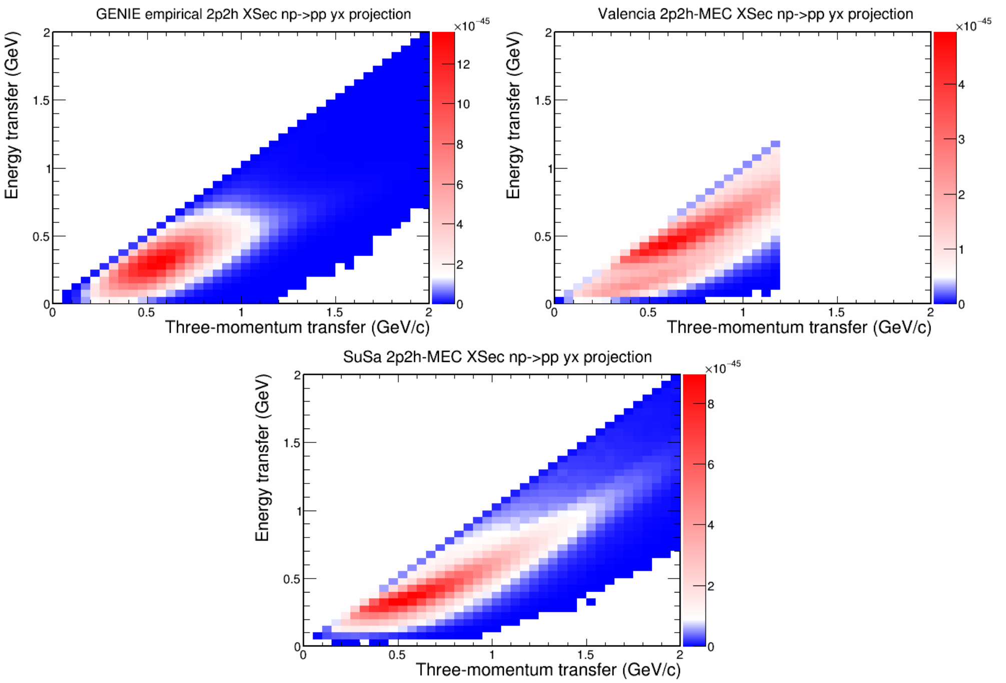


Figure 8.4: Predictions of the pp final state cross sections for GENIE (top left), Valencia (top right), and SuSa 2p2h-MEC models (bottom). The shape predicted by GENIE empirical 2p2h is independent of the identity of the final state nucleon pair. Valencia actually gives a double peak however the second peak is not discernible in this figure. SuSa has a single peak in the pp final state.

Figures 8.6, 8.7, and 8.8 show the cross section predicted by each model for q_0, $|\vec{q}|$, and E_ν respectively, for the two pp final states (left-side plots), and np final states (right-side plots). In these single-variable histograms the Gaussian shape assumed for empirical MEC in q_0 and $|\vec{q}|$ is clearly visible. The Valencia double peak is a feature of the q_0 distributions (Fig. 8.6) but not of $|\vec{q}|$ distributions (Fig. 8.7). It is easiest to see the relative strengths of the models in Fig. 8.8 as all three models have roughly the same shape. Empirical MEC has the largest cross section, the SuSa model cross section is roughly 80% of empirical MEC, while the Valencia model cross section is 60% of the empirical MEC.

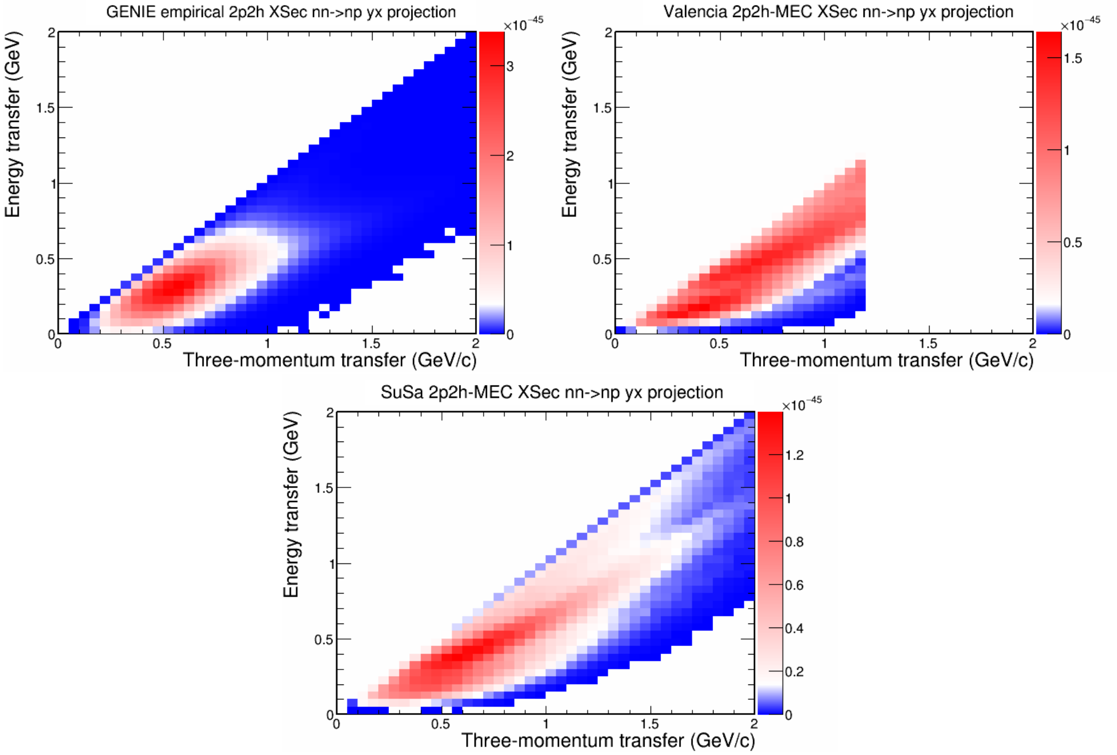


Figure 8.5: Predictions of final state cross sections for GENIE empirical 2p2h (top left), Valencia 2p2h (top right), and SuSa 2p2h-MEC models (bottom). GENIE empirical has the same shape for *np* as for *pp* final states, but a smaller rate. Valencia 2p2h has a double peak and broad coverage of the phase space. SuSa 2p2h-MEC has a secondary peak at lower E_{avail}, which is absent in the *pp* final states.

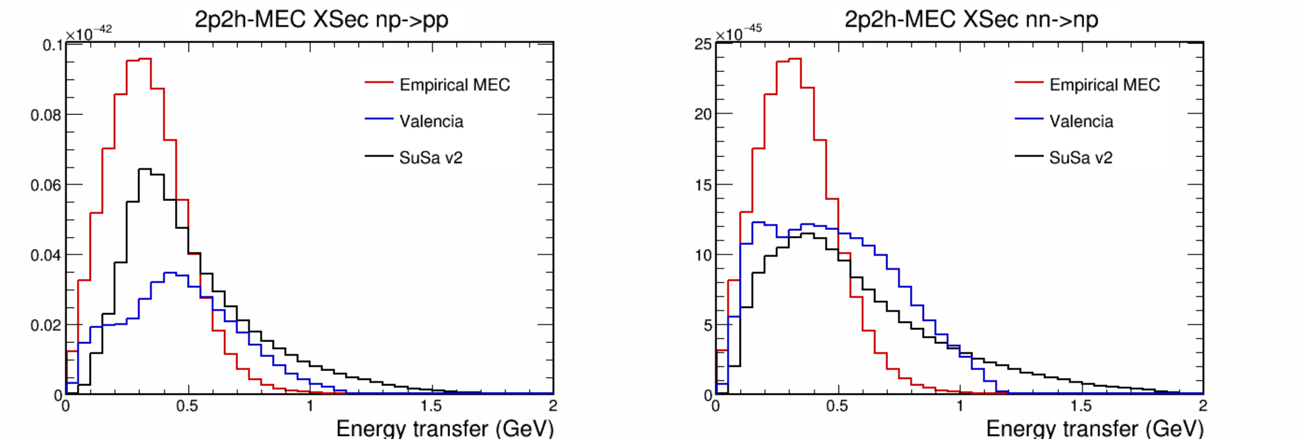


Figure 8.6: Flux-integrated cross sections for 2p2h versus energy transfer predicted by GENIE empirical (red), Valencia (blue), and SuSa (black). The *pp* final state projection (left) shows that empirical 2p2h and SuSa peak at a similar q_0 but SuSa has much lower strength, while the Valencia model peaks higher and has a shoulder at lower q_0. For *np* final states (right) empirical 2p2h has a more pronounced peak than the other two models.

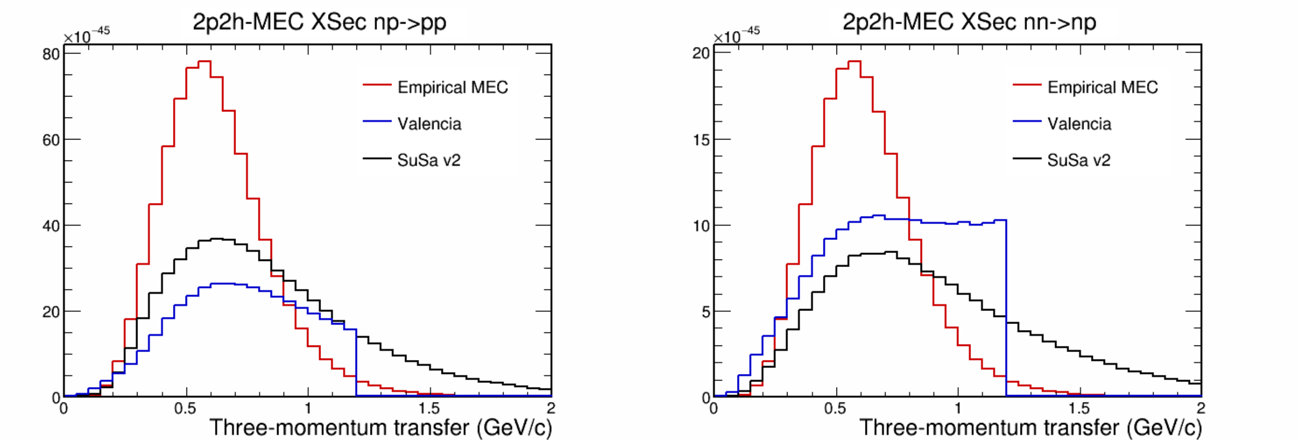


Figure 8.7: Flux-integrated cross sections for 2p2h versus three-momentum transfer predicted by GENIE empirical 2p2h (red), Valencia (blue), and SuSa (black). Empirical 2p2h peaks higher and has a narrower distribution in $|\vec{q}|$ in both the pp (left) and np (right) final states. The other two models predict broader distributions with smaller peaks.

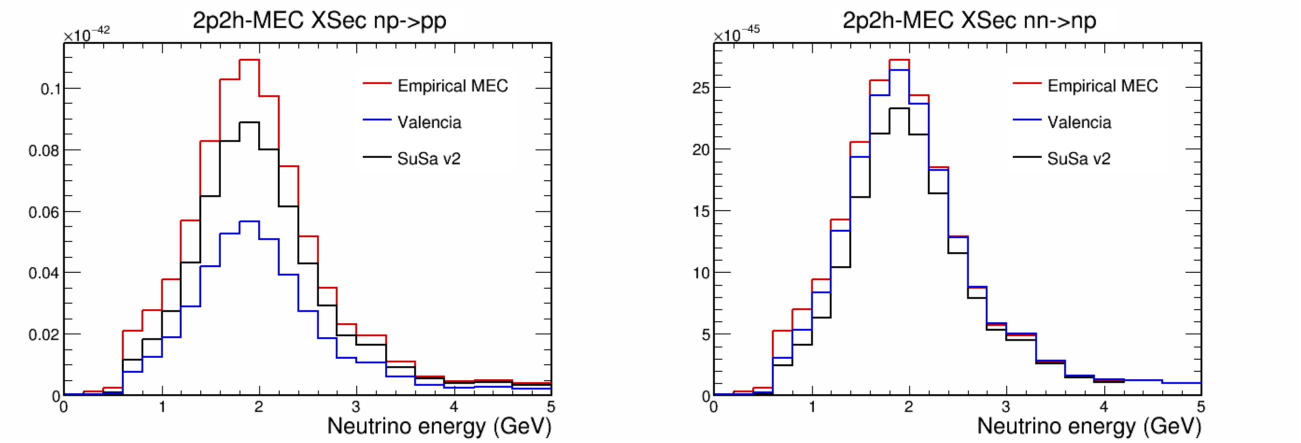


Figure 8.8: Flux-integrated cross sections for 2p2h versus neutrino energy predicted by GENIE empirical 2p2h (red), Valencia (blue), and SuSa (black). Empirical 2p2h peaks higher and has a broader distribution in E_ν for both the pp and np distributions. All three distributions have a similar shape.

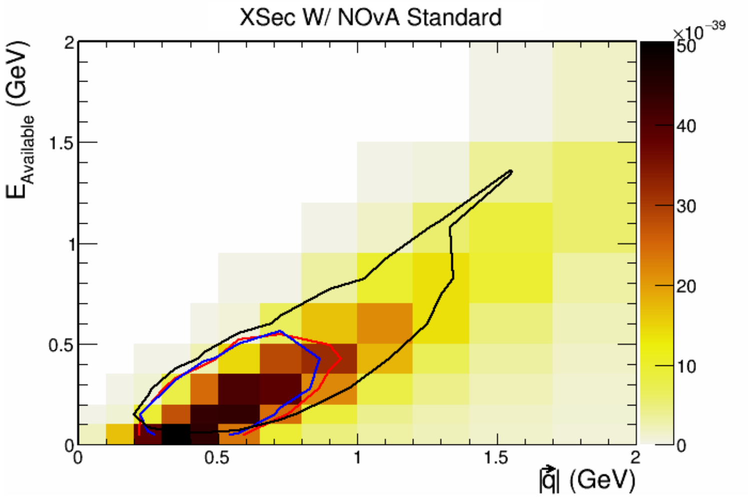


Figure 8.9: Contours containing 75% of the total 2p2h-MEC rate, for of the empirical 2p2h, Valencia 2p2h, and SuSa 2p2h-MEC overlaid onto the inclusive cross section. The empirical 2p2h (red) and Valencia 2p2h (blue) have similar contours. The SuSa MEC-2p2h has a much larger contour with an irregular shape in the regions of higher phase space.

The models cover different regions of $|\vec{q}|$ and E_{avail}. Figure 8.9 shows the inclusive cross section with contours outlining the 75% level of the 2p2h cross section predictions. All three of the cross section models peak roughly the same region of phase space in the vicinity of 0.5 GeV/c of $|\vec{q}|$ and 0.3 GeV of E_{avail}. Empirical 2p2h (red), and Valencia 2p2h (blue) have similar contours that extend from 0.2 GeV/c up to 0.8 GeV/c in $|\vec{q}|$ and 0.1 GeV to 0.5 GeV in E_{avail}. The SuSa 2p2h-MEC model (black) extends much further in both $|\vec{q}|$ and E_{avail} up to 1.6 GeV/c and 1.4 GeV respectively. The Superscaling model predicts a low cross section between 0.3 GeV/c and 0.6 GeV/c of $|\vec{q}|$ below 0.1 GeV of E_{avail} as shown by the closing of the contour in that region, which the other two models do not do. The SuSa model also has a more irregularly shaped distribution at higher $|\vec{q}|$ and E_{avail} .

8.2.1.1 Use of weights to render models

Weights are created to mold the GENIE empirical 2p2h of the NOvA Monte Carlo into the other models. Weights in q_0 and $|\vec{q}|$ and also in E_ν as well, are required to model all of the relevant variables and to correctly reproduce the input GENIE splines. The required behavior in E_ν is illustrated in Fig. 7.2 for the GENIE empirical MEC model. Figure 8.10 shows the

3D distributions of weights in the variables, $|\vec{q}|$, q_0, and E_ν. The plot attempts to use color to indicate the strength of the weight at a given value of the input variables, but that aspect is obscured by the density of event points in the plot.

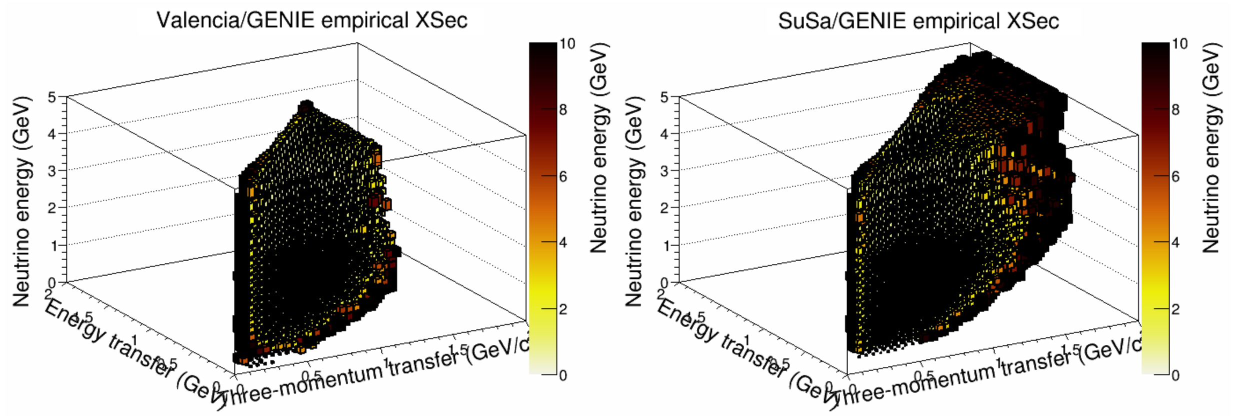


Figure 8.10: Cross-section weights that change the empirical 2p2h model to the Valencia (left) and Susa (right) 2p2h-MEC models. The vertical axis displays the E_ν coordinate, while the horizontal coordinates are q_0 and $|\vec{q}|$. The analysis accesses the weights in lookup table fashion.

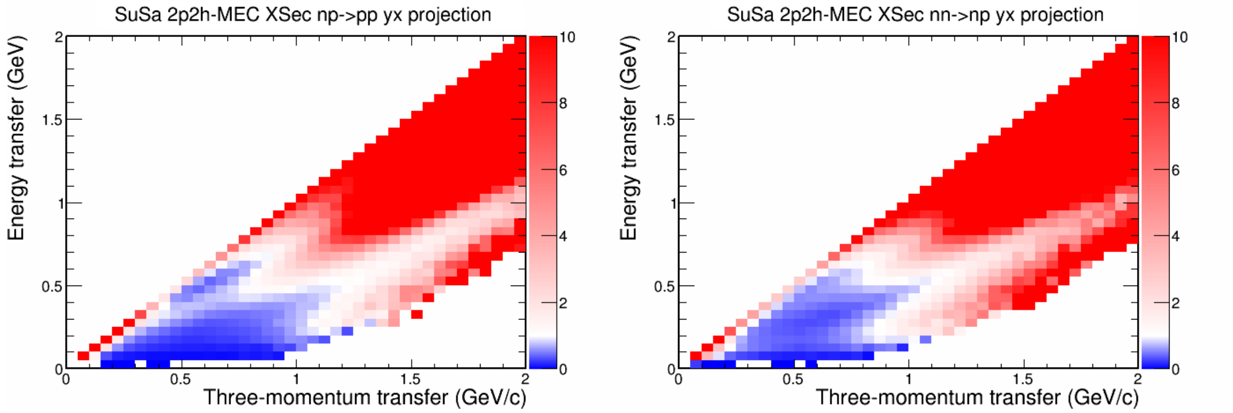


Figure 8.11: Cross section weights in only q_0 and $|\vec{q}|$ that change the empirical 2p2h model to the SuSa 2p2h-MEC model. The *pp* final state (left) is down-weighted at lower q_0 and is weighted upwards at higher q_0. The *np* final state weight (right) has a similar shape but the down weighting is weaker.

Trends can be seen by viewing the projections of the weights into the $|\vec{q}|$ and q_0 plane. The cross-section weights developed for this analysis are illustrated using the SuSa model in Fig. 8.11. The left-hand plot shows the weights for the *pp* final state. The final state is weighted down at low energy transfer and weighted upwards at energy transfers greater than 1 GeV. Between 0.4 and 0.8 GeV of energy transfer the weight is nearly 1.0. Figure 8.11 (right) shows the cross section weight for the np final state. The cross section is down-weighted in the region with less than 0.6 GeV/c of three-momentum transfer and less than 0.5 GeV of energy transfer. However events having more than 0.8 GeV of q_0 are weighted upwards. The Gaussian shape of empirical MEC gives a rapid falloff of cross section as one moves away from the central peak, while the other models retain their strength.

It is important to note how the weighting from empirical MEC to other models and cross-section tunes effects the cross section in neutrino energy, which is shown in Fig. 8.12. The cross section of the various models and tunes being reconstructed from weighting GENIE empirical MEC are shown as crosses, while the splines for the input models are shown as solid lines. Without the E_ν weighting the cross sections tend to rise in an unphysical way with energy. The GENIE empirical MEC (red), reproduces the spline that it was made from. The Valencia model (blue), which also reproduces its spline, has the smallest overall cross section, and reaches its turn over in E_ν more gradually than the empirical MEC. The SuSa model (black) turns over the slowest and has a cross section between the Valencia, and empirical models. The SuSa model similarly lines up with the input spline. Both the NOvA (green) and MINERvA (pink) experimental tunes increase the overall cross sections of their base models, but retain their shapes.

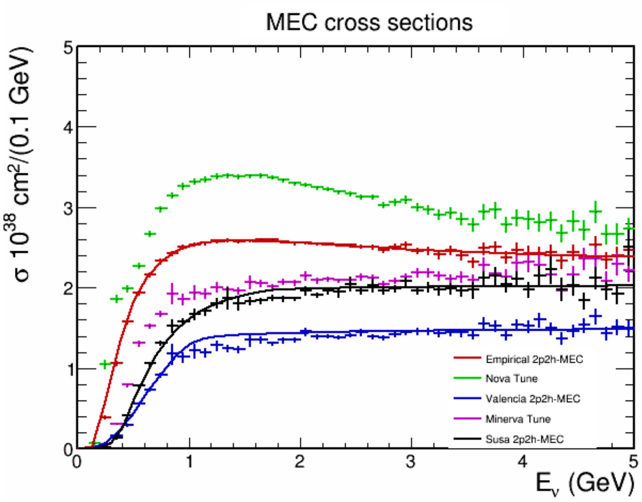


Figure 8.12: The various MEC models and tunes reconstructed from weighting the GENIE empirical MEC (crosses) and the GENIE splines for the models (solid lines). The empirical MEC (red) lies directly on the spline. The SuSa (Black) and Valencia (blue) models come close to their splines. Both the NOvA (green) and MINERvA (pink) tunes increase the overall cross section of their base models.

8.2.2 Establishing the systematic uncertainty for 2p2h-MEC

To assign the systematic uncertainty, the inclusive cross section is estimated using the NOvA -tuned GENIE empirical 2p2h standard Monte Carlo as pseudo-data and a statistically independent sample of Monte Carlo as a reference simulation. The cross section is then calculated using a different MEC model as the reference Monte Carlo. Figure 8.13 shows the cross section predictions made using the NOvA -tuned empirical 2p2h (top left), MINERvA -tune Valencia 2p2h (top right), and SuSa 2p2h-MEC (bottom). The predicted cross sections are similar in the regions of relatively high cross section, but vary greatly along the kinematic ridge, as shown by the cross section ratios to NOvA -tuned empirical 2p2h in Fig. 8.14.

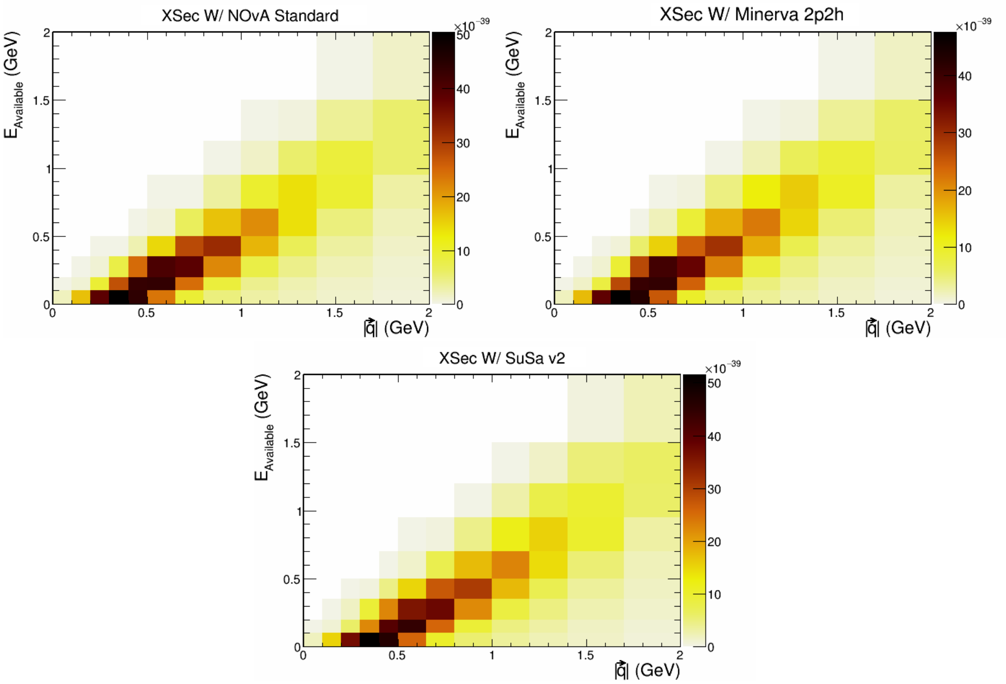


Figure 8.13: Inclusive cross section estimated using NOvA tuned MEC (top left), Valencia 2p2h (top right), and SuSa 2p2h-MEC (bottom) in the reference Monte Carlo. All three estimated cross sections have a peak around 0.1 GeV of E_{avail} and 0.4 GeV of $|\vec{q}|$. The high cross section ridge extend from the peak up to 0.3 GeV of E_{avail} and 0.7 GeV of $|\vec{q}|$.

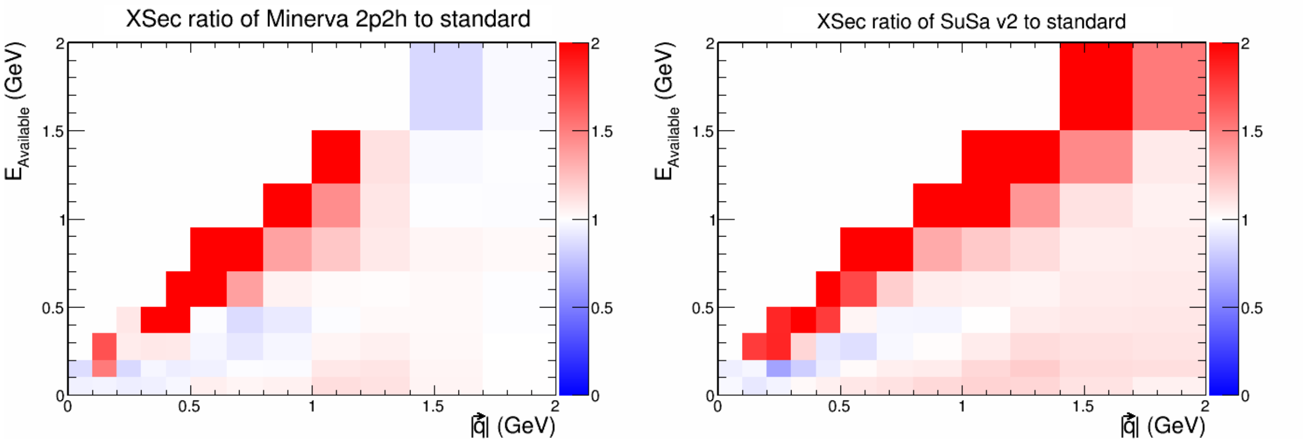


Figure 8.14: Ratios of cross sections for the MINERvA-tuned Valencia model (left) and for the SuSa model (right) to the cross section predicted by NOvA -tuned empirical 2p2h. The ratios are nearly 1.0 in all regions except along the kinematic ridge where the ratios exceed 2.0.

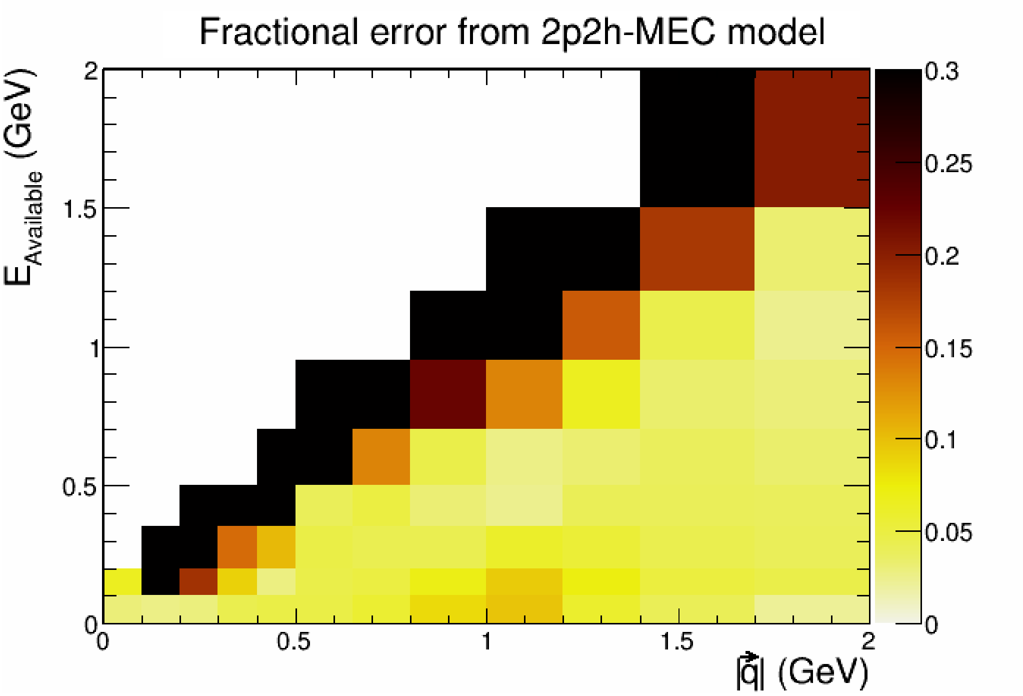


Figure 8.15: Fractional systematic error from 2p2h-MEC modeling for the predicted cross section. The uncertainty is around 10% in areas away from the kinematic ridge. Along the kinematic ridge the uncertainty exceeds 30%.

In view of current world knowledge about 2p2h-MEC, the MINERvA tune to Valencia 2p2h, which is a data-driven construction, and the SuSa model, the most developed theoretical model, are likely to offer the best predictions. Consequently, on a bin-by-bin basis, the largest excursion from nominal, predicted by either model, is taken as a conservative estimate of the error. This error is taken to be symmetric about the nominal, meaning that the larger error is taken as both positive and negative uncertainty. The 2p2h-MEC modeling error is added in quadrature, bin-by-bin with the other sources of systematic error to get the total systematic error. The bin-by-bin estimated systematic error shown by Fig. 8.15 is very large in the region of phase space near the kinematic ridge where E_{avail} is approximately equal in magnitude to $|\vec{q}|$. In other regions, where E_{avail} is less than $|\vec{q}|$ in magnitude, the systematic error is estimated to be less than fifteen percent of the estimated cross section.

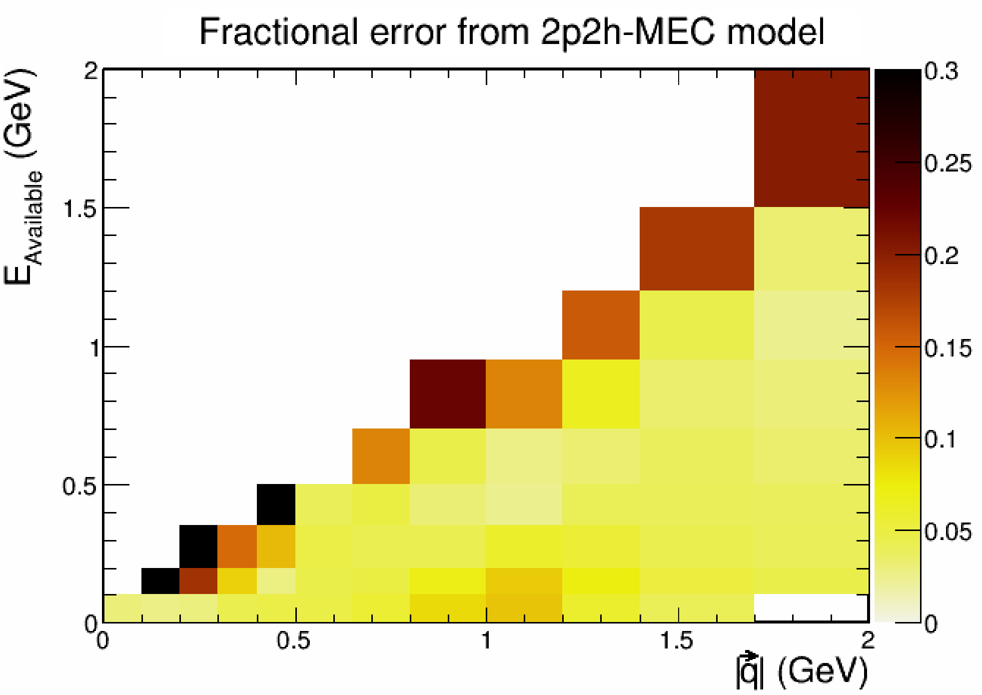


Figure 8.16: Fractional systematic error from MEC modeling on the predicted cross section after low pseudo-data bins are removed. After masking off the low pseudo-data bins the total systematic error is more in-line with errors from other sources. A few bins with large uncertainties arising from 2p2h-MEC uncertainties remain. The bins masked off are excluded from the analysis.

The 2p2h-MEC systematic is implemented as follows: The 2p2h-MEC events are weighted so as to map the NOvA nominal 2p2h-MEC to represent each of the above two models. Each of the latter 2p2h-MEC models is then used to calculate two new cross sections, each of which includes new efficiency, background, and unfolding calculations to obtain a systematic for uncertainties associated with 2p2h-MEC. The larger deviation from nominal is chosen bin-by-bin to give a double-sided error arising from 2p2h-MEC modeling uncertainties.

Most of the bins with large systematic errors from the 2p2h-MEC model are bins that contain very few reconstructed pseudo-data events. These bins are not be reported in the final analysis. Figure 8.16 shows the predicted systematic error after removing the bins with low data statistics. A few bins with large systematic error still remain along the kinematic ridge, however the largest ones have been removed.

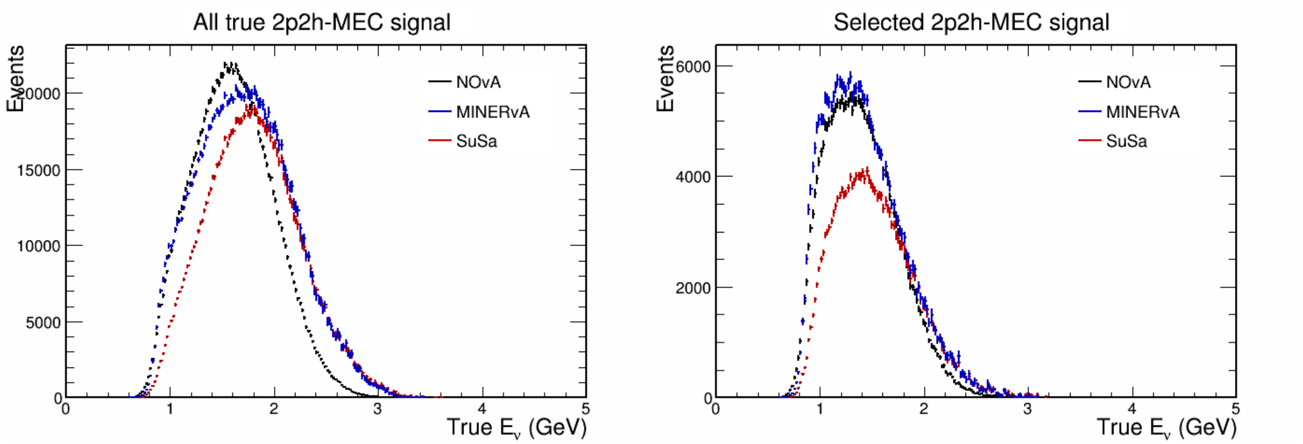


Figure 8.17: Distributions for 2p2h-MEC. Left: Distributions of true 2p2h-MEC signal events in true E_ν as predicted by the models. Right: Distributions of true 2p2h-MEC in true E_ν after all selection cuts. The SuSa model (red) gives a relatively high distribution in E_ν, consequently containment selection gives a relatively large reduction. The MINERvA tune (blue) has relatively more events along the high efficiency kinematic boundary, consequently it incurs a smaller depletion from selections than does the NOvA tune (black).

8.2.3 Neutrino energy spectrum of 2p2h-MEC

The changes in the 2p2h-MEC neutrino energy spectrum due to weighting the 2p2h-MEC models are shown in Fig 8.17. The left-hand distributions show 2p2h-MEC events that are part of the true signal, and the right-hand plot are the selected 2p2h-MEC events that are signal. The SuSa model (red) gives a lower total rate than the experimental tunes, and it the selection cuts induce a noticeable rate reduction. This rate reduction is due to the SuSa prediction of more events having large muon angles that are less likely to be contained. The MINERvA tune (blue) predicts more signal events than the SuSa model. It also predicts more events along the kinematic boundary and therefore has more forward-going muons that pass the containment cut, thus a lesser reduction in rate due to selection cuts. The NOvA tune (black) predicts the largest number of true signal events but predicts fewer forward going muons than MINERvA and incurs more of a depletion due to selection cuts.

8.2.4 2p2h-MEC systematic summary

Four 2p2h-MEC models plus two variants were investigated to estimate the systematic error due to 2p2h mismodeling. The SuSa model and the Minerva tuned Valencia model are used to create the final error band as these two models have the most theoretical and experimental support among the models considered. The error band was assigned bin-by-bin by taking the model with the largest excursion from nominal among from either of the two models and using that excursion as the uncertainty. These models imply large systematic errors in regions of phase space where few events are reconstructed, but where events are placed after unfolding. Bins with fewer than four reconstructed pseudo-data events are masked out of the analysis, but are still used for unfolding purposes. After masking out the large systematic bins, the systematic uncertainty from 2p2h-MEC is everywhere lower than the flux uncertainty except for a bin at $E_{avail} \geq 1.5$ GeV which is sparsely populated. Only 3.2% of reconstructed data events fall in the $E_{avail} \geq 1.5$ GeV region.

8.2.5 Composition of analysis total systematic uncertainty

For this analysis, the fractional uncertainty from all sources of systematic error were initially studied using simulations entirely; no real data was used. These studies were conducted without imposing selection cuts for muon kinetic energy versus muon production angle. The latter selections were finalized at a later time. Consequently the fractional uncertainties presented in this Section differ from those that are finally obtained with the data and with all data selections applied. The latter, final fractional uncertainties are reported in Chapter 8.

The fractional uncertainty from all sources of systematic error are shown in Fig. 8.18 in bins of $|\vec{q}|$ (top) and of E_{avail} (bottom). The uncertainty due to the flux (green) is the largest source of uncertainty in most bins and is relatively constant at 11.4%. The light calibration uncertainties (purple) are large at low and high values of $|\vec{q}|$ and E_{avail}. The cross section uncertainties (cyan) have a significant presence especially at higher values of E_{avail}. Smaller uncertainties arise from 2p2h-MEC (gold) and energy scale (blue). The statistical uncertainty (red) is small compared to the others. The total uncertainty (black) is less than 15% for $E_{avail} \leq 1.25$ GeV and $|\vec{q}| \geq 0.2$ GeV/c. The fractional uncertainty from each source for the integrated cross section is shown in Table 8.1. The flux uncertainty is the largest at 11.4%, followed by cross section modeling uncertainties at 2.3% , and the 2p2h-MEC modeling uncertainty at 0.7%. The fractional uncertainties for the integrated cross section shown in Table 8.1 represent event weighted averages over all bins shown in Fig. 8.18.

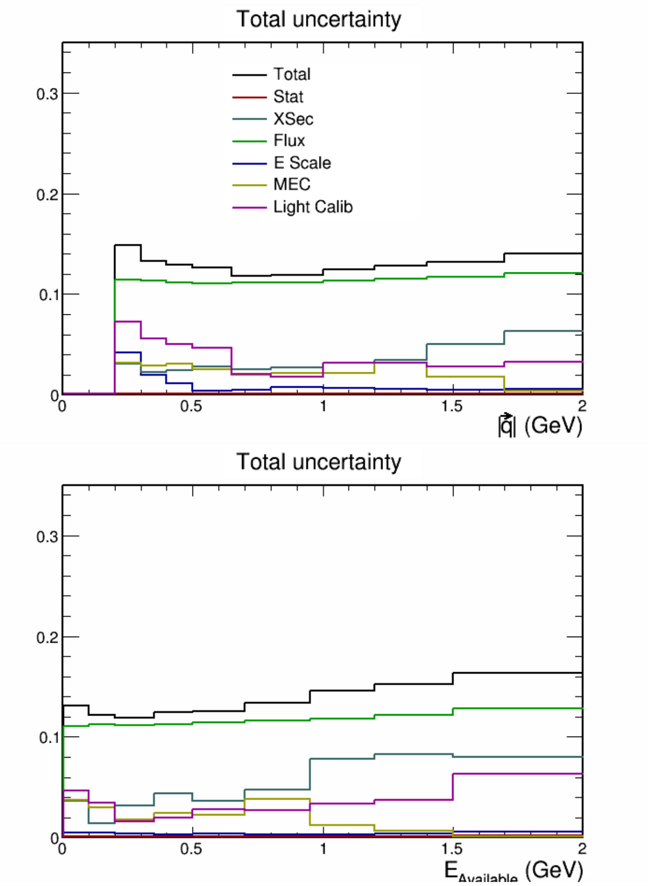


Figure 8.18: Fractional uncertainty from all sources for $|\vec{q}|$ (top) and E_{avail} (bottom). The total uncertainty (black) is mostly $\leq 13\%$ but becomes larger at high E_{avail}. The flux (green) is the largest single contributor to the uncertainty in every bin.

Table 8.1: Fractional uncertainty on the integrated cross section arising from all sources of systematic uncertainty. The total systematic uncertainty (quadrature sum) is 11.6%. The flux is the leading source of uncertainty; it contributes a fractional error of 11.4%.

Source of uncertainty	Fraction error on the cross section
Flux	11.4%
Cross section model	2.3%
Light calibration	0.7%
2p2h-MEC model	0.7%
Energy scale	0.3%
Total	11.6%

Chapter 9

The Cross-Section Measurement

9.1 Measurement of the double-differential cross section

The data that passes the event selections defined in Sec. 4.1.0.2, prior to background subtraction, efficiency corrections and unfolding, are shown in Fig. 9.1. The selected sample contains a total of 457,000 data events that occur in the kinematic domain of $|\vec{q}| \leq 2$ GeV/c and $E_{avail} \leq 2$ GeV. The data exhibits an elevated ridge structure of roughly 40,000 events per bin in the region 0.6 $\leq |\vec{q}| \leq 1.2$ GeV/c and $0.2 \leq E_{avail} \leq 0.6$ GeV.

The estimated background that is included in the selected CC-inclusive sample is shown in Fig. 9.2. The background is mostly comprised of events that do not pass the muon phase space requirements in the signal definition (45.5 %), together with CC-neutrino interactions whose vertices are reconstructed inside the fiducial volume but are actually outside the fiducial volume (23.0%). Smaller background contributions arise from $\bar{\nu}_{\mu}$-CC interactions (18.2%) and NC interactions (12.7%). The background peaks at higher values of $|\vec{q}|$ and E_{avail} than does the selected data events, with $|\vec{q}| \approx 1.5$ GeV/c with $E_{avail} \geq 1$ GeV. The number of background events is estimated to be 10% of the data events. The estimate is based on the reference MC of this analysis which uses version 2.12 of the GENIE generator.

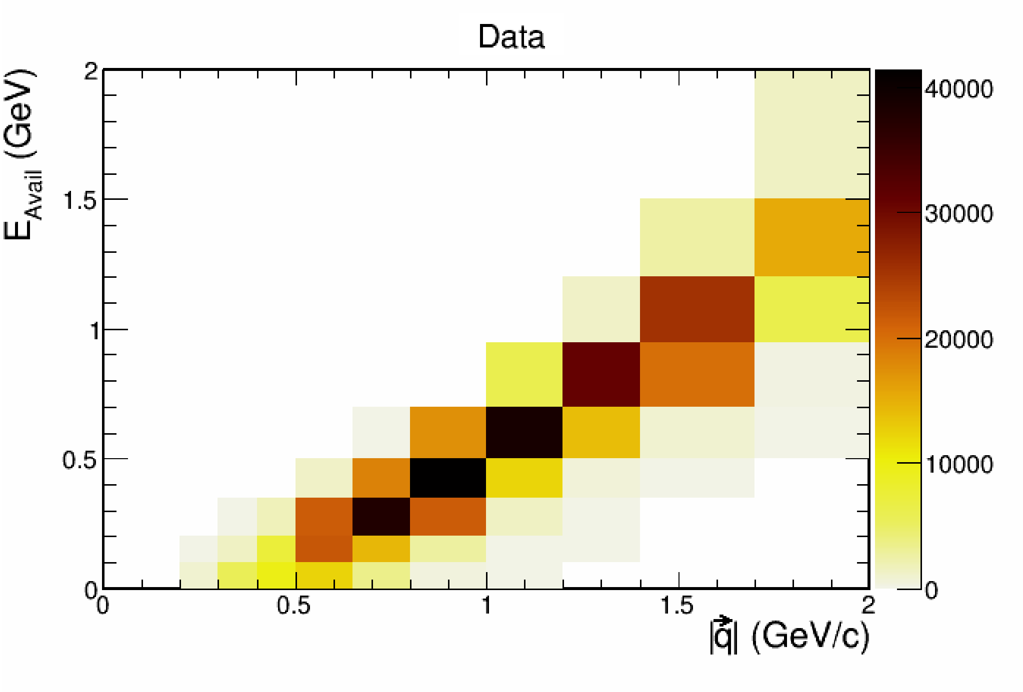


Figure 9.1: The distribution in $|\vec{q}|$ versus E_{avail} of events selected by this analysis prior to corrections for detector efficiency and resolution effects.

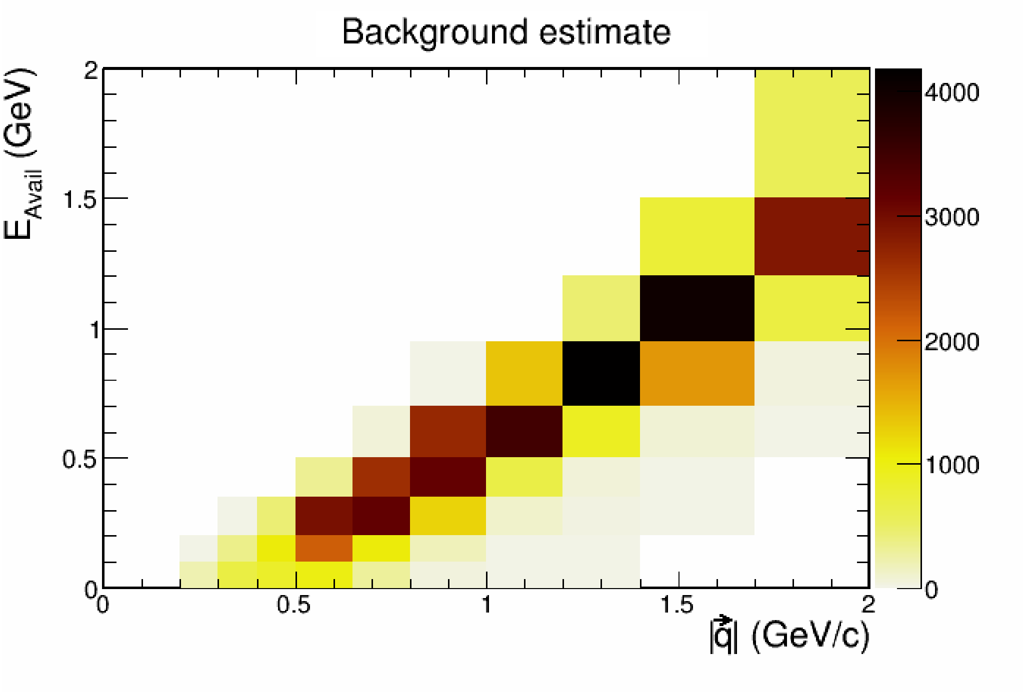


Figure 9.2: Distribution of the estimated background included in the candidate ν_μ-CC inclusive event sample of Fig. 9.1.

Figure 9.3 shows the inclusive CC signal obtained by subtracting the estimated background from the selected data. The shape of the signal distribution is not significantly distorted by the background subtraction. As previously, the CC-inclusive signal is peaked in bins between $0.6 \leq |\vec{q}| \leq 1.2$ GeV/c and $0.2 \leq E_{avail} \leq 0.6$. The most populated bin contains 36,000 events.

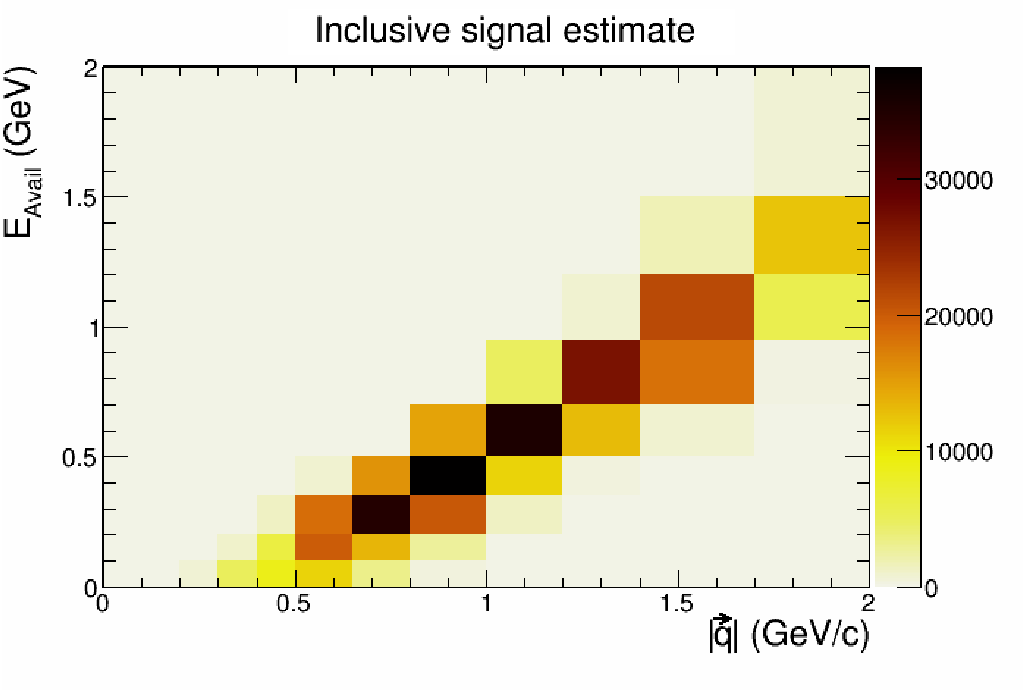


Figure 9.3: Distribution of the signal sample after subtraction of the estimated background (Fig. 9.2) from the initially selected sample (Fig. 9.1).

Figure 9.4 shows the unfolding matrix used by the analysis. The unfolding matrix relates the reconstructed values for the analysis observables to the true values of the observables. It is a four-dimensional matrix in which the outer blocks are bins that map reconstructed to true E_{avail}, within which are distributions that map reconstructed to true $|\vec{q}|$ in that particular mapping of E_{avail}. The matrix is nearly diagonal; that is, bins that are away from the diagonal have relatively small event populations.

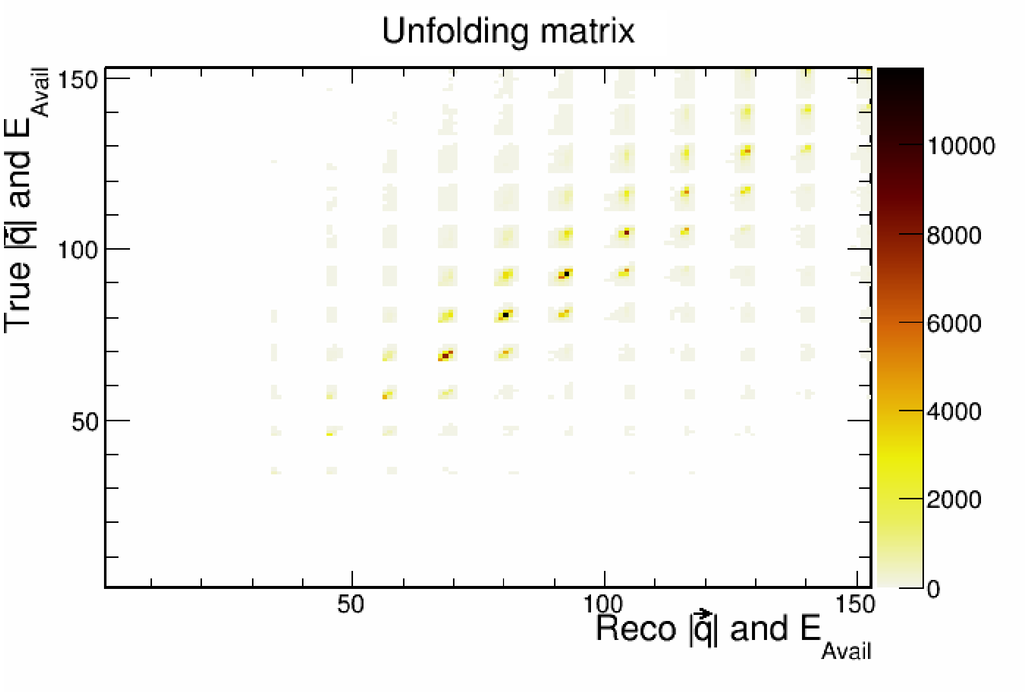


Figure 9.4: The 4D unfolding matrix provides the mapping from reconstructed $|\vec{q}|$ and E_{avail} to true $|\vec{q}|$ and E_{avail}.

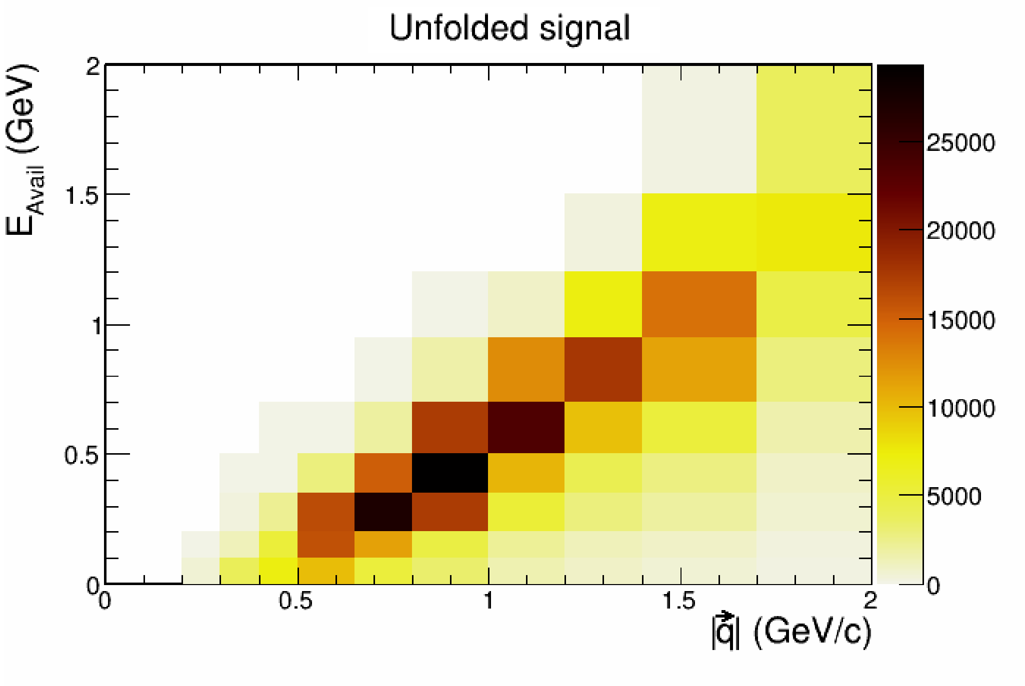


Figure 9.5: Distribution of the signal events after unfolding. The unfolding procedure moves events from the peak region into regions that are away from the central ridge.

Figure 9.5 shows the result of unfolding the detector and reconstruction effects to obtain values that are closer to the true values. Comparing with Fig. 9.3, the unfolding results in the migration of events from the $|\vec{q}| \approx E_{avail}$ ridge to regions where $|\vec{q}|$ exceeds E_{avail} in magnitude. The result is a kinematic distribution which is broader than that observed before unfolding.

Figure 9.6 shows the detection efficiency for this analysis as a function of the two kinematic variables. The efficiency is highest along the kinematic boundary where the magnitudes of E_{avail} and $|\vec{q}|$ are comparable. The efficiencies diminish in regions that are remote from the kinematic edge tending toward 20% for the much of the plot.

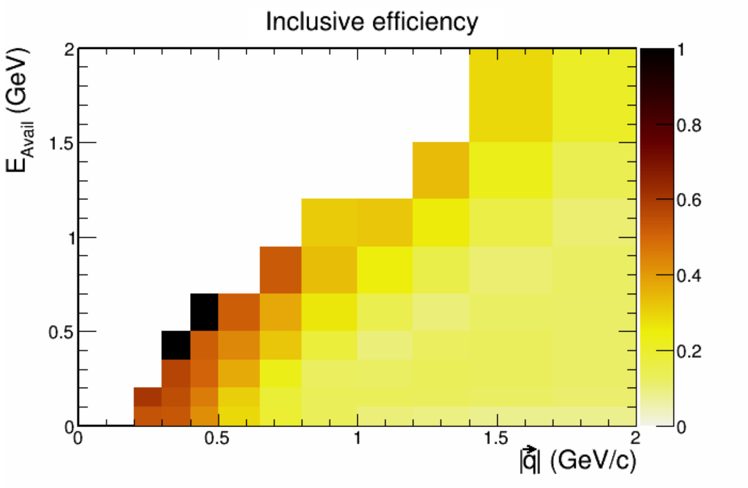


Figure 9.6: Event detection efficiency plotted over the kinematic plane. The efficiency peaks along the kinematic boundary and diminishes in regions away from the boundary.

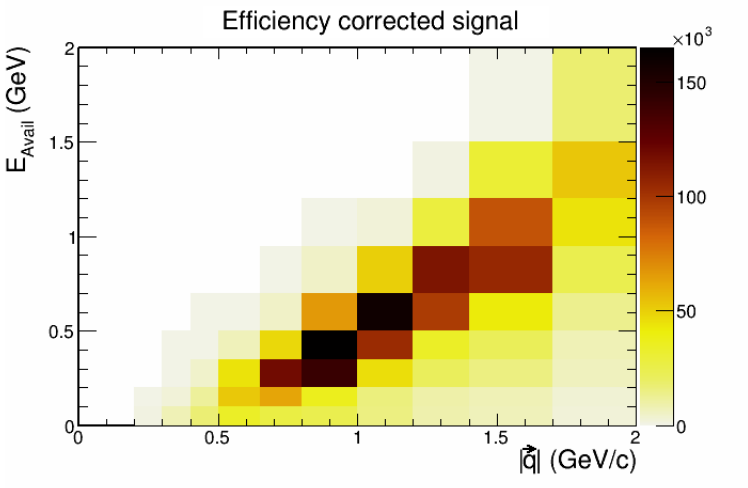


Figure 9.7: Distribution of signal events after unfolding and efficiency correction. The efficiency correction shifts the distribution towards higher $|\vec{q}|$ and E_{avail}.

Figure 9.7 shows the signal after correction for detection efficiency. The distribution shape is modestly changed. It now peaks more sharply between $0.8 \leq |\vec{q}| \leq 1.2$ GeV/c and $0.2 \leq E_{avail} \leq 0.7$ GeV.

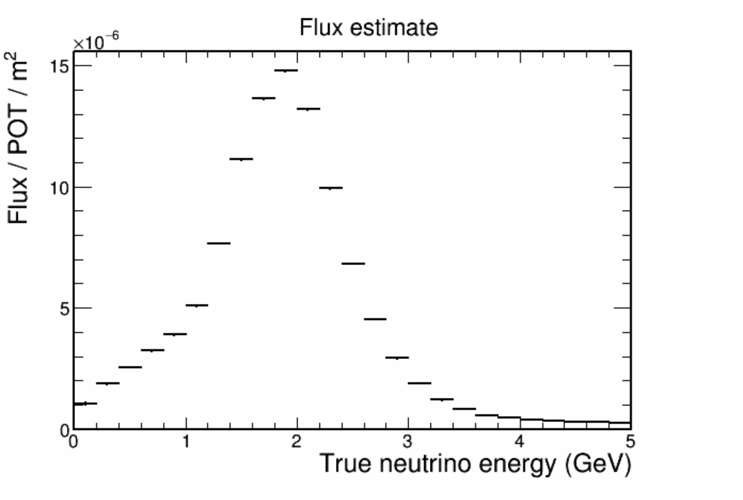


Figure 9.8: The NOvA neutrino flux distribution incident on the Near Detector when the NuMI beam is operated in forward-horn-current (FHC) mode.

Table 9.1: Bin-by-bin display of the double-differential cross section shown in Fig. 9.9. The tabled values represent $d^2\sigma/d|\vec{q}|dE_{avail}$ per bin in units of cm^2/GeV/GeV/c/nucleon. Note that, in contrast to Fig. 9.9, E_{avail} is denoted on the horizontal axis, while $|\vec{q}|$ is given by the vertical axis.

$\lvert\vec{q}\rvert \backslash E_{avail}$	0.00-0.10	0.10-0.20	0.2-0.35	0.35-0.50	0.50-0.70	0.70-0.95	0.95-1.20	1.20-1.50	1.50-2.00
0.00-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.10-0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.20-0.30	2.79e-40	1.71e-41	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30-0.40	2.15e-39	6.15e-40	8.16e-41	6.12e-44	0.00	0.00	0.00	0.00	0.00
0.40-0.50	4.94e-39	3.38e-39	8.51e-40	1.86e-41	6.45e-44	0.00	0.00	0.00	0.00
0.50-0.65	6.66e-39	1.02e-38	5.72e-39	8.52e-40	8.68e-42	0.00	0.00	0.00	0.00
0.65-0.80	5.24e-39	1.22e-38	1.54e-38	6.19e-39	4.95e-40	4.62e-43	0.00	0.00	0.00
0.80-1.00	3.58e-39	5.34e-39	1.36e-38	1.61e-38	4.84e-39	2.93e-40	1.04e-43	0.00	0.00
1.00-1.20	2.18e-39	2.36e-39	4.41e-39	1.01e-38	1.15e-38	2.93e-39	1.51e-40	0.00	0.00
1.20-1.40	1.28e-39	1.37e-39	2.07e-39	3.28e-39	7.09e-39	6.67e-39	1.64e-39	4.48e-41	0.00
1.40-1.70	6.33e-40	7.06e-40	1.00e-39	1.42e-39	2.05e-39	4.10e-39	3.46e-39	9.99e-40	1.23e-41
1.70-2.00	2.30e-40	2.71e-40	3.84e-40	5.01e-40	6.55e-40	9.34e-40	1.73e-39	1.71e-39	3.50e-40

Figure 9.8 shows the estimated neutrino flux at the Near Detector. By design, the neutrino flux for the NOvA experiment is that produced by a narrow-band beam. The flux peaks at 1.84 GeV and has a FWHM of 1 GeV. The result of this narrow-band flux is that the selected inclusive

sample is dominated by CC-QE, 2p2h-MEC and RES events together with a modest contribution from DIS.

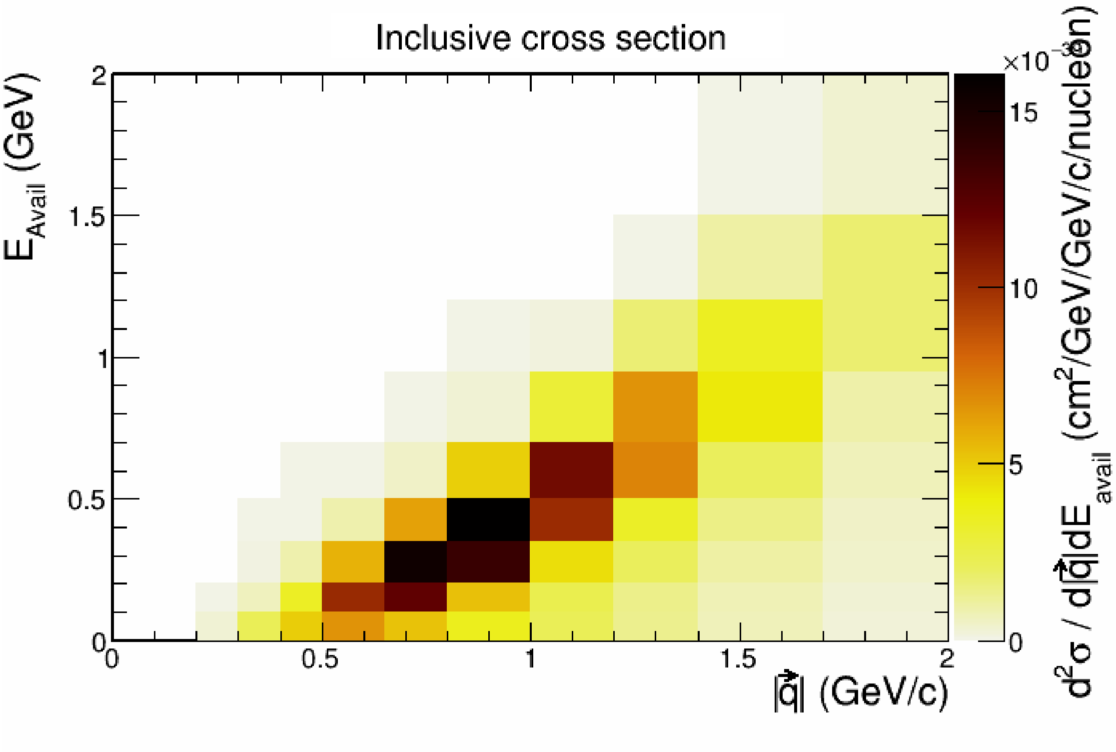


Figure 9.9: The double-differential inclusive cross section obtained by this analysis.

The double-differential inclusive cross section is shown in Fig. 9.9. It is the first measurement of a double differential distribution in $|\vec{q}|$ and E_{avail} obtained by the NOvA experiment. The double-differential cross section retains the ridge structure exhibited by the previous plots. The cross-section strength is greatly diminished at the outer boundaries of the analyzed kinematic region. A bin-by-bin breakout of the measured cross section is presented in Table 9.1. Due to the muon phase space restriction, the measured range of $|\vec{q}|$ starts at 0.20 GeV/c. The cross section peaks at 0.9 GeV/c in $|\vec{q}|$ and 0.4 GeV in E_{avail} with a value of 1.61×10^{-38} cm^2/GeV/GeV/c/nucleon.

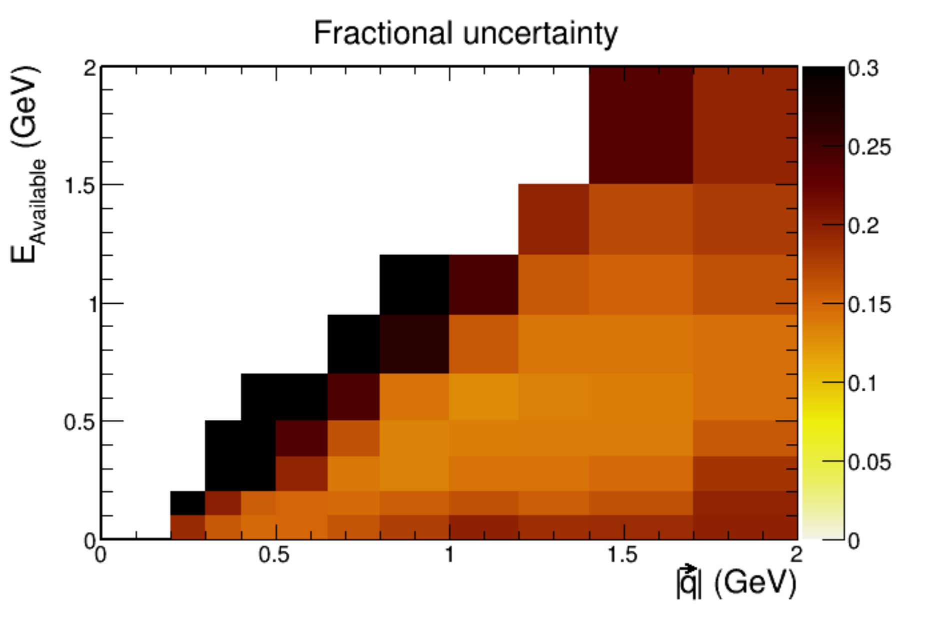


Figure 9.10: The fractional uncertainty on the double-differential cross section shown in Fig. 9.9.

Table 9.2: Bin-by-bin display of fractional uncertainty for the double differential cross section shown in Fig. 9.9. Note that, in contrast to Fig. 9.9, E_{avail} is denoted on the horizontal axis, while $|\vec{q}|$ is given by the vertical axis.

$\vert\vec{q}\vert\backslash E_{avail}$	0.00-0.10	0.10-0.20	0.2-0.35	0.35-0.50	0.50-0.70	0.70-0.95	0.95-1.20	1.20-1.50	1.50-2.00
0.00-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.10-0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.20-0.30	0.19	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30-0.40	0.16	0.20	0.40	0.63	0.00	0.00	0.00	0.00	0.00
0.40-0.50	0.15	0.16	0.39	0.37	0.93	0.00	0.00	0.00	0.00
0.50-0.65	0.15	0.15	0.20	0.24	0.60	0.00	0.00	0.00	0.00
0.65-0.80	0.16	0.15	0.14	0.16	0.24	4.89	0.00	0.00	0.00
0.80-1.00	0.18	0.15	0.13	0.13	0.14	0.27	38.73	0.00	0.00
1.00-1.20	0.20	0.16	0.14	0.14	0.13	0.16	0.24	0.00	0.00
1.20-1.40	0.19	0.16	0.14	0.14	0.13	0.14	0.16	0.20	0.00
1.40-1.70	0.19	0.16	0.15	0.14	0.14	0.14	0.15	0.17	0.23
1.70-2.00	0.20	0.20	0.18	0.16	0.164	0.14	0.16	0.18	0.19

The bin-by-bin fractional uncertainty is shown in Fig. 9.10 and in Table 9.2. The bin-by-bin uncertainty is 0.15-0.20 throughout most of the phase space and increases near the kinematic boundary. The covariance matrix representing the covariance between bin contents is shown in Fig. 9.11. The covariance matrix shows how bins react to changes in other bins due to uncertainty and is one of the deliverables of this analysis.

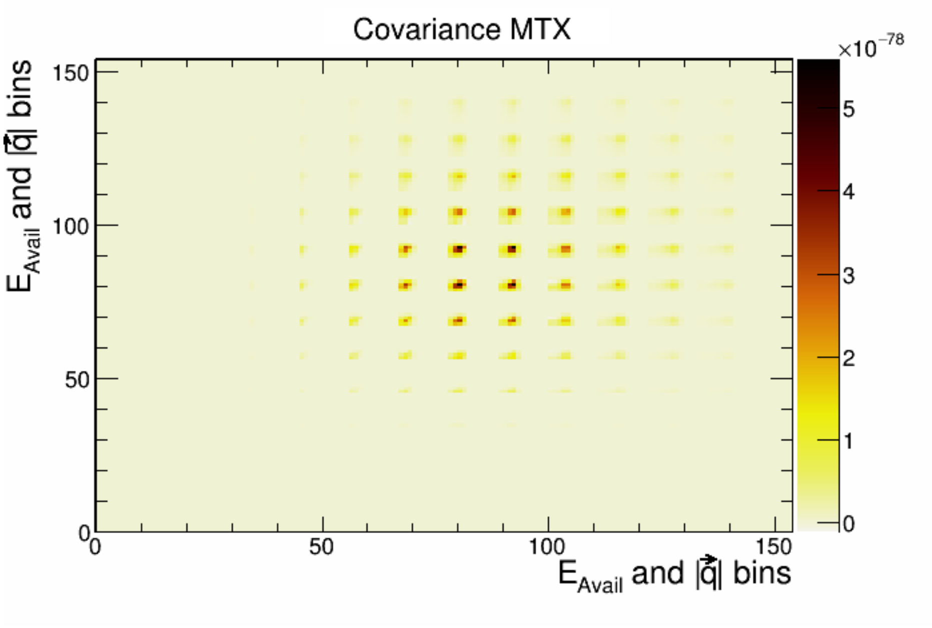


Figure 9.11: The covariance matrix for the double-differential inclusive cross section in E_{avail} and $|\vec{q}|$ bins. The large blocks depict bins in E_{avail} with bins of $|\vec{q}|$ within them.

Figure 9.12 shows the fractional systematic uncertainties as functions of 3-momentum transfer. The uncertainties here are calculated using multiverse cross sections with the data as input, in contrast to the estimation in Fig. 8.18 that refer to a Monte Carlo pseudo-data sample. The total uncertainty (black curve) is 20% at $|\vec{q}|$ of 0.25 GeV/c and is a fairly consistent 15% in other bins. The uncertainty associated with the neutrino flux (green curve) is highest in the first two populated bins of $|\vec{q}|$ below 0.4 GeV/c. At higher values it flattens to 13% across the rest of the range. The light calibration systematic (purple curve) is the next largest in bins below 1.0 GeV/c, while the cross section systematic (teal curve) is second-largest in bins greater than 1.0 GeV/c. Figure 9.13 shows the fractional uncertainty breakout for hadronic available energy. Here, the flux uncertainty is fairly constant at 13%. Light calibration is the second highest systematic in bins below 0.35 GeV. Above 0.35 GeV it is overtaken by the cross-section systematic.

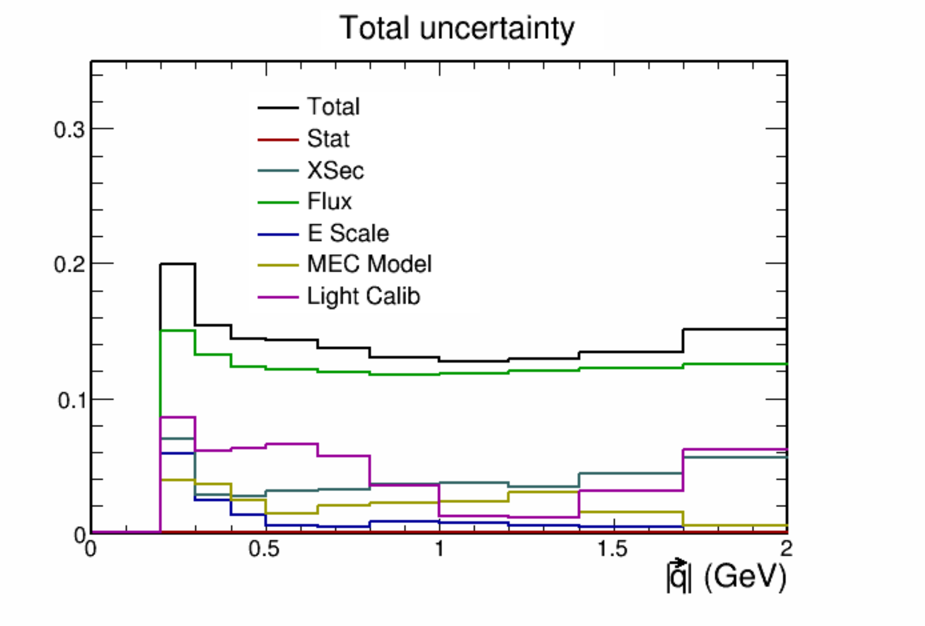


Figure 9.12: Breakout of fractional uncertainties for the systematic error sources as a function of $|\vec{q}|$. The uncertainties are estimated using the data, whereas previous estimates (Fig. 8.18) were based on Monte Carlo simulation.

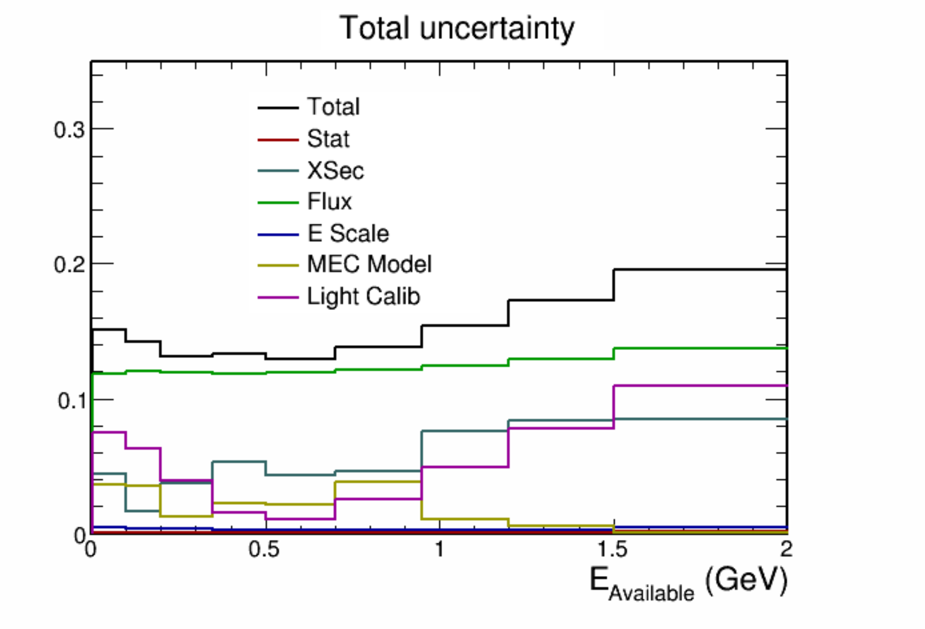


Figure 9.13: Fractional uncertainties from the systematic error sources, as a function of E_{avail}. The uncertainties are estimated using the data.

Table 9.3: Fractional uncertainty on the integrated cross section arising from all sources of systematic uncertainty. The total systematic uncertainty (quadrature sum) is 12.4%. The flux is the leading source of uncertainty; it contributes a fractional error of 12.0%.

Source of uncertainty	Fraction error on the cross section
Flux	12.0%
Cross section model	2.65%
Light calibration	0.94%
2p2h-MEC model	0.56%
Energy scale	0.28%
Total	12.4%

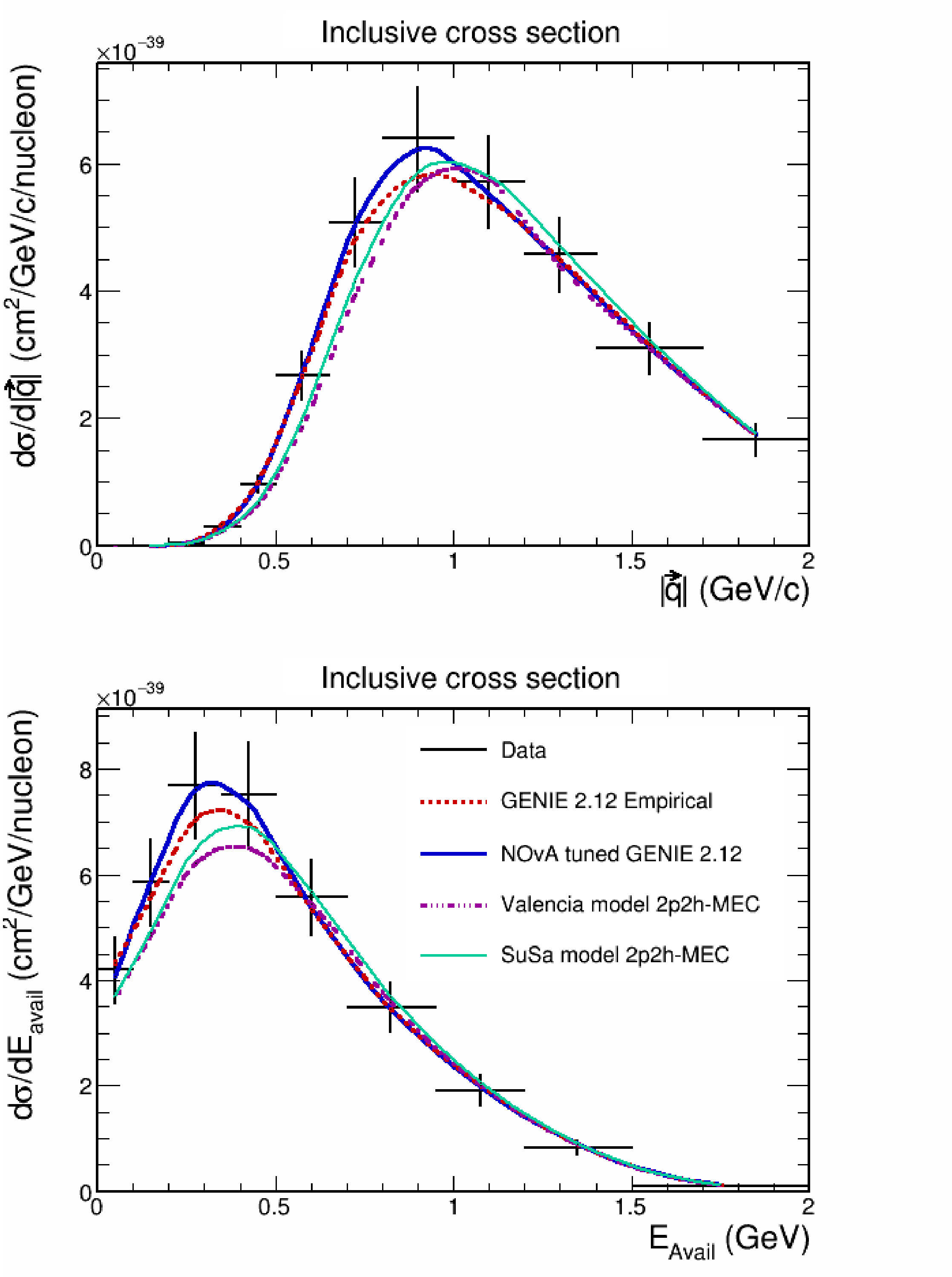


Figure 9.14: Inclusive differential cross sections in $|\vec{q}|$ (top) and E_{avail} (bottom). The data (black crosses) is compared to MC predictions that use different 2p2h models: the GENIE 2.12 default model using GENIE empirical MEC (red), GENIE 2.12 with NOvA cross section tunes (blue), GENIE 2.12 using the Valencia 2p2h-MEC model (magenta), and GENIE 2.12 using the SuSa v2 2p2h-MEC model (teal).

Figure 9.14 shows the projections of the double differential cross section onto the each of the kinematic variables. The distribution of $|\vec{q}|$ (top plot) is observed to rise smoothly from a threshold at 0.2 GeV/c to peak at 0.9 GeV/c, and then falls off nearly linearly beyond the peak. The distribution of E_{avail} (bottom plot) rises from 0 GeV to 0.3 GeV and subsequently falls off roughly exponentially. The data (black crosses) are compared to predictions for the differential cross sections that utilize different models for 2p2h-MEC. The blue histograms show the predictions based on the NOvA tuned GENIE 2.12 generator. This tune has been fit to NOvA ND data so it is not surprising that it gives a good description. The remaining three simulations are based GENIE 2.12 with different 2p2h-MEC models invoked. One observes that the GENIE empirical model (red histogram), the Valencia model (blue histogram), and the SuSa v2 model (Green histogram) gives predictions that fall below the data in regions around the peaks of the distribution. Additionally, the Valencia and SuSa models under-predict the rising edge of both distributions.

Figure 9.15 shows the differential cross section versus E_{avail} in bins of $|\vec{q}|$. The data is shown by the black crosses. The histograms show the component reaction categories as estimated by GENIE 2.12. In most of the plots QE (blue) and RES (green) give significant contributions. The MEC contribution (brown) is sizable in the bins from $0.5 < |\vec{q}| < 1.0$. The total cross section estimated by GENIE 2.12 (red) underestimates the data in most bins. An enhancement of 2p2h-MEC in these bins would likely correct the discrepancy. The χ^2/DoF between the inclusive cross section and empirical and theoretical models is shown in Table. 9.4. The GENIE empirical model has the best agreement with the data with a χ^2/DoF of 0.58, the SuSa-v2 model has a χ^2/DoF of 0.90, and the Valencia model has a χ^2/DoF of 1.29. All three models show good agreement with the data.

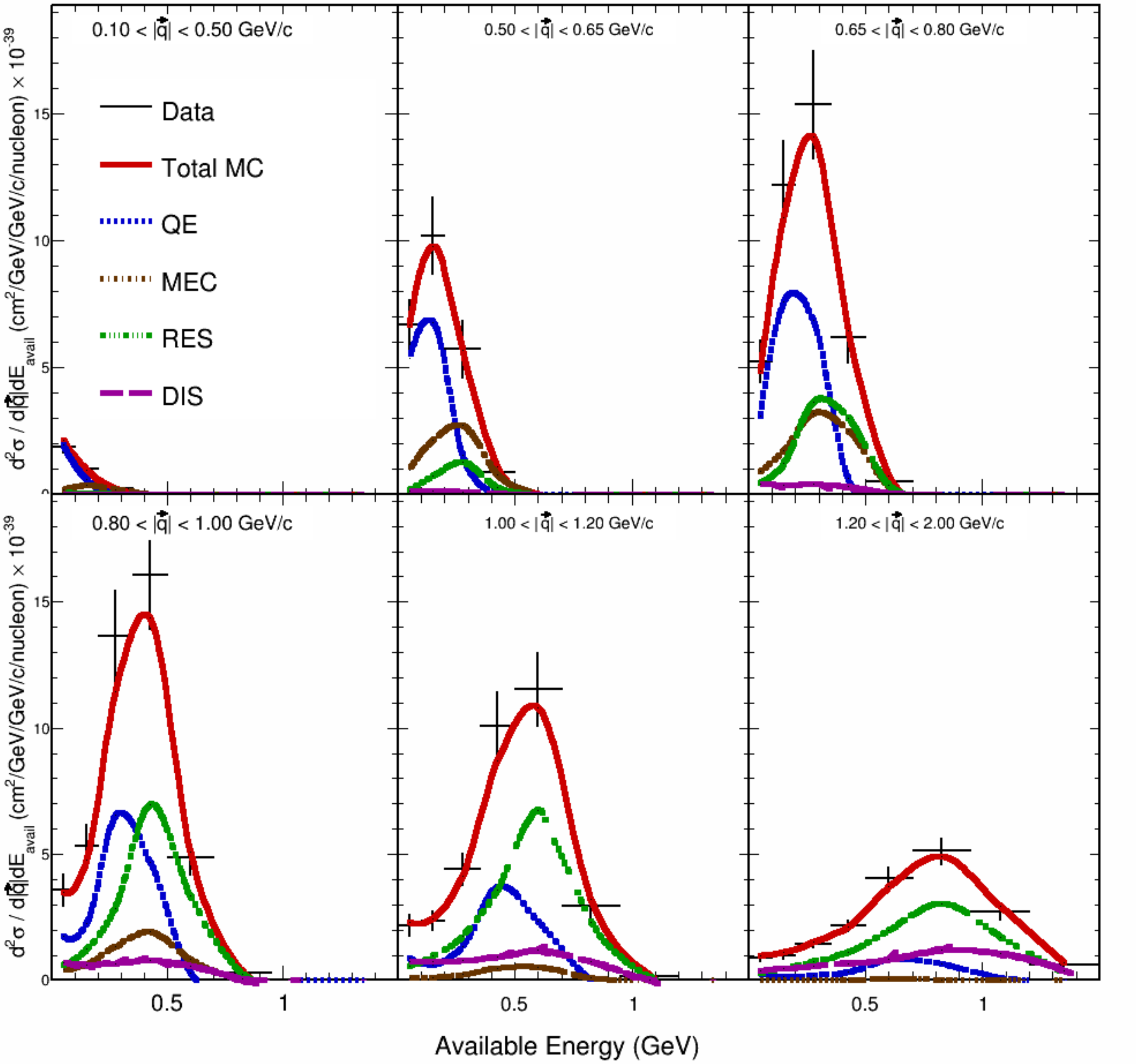


Figure 9.15: TThe inclusive cross section in bins of E_{avail}, for six different slices of $|\vec{q}|$. The excess of data over GENIE 2.12 MC is most pronounced in the peak regions for bins in the range 0.5 $< |\vec{q}| <$ 1.0 GeV/c.

Table 9.4: The χ^2/DoF between the inclusive double-differential cross section and various models.

	χ^2/DoF
GENIE empirical 2p2h-MEC	0.58
Valencia 2p2h-MEC	1.29
SuSa-v2 2p2h-MEC	0.90

The apparent agreement between the data (crosses) and the reference MC (solid line, red) shown in Fig. 9.15 hinges on the verity of the predicted contributions from component CC processes. These processes include 2p2h-MEC (dot-dash curve, brown), estimated here using the GENIE empirical model. The 2p2h-MEC contribution is estimated to be significant in the range $0.5 \leq |\vec{q}| \leq 1.0$ GeV/c.

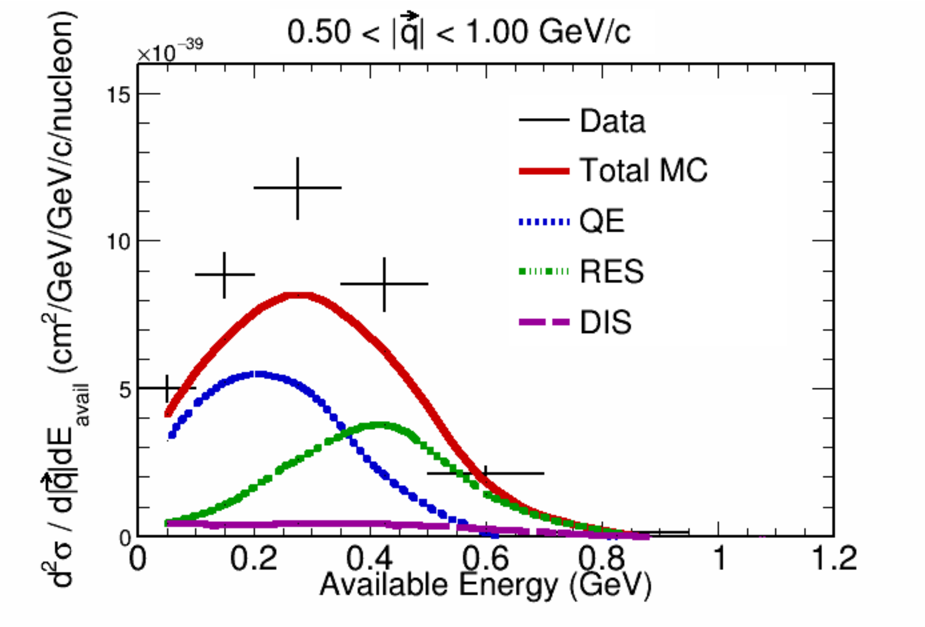


Figure 9.16: Cross section versus E_{avail} for the CC inclusive cross section restricted to the range in three-momentum transfer where the 2p2h-MEC contribution is estimated to be large. The data (crosses) is compared to the sum of contributions from CCQE, RES, and DIS reactions.

The cross section versus E_{avail} for this restricted range of three-momentum transfer is displayed in Fig. 9.16. Here the data is compared to the contributions from QE, RES, and DIS, without an estimation for 2p2h-MEC. The apparent excess in the data, when plotted in this way, is observed to be largest in the region of Eavail that lies between the QE and RES contributions, where the latter arises predominantly from $\Delta(1232)$ baryon resonance production. This situation is to be expected if the excess does indeed reflect the presence of 2p2h-MEC in the data, for prevalence of 2p2h processes in the kinematic region between elastic scattering and $\Delta(1232)$ production is well-established in electron-nucleus scattering.

The present NOvA analysis confirms key aspects of 2p2h-MEC neutrino-production in carbon nuclei that were first reported by the MINERvA collaboration [5]. Indeed, our Fig. 9.15 was inspired by the MINERvA plot reproduced below as Fig 9.17. The MINERvA figure indicates the existence of an event excess (data points with errors) that lies over and above a simulation that includes a 2p2h-MEC contribution based upon the Valencia model. The excess is interpreted as indicating additional event rate into 2p2h-MEC that is not captured by the model. It should be noted that MINERvA data are obtained with a wide-band neutrino flux that peaks near 3.0 GeV, hence the MINERvA analysis is based on data from a higher E_ν range than the present analysis. Nevertheless, trends in data excess relative to simulation that does not fully account for 2p2h as exhibited by the NOvA data of Fig. 9.15, are rather similar to those reported by MINERvA. The MINERvA distributions of Fig. 9.17, indicate significant 2p2h contributions to start around 0.3 GeV/c and to extend to above 0.8 GeV/c. In the NOvA data of Fig. 9.15, the 2p2h contribution starts at a higher $|\vec{q}|$ value, but otherwise the $|\vec{q}|$ response is similar.

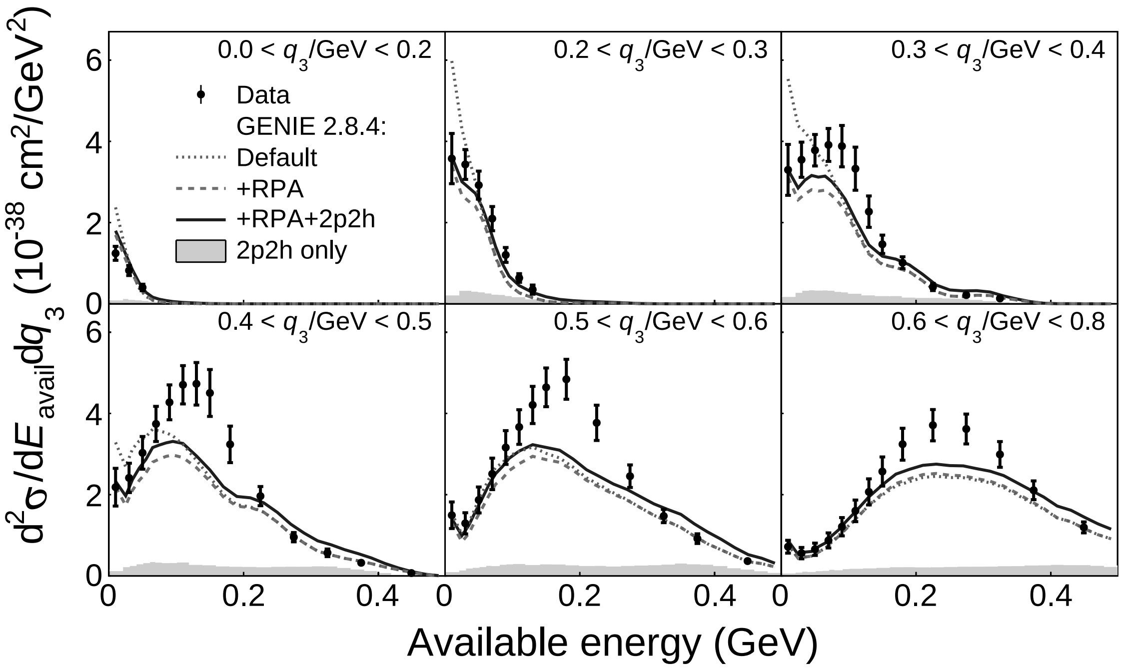


Figure 9.17: Cross section versus available energy in bins of |q|, reported by the MINERvA experiment [5].

The MINERvA experiment also demonstrated that the 2p2h contribution peaks above the CCQE peak and below the RES peak. This situation is shown clearly with distributions reported by them, which are reproduced here in Fig. 9.18.

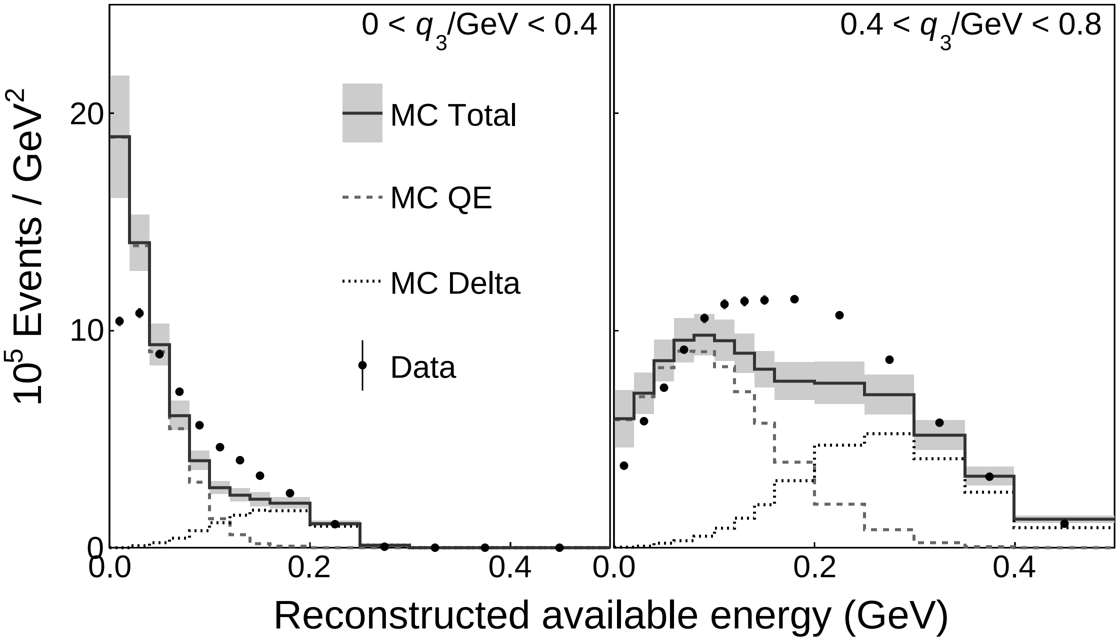


Figure 9.18: Event distributions in E_{avail} for two regions of three-momentum transfer, reported by MINERvA [5]. The event excess attributed to 2p2h peaks in a kinematic region that lies between the quasielastic and Δ resonance contributions.

9.2 Summary of the inclusive cross-section measurement

We have obtained for the first time the CC-inclusive double-differential cross section in the variables three-momentum transfer and available energy using the NOvA near detector data. The measurement is a high-statistics probe of CC inclusive processes in the 1.0 to 3.0 GeV region of neutrino energy. The measured cross section modestly exceeds most neutrino generator predictions for kinematic ranges in these variables from threshold values to the peak values. A decomposition of the differential cross section for E_{avail} in bins of $|\vec{q}|$ suggests that 2p2h-MEC is being underestimated in the Valencia and SuSa phenomenological models. The contribution of 2p2h-MEC processes to the CC inclusive sample is estimated in the second stage of this Thesis.

Chapter 10

Methodology of the 2p2h-MEC Determination

10.1 Overview of 2p2h-MEC estimation

The analysis, upon determining the double-differential ν_μ-CC inclusive cross section in E_{avail} and $|\vec{q}|$ as described above, now undertakes to isolate the contribution arising from 2p2h nuclear medium processes and to convert the observed rate into a cross section. The delineation of the 2p2h excess is carried out for the analysis domain $0.2 \leq |\vec{q}| \leq 2.0$ GeV/c and $0.0 \leq E_{avail} \leq 2.0$ GeV. The general approach is illustrated by Fig. 10.1, which indicates that 2p2h (modeled here using the empirical Dytman model [45]) occurs in regions of E_{avail} (Fig. 10.1 right) and $|\vec{q}|$ (Fig. 10.1 left) that are somewhat localized in these variables and are distinctive to a useful degree from regions populated by other, known CC interaction processes. Figure 10.2 shows that this localization aspect allows a search region to be defined on the kinematic plane of $|\vec{q}|$ versus E_{avail}. Figure 10.3 shows 2p2h-MEC as a fraction of the total CC-inclusive event distribution over the kinematic phase space of the analysis.

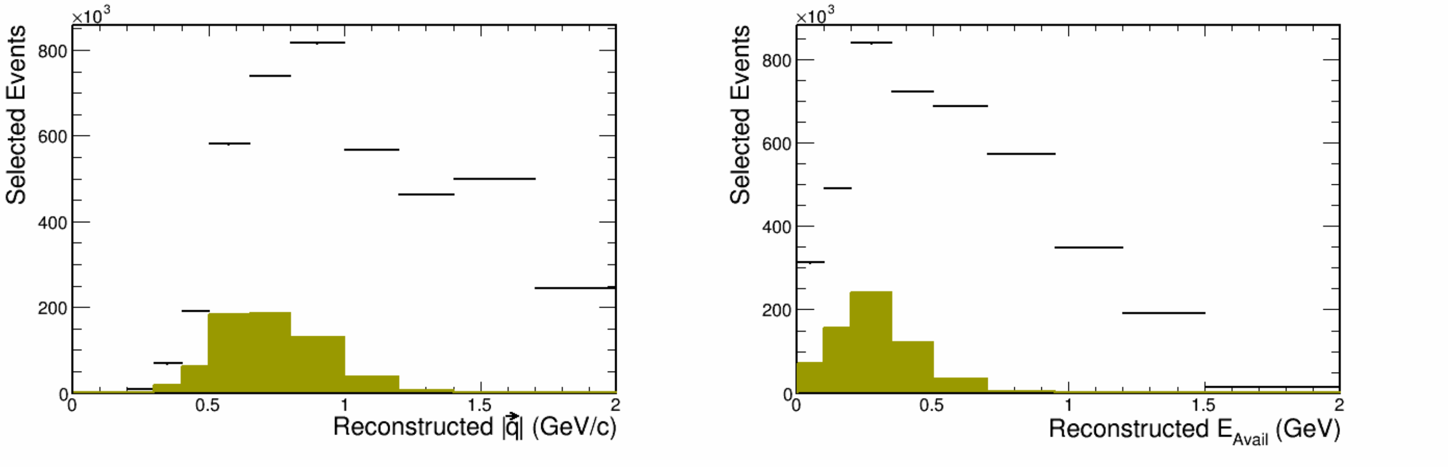


Figure 10.1: Event distribution in $|\vec{q}|$ and E_{avail}. The black bars show the inclusive sample of selected events, while the gold distributions shows Dytman MEC events. The MEC distribution peaks at 0.5 GeV/c and extends to 1.2 GeV/c in $|\vec{q}|$. It makes a significant contribution to the event rate up to 0.6 GeV in E_{avail}.

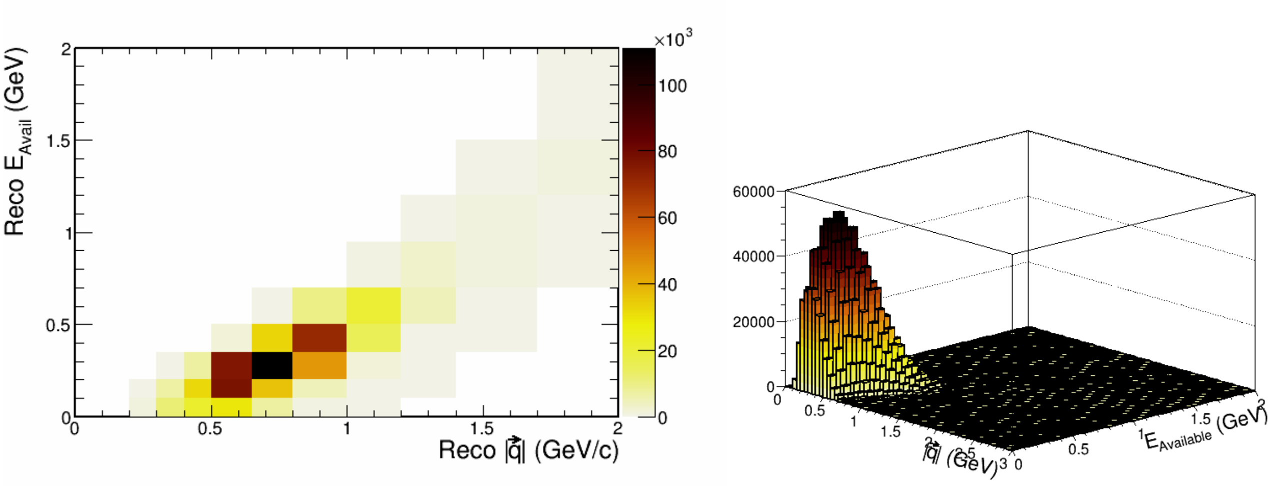


Figure 10.2: Two-dimensional empirical MEC distribution in $|\vec{q}|$ versus E_{avail}. The left-hand plot is a distribution with bins colored by bin content, while the right-hand plot is a three dimensional view of the same distribution where the z-axis is the number of events. The majority of events lie in the area of $|\vec{q}| < 0.9$ GeV/c and $E_{avail} <$ 0.5 GeV.

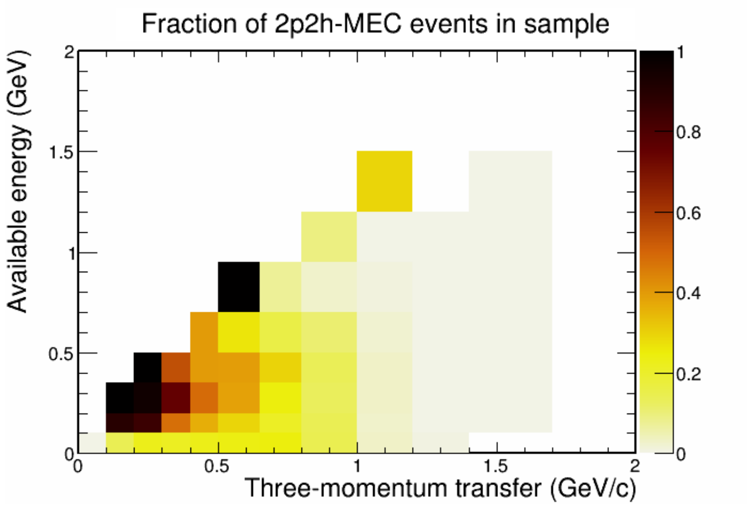


Figure 10.3: The bin-by-bin fraction of 2p2h-MEC events versus CC-inclusive selected events. According to the GENIE empirical MEC model, 2p2h-MEC dominates some of the border bins of the CC-inclusive distribution.

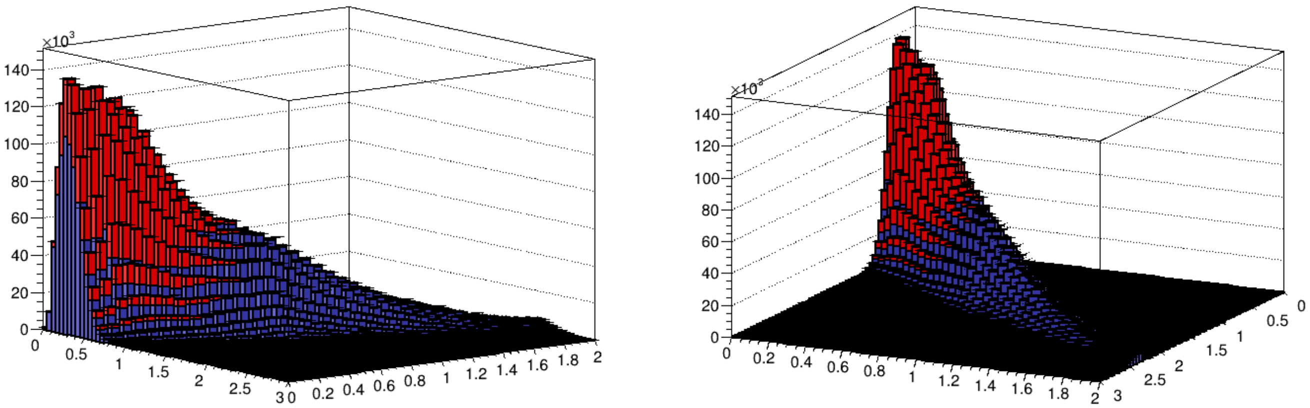


Figure 10.4: Two-dimensional empirical MEC distribution (red) on top of all other ν_μ-CC events (blue). The MEC distribution above the non-MEC distribution is concentrated in areas where $|\vec{q}|$ and E_{avail} are similar in magnitude. MEC makes non trivial contributions to the overall rate when $|\vec{q}|$ is less than 1.2 GeV and E_{avail} is less than 0.6 GeV. Events where either $|\vec{q}|$ or E_{avail} are larger or very different in magnitude contain trace amounts of MEC.

Charged current events whose topology differ from those exhibited by 2p2h-MEC provide a control sample. The control sample is used to anchor the normalizations of distributions predicted by the reference simulation for CC baryon resonance production (RES) and deep inelastic (DIS) scattering [19]. The contribution from CC quasielastic scattering is 'dead-reckoned' using the standard Lewellyn Smith phenomenology [53] with an axial mass of 1.0 GeV; this is the default

GENIE model for quasielastic scattering.

In brief, we construct distribution templates for the known CC interaction processes, namely CCQE, RES, DIS, and CC coherent scattering. The template shapes are based on the reference simulation; the normalizations for the RES and DIS templates are set by fitting to the a control sample, while the CCQE and CC-coherent templates are based on phenomenology and – for CC coherent – the current world measurements. The event rate into 2p2h processes are taken to be the data excess that is observed to lie above the event rate predicted by the template distributions extrapolated over the search region in the plane of $|\vec{q}|$ versus E_{avail}. The localization in these kinematic variables is suggested by the empirical Dytman MEC [45], as well as the Valencia 2p2h model [54]. This method is similar to the that used by MINERvA to isolate the 2p2h contribution to the cross section [25, 26]. Figure 10.4 illustrates where the excess is expected to lie. The blue distribution is made from all non MEC events that pass the selection cuts, while the red distribution is the MEC events. MEC events makes significant contributions to the overall rate at low $|\vec{q}|$ and E_{avail} and in regions where the two variables are similar in magnitude.

Theoretical models and experimental tunes that attempt to describe multi-nucleon 2p2h interactions exhibit large degrees of variability among themselves. The extent to which the models agree and disagree can be seen in Fig. 10.6. The top-left plot shows the GENIE empirical model [45], which describes the excess using a two-dimensional Gaussian in $|\vec{q}|$ and q_0. This empirical model predicts the largest total cross section. The top-right plot shows the Valencia model which predicts a fairly flat cross section over the phase space. Unfortunately this model is based on a calculation that is not fully relativistic and does not extend past $|\vec{q}|$ 1.2 GeV/c. The lower plot shows the SuSa-v2 model [24, 18] which is currently the most well-developed theoretical model available in GENIE version 3.0. The SuSa model is seen to predict higher cross section over a larger extent of phase space.

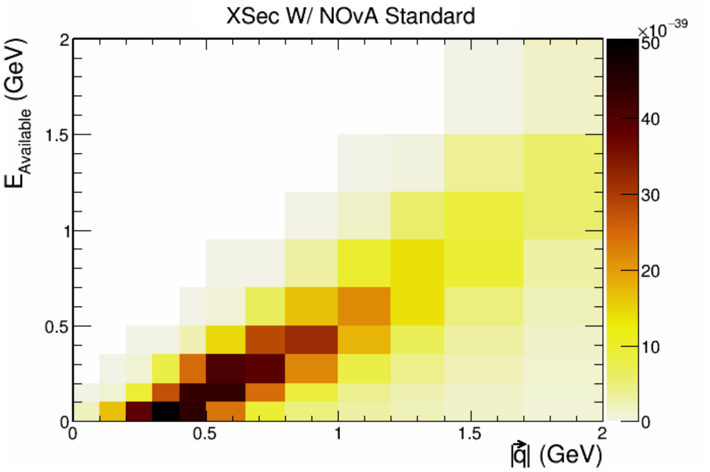


Figure 10.5: Estimation of the double differential CC-ν_μ cross section based on the nominal NOvA Monte Carlo. The cross section is seen to form a ridge is peaked at low values of $|\vec{q}|$ and E_{avail} which gradually diminishes as these variables are increased.

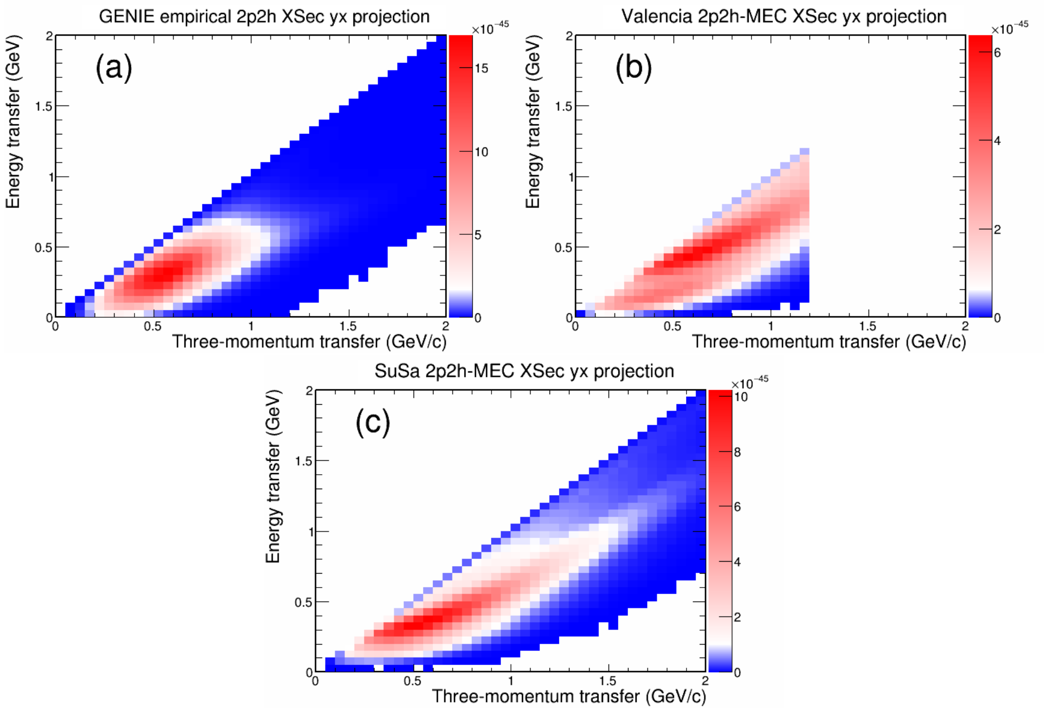


Figure 10.6: Differential cross sections predicted by 2p2h-MEC models. These models include (a) the GENIE empirical model, (b) the Valencia model, and (c) the SuSa-v2 model. The models are seen to cover somewhat different areas of phase space.

The determination of the excess in this Thesis is carried out using event counts. The same methods are then used to carry out a determination in terms of cross section. Figure 10.7 shows the total population of the selected events for this analysis. This distributions serves as the starting point for our determination of the 2p2h-MEC contribution.

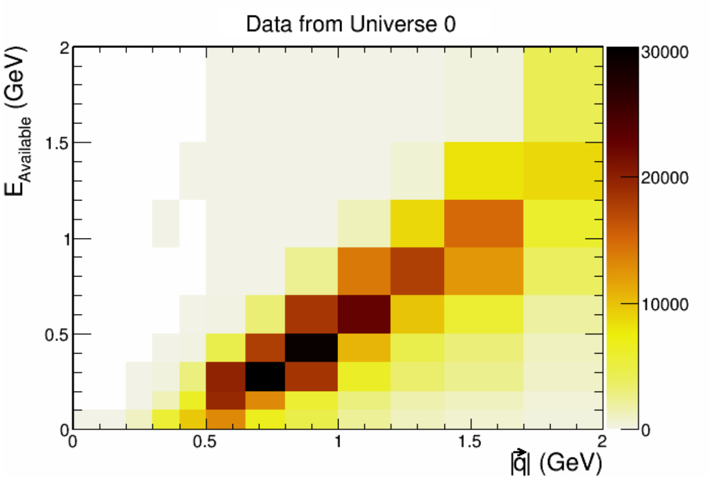


Figure 10.7: Distribution of total selected events. The event rate follows the general ridge structure observed with the inclusive differential cross section of Fig. 10.5.

10.1.1 Establishing templates for known processes

To estimate the data excess, the contributions from known neutrino-nucleon interaction processes in a Fermi-model nucleus need to be quantified. The known processes include baryon resonance production (RES), shallow and deep inelastic scattering (DIS), and quasi-elastic scattering (QE). Smaller contributing processes and backgrounds include neutral current reactions (NC), $\bar{\nu}_\mu$ CC events, ν_e CC events, neutrino-nucleus CC coherent scattering, and a 2p2h-MEC contribution that is not part of the signal definition. All of the latter processes are included in an Other category.

The four templates that are used in the analysis presented here are based on GENIE version 2.12 [19]. The templates for the four processes are shown in Fig 10.8. The templates are as follows: CC-RES (a), CC-DIS (b), CC-QE (c), and Other (d). The largest population is with RES events as shown in (a). DIS events (b) populate relatively remotes regions of the phase space. CCQE events (c) populate a region in which 2p2h-MEC is likely to be abundant. The distribution of events in the Other category (d) contains relatively few events that mostly populate the same region of phase space as QE.

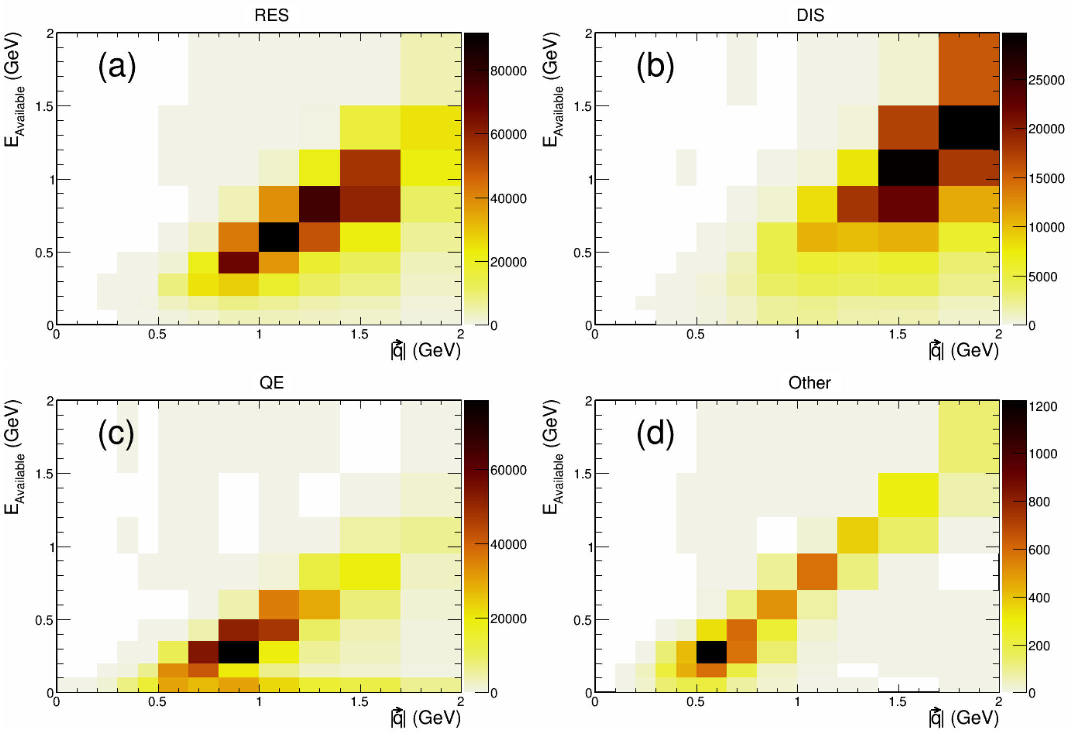


Figure 10.8: Event distributions for the four templates that characterize all non-2p2h-MEC processes, namely RES, DIS, QE, and Other. The RES template makes up the majority of selected events.

10.1.2 Construction of a control subsample

A control subsample, designed to be as devoid of 2p2h-MEC as possible, is defined for the purpose of constraining the rate normalizations of the templates. The RES and DIS templates have their normalizations constrained via a fit to data in the subsample. The CCQE template cannot be sufficiently separated from 2p2h-MEC and as such has a fixed normalization, but still participates as a static background in the fit. Similarly, the Other contribution is too small to enable a data-driven normalization to be obtained and so the normalization is set according to the GENIE prediction.

For the purpose of defining a control sample, we first define an “event prong”. Event prongs are clusters of hits, some of which lie in close proximity to the event vertices. Furthermore we require the visible energy of a reconstructed prong to be larger than a selected value. With this definition in mind, events of the control sample are required to satisfy any one of the following criteria:

- The event has a non-muon prong of length $>$ 100 cm;

- the event has more than 2 reconstructed prongs from the primary vertex. We define a reconstructed prong to be a cluster of hits of reconstructed calorimetric energy greater than 50 MeV, that has one end in close proximity to the event primary vertex;

- the event has a reconstructed photon.

These criteria define properties that are not expected to be possessed by 2p2h-MEC or of CCQE as shown by the plots in Fig. 10.9. The left-hand plot shows the length of non muon prongs in various reaction categories, and the right-hand plot shows the number of reconstructed prongs. There are relatively few 2p2h-MEC events that have prong lengths greater than 100 cm, or that have more than two reconstructed prongs.

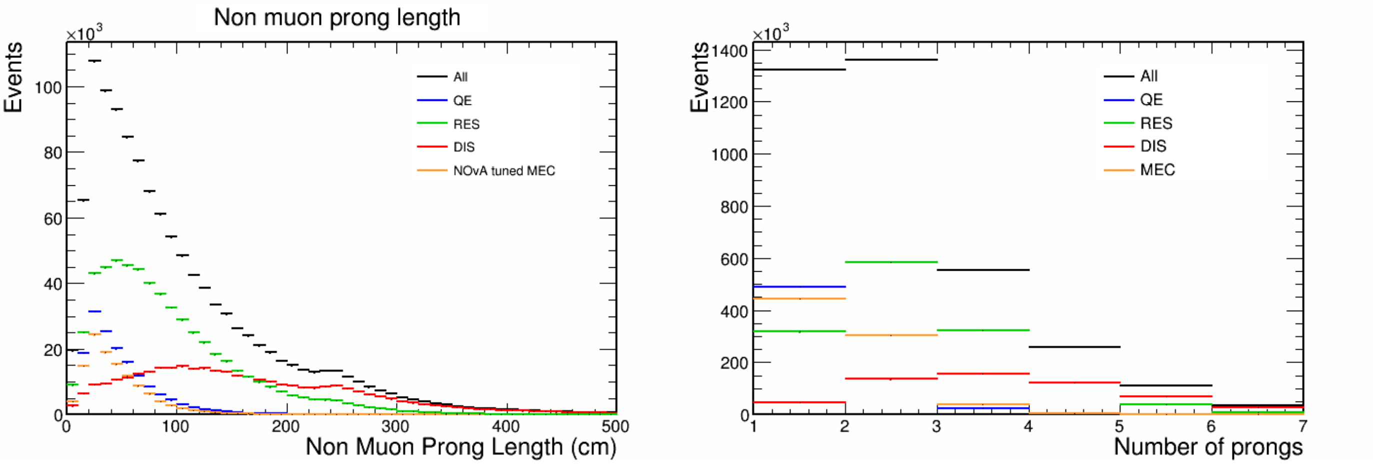


Figure 10.9: The distribution of non-muon prong length is shown on the left, where it seen that 2p2h-MEC events as predicted by the NOvA tune have relatively few prongs with length greater than 100 cm. The plot on the right shows the number of prongs broken out by reaction category; the 2p2h-MEC (gold histogram) have few events with more than two prongs.

Prong length distributions using alternate 2p2h-MEC models are shown in Fig. 10.10. The analysis has considered five 2p2h-MEC models in total including MINERvA and GENIE empirical models in addition to those displayed in Figs. 10.9 and 10.10. In all cases relatively few non-muon prongs of length greater than 100 cm are predicted for 2p2h-MEC processes.

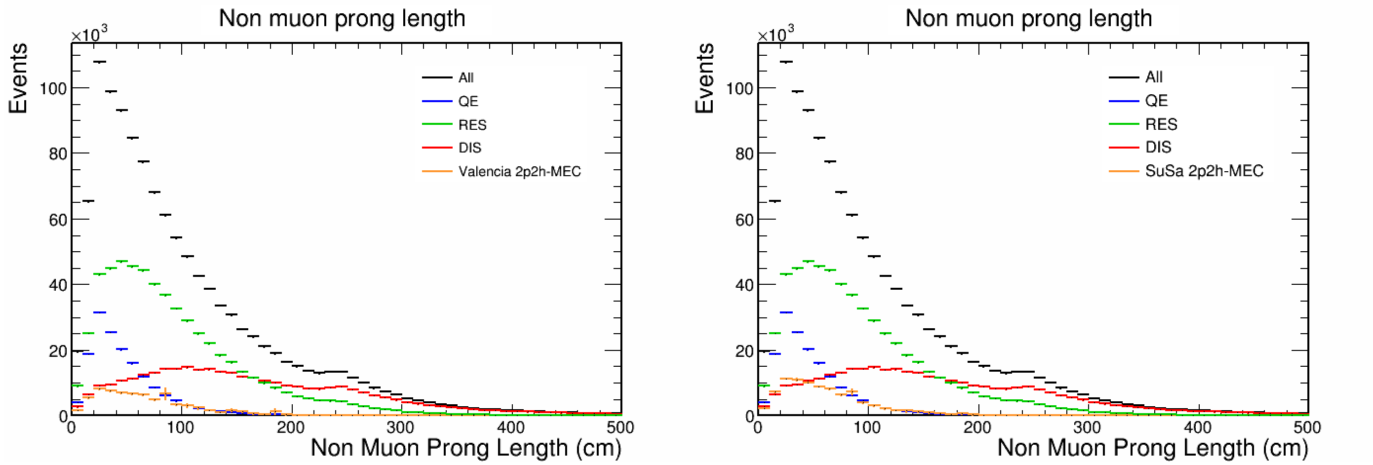


Figure 10.10: Distribution of non-muon prong length obtained using different 2p2h-MEC models. As is the case for 2p2h-MEC predicted the reference MC (Fig. 10.9 left) the distributions here (Valencia (left) and SuSa (right)) show that other 2p2h-MEC models also predict relatively few prongs to exceed 100 cm in length.

Additionally we require that the template events that satisfy the control subsample criteria be normalized to a region that is likely densely populated by 2p2h-MEC events. The kinematic region is taken to be the interval in $|\vec{q}|$ 0.5 GeV/c< $|\vec{q}|$ <1.2 GeV/c.

The signal and background templates that pass the control subsample selections are shown in Fig. 10.12 and Fig. 10.11 respectively. Comparison of the five templates in these two figures shows that the RES and DIS templates contribute a significantly higher event rate than do the QE, Other, and Signal templates.

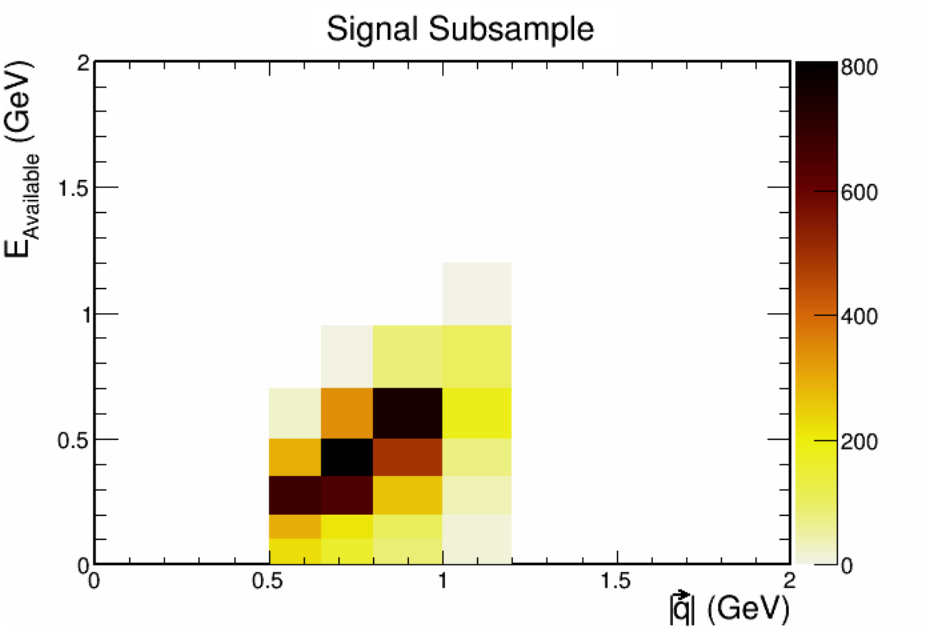


Figure 10.11: The true 2p2h-MEC signal events that pass the control subsample selections. The events concentrate around 0.5 $\leq |\vec{q}| \leq$ 1.0 GeV/c and 0.2 $\leq E_{avail} \leq$ 0.7 GeV.

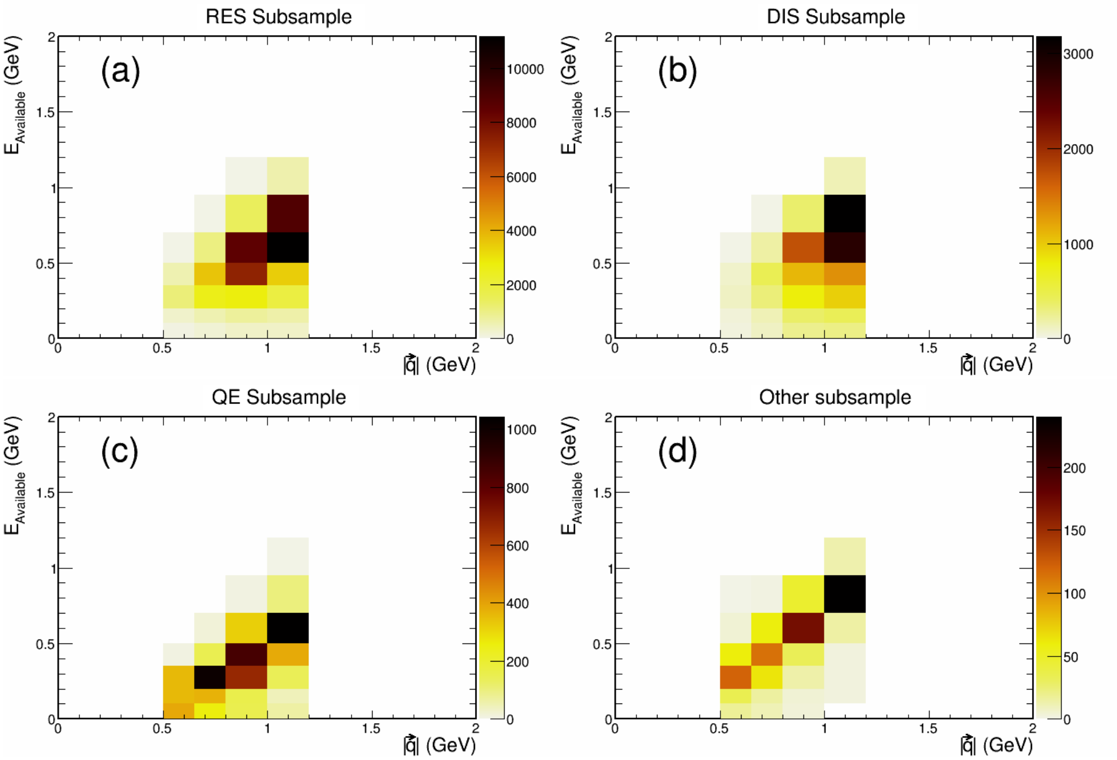


Figure 10.12: Templates for the contributing processes that pass the control subsample selections. Note the differences in event rates as displayed by the right hand scales. By design the RES (a) and DIS (b) templates contribute a significantly higher event rate than do the CCQE (c), Other (d), and residual signal (Fig. 10.11).

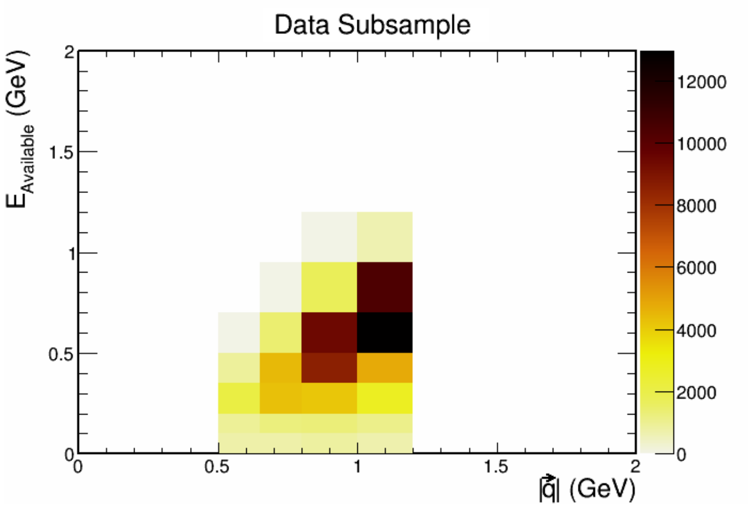


Figure 10.13: Distribution of pseudo-data events that satisfy the control subsample criteria and lie within the restricted kinematic domain. Notice that the event populations here (right-side axis) are much higher than those of the residual signal, CCQE, and Other templates of Figs. 10.11, 10.12(c), and 10.12(d).

10.1.3 Fitting procedure

The procedure to determined the data excess over single nucleon scattering is as follows: Initially the RES, DIS, and 2p2h-MEC events are fitted in the control subsample kinematic region to the data. Figures 10.11 and 10.12 show the templates that are used. The Monte Carlo pseudo-data is displayed in Fig. 10.13; the templates are fit to distributions of this kind to establish their normalizations. The QE and Other templates are also present in the fit however the normalizations for those templates are held fixed. The normalizations of the RES and DIS templates of the main sample are then set according to the normalizations obtained in the fit. With the normalizations for the known processes thusly established, the templates (RES, DIS, QE, and Other) are subtracted from the pseudo-data to obtain the signal excess relative to conventional neutrino-nucleon scattering.

To test the fitting procedure, fits to the control sample are carried out using many different systematically independent Monte Carlo universes. The ensemble of universes represents the variations allowed to the reference Monte Carlo. Each reference MC is then compared to the a single Monte Carlo universe used as pseudo-data. The pseudo-data MC is statistically and systematically independent from reference MCs of the various universes. The extracted data excess is then compared to the true 2p2h-MEC from the pseudo-data to see how well the fit estimates the true data excess. Figure 10.14 shows the χ^2 between the extracted signal and the true signal for 100 independent universes. The χ^2 prior fitting is displayed in red, and the χ^2 after fitting is shown in black. The χ^2 before and after fitting is similar in most cases, but the fitting eliminates occurrences of large χ^2. The ratio of predicted events to true signal events, however greatly improves on average upon fitting. The extend of the improvement is seen by the before and after distributions shown in Fig. 10.15.

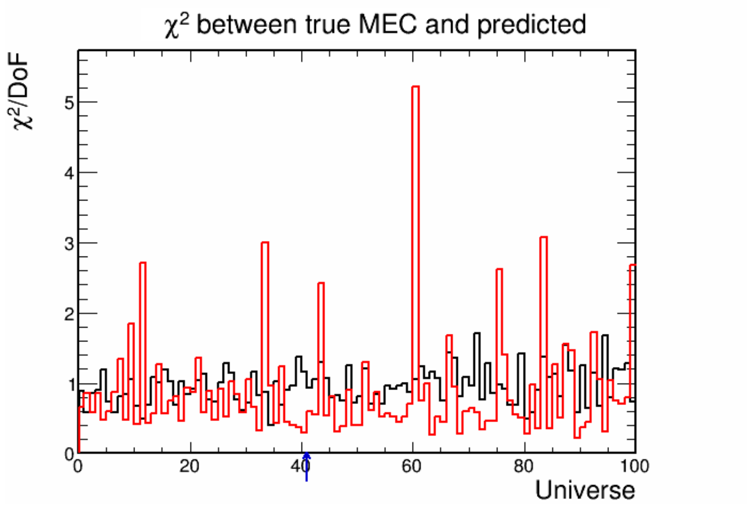


Figure 10.14: Comparison of χ^2/DoF values for the extracted versus the true 2p2h-MEC signal for each of 100 universes. The χ^2/DoF before and after fitting is shown by (red) and (black) distributions respectively. Extractions with large χ^2/DoF are seen to be eliminated by the fit.

The measurement demonstration carried out here uses the default NOvA tuned Monte Carlo [33] as pseudo-data, and an error band is established using the multi-universe technique [44] as illustrated in Fig. 10.15. Figure 10.16 shows the estimated signal for the central value (left) compared to the true signal from the pseudo-data (right). The ratio of the estimated signal divided by the true signal is shown in Fig. 10.17. The estimated-to-true-signal ratio approximates 1.0 in the kinematic regions populated by 2p2h-MEC, however significant deviations are seen in the outer kinematic regions.

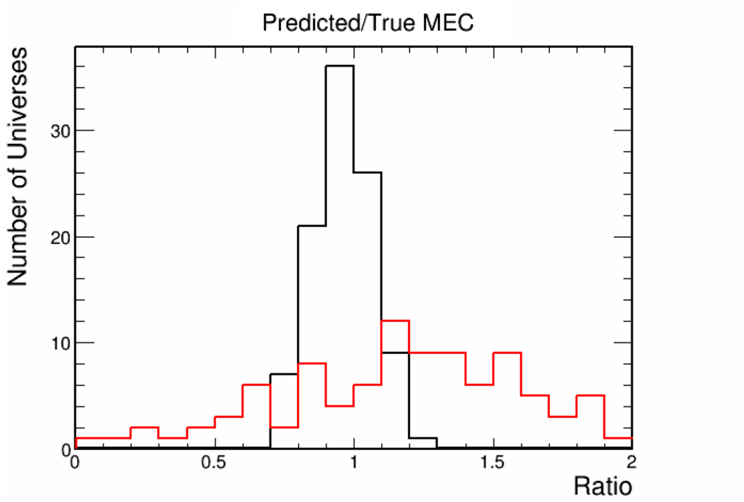


Figure 10.15: Distributions of ratios for (total estimated signal events)/(total true signal events) determined before fitting (red), and after fitting (black). The overall excess event rate due to 2p2h-MEC after fitting is determined to within 30% (FWHM).

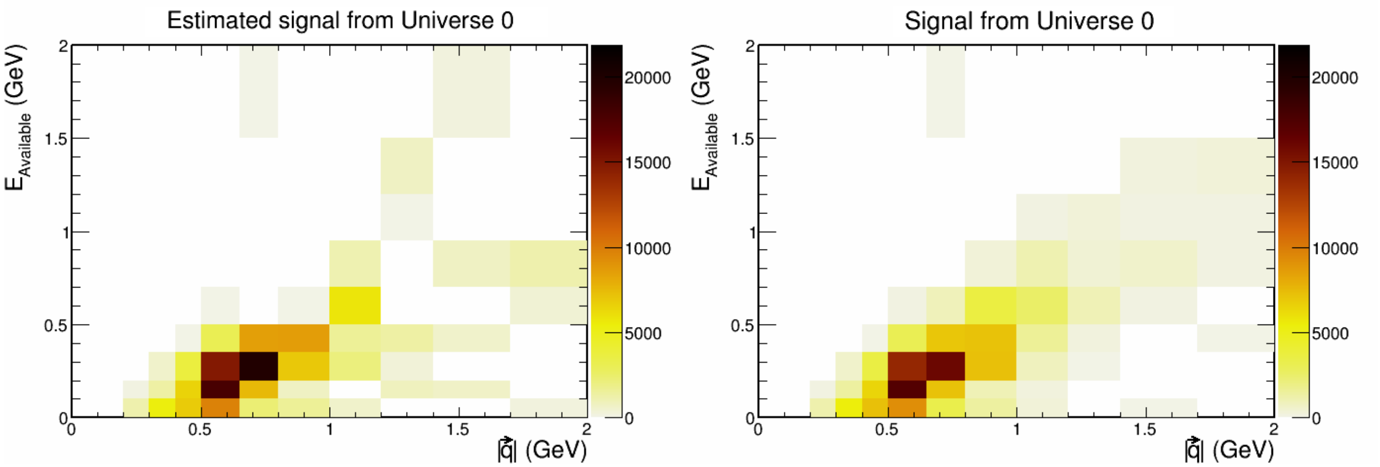


Figure 10.16: Estimated signal (left) and true signal (right) for the central value excess estimation. The estimated signal is similar to the true Monte Carlo signal in the regions where 2p2h-MEC is concentrated.

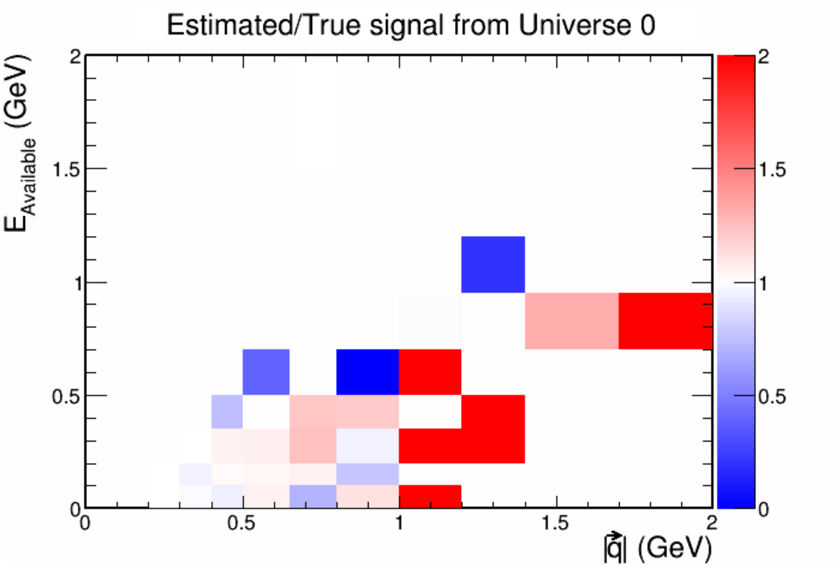


Figure 10.17: Ratio of estimated signal to the true signal. The ratios are of the order 1.0 in the regions of dense 2p2h-MEC populations, however the ratios are large in some bins of the outer regions of $|\vec{q}|$ and E_{avail}.

The fractional error on the signal estimate from the multi-universe technique is shown in Fig. 10.18. The fractional error is around 30% in areas where the signal is concentrated, $|\vec{q}| < 1.2$ GeV/c and $E_{avail} < 0.5$ GeV, but the fractional error can exceed 200% in the outer kinematic regions. Fortunately this situation can be addressed by using additional template normalization regions with respect to the control subsample.

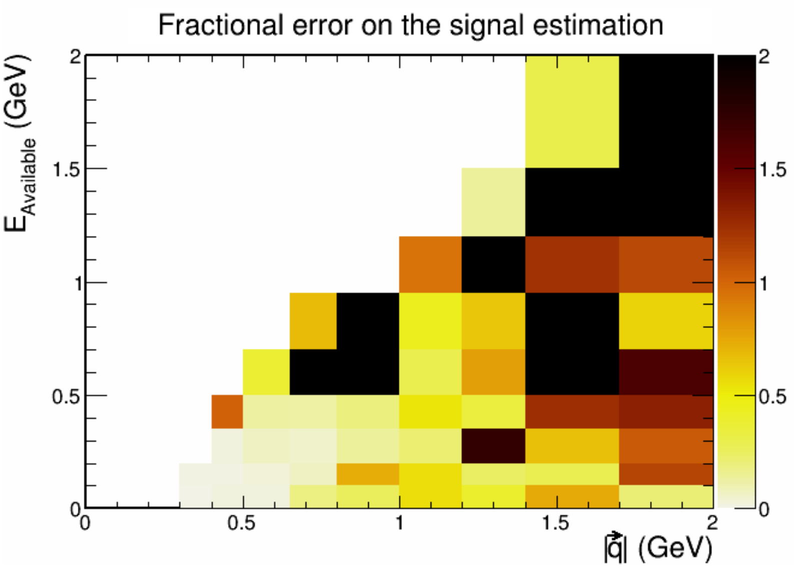


Figure 10.18: Fractional error on the signal estimation. The error is around 30% in the phase space where 2p2h-MEC is predicted, but is much higher elsewhere. The large uncertainties for the kinematic regions distant from the main signal region can be addressed by subdividing the templates.

10.1.4 Fitting to multiple regions

It was found that fitting the control sample to just the signal regions gave poor results outside the signal region. With the previously described approach only the template normalizations are adjusted; there is no adjustment allowed for template shapes in the kinematic phase space. In order to introduce a capability to modify template shapes, we split the phase space into three regions and perform the subsample fits within those regions. The three regions are defined as:

1. The signal region (I): $|\vec{q}| \leq 1.2$ GeV/c;
2. the intermediate region (II): $1.2 < |\vec{q}| < 1.4$ GeV/c;
3. the outer region (III): $1.4 \leq |\vec{q}|$ GeV/c.

The three regions were selected according to relative contributions among the component processes. Region (III) is populated by RES and DIS at nearly equal strength; region (II) is dominated by RES; region (I) is predominantly RES and QE. Our approach is to fit RES, DIS, and Residual Signal in the signal region (I) as previously described in Sec. 10.1.3. In the outer region

(III), we fit the same set of three templates. However in the intermediate region (II), the DIS, RES, and Residual Signal normalizations are taken to be the average of the fit normalizations from the signal and remote regions. This is done to ensure continuity between the different regions. Fitting the RES template in the intermediate region was explored and ultimately rejected due to lack of improvement in the resulting background estimate.

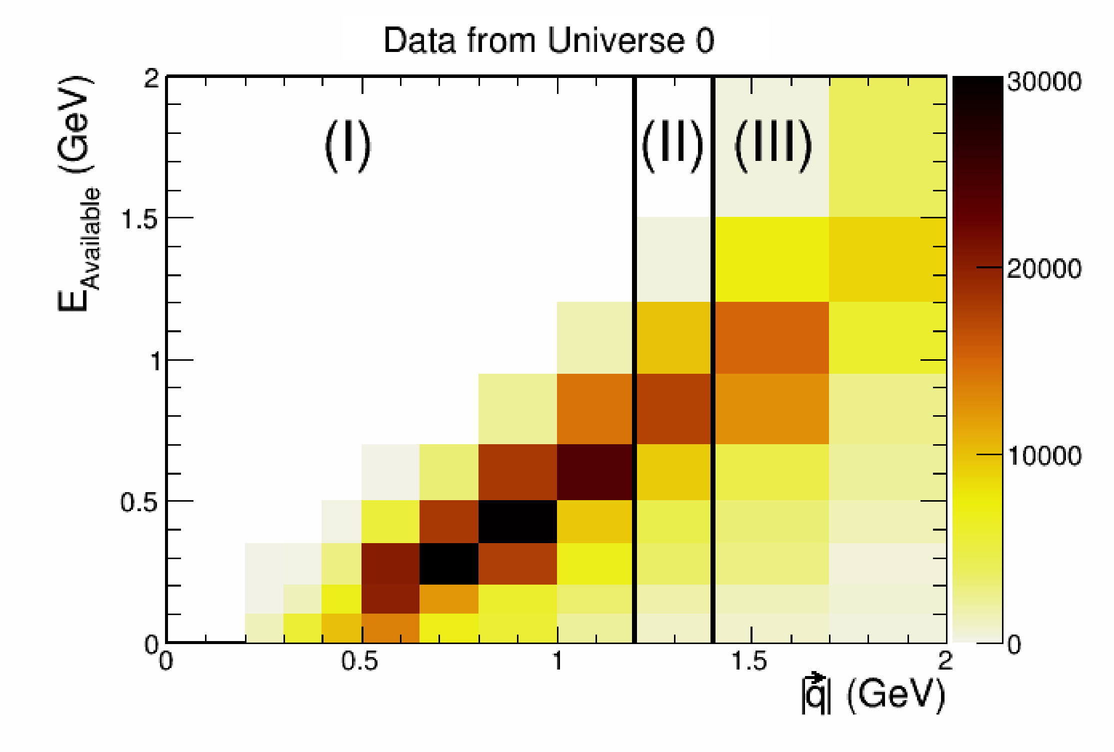


Figure 10.19: Division of the kinematic plane into three regions defined in the text. Template normalizations are fit to the control sample in region I and III; their averages establish the normalizations in region II.

The template distributions over the full phase space are shown in Fig. 10.20. One observes that the RES and QE templates have sizable populations that lie closer to the signal region than do the DIS and other templates.

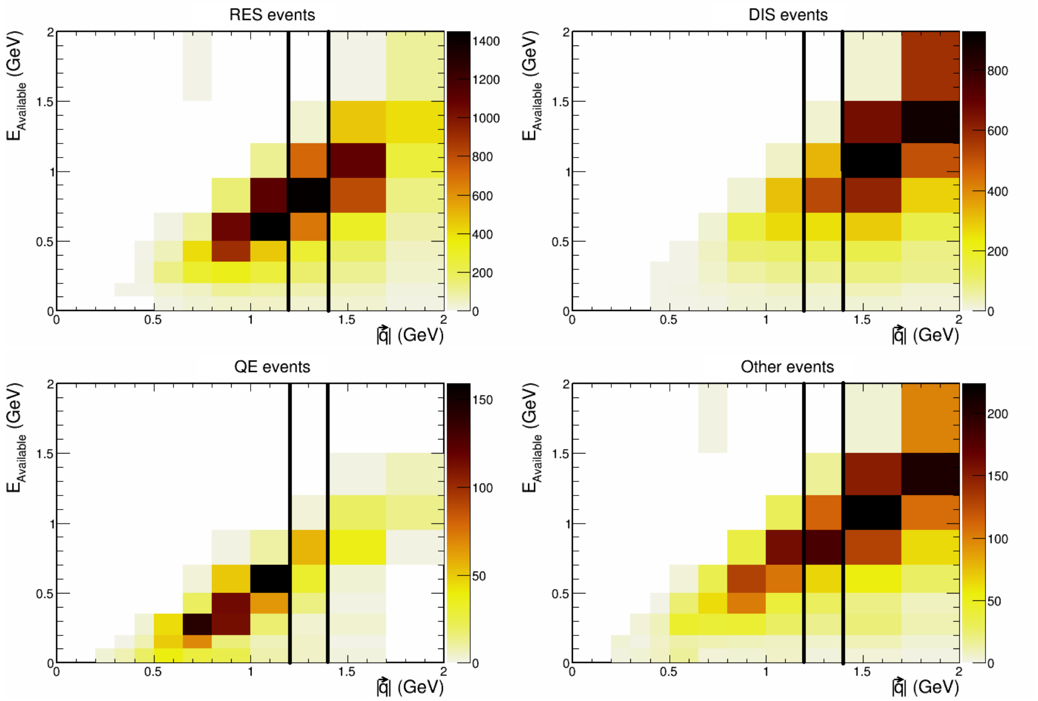


Figure 10.20: Templates over the full phase space for the component processes that pass the control subsample selections. The templates are fit separately to regions I and III. Note the differences in event rates as displayed by the right hand scales.

The description of the data in the control sample is greatly improved by fitting the template normalizations to that sample. The red histograms in Fig. 10.21 show the χ^2/DoF for the template description of the control sample before fitting. As described above, the templates are fit separately in regions I and III and the average of those normalizations for each template is used for the normalizations in region II. The black histograms in Fig. 10.21 show the χ^2/DoF after fitting in each of 100 universes (left-hand plot), and the distribution of the χ^2/DoF for the ensemble (right-hand plot). The figure clearly demonstrates that fitting of template normalizations greatly improves the description of the control sample over the full kinematic phase space. Figure 10.22 provides similar comparisons of the fit results when the fitting is confined to signal region, that is, region I. The χ^2/DoF is observed to have values that in many cases are below 1.0. This is in contrast to the situation before fitting (red histograms) where large swings in the χ^2/DoF are observed.

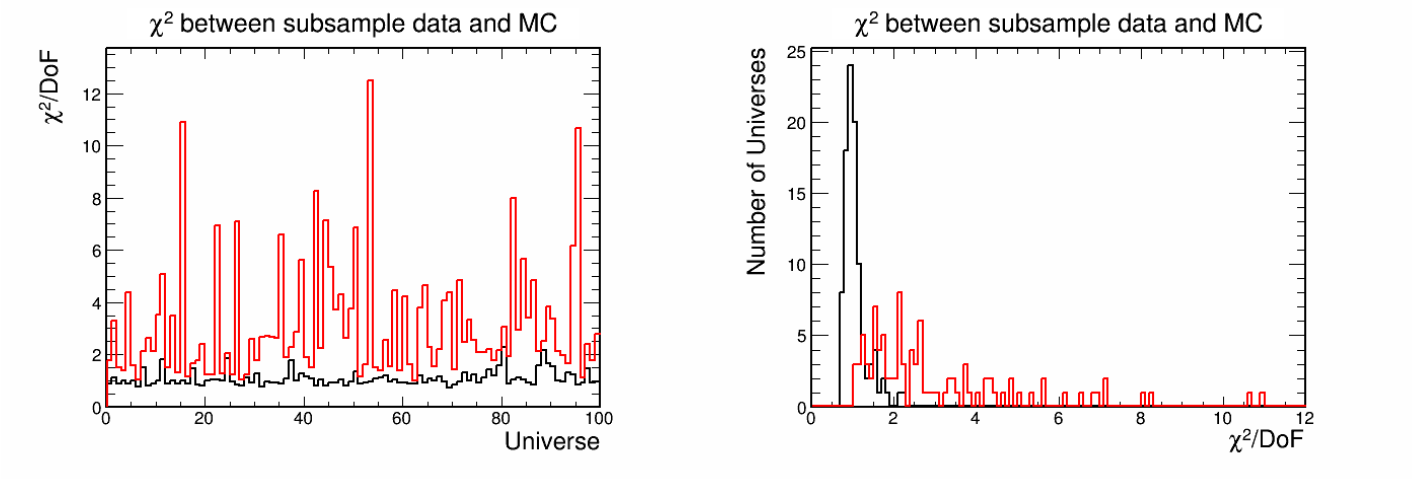


Figure 10.21: Summary of template fits to the control region taken over the full kinematic phase space. Fitting is done separately to regions I and III and the average of the normalizations is used in region II. Left: Histograms show the χ^2/DoF from matching the templates to the control sample before fitting (red) and after fitting (black) for each universe of a 100 universe ensemble. Right: Histograms show the distributions of χ^2/DoF before (red) and after (black) fitting. Clearly the fit improves the template characterization of the control sample.

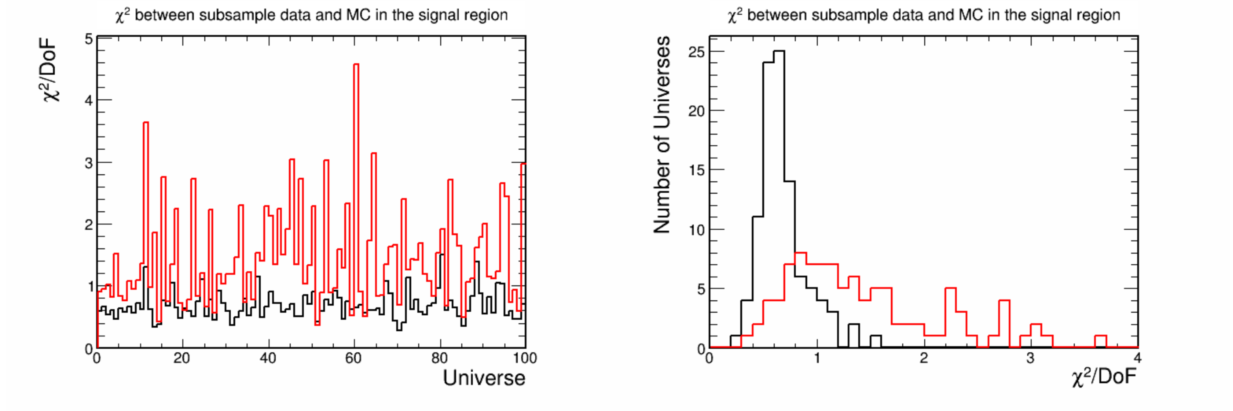


Figure 10.22: Plots show χ^2/DoF distributions when fitting is restricted to the signal region (region I). Left: Histograms show χ^2/DoF before (red) and after (black) the fit in each of 100 universes. Right: Histograms show the distributions of χ^2/DoF before (red) and after (black) fitting of the templates to the signal region.

10.1.5 Fit Performance

In this section we first show how the fit performs using the Nominal Universe. We then summarize the fit performance for an ensemble of universes. Figure 10.23 shows the background, that is, the non-2p2h-MEC events, estimated after fitting to the three kinematic regions. The background is dominated by the RES contribution.

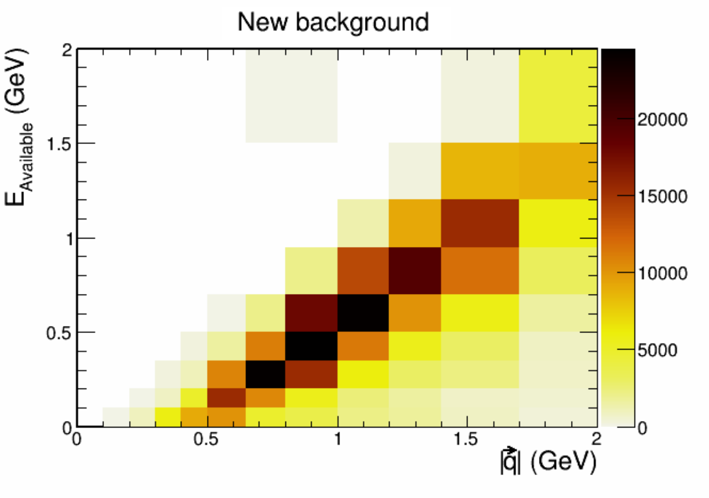


Figure 10.23: Background from the multiple-region fit as obtained by summing over the fit templates RES, DIS, Residual Signal plus the two templates that are taken straightaway from GENIE, namely the CCQE and Other templates.

The 2p2h-MEC excess estimated by subtracting the background (Fig. 10.23) from the pseudo-data (Fig. 10.7) is shown in Fig. 10.24 (left) and is compared to the true signal (right). Figure 10.25 shows the difference between the estimated signal and the true signal. The difference is centered around 0 events but extends to ± 2000 events; these swings are ≈ 25% of a peak rate of the true signal.

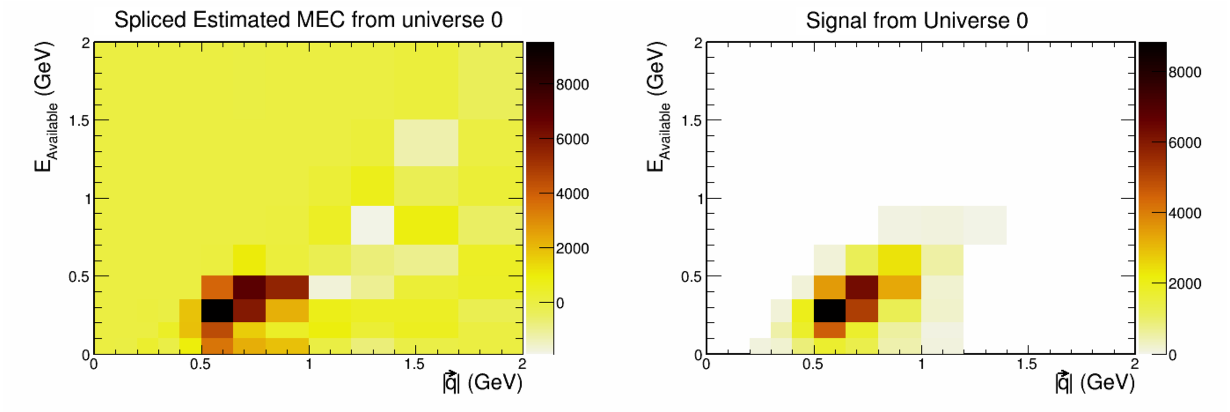


Figure 10.24: The candidate 2p2h signal obtained as the data excess that lies above the sum over templates for conventional physics processes. The signal, which is derived from the GENIE empirical MEC model, peaks around 0.7 GeV/c in $|\vec{q}|$ and 0.5 GeV in a E_{avail}. Left: Plot shows the signal after subtracting the background. This is to be compared with the plot (Right) which shows the True Signal.

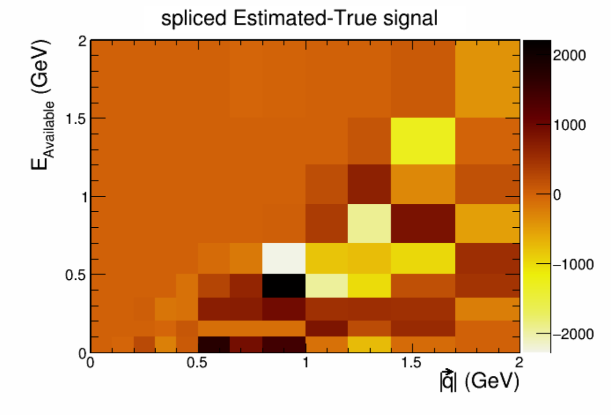


Figure 10.25: The bin-by-bin difference (Estimated-True) 2p2h-MEC. varies from being excessive to being too weak, however the general location of the signal appears to be ascertained. Note that the fit to the outer region is much improved.

The projections of the estimated signal are shown in Fig. 10.26, with $|\vec{q}|$ shown on the left, and E_{avail} shown on the right. The $|\vec{q}|$ distribution peaks at 0.5 GeV/c and has a FWHM of 0.3 GeV/c, and is nearly zero outside of the central peak. The distribution of E_{avail} peaks sharply at 0.3 GeV with a FWHM of 0.4 GeV and is also zero away from the central peak structure.

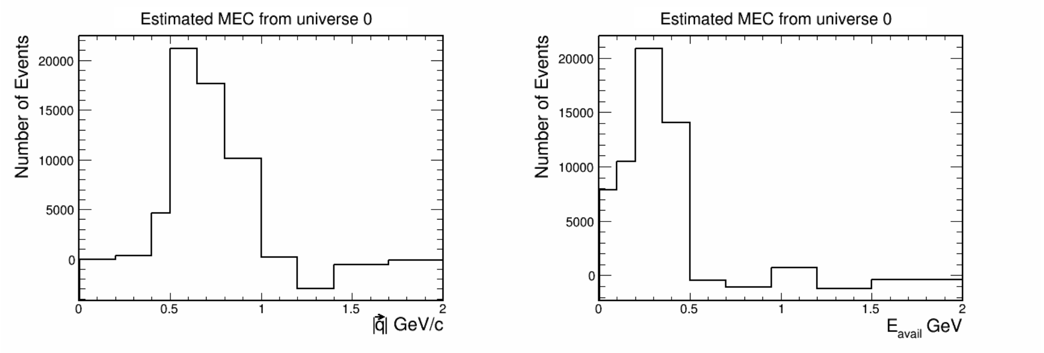


Figure 10.26: Projections of the estimated signal in each two kinematic variables. For $|\vec{q}|$, the distribution peaks at 0.5 GeV/c and is skewed toward higher values; it has a FWHM of 0.3 GeV/c. The distribution of E_{avail} peaks at 0.3 GeV and has a FWHM of 0.4 GeV. The distribution effectively cuts off at 0.5 GeV.

The above plots were obtained using the Nominal Universe (Universe 0). We now report our multi-region template fit to large numbers of universes. Each fit is to the same altered data, but uses a systematically different GENIE MC obtained by altering the GENIE systematic parameters.

The χ^2/DoF between the true and signal MEC is displayed in Fig. 10.27. The left-hand plot shows the universe-to-universe χ^2/DoF, where the red distribution is the χ^2/DoF without any fitting and the black distribution is after the fit. The right-hand plot shows the distribution of χ^2/DoF over the universes. The pre-fit distribution (red) peaks lower than the post-fit distribution (black) but has a much larger tail. Overall, fitting reduces the average χ^2/DoF. Figure 10.28 shows similar distributions to Fig. 10.27, but is restricted to just the signal region (I). The χ^2/DoF in the signal region in most universes is lower than the total χ^2/DoF, and the post fit χ^2/DoF peaks around 1.8.

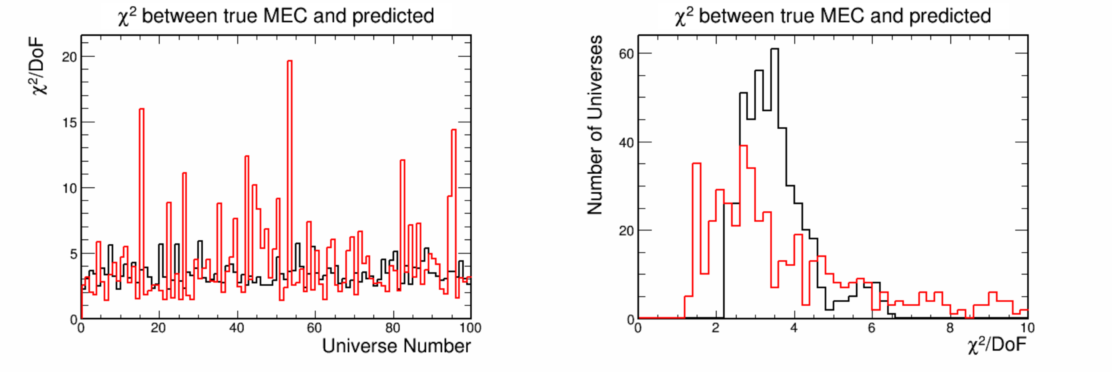


Figure 10.27: Left: The red histogram shows the χ^2/DoF for each universe without fitting, while the black histogram shows the χ^2/DoF after the fit. One observes that large swings in χ^2/DoF are dampened by the fit. Right: The distribution of χ^2/DoF (the mean value is 3.5) peaks higher after the fit but has fewer results with high χ^2/DoF.

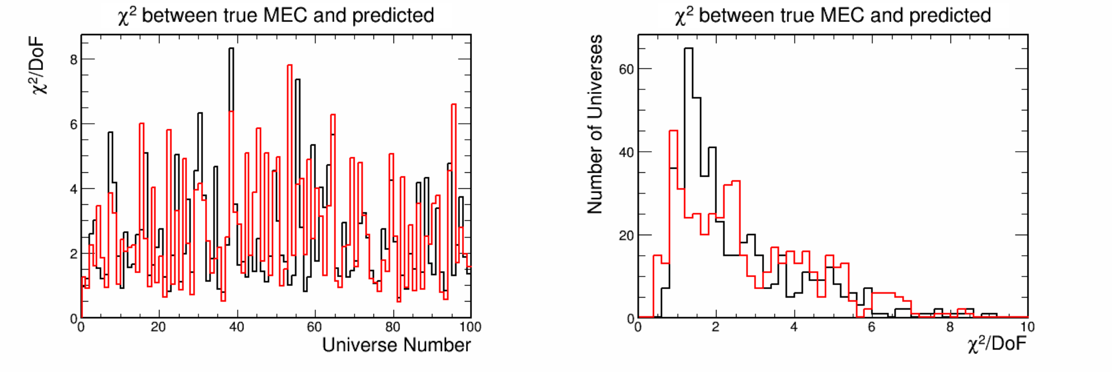


Figure 10.28: Plots show χ^2/DoF distributions as restricted to the signal region. From the right-hand plot one surmises that fitting to the control sample gives improvement over GENIE nominal, however the improvement is modest.

The ratios of predicted to true 2p2h-MEC is shown in Fig. 10.29 with the universe-to-universe ratios on the left, and the frequency of each ratio on the right. The fit greatly reduces the swings, universe-to-universe, in the ratio values and tends to give ratios in the vicinity of unity. It is clear that the fit is required to estimate the 2p2h contribution; the nominal GENIE prediction (red histogram) is incapable of identifying the signal amount.

Figure 10.30 is the ratio of predicted to true signal events that occur in the signal region. Once again the fit yields an improvement over the estimation obtained with just the nominal MC without template normalization adjustments. The fit distribution is narrower than the nominal MC prediction. The latter has a FWHM of 0.7 compared to the fit result of 0.5.

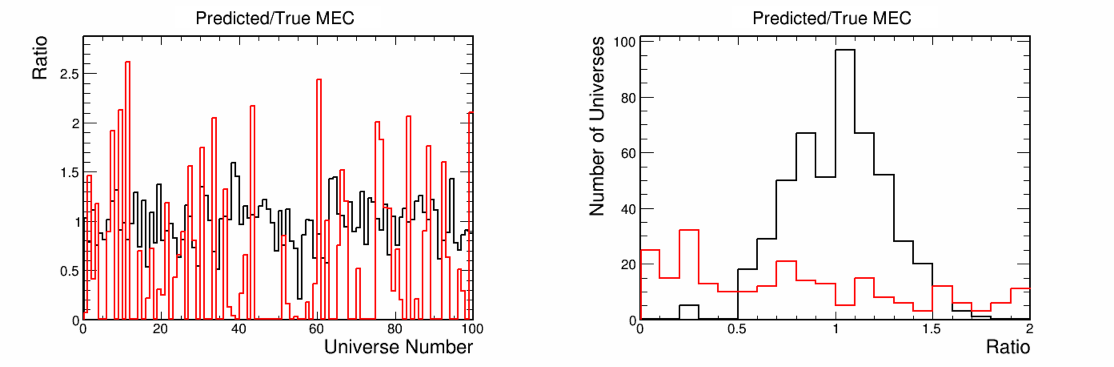


Figure 10.29: The left-hand plot shows the ratio of predicted 2p2h-MEC events divided by true 2p2h-MEC events, for each universe. The right-hand plot is the distribution of ratios from an ensemble of 500 universes. The distribution peaks at 1.0 and the FWHM is 0.5; the latter spans the range 0.7 to 1.2.

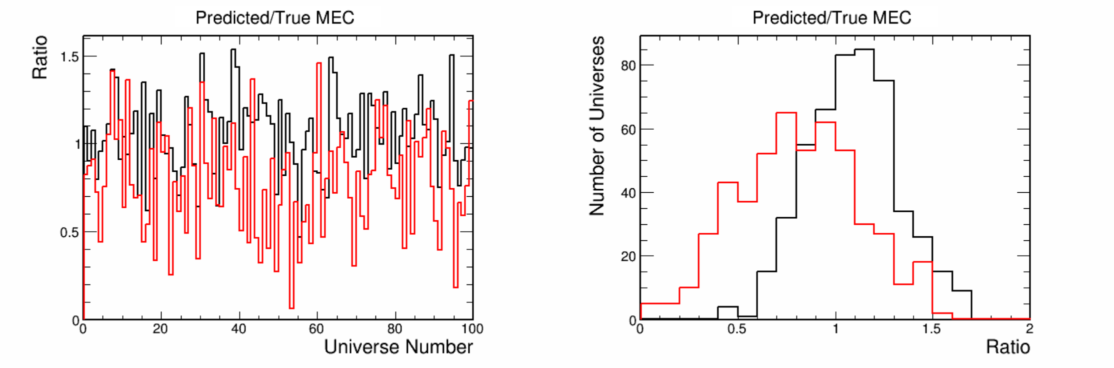


Figure 10.30: Left: the event ratios in the signal region only. Right: The prediction obtained with nominal MC (red) is observed to under-predict on average, that is, the per universe values tend to lie below 1.0. The fit brings the ratio values closer to 1.0 on average. Right: The fit value tends to overshoot, peaking at 1.1, with the distribution FWHM extending from 0.8 to 1.3.

The deviation of the predicted 2p2h-MEC from the true signal 2p2h-MEC is shown in Figs. 10.31 and 10.32 for the entire kinematic plane and for the signal region respectively. In the full kinematic space the deviation without fitting (red) is not concentrated at any particular value, here no inference can be made about 2p2h-MEC. After fitting (black) the deviation is concentrates around zero and has a FWHM of 0.5.

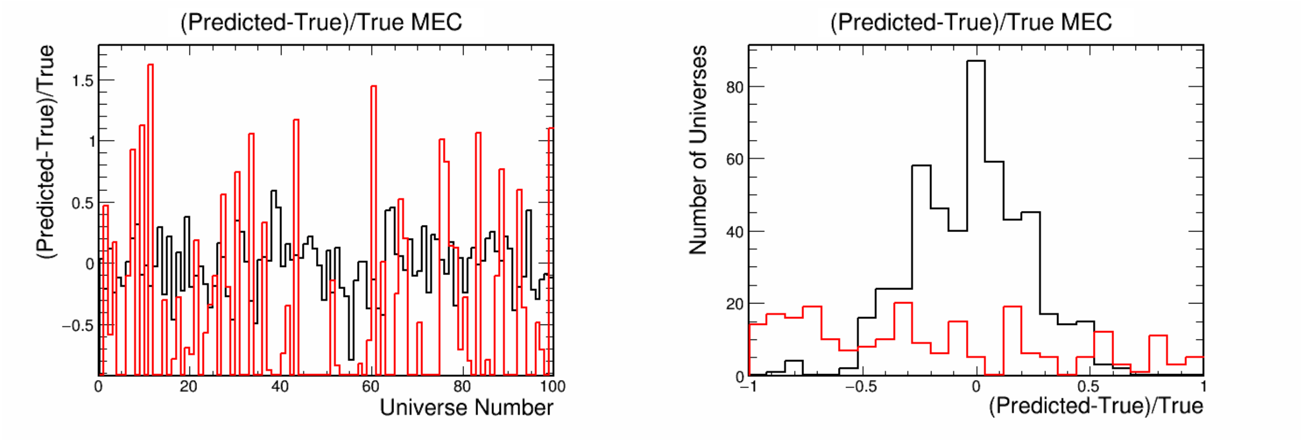


Figure 10.31: Left: Plot shows the deviation in each universe; the fit greatly reduces the range that the values populate. Right: The fit results give a broad, block-shaped distribution that is centered at zero and ranges from -0.3 to +0.2.

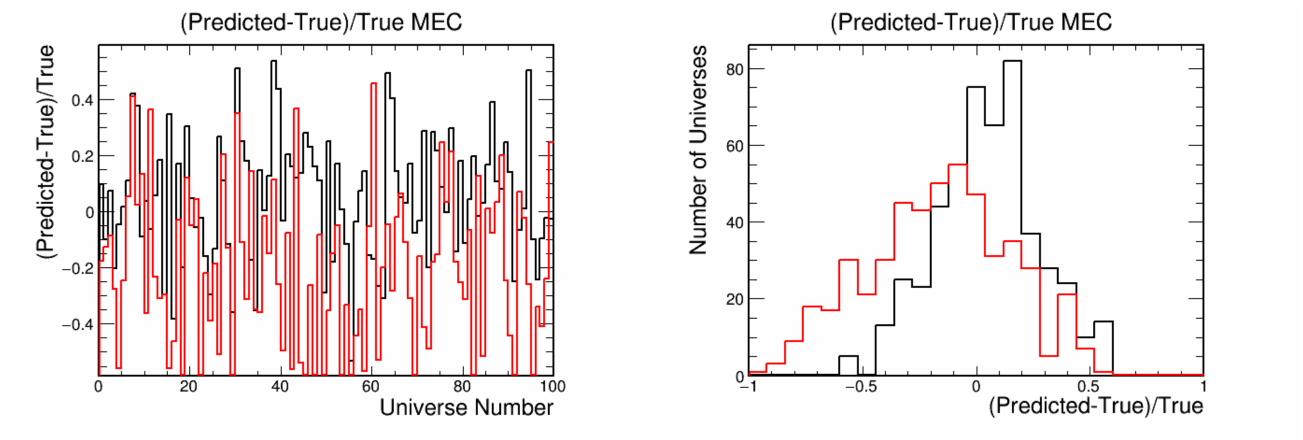


Figure 10.32: Plots show the deviation in the signal region. Left: As in Fig. 10.31, the fit tends to cluster the results in the vicinity of 0.0. Right: The peak post fit is shifted from 0.8 to 1.0 and the FWHM is reduced from 0.7 to 0.4.

Figure 10.33 shows the average bin-by-bin residual for all bins in a universe, for an ensemble of 100 universes. Recall that the residual is defined as:

$$Residual = (\text{Extracted-signal - True-SuSa-MEC})/\text{True-SuSa-MEC}. \tag{10.1}$$

Figure 10.33 (left) shows the residual over the full phase space. The distribution is centered at -0.30 indicating that the extraction method tends to underestimate the signal. However, when the extraction is restricted to the signal region, the distribution of the residual (Fig. 10.33 (right)) is centered at 0 and exhibits a FWHM of 0.20. The improvement of the residual when the extraction is restricted to the signal region indicates a tendency of the templates to overestimate the background when anchored to the kinematic outer regions.

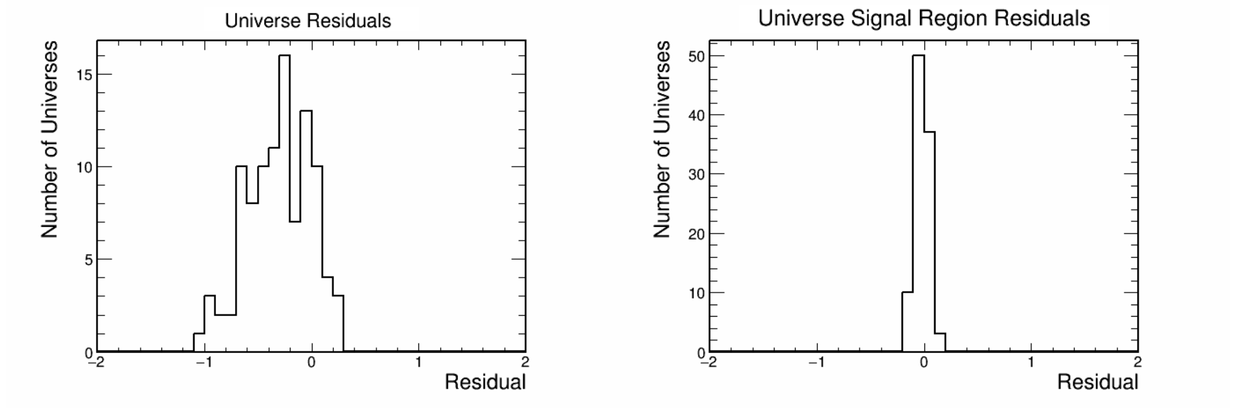


Figure 10.33: The average of the bin-by-bin residuals over the whole phase space (left) versus restricted to the signal region (right). Over the whole phase space the extracted signal tends to underestimate the amount of signal.

We now summarize the behavior of the template normalizations that are obtained from the fit, it is important to note that no penalty terms were applied to constrain the normalizations to within "reasonable" values (e.g. normalization changes less than 50%). The normalization of the RES template is shown in Fig. 10.34, where the black histogram shows the RES normalization in the signal region, the blue histogram shows the RES normalization in the intermediate region, and the red histogram gives the RES normalization in the outer region. The normalizations show variations of order 30% from universe nominal values. Figure 10.35 compares the RES normalizations (left) with the DIS normalizations (right). The normalizations for each process and for all three kinematic regions tend to cluster around 1.0, which indicates little change from the nominal MC predication, and for the most part they stay within the range 0.8 to 1.2.

Note that the normalizations in the intermediate region (II) are set to be the average of the normalizations in the signal region (I) and outer region (III).

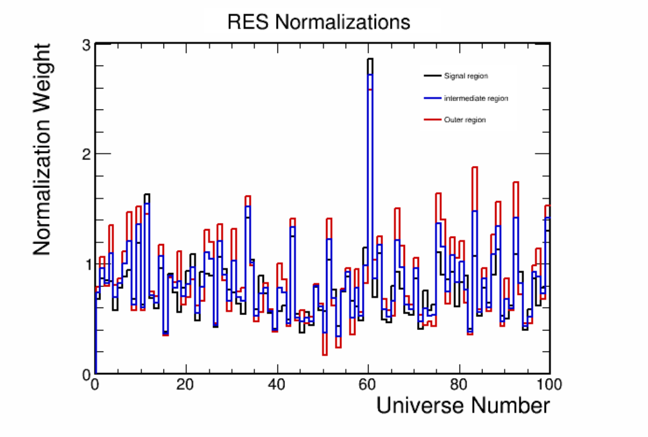


Figure 10.34: The change from the nominal normalization (MC default value) to the fit outcome for the RES template. The black histogram shows RES normalization in the signal region, while blue and red show the normalizations of the RES template in fitting to the Outer region and to the Intermediate region respectively.

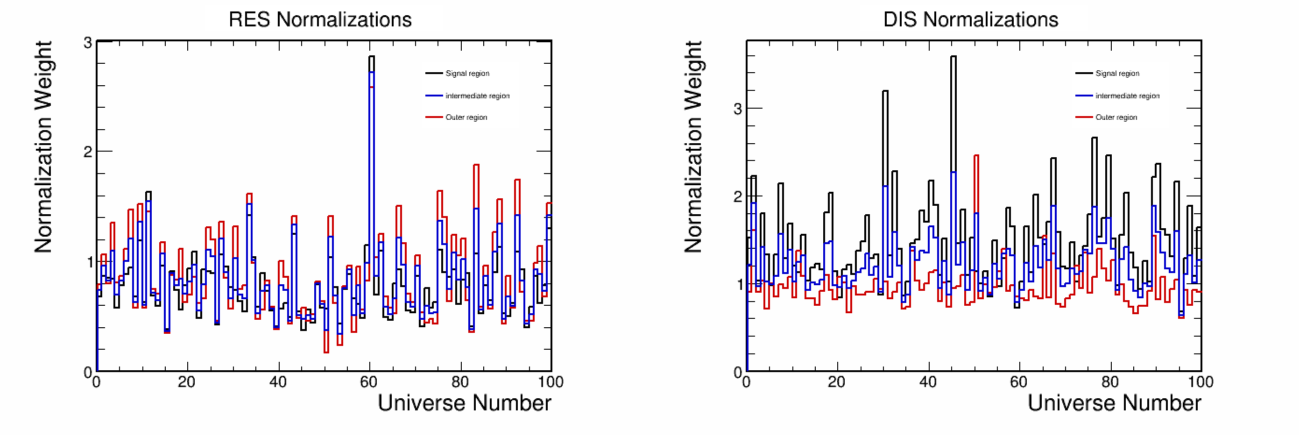


Figure 10.35: Changes in RES normalizations (Left) compared to those of DIS normalization (Right). Roughly speaking, the fit tends to lower the RES normalizations while increasing (on average) the DIS normalizations.

The normalizations for the residual signal template are shown in Fig. 10.36. The results, averaged over the ensemble of 100 universes, indicate that the control sample has approximately 20% more of residual signal than the amount estimated as nominal. There is no predicted signal in the outer region and therefore no normalization is applied.

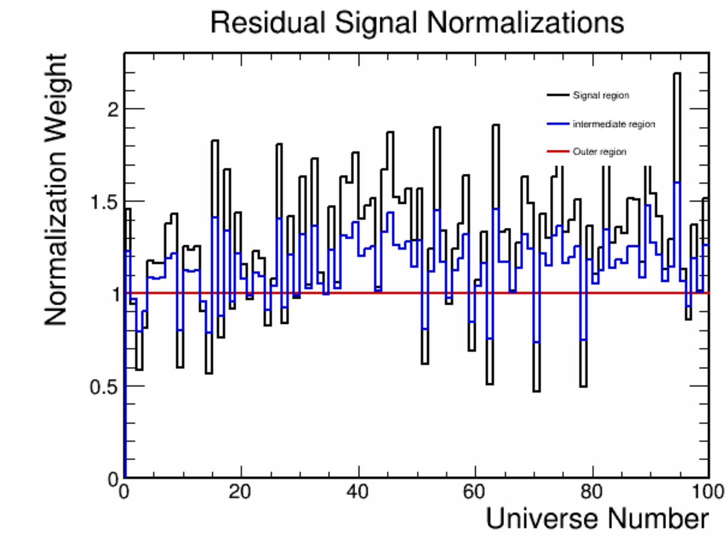


Figure 10.36: The change from nominal of the residual signal in the control sample. The results are for the signal region (black) and the Intermediate region (blue); there is no residual signal in the outer region (a property of the GENIE empirical 2p2h-MEC model).

The universe-to-universe projections are shown in Figs. 10.37 and 10.38 with the $|\vec{q}|$ distribution displayed on the left and the E_{avail} distribution shown on the right. The projections show that the kinematic distributions have shapes that are common to most universes, with $|\vec{q}|$ peaking near 0.5 GeV/c and E_{avail} peaking near 0.3 GeV.

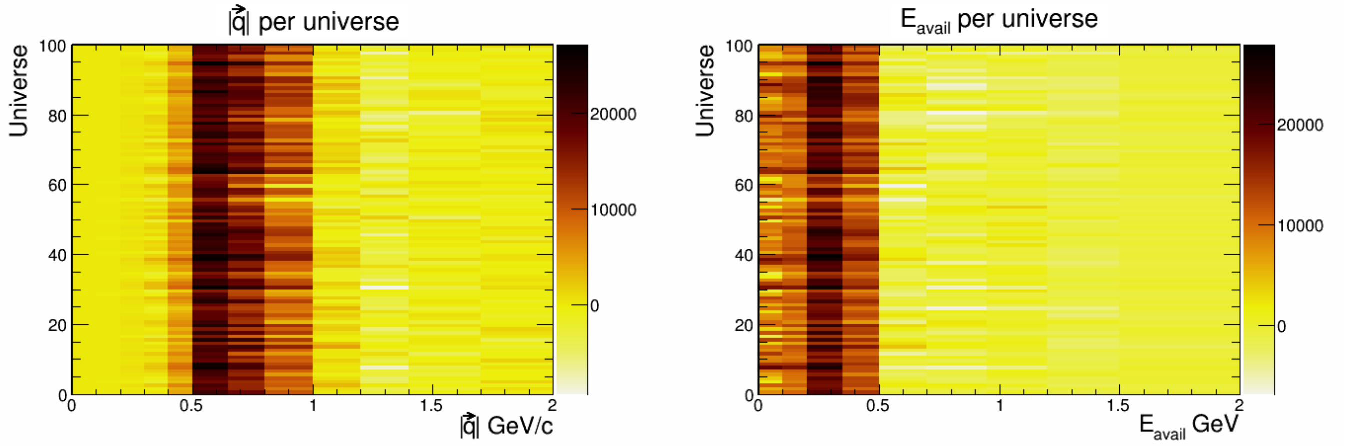


Figure 10.37: The distribution of the estimated signal in each of the kinematic variables, as obtained from the ensemble of 100 universes. The plots indicate that the shapes of the distributions shown in Fig. 5 for $|\vec{q}|$ and E_{avail} are typical for nearly all members of the ensemble.

The distributions of estimated signal from the ensemble of universes is displayed in Fig. 10.38 for $|\vec{q}|$ (left) and for E_{avail} (right) using a 3D perspective. These plots show more clearly that the templates sometimes overestimate the background event rates giving negative contributions to individual bins. This aspect is most pronounced in the intermediate region.

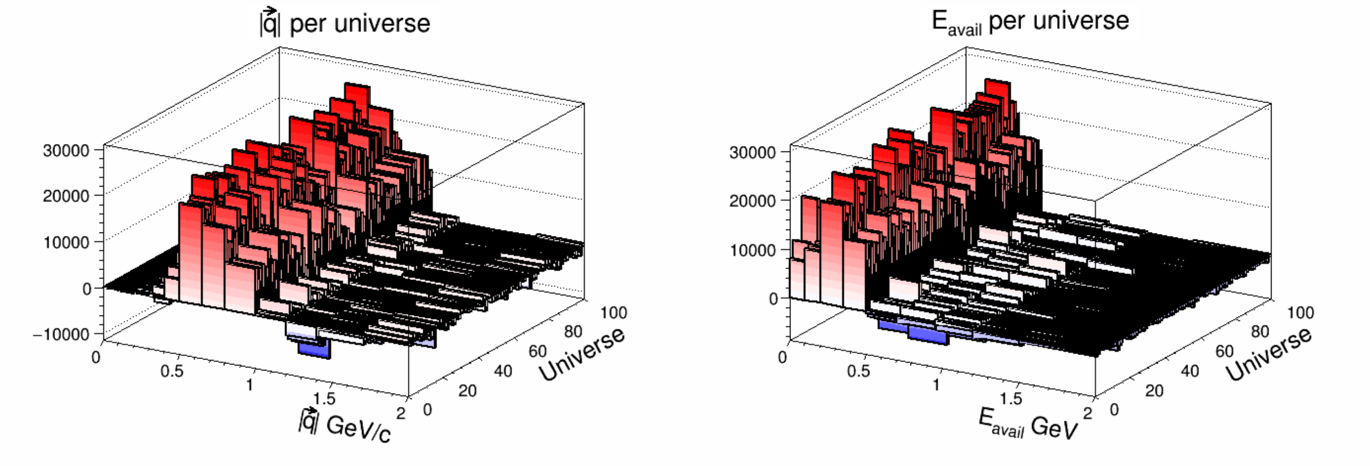


Figure 10.38: Distributions of Fig. 10.37, displayed with a 3D perspective.

Distributions for the mean values of $|\vec{q}|$ and E_{avail} in each of 100 universes are displayed in Fig. 10.39. The mean values for $|\vec{q}|$ peak at 0.6 GeV/c and have a FWHM of 0.38 GeV/c. The mean values for E_{avail} are strongly peaked at 0.35 GeV and have a FWHM of 0.2 GeV. The mode of the projections for each universe is almost always in the same bin, which is 0.55 GeV/c for $|\vec{q}|$ and 0.25 GeV for E_{avail}.

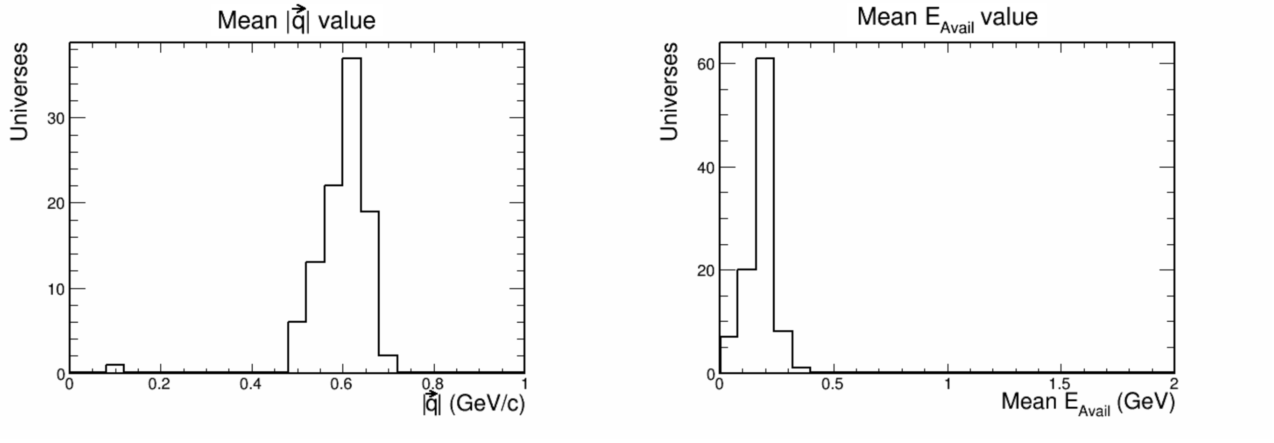


Figure 10.39: The distributions of mean values for $|\vec{q}|$ (left) and E_{avail} (right) over 100 universes. The distributions have FWHMs of 0.38 GeV/c for $|\vec{q}|$ and 0.2 GeV for E_{avail}.

10.1.6 Performance with alternative 2p2h-MEC samples

As further verification of the method we examined signal extraction from universes in which the 2p2h-MEC samples were altered. In the first test the amount of 2p2h-MEC the GENIE empirical model has been doubled, which we have used throughout our development work. In each universe has been carried out an extraction, first with nominal 2p2h-MEC then with doubled 2p2h-MEC rate. Figure 10.40 (left) shows the ratio of the signal extracted for the doubled-rate case versus the nominal rate case. The ratio is centered at 2 with small variations. The extent of the variations can be seen in Fig. 10.40 (right) for which the FWHM about the mean of 2.0 is 0.2.

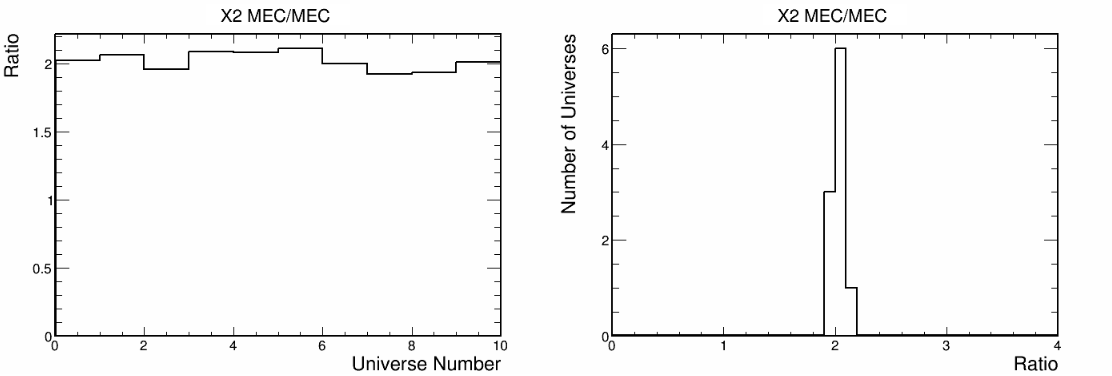


Figure 10.40: The ratio of extracted signal for universes with 2p2h-MEC doubled versus their nominal 2p2h-MEC amounts. The left-hand plot shows the ratio to be very nearly 2.0 in each universe. The right-hand plot shows the distribution of the ratios which has a FWHM of 0.2. The effect of doubling the 2p2h-MEC doubles the extracted signal with very little change, as expected.

Similar good performance is observed when the signal extraction is restricted to the signal region. Figure 10.41 on the left and right displays the ratio and spread about the mean respectively. Figure 10.41 (right) shows that the FWHM has been reduced by 50% when compared to the extraction over the full kinematic region.

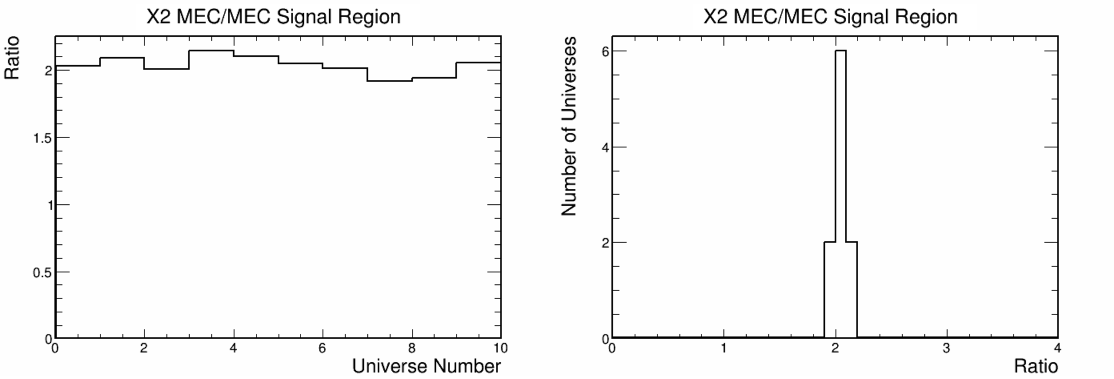


Figure 10.41: The ratio between the extracted signal with a doubled 2p2h-MEC sample and a standard sample when restricted to the signal region. The plot on the left gives the ratio in each universe. The plot on the right shows the distribution of ratios. It has a FWHM of 0.1.

For the second test the GENIE empirical 2p2h-MEC is replaced by the prediction of the SuSa model. Note that the SuSa model has a wider extent over the $|\vec{q}|$ and E_{avail} phase space.

Here the residual is plotted as a measure of the difference between the extracted signal and MC truth divided by MC truth as defined in Eq. 10.1.

Figure 10.42 (left) shows the residual for an extraction that uses the entire phase space. The residual is close to 0 however the extraction tends to underestimate modestly, by approximately 15%. Figure 10.42 (right) shows the residual when the extraction is restricted to the signal region. In this case the distribution is centered at 0.0 and has a FWHM of 0.3.

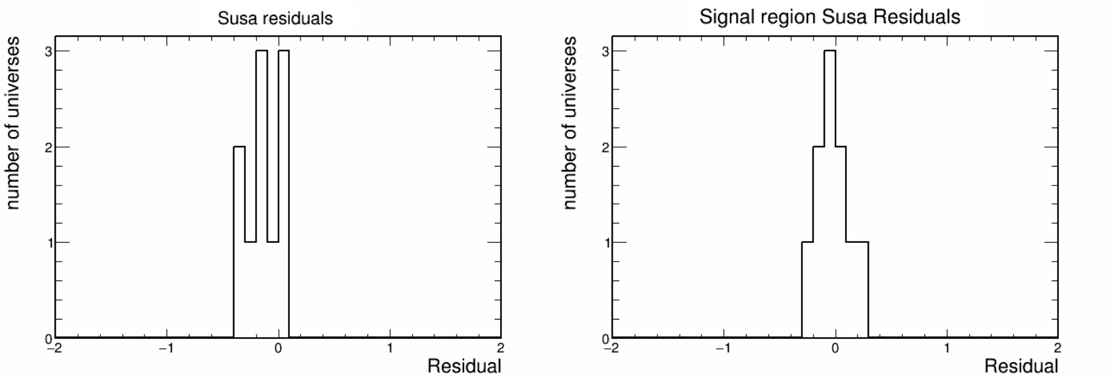


Figure 10.42: The extracted residuals when the SuSa-v2 2p2h-MEC model is used as data instead of the GENIE empirical 2p2h-MEC. The left-hand plot shows the number of universes as a function of residual over the whole phase space. The right-hand plot shows the number of universes as a function of residual only in the signal region. The residual in the signal region is now centered at zero with a FWHM of 0.3.

Chapter 11

Determination of the 2p2h-MEC contribution

11.1 Estimation based on nominal GENIE reaction-process templates

While the CC double-differential cross section can be approximately described in terms of scattering processes that involve single struck nucleons, namely CCQE, RES, and DIS, it is also known that scattering processes involving interactions with two or more nucleons are present. The goal of the second stage of this Thesis is to determine the amount of the 2p2h-MEC contribution to the inclusive cross section and to identify the regions of the $|\vec{q}|$ versus E_{avail} phase space that are populated by 2p2h-MEC events.

To estimate the data excess over the standard GENIE neutrino single-nucleon predictions, templates that embody the latter predictions are subtracted from the measured inclusive CC inclusive cross section. Figure 11.1 shows the cross-section templates for the three major processes to be subtracted, namely RES, DIS, CCQE, and Other. In NOvA data, the RES and DIS processes occur at roughly twice the rate of CCQE scattering and at roughly four times the rate of Other processes. The Other processes include a sum

Subtraction of the cross-section templates from the double-differential cross section shown in Fig. 9.9 and a table of the residual cross section is presented in Table 11.1, yields an initial estimate for the data excess relative to the prediction by the standard GENIE 2.12 Monte Carlo for CC neutrino-nucleon scattering within nuclei modeled as a relativistic Fermi gas. Figure 11.2 shows the result obtained by carrying out this subtraction. The excess thereby identified is maximal between 0.5 and 1.0 GeV/c in $|\vec{q}|$ and between zero and 0.5 GeV in available energy. The distribution peaks at 0.65 GeV/c in $|\vec{q}|$ and at 0.3 in E_{avail}, with a value of $4.0 \times 10^{-39}\,\mathrm{cm}^2$/GeV/c/GeV/nucleon. Smaller contributions to the excess are indicated at larger $|\vec{q}|$ and E_{avail} values. The cross section for the excess sample averaged over the analysis region is $(1.08 \pm 0.55) \times 10^{-39}\,\mathrm{cm}^2$/GeV/GeV/c/nucleon. This rate represents $16.0 \pm 9.2\%$ of the observed CC inclusive cross section; its cross section ratio relative to CCQE scattering is estimated to be $66.5 \pm 33.9\%$.

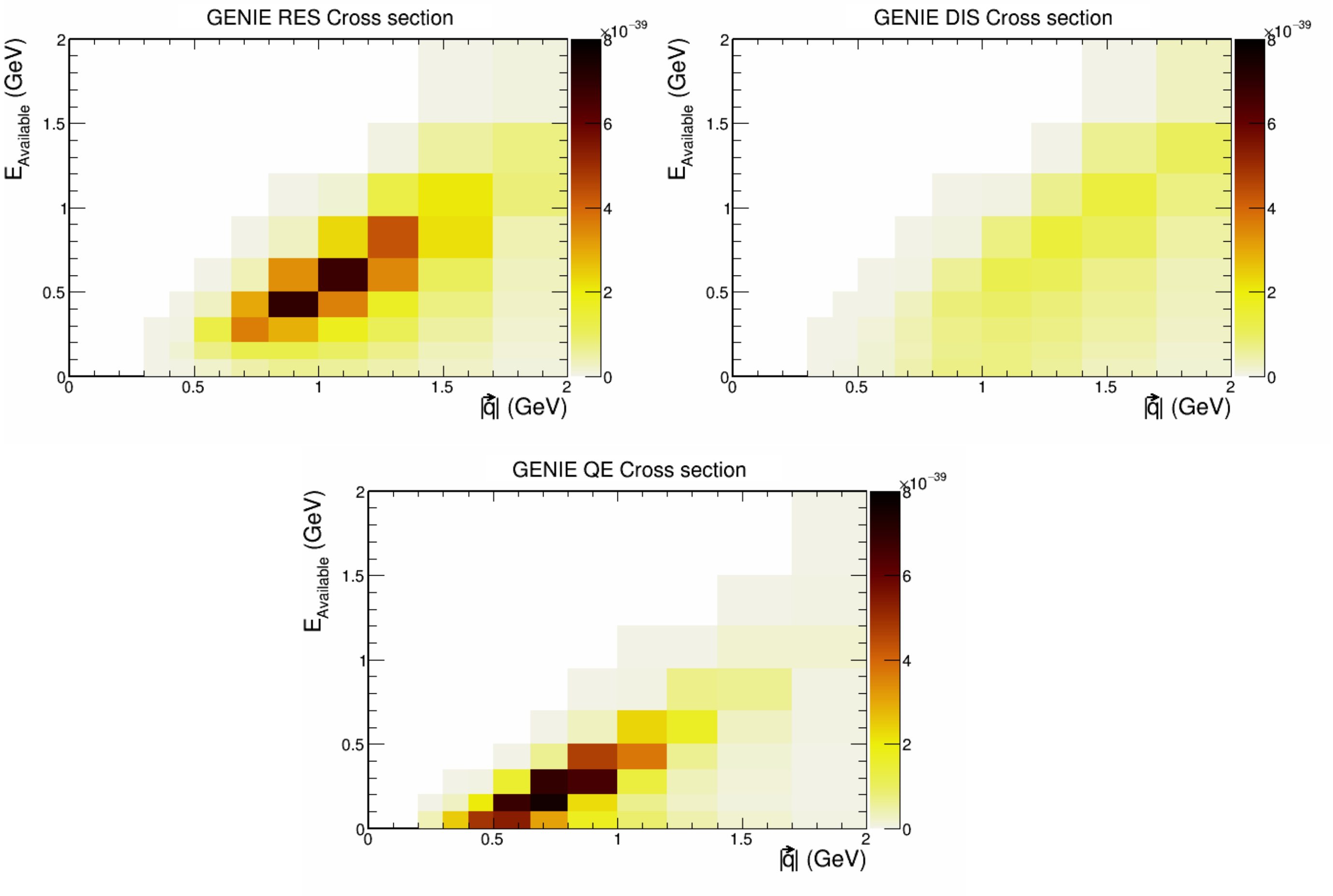


Figure 11.1: Cross-section templates for the three major background categories of CC processes that contribute to the measured double-differential cross section. The relative rates of occurrence follow the sequence of the plots, with RES and DIS processes being the most abundant.

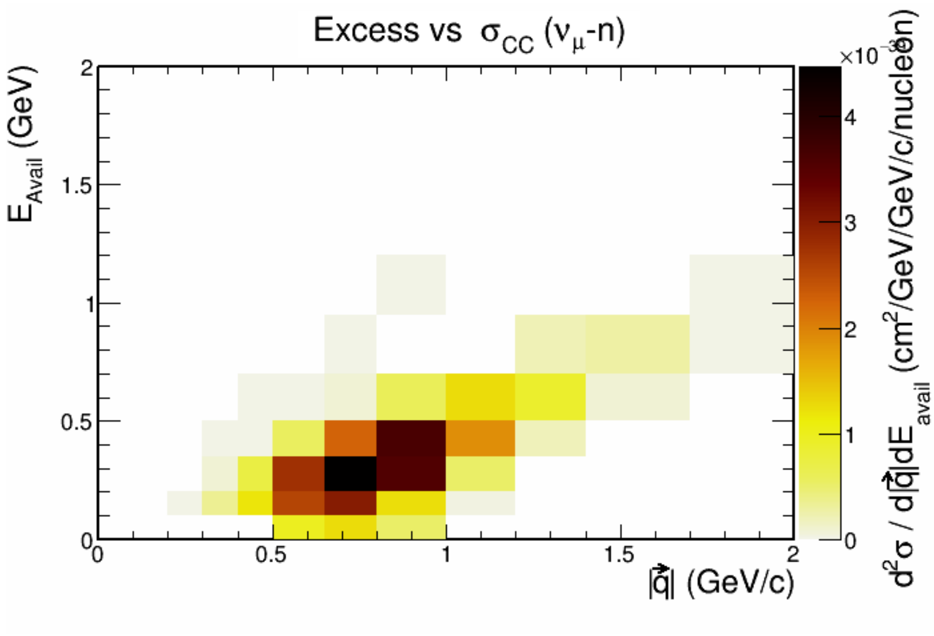


Figure 11.2: Excess events compared to GENIE-based estimation of conventional CC neutrino-nucleon scattering. The excess cross section is attributed to scattering on multi-nucleons as described by 2p2h-MEC models.

Table 11.1: Bin-by-bin display of the double differential cross section shown in Fig. 11.2. The tabled values represent $d^2\sigma/d|\vec{q}|dE_{avail}$ per bin in units of cm^2/GeV/GeV/c/nucleon. Note that, in contrast to Fig. 11.2, E_{avail} is denoted on the horizontal axis, while $|\vec{q}|$ is given by the vertical axis.

$\lvert\vec{q}\rvert \backslash E_{avail}$	0.00-0.10	0.10-0.20	0.2-0.35	0.35-0.50	0.50-0.70	0.70-0.95	0.95-1.20	1.20-1.50	1.50-2.00
0.00-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.10-0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.20-0.30	-8.63e-41	1.45e-41	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30-0.40	-3.61e-40	3.44e-40	8.07e-41	6.06e-44	0.00	0.00	0.00	0.00	0.00
0.40-0.50	-5.02e-42	1.19e-39	7.31e-40	1.60e-41	6.41e-44	0.00	0.00	0.00	0.00
0.50-0.65	9.67e-40	2.55e-39	2.77e-39	5.44e-40	4.55e-42	0.00	0.00	0.00	0.00
0.65-0.80	1.26e-39	3.02e-39	4.48e-39	2.25e-39	8.20e-41	3.16e-43	0.00	0.00	0.00
0.80-1.00	5.22e-40	1.23e-39	3.55e-39	3.62e-39	5.91e-40	-3.69e-41	4.58417e-44	0.00	0.00
1.00-1.20	-2.33e-41	3.02e-41	5.14e-40	1.88e-39	1.26e-39	-2.09e-40	-4.60887e-41	0.00	0.00
1.20-1.40	-1.17e-40	-1.41e-40	-4.93e-41	1.93e-40	8.90e-40	2.00e-40	-2.81125e-40	-1.09e-41	0.00
1.40-1.70	-7.89e-41	-1.02e-40	-8.55e-41	-2.86e-41	8.62e-41	2.64e-40	-1.39859e-40	-1.31e-40	-8.26e-43
1.70-2.00	-4.06e-41	-5.16e-41	-4.97e-41	-3.64e-41	-1.50e-41	4.15e-42	1.47053e-41	-1.12e-40	-2.48e-41

The fractional uncertainty on the excess cross section is displayed in Fig. 11.2 and shown in Table 11.2. The cross section uncertainty is 0.4 to 0.6 in the region of $0.2 \leq |\vec{q}| \leq 1.0$ GeV/C and $0.1 \leq E_{avail} \leq 0.5$ GeV. The uncertainty is greater than 1.0 in other regions of phase space.

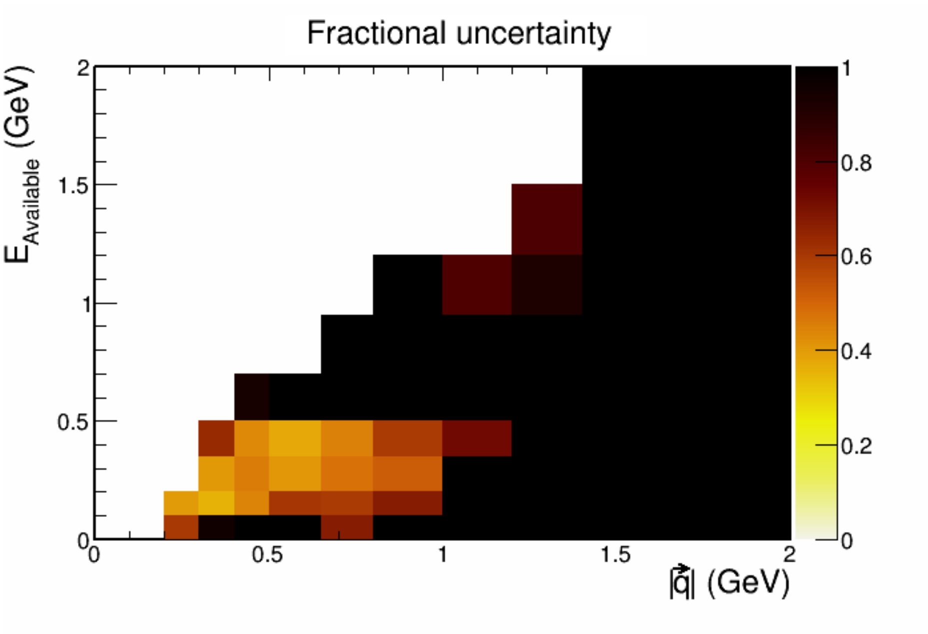


Figure 11.3: The fractional uncertainty on the excess cross section obtained by subtracting GENIE nominal cross sections from the inclusive measurement.

Table 11.2: Bin-by-bin display of fractional uncertainty on the double differential cross section shown in Fig. 11.3. Note that, in contrast to Fig. 11.3, E_{avail} is denoted on the horizontal axis, while $|\vec{q}|$ is given by the vertical axis.

$\lvert\vec{q}\rvert \backslash E_{avail}$	0.00-0.10	0.10-0.20	0.2-0.35	0.35-0.50	0.50-0.70	0.70-0.95	0.95-1.20	1.20-1.50	1.50-2.00
0.00-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.10-0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.20-0.30	0.596517	0.399469	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30-0.40	0.954698	0.355375	0.402612	0.634189	0.00	0.00	0.00	0.00	0.00
0.40-0.50	243.821	0.440911	0.45762	0.429834	0.93522	0.00	0.00	0.00	0.00
0.50-0.65	1.03833	0.602002	0.40587	0.373031	1.14104	0.00	0.00	0.00	0.00
0.65-0.80	0.671928	0.592484	0.478352	0.447658	1.45986	6.91928	0.00	0.00	0.00
0.80-1.00	1.20136	0.67338	0.519405	0.592284	1.16105	2.12311	87.9378	0.00	0.00
1.00-1.20	18.3127	13.4803	1.21795	0.72223	1.15665	2.21578	0.799034	0.00	0.00
1.20-1.40	2.06501	1.51448	6.19536	2.33924	1.0734	4.69425	0.920964	0.801399	0.00
1.40-1.70	1.52566	1.12184	1.72571	6.78307	3.20407	2.18527	3.75763	1.29563	3.60567
1.70-2.00	1.11505	1.03164	1.40988	2.18606	6.11053	40.1588	20.5543	2.70974	2.7649

The covariance matrix representing the covariance between bins in shown in Fig. 11.4.

The fractional statistical and systematic errors for the estimated excess are shown in Fig. 11.5. The large fractional errors are are inherent to subtracting a large unconstrained background in order to estimate a small signal.

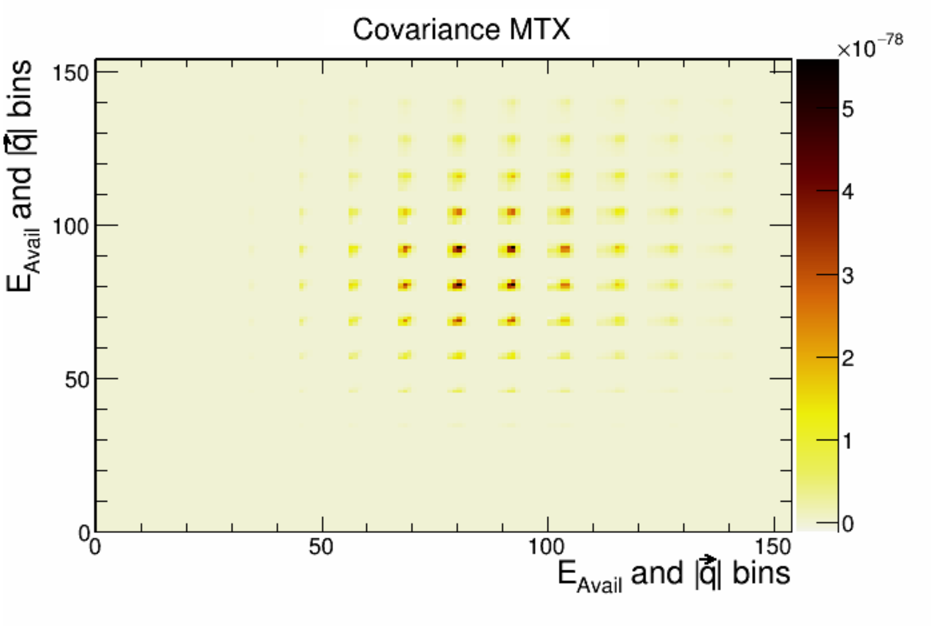


Figure 11.4: The covariance matrix for the excess cross section estimation in E_{avail} and $|\vec{q}|$ bins. The large blocks indicate bins in E_{avail} with bins of $|\vec{q}|$ within them.

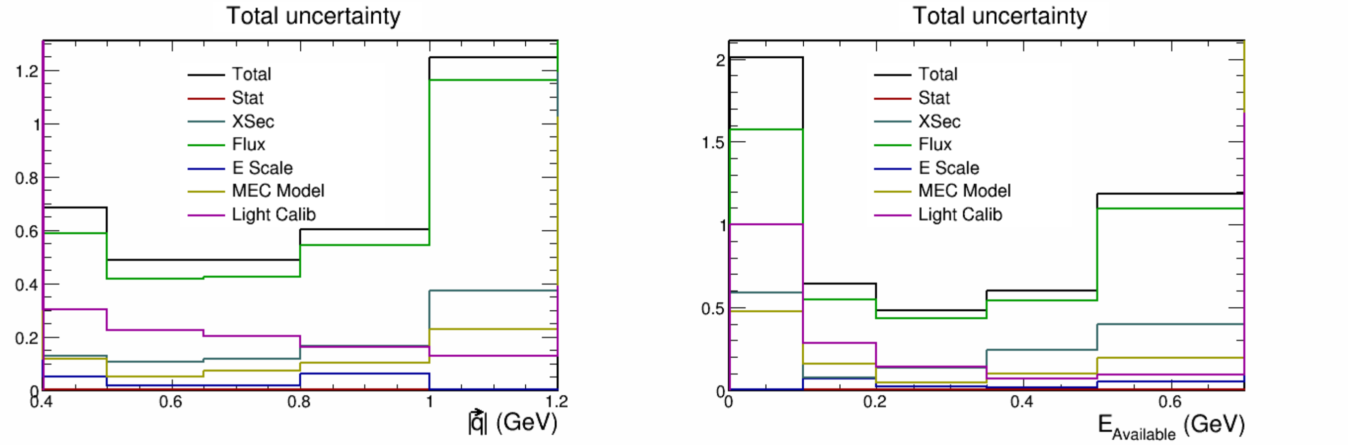


Figure 11.5: Fractional errors on the estimated excess attributed to 2p2h-MEC processes, including statistical and systematic sources.

Table 11.3: Fractional uncertainty on the integrated excess cross section arising from all sources of systematic uncertainty. The total systematic uncertainty (quadrature sum) is 51.2%. The flux is the leading source of uncertainty; it contributes a fractional error of 44.3%.

Source of uncertainty	Fraction error on the cross section
Flux	44.3%
Cross section model	10.5%
Light calibration	22.9%
2p2h-MEC model	4.8%
Energy scale	0.5%
Total	51.2%

Figure 11.6 shows the projections of the excess cross section in $|\vec{q}|$ and in E_{avail}, compared to 2p2h-MEC models. These include the GENIE 2.12 empirical model, the NOvA cross-section tune of GENIE 2.12, the Valencia model as implemented in GENIE 2.12, and the SuSa model. By design the NOvA-tune GENIE 2.12 model agrees well with the data. The GENIE empirical 2.12 model peaks in the same region as the data, namely at ~0.6 GeV/c of $|\vec{q}|$ and at ~0.2 GeV in E_{avail}, and it has roughly the same shape as the data. The SuSa and the Valencia models both peak in regions that are higher than the data, and they underestimate the data in regions where the data peaks.

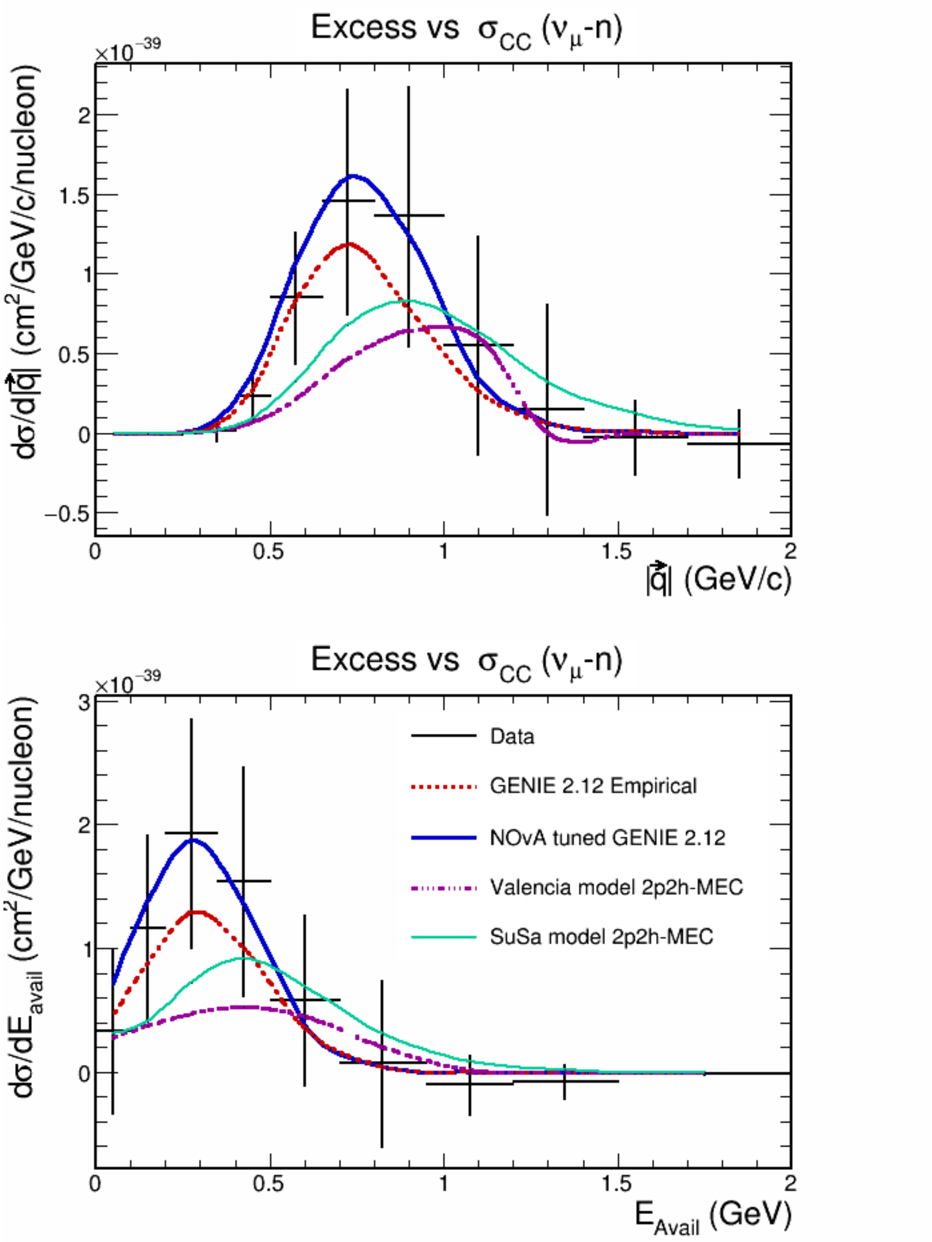


Figure 11.6: Distribution of the excess cross section relative to conventional CC neutrino-nucleon scattering, projected onto three-momentum transfer (top) and available hadronic energy (bottom). The excess is compared to predictions from data-based and phenomenological models for 2p2h-MEC.

It is of interest to examine the data excess versus E_{avail}, in slices of $|\vec{q}|$, as displayed in Fig. 11.7. The data excess grows with increasing $|\vec{q}|$ transfer up to 0.8 GeV/c, whereupon the excess subsequently falls off. The data excess is compared to GENIE-based predictions of the various 2p2h-MEC models. The excess over all of the models is greatest between 0.65 GeV/c and 1.2 GeV/c in $|\vec{q}|$. The excess cross section is compared to various 2p2h-MEC models in Table 11.4.

The data has the greatest agreement with the GENIE empirical model with a χ^2/DoF of 0.58, the SuSa model has a χ^2/DoF of 1.17, and the Valencia model has a χ^2/DoF of 1.68.

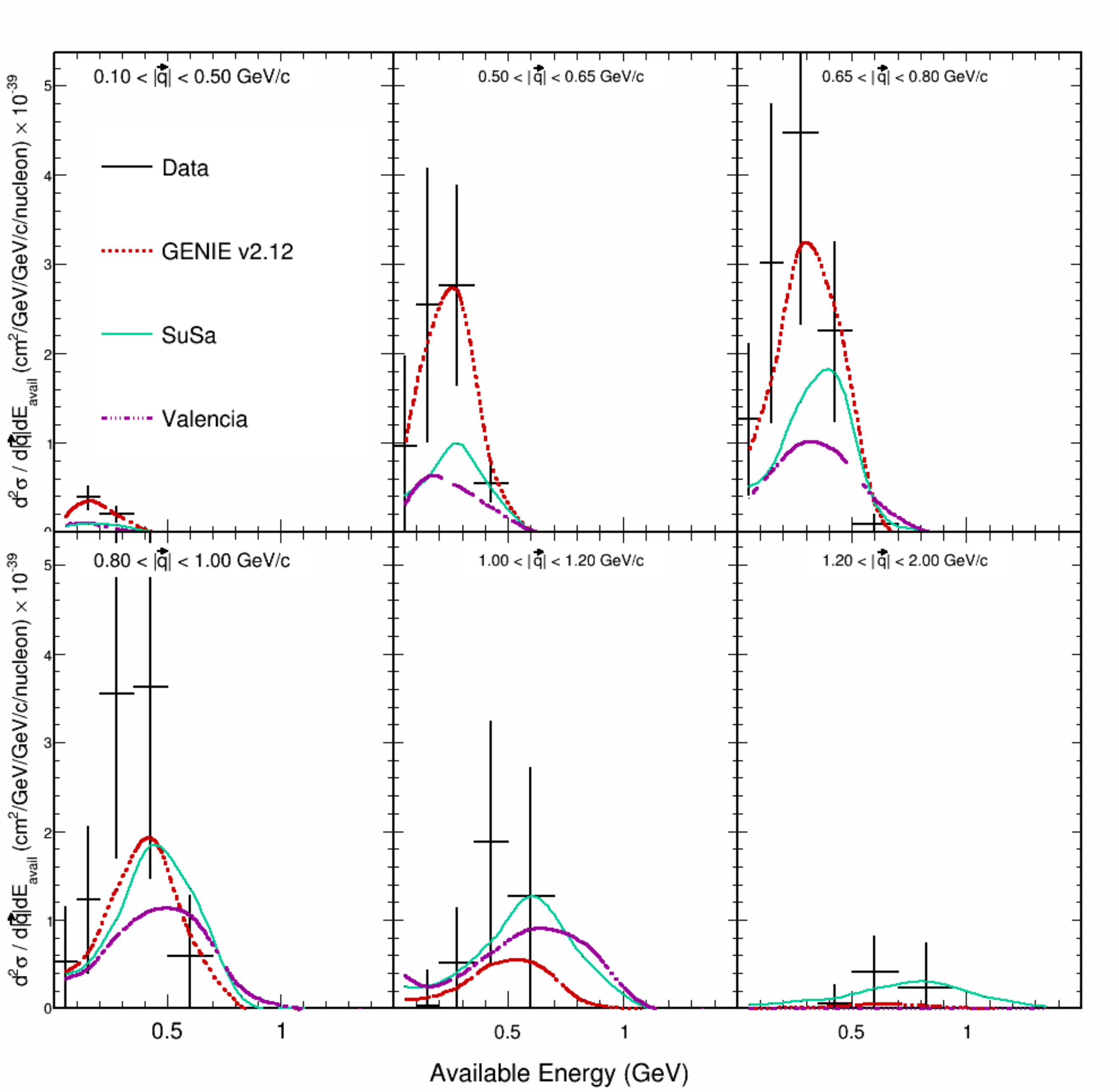


Figure 11.7: Distribution of the candidate 2p2h-MEC excess plotted in bins of E_{avail} for slices of $|\vec{q}|$. The general trend is predicted by the models, however the magnitude is underestimated, especially so for the Valencia and SuSa models.

Table 11.4: The χ^2/DoF between the extracted excess cross section and various models.

	χ^2/DoF
GENIE empirical 2p2h-MEC	0.58
Valencia 2p2h-MEC	1.68
SuSa-v2 2p2h-MEC	1.17

11.2 Determination of 2p2h-MEC using templates normalized to control sample

In order to improve the estimate of the excess, the sum of the GENIE based single-nucleon scattering templates is fitted to a data control sample as outlined in Sec. 10.1.2, and the excess cross section is then re-estimated. Here, the subtraction is carried out in terms of observed events; the excess event rate is then converted to cross section.

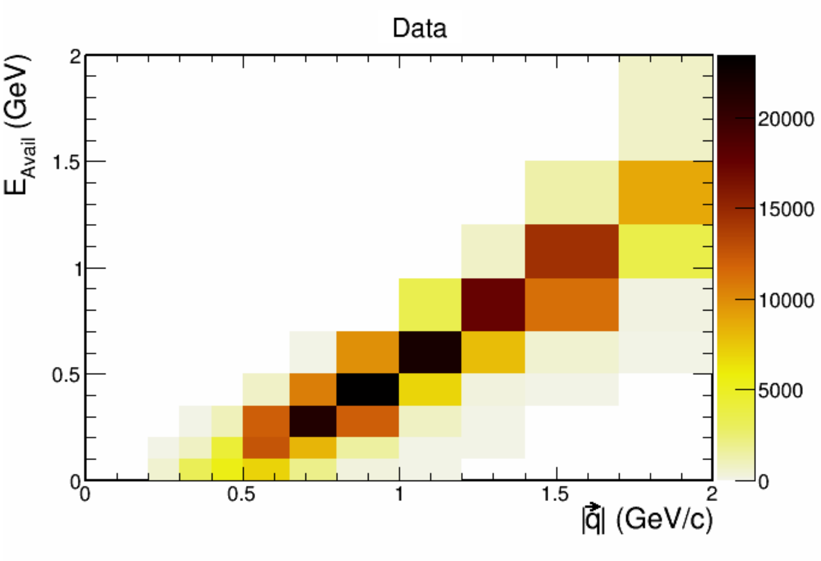


Figure 11.8: Distribution of selected data events displayed in terms of events per bin rather than in terms of cross section. (The latter is shown in Fig. 9.9.) The number of events arising from conventional scattering is estimated using constrained templates, which are then subtracted from this distribution to obtain the 2p2h estimate.

Figure 11.8 shows the number of data events that pass the CC event selections of this analysis. The events mostly populate the region around 0.65 GeV/c in $|\vec{q}|$ transfer and 0.22 GeV of E_{avail}; the event rate rapidly falls off as one moves away from those values. This distribution presumably includes both single-nucleon and multi-nucleon scattering events. By fitting the templates for neutrino processes to regions of the control sample, the template normalizations are adjusted, providing a more accurate estimate of the single-nucleon background contribution to Fig. 11.8. Subtraction of the normalization-adjusted templates from the distribution of Fig. 11.8 provides an estimate of the multi-nucleon scattering contribution.

Figure 11.9 shows the data to which the Monte Carlo templates are fitted in the control sample. The control sample is divided into three regions: Region I, defined by $|\vec{q}| \leq 1.2$ GeV/c; region II, $1.2 \leq |\vec{q}| \leq 1.4$ GeV/c, and region III, $|\vec{q}| \geq 1.4$ GeV/c. The control subsamples are fitted in regions I and III and the average of those fit normalizations is used in region II. This procedure has been previously described in Sec. 10.1.

Figure 11.10 shows the reference MC predictions for the templates in the control sample where RES and DIS are the largest contributors in the both region I and region III, while CCQE and Other background and Residual Signal contribute much less. Both RES and DIS contribute much more in region III than in region I. Because the Other and CCQE templates are small their normalizations are held fixed, however their templates still have a presence in the fit.

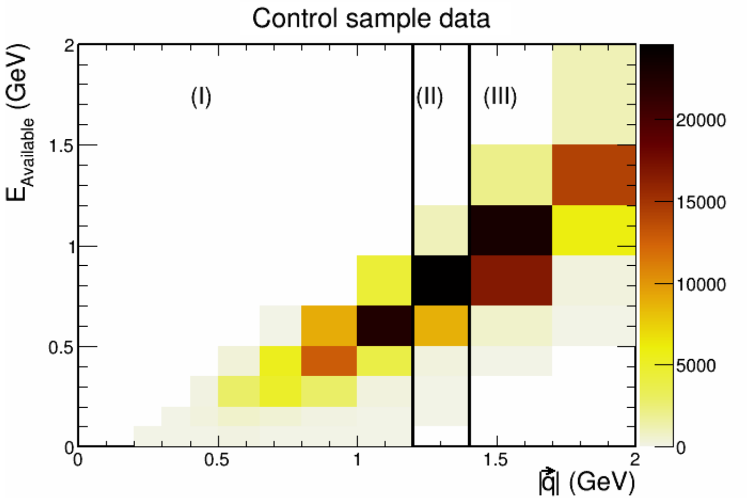


Figure 11.9: Data events of the control sample, plotted over the analyzed kinematic domain. The MC templates are fitted separately in regions I and III.

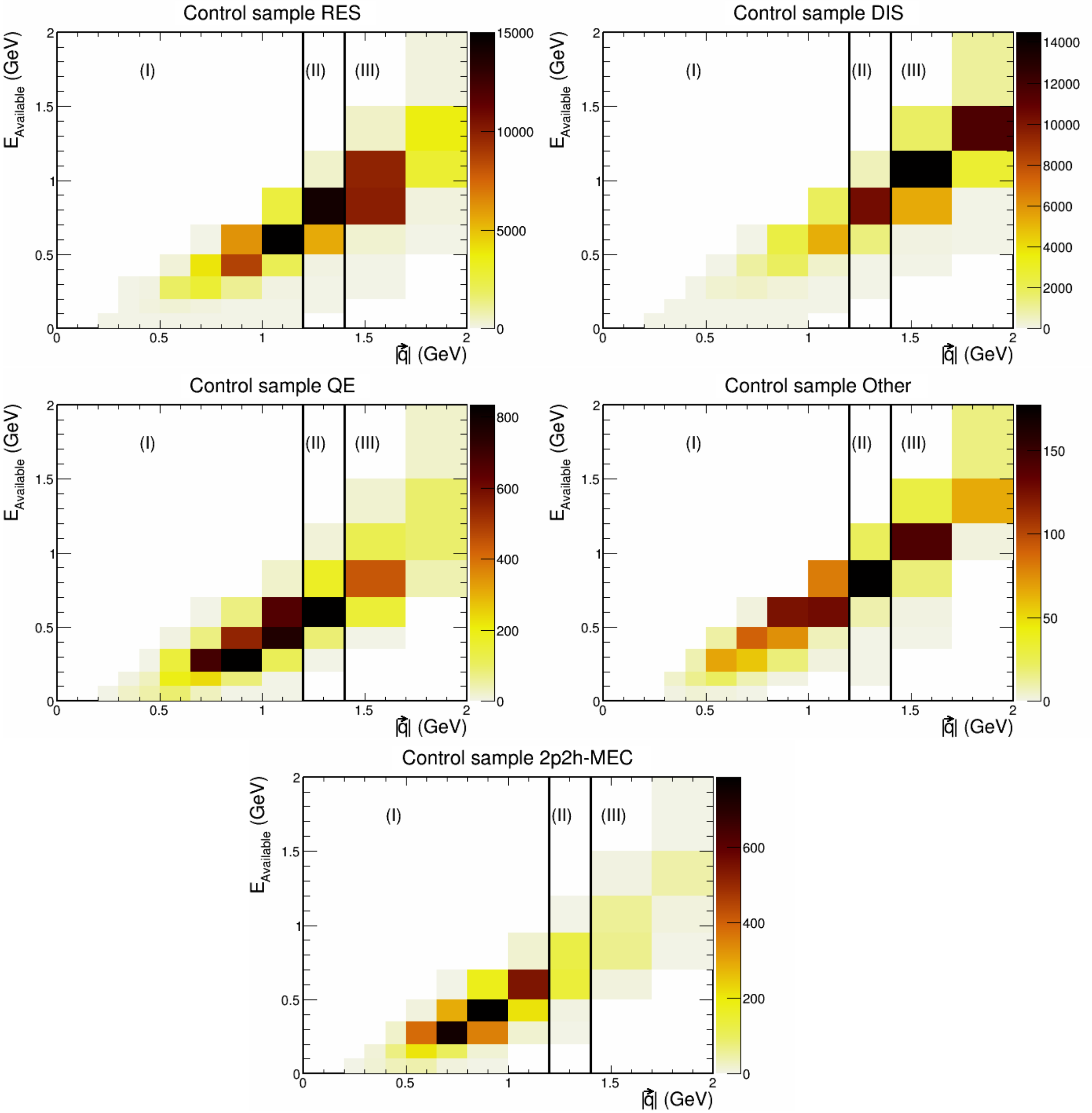


Figure 11.10: Input template distributions based on the MC that predict the data distribution in the control sample. Fitting is used to adjust the normalizations of the dominant templates in regions I and III.

In the fits performed in regions I and III, the RES, DIS, and residual MEC templates are allowed to vary, while the CCQE, and other templates are held fixed. The result of fitting to the control sample are summarized in Table 11.5. In region III, the normalization of the RES template is modestly increased and DIS is slightly decreased. The residual signal estimate is decreased to zero. In region I, the RES normalization is increased by 21% relative to the GENIE nominal setting. In region II, the normalizations are set to the average of the fitted normalizations obtained in regions I and III. Consequently RES is increased by 13% , DIS is decreased by 23%, and residual MEC is decreased by 84%. The χ^2/DoF between the Monte Carlo and the data in the control sample before the fit is 11.98 and the χ^2/DoF after the fit is reduced to 5.50.

Table 11.5: Summary of normalization adjustments relative to the GENIE nominal normalization of 1.0. The fits to regions III and I are observed to increase RES, decrease DIS, with the residual signal estimate remaining nearly constant.

	RES	DIS	QE	Other	Residual signal
Region I	1.21	0.60	[1.00]	[1.00]	0.63
Region II	1.13	0.77	[1.00]	[1.00]	0.37
Region III	1.05	0.93	[1.00]	[1.00]	0.00

Figure 11.11 shows the sum of events from the estimated single-nucleon scattering templates. The total amounts are subtracted bin-by-bin from the data shown in Fig. 11.8, yielding an estimate of the excess above the standard GENIE scattering prediction.

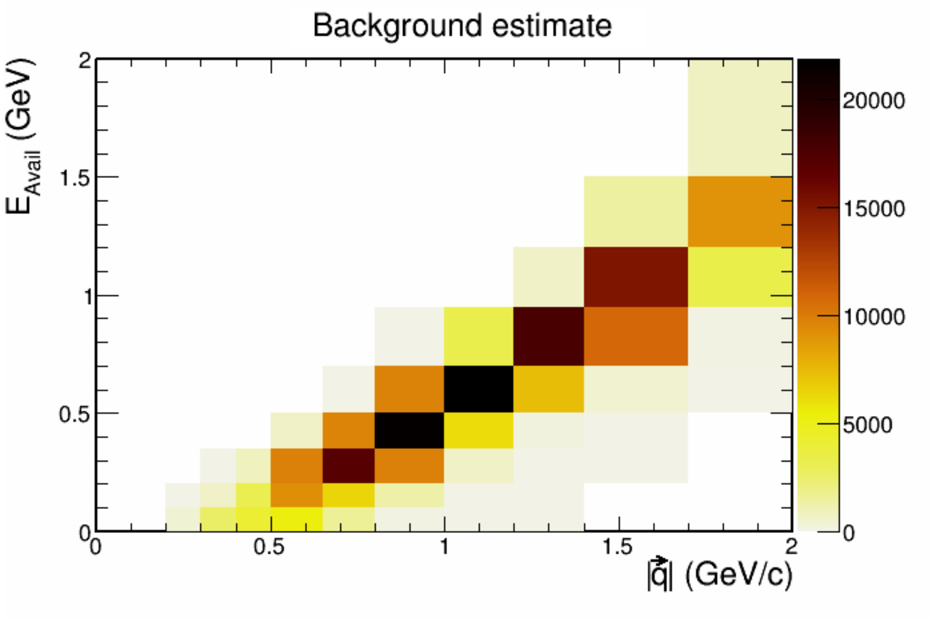


Figure 11.11: Event distribution predicted by the sum over the normalization-adjusted MC templates describing the individual neutrino-nucleon scattering processes.

Figure 11.12 shows the result of the subtraction of the GENIE nominal templates in Fig. 11.10

from the data in Fig. 11.8.The signal is observed to be maximum around 0.65 GeV/c and 0.3 GeV in $|\vec{q}|$ and E_{avail} respectively. The negative bins that appear at higher three-momentum transfer and available energy indicate an overestimation of RES and DIS in in region II.

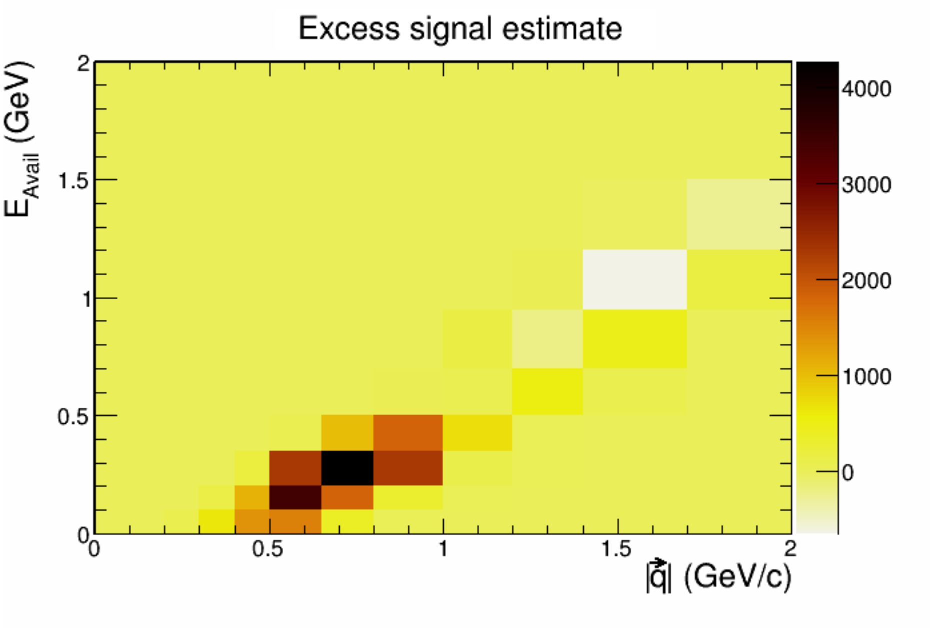


Figure 11.12: Distribution of excess events relative to the MC estimation of conventional neutrino single-nucleon scattering, as obtained by subtracting the distribution of Fig. 11.11 from the data distribution of Fig. 11.8. As expected on the basis of 2p2h-MEC models, the excess is predominantly found at $|\vec{q}| \leq 1.0$ GeV/c and $E_{avail} \leq 0.5$ GeV.

The signal events in Fig. 11.12 are then unfolded from reconstructed to “true” variables in the same way as described in Sec. 9.1. The unfolded signal events are shown in Fig. 9.5. The unfolded distribution peaks in the same region as the raw signal estimate, which is at 0.65 GeV/c in $|\vec{q}|$ and 0.3 GeV in E_{avail}. However the process of unfolding has the effect of spreading out the events into a broader region of the phase space. Note that the negative event regions are moved from higher available energy to lower E_{avail} but are still in the same regions of $|\vec{q}|$.

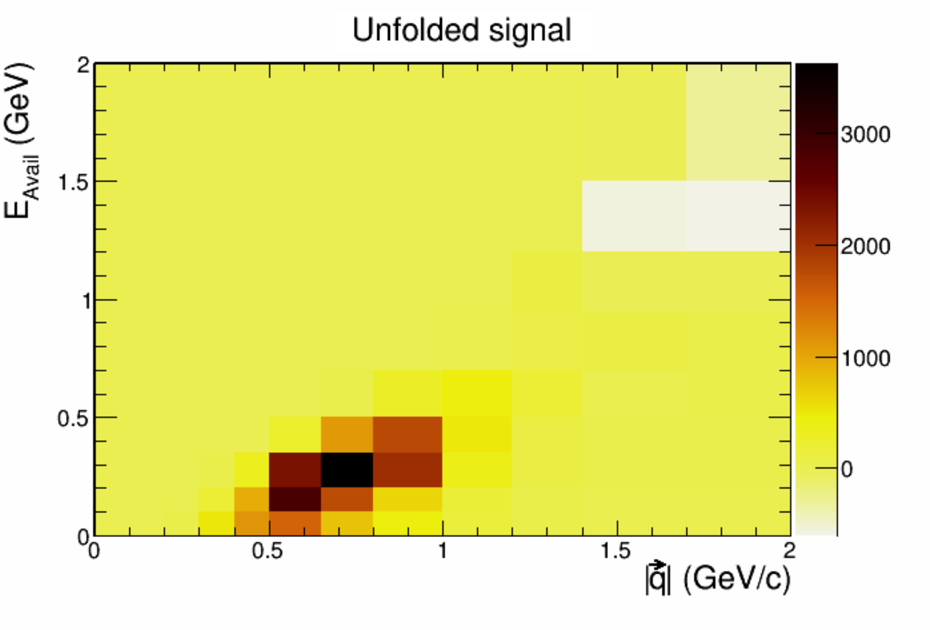


Figure 11.13: Distribution of 2p2h-MEC event excess after data unfolding removes detector effects from the variables. The unfolding process is observed to broaden the raw distribution of Fig. 11.12.

The signal is then corrected for detection inefficiencies by applying by the efficiency outlined in Fig. 9.6. The efficiency-corrected signal is shown in Fig. 11.14. The correction has the effect of increasing the amount of events in the lower region by a factor of three and increasing the number of events in the upper region by roughly a factor of five, including the negative counts in bins that are estimated to have negative counts.

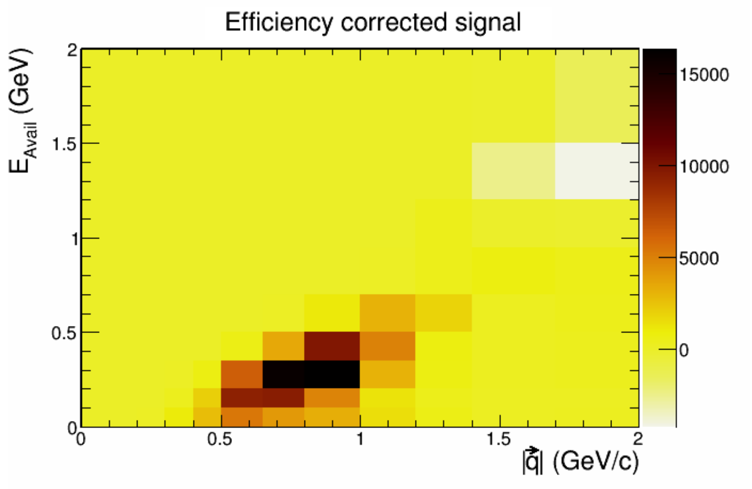


Figure 11.14: Distribution of the event excess corrected for detection inefficiency.

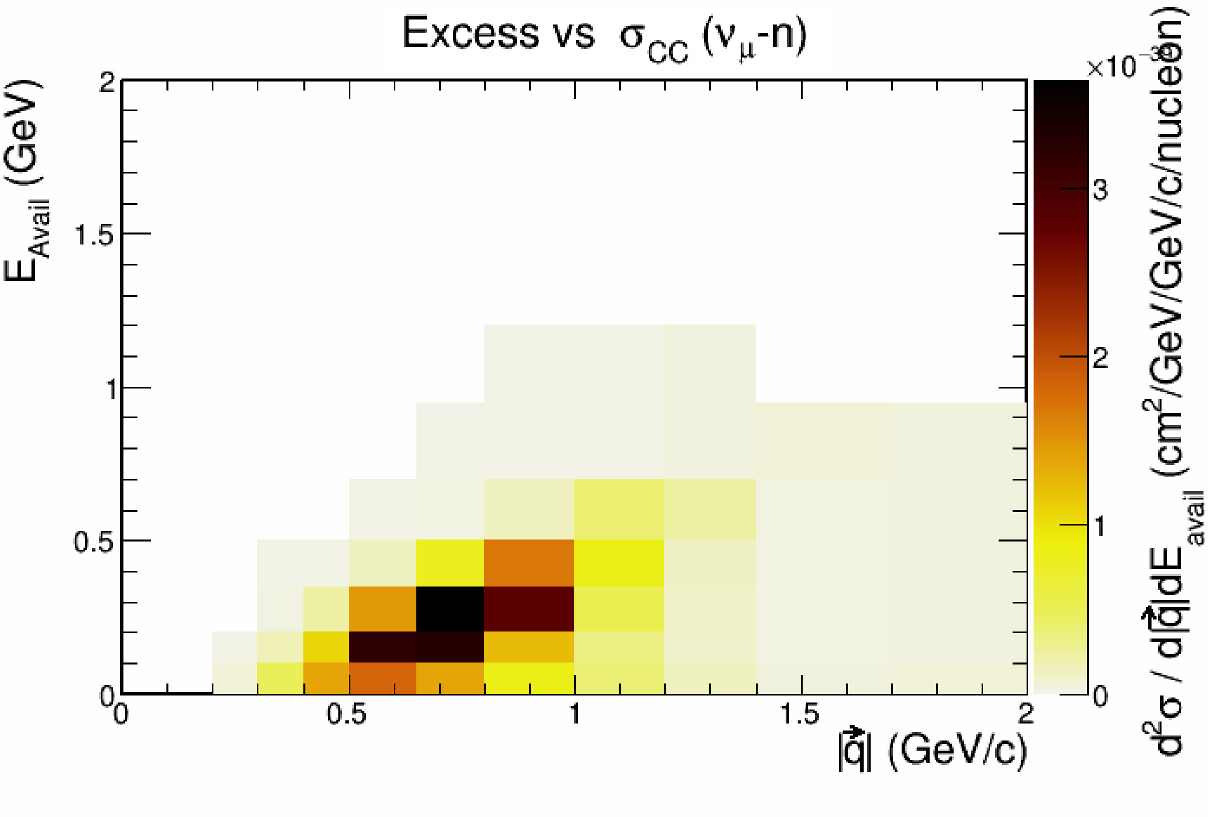


Figure 11.15: The double-differential cross section for the excess event sample relative to conventional neutrino-nucleon scattering, which is attributed to 2p2h-MEC processes.

The efficiency corrected signal is then divided by the integrated flux shown in Fig. 9.8 and by the number of target nucleons to obtain a total cross section. Figure 11.15 and Table 11.6 show the estimate of the excess cross section. The cross section for the excess sample averaged over the analysis region is $(0.72 \pm 0.39) \times 10^{-39}\,\text{cm}^2/\text{GeV}/\text{GeV}/\text{c}/\text{nucleon}$. This rate represents $12.0 \pm 6.5\%$ of the observed CC inclusive cross section; its cross section ratio relative to CCQE scattering is estimated to be $44.3 \pm 23.9\%$. The excess is largest at 0.65 GeV/c of $|\vec{q}|$ and at 0.3 GeV in E_{avail} at a value of $2.5 \times 10^{-39}\,\text{cm}^2/\text{GeV}/\text{GeV}/\text{c}/\text{nucleon}$. The bulk of the cross section occurs between 0.4 - 1.2 GeV/c in $|\vec{q}|$ and 0.0 - 0.5 GeV in E_{avail}. There is a region of negative cross section at higher $|\vec{q}|$ and E_{avail} which is caused by an overestimate of a RES and DIS in that region.

Table 11.6: Bin-by-bin display of the double differential cross section shown in Fig. 11.2. The tabled values represent $d^2\sigma/d|\vec{q}|dE_{avail}$ per bin in units of cm^2/GeV/GeV/c/nucleon. Note that, in contrast to Fig. 11.2, E_{avail} is denoted on the horizontal axis, while $|\vec{q}|$ is given by the vertical axis.

$\lvert\vec{q}\rvert\backslash E_{avail}$	0.00-0.10	0.10-0.20	0.2-0.35	0.35-0.50	0.50-0.70	0.70-0.95	0.95-1.20	1.20-1.50	1.50-2.00
0.00-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.10-0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.20-0.30	5.51e-41	3.91e-42	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30-0.40	4.83e-40	1.65e-40	2.05e-41	2.79e-44	0.00	0.00	0.00	0.00	0.00
0.40-0.50	1.38e-39	1.06e-39	2.23e-40	3.07e-42	0.00	0.00	0.00	0.00	0.00
0.50-0.65	1.81e-39	3.18e-39	1.46e-39	1.30e-40	4.88e-43	0.00	0.00	0.00	0.00
0.65-0.80	1.38e-39	3.26e-39	3.63e-39	7.90e-40	1.87e-41	1.27e-44	0.00	0.00	0.00
0.80-1.00	8.33e-40	1.23e-39	2.80e-39	1.69e-39	1.31e-40	2.33e-42	1.01e-44	0.00	0.00
1.00-1.20	3.74e-40	3.41e-40	5.32e-40	8.39e-40	3.97e-40	5.88e-42	2.33e-42	0.00	0.00
1.20-1.40	1.41e-40	9.89e-41	1.10e-40	1.19e-40	2.42e-40	4.01e-41	4.13e-41	-9.45e-43	0.00
1.40-1.70	3.19e-41	2.51e-41	2.74e-41	2.62e-41	1.82e-41	5.16e-41	-7.35e-42	-1.44e-40	-1.40e-42
1.70-2.00	4.41e-41	3.17e-41	3.59e-41	3.83e-41	3.67e-41	3.11e-41	-1.20e-41	-2.41e-40	-4.90e-41

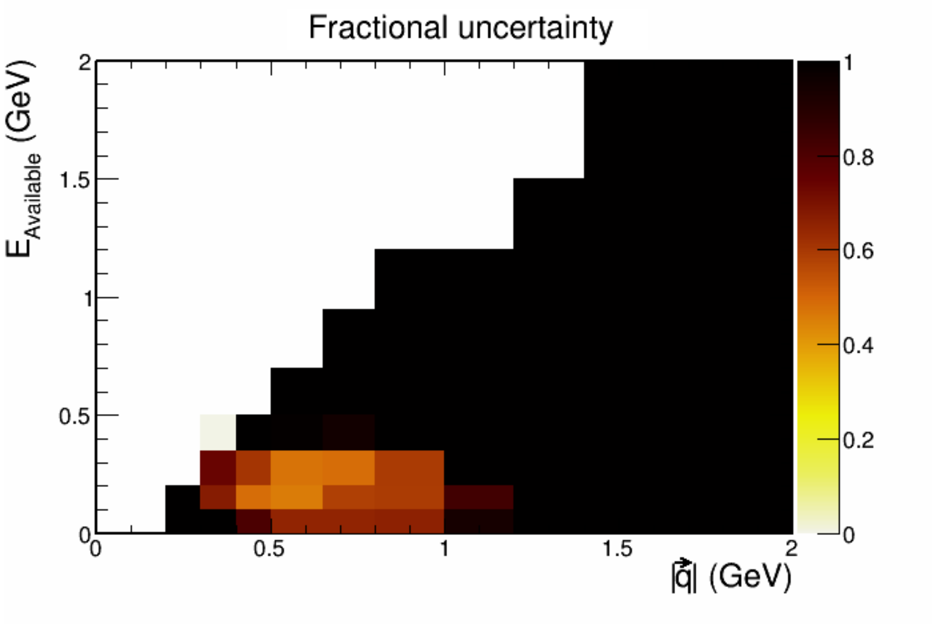


Figure 11.16: The fractional uncertainty on the excess cross section relative to conventional neutrino-nucleon scattering.

The fractional uncertainty on the estimate of the excess cross section is shown in Fig. 11.16 and Table 11.7. The cross section uncertainty is 0.5 to 0.6 in the region of $0.4 \leq |\vec{q}| \leq 1.0$ GeV/C and $0.0 \leq E_{avail} \leq 0.4$ GeV. In other regions the uncertainty exceeds 1.0.

Table 11.7: Bin-by-bin display of fractional uncertainty on the double differential cross section shown in Fig. 11.3. Note that, in contrast to Fig. 11.3, E_{avail} is denoted on the horizontal axis, while $|\vec{q}|$ is given by the vertical axis.

$\lvert\vec{q}\rvert\backslash E_{avail}$	0.00-0.10	0.10-0.20	0.2-0.35	0.35-0.50	0.50-0.70	0.70-0.95	0.95-1.20	1.20-1.50	1.50-2.00
0.00-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.10-0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.20-0.30	5.01	1.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.30-0.40	1.98	0.67	0.74	37.41	0.00	0.00	0.00	0.00	0.00
0.40-0.50	0.81	0.48	0.60	1.33	0.00	0.00	0.00	0.00	0.00
0.50-0.65	0.65	0.46	0.47	0.99	3.11	0.00	0.00	0.00	0.00
0.65-0.80	0.65	0.58	0.49	0.95	3.84	33.67	0.00	0.00	0.00
0.80-1.00	0.66	0.59	0.59	1.03	3.22	22.98	151.63	0.00	0.00
1.00-1.20	0.94	0.83	1.05	1.49	4.21	85.97	76.38	0.00	0.00
1.20-1.40	3.00	5.41	7.34	8.84	13.33	72.91	40.84	90.35	0.00
1.40-1.70	17.55	26.59	25.09	29.25	54.09	40.19	280.7	10.29	17.95
1.70-2.00	8.17	8.01	9.04	8.62	9.50	12.87	78.76	7.56	14.55

The covariance matrix for the excess cross section estimation is shown in Fig. 11.17.

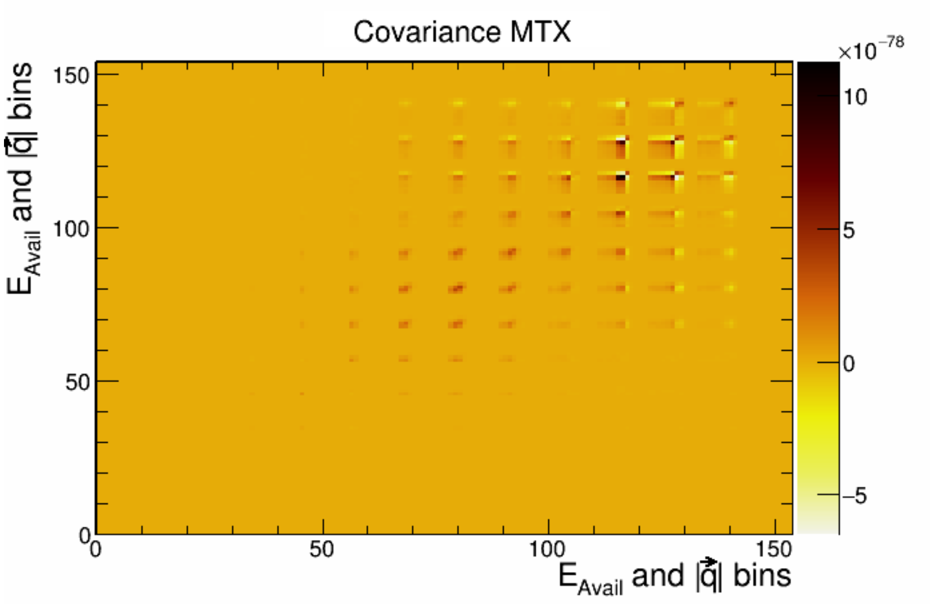


Figure 11.17: The covariance matrix for the excess cross section estimation relative to conventional neutrino-nucleon scattering in E_{avail} and $|\vec{q}|$ bins. The large blocks indicate bins in E_{avail} with bins of $|\vec{q}|$ within them.

Projections of the excess cross section in $|\vec{q}|$ and E_{avail} are shown in Fig. 11.18. The distributions are compared to the GENIE empirical model, the Valencia model, and the SuSa model. The 2p2h-MEC cross section is lower than that obtained by subtracting GENIE nominal templates from the inclusive cross section. Consequently the distributions in Fig. 11.18

are smaller than corresponding distributions show in Fig. 11.6 ($7.53 \times 10^{-39}\,\mathrm{cm}^2$/nucleon versus $5.70 \times 10^{-39}\,\mathrm{cm}^2$/nucleon). In Fig. 11.18 the data shows good agreement with the GENIE empirical model; as also observed in the estimation without fitting. Both the Valencia and SuSa phenomenological models under-predict the extracted 2p2h-MEC cross section. The position of the peak in the data occurs at points that are lower in both variables compared to the peak positions predicted by the phenomenological models. The errors on the data points in the peak region of $|\vec{q}|$ are modestly reduced by the fit. Large errors are incurred in the regions above the central peak for both variables reflecting the lack of signal counts in the regions.

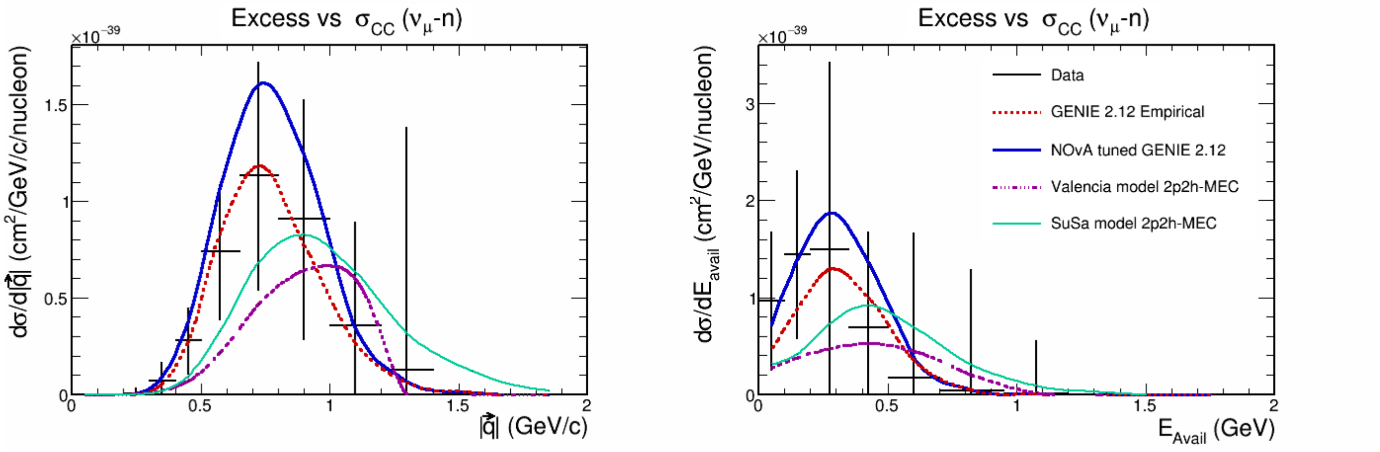


Figure 11.18: Projections of the excess cross section that lies above the fitted templates, in the variables $|\vec{q}|$ and E_{avail}. The projected distributions are compared to the empirical and phenomenological models for 2p2h-MEC.

The statistical and systematic errors are shown in Fig. 11.19 with $|\vec{q}|$ on the left and E_{avail} on the right. As is the case for the determination without fitting, the flux uncertainty is dominant with cross section and light calibration uncertainties also contributing significantly. The very large errors in high bins reflect the circumstance that very little 2p2h-MEC cross section is estimated for those bins.

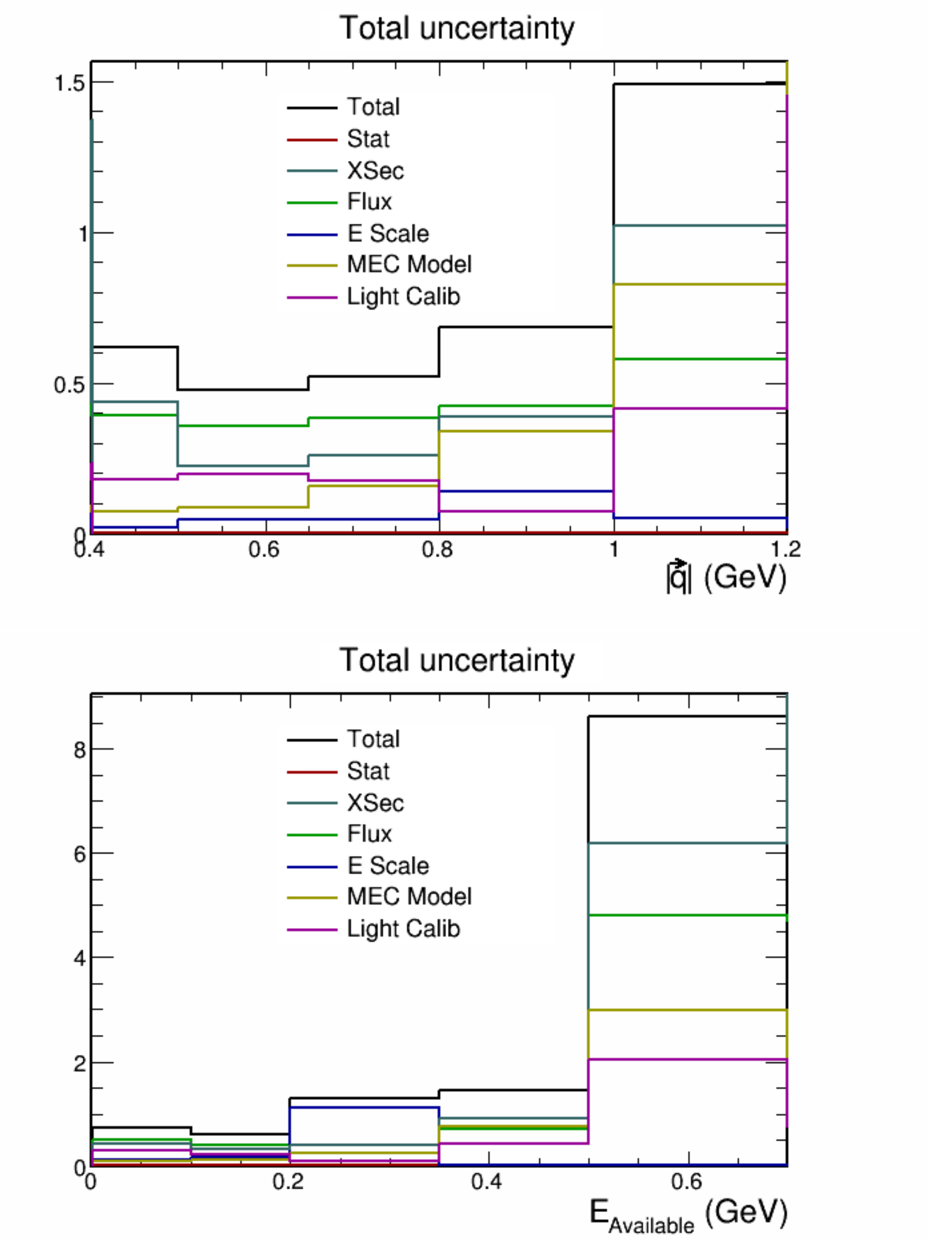


Figure 11.19: Systematic and statistical uncertainties for the fit determination of 2p2h-MEC in the analysis sample, for $|\vec{q}|$ (left) and E_{avail} (right).

Table 11.8: Fractional uncertainty on the integrated excess cross section from fitting arising from all sources of systematic uncertainty. The total systematic uncertainty (quadrature sum) is 54.0%. The flux is the leading source of uncertainty; it contributes a fractional error of 39.8%.

Source of uncertainty	Fraction error on the cross section
Flux	39.8%
Cross section model	24.2%
Light calibration	26.1%
2p2h-MEC model	7.2%
Energy scale	2.8%
Total	54.0%

Figure 11.20 displays the excess cross section in six slices of $|\vec{q}|$, for a range of E_{avail} in each slice. The excess is observed to be sizable between 0.5 - 1.0 in $|\vec{q}|$, and to fall off on either side of that range. It peaks in the region 0.65 - 0.8 GeV/c in $|\vec{q}|$, where it is large at 0.3 to 0.4 GeV in E_{avail}. The excess is compared to several models of 2p2h-MEC in Table 11.9. There is a χ^2/DoF of 0.65 between the GENIE empirical model and data, a χ^2/DoF of 0.87 between SuSa-v2 and data, and a χ^2/DoF of 1.65 between the Valencia model and data.

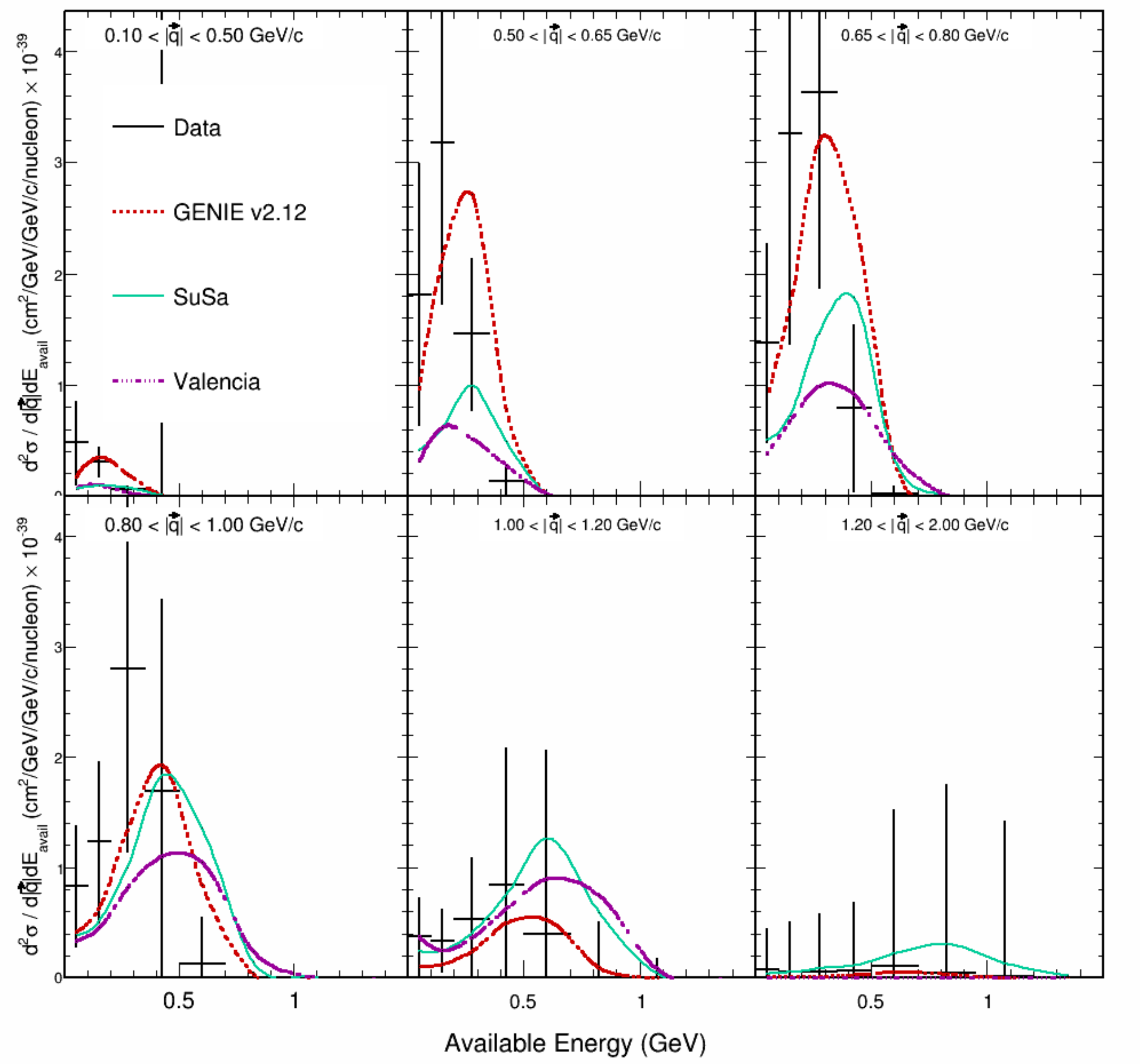


Figure 11.20: Event excess attributed to 2p2h-MEC processes plotted in bins of E_{avail}, for six intervals of $|\vec{q}|$. The excess here is estimated using reaction templates whose normalizations are adjusted by fitting to the control sample. The trends exhibited by the excess are similar to those observed when the excess is estimated using templates with GENIE-nominal normalizations. See Fig. 11.7.

Table 11.9: The χ^2/DoF between the extracted 2p2h signal and various models.

	χ^2/DoF
GENIE empirical 2p2h-MEC	0.65
Valencia 2p2h-MEC	1.65
SuSa-v2 2p2h-MEC	0.87

Chapter 12

Search for 2p2h structure using muon kinematics

In the Valencia model for 2p2h processes, CC events that arise from scattering on a two-nucleon system are predicted to distribute more broadly in muon kinetic energy, KE_μ and in muon production angle, $(\cos\theta_{\mu\nu})$, than do CCQE events. Figure 12.1 shows the Valencia predictions for incident neutrino energies below and above the NOvA flux spectral peak, at 1.0 GeV (upper plot) and at 3.0 GeV (lower plot). The black-elliptical contours indicate that CCQE is relatively localized and KE_μ and at very forward $\cos\theta_{\mu\nu}$, in contrast with the gray-scale contours for 2p2h-MEC, which are distributed more broadly in both variables. In view of these predictions, it is of interest to see whether such features can be detected in NOvA reconstructed events.

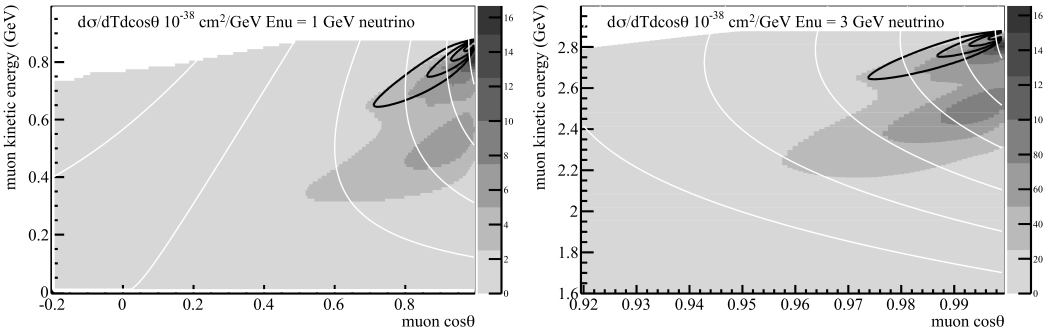


Figure 12.1: Valencia model comparison of distributions of 2p2h-MEC versus CCQE in muon kinetic energy and muon production angle relative to the beam direction.

To search for these kinematic effects, a subsample of events that are enriched in 2p2h has been selected as follows: Firstly, selected CC inclusive events are required to lie within a range of $|\vec{q}|$ between 0.4 and 1.2 GeV/c, with $E_{avail} \leq 0.7$ GeV. This subsample is 2p2h enriched but it also contains significant numbers of CCQE, RES, and DIS events. For purposes of this search, it is highly desirable to minimize the RES and DIS contributions. A selection as required for a "CC zero pion" analysis would be very useful here, however such a selection is not readily available within NOvA at the present time. As a fallback, selections are chosen that are the opposite of those that were used to define the control sample of this analysis. Specifically, events are required to have fewer than three non-muon prongs, with the longest non-muon prong being less than 100 cm in length. Additionally, events with reconstructed gamma showers are excluded. These selections mostly remove DIS events, however abundant RES events remain, together with 2p2h and CCQE events. Figure 12.2 shows the subsample selected in this way, as distributed over the plane of KE_μ and $\cos(\theta_{\mu\nu})$. The distribution is largely populated in the range $0.7 \leq \mathrm{KE}_\mu \leq 1.5$ GeV, with production cosines mostly 0.83 between 0.8 and 1.5 in muon KE, and at fairly forward production angle cosine values. The distribution unfortunately appears to be fairly devoid of distinguishing structural features.

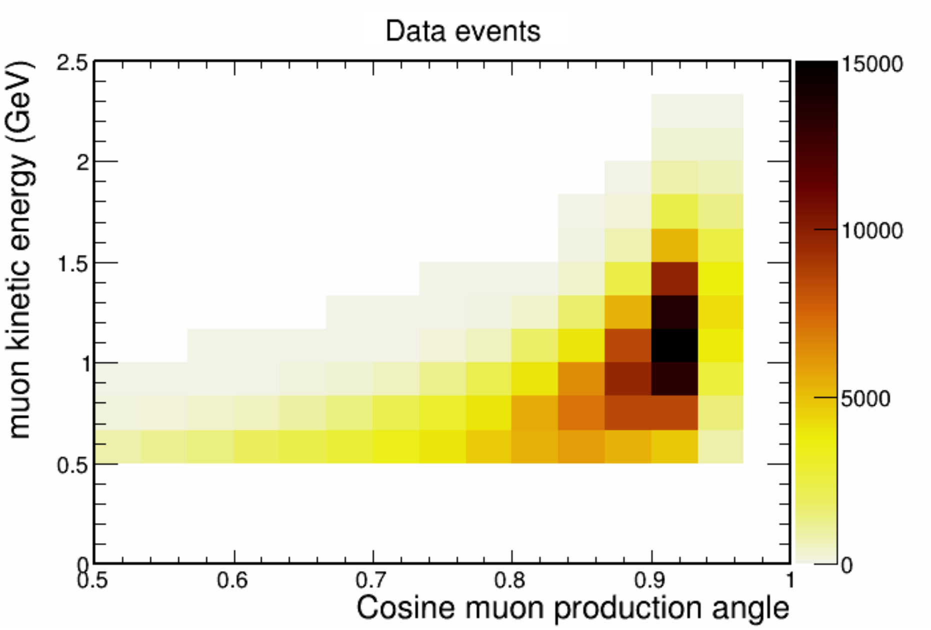


Figure 12.2: Distribution of data events of the search subsample, over the plane of muon kinetic energy and muon production-angle cosine.

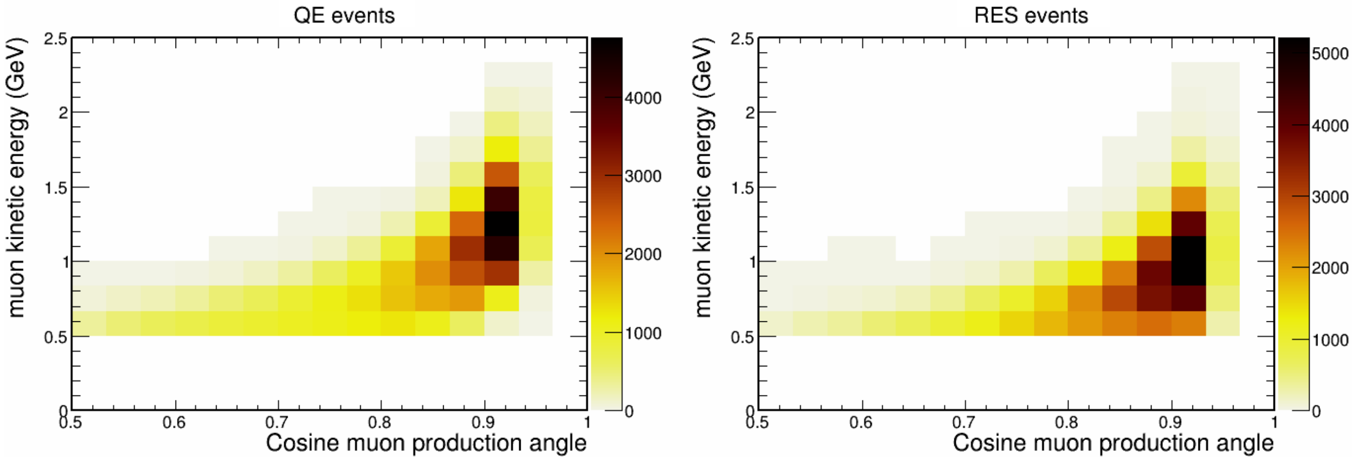


Figure 12.3: Distributions as predicted by the reference MC, for the CCQE and RES contributions to the search subsample. The distributions are fairly similar to each other, and to the data distribution displayed in Fig. 12.2.

As noted above, the search subsample includes CCQE and RES events. Figure 109 shows the distributions of the CCQE and RES processes that are predicted by the reference MC to occur in the search subsample. The contribution from DIS is predicted to be an order of magnitude smaller than the 2p2h, CCQE, and RES contributions and is not shown. One observes in Fig. 12.3 that neither CCQE or RES contributions exhibit features that distinguish them from the 2p2h-enriched distribution of Fig. 12.2.

The Valencia model predictions shown in Fig. 12.1 are presumably for muons that are perfectly reconstructed. It is of interest to see whether kinematic features are expected to survive muon track reconstruction. Figure 12.4 shows the distributions of reconstructed 2p2h-MEC events as predicted by the Valencia (left) and by the SuSa model (right). The distributions predicted by these two models are rather similar and do not exhibit distinctive features.

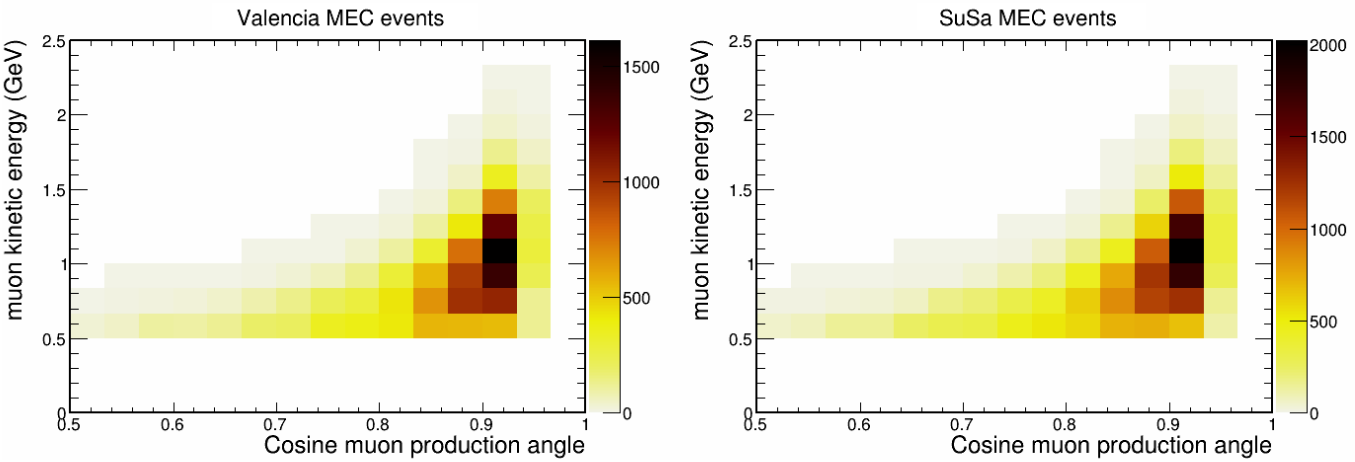


Figure 12.4: Distributions of 2p2h-MEC over the KE_μ versus $\cos\theta_{\mu\nu}$ plane, as predicted by the Valencia (left) and SuSa (right) models.

To conclude, the data distribution in muon kinematic variables displayed in Fig. 12.2 does not have features that suggest kinematic differences between the 2p2h component and the single-nucleon scattering processes of Fig. 12.3. There is no contradiction here with the phenomenological models, since predictions for reconstructed events derived from the models indicate that discernible differences are not be present in reconstructed events.

Chapter 13

Summary and Conclusion

This Thesis reports a new measurement of the CC-inclusive double-differential cross section $d^2\sigma/d|\vec{q}|dE_{avail}$ for neutrino interactions in the NOvA detector medium. This medium is predominately carbon but includes mixtures of heavier nuclei. The measurement is obtained at a median neutrino energy of 1.86 GeV. The cross-section distribution displayed over the plane of $|\vec{q}|$ and E_{avail} is shown in Fig. 9.9 and tabulated in Table 9.1. The double-differential cross section value that obtains when the bin widths are taken to be the dimensions of the entire analysis domain, e.g. $0.2 \leq |\vec{q}| \leq 2.0$ GeV/c and $0.0 \leq E_{avail} \leq 2.0$ GeV, is

$$(6.05 \pm 0.75) \times 10^{-39}\ \text{cm}^2/\text{GeV}/\text{GeV/c}/\text{nucleon}.$$

The measurement presented in this Thesis of the inclusive cross section compliments the MINERvA which is at a higher mean neutrino energy of 3.9 GeV. The MINERvA measurement is based on 74,949 selected signal events, while this measurement uses 394,101 selected signal events. The inclusive cross section here enables new comparisons with current models for the CC inclusive cross section. Such measurements are important for improving model descriptions that pave the way for higher precision in neutrino oscillation measurements. This measurement covers the region in E_ν from 0.8 to 3.2 GeV, the region most relevant to the NOvA neutrinos oscillation measurements. This region lies above the sub-GeV region of E_ν analyzed by T2K, while being moderately below the 2 to 6 GeV examined by MINERvA. It covers the lower half of the high-flux plateau in the $\nu_\mu$energy spectrum planned for the DUNE experiment [28].

The CC-inclusive cross section in the NOvA data includes contributions from processes that involve multi-nucleon target systems referred to as 2p2h-MEC. The contribution from these processes is not well known, a situation that has motivated the determination obtained by this Thesis. The analysis reports an estimation for this these processes by subtracting the known single nucleon cross section as estimated by GENIE v2.12. The cross section that remains after the subtraction provide a measure of the multi-nucleon scattering process. The cross section over the whole analysis region obtained by this method is

$$(1.08 \pm 0.55) \times 10^{-39}\ \mathrm{cm}^2/\mathrm{GeV}/\mathrm{GeV/c}/\mathrm{nucleon}.$$

The analysis also uses a data control sample to set the normalizations of simulation templates describing the well-known neutrino single-nucleon reaction types. The templates are used to estimate the cross section of these single-nucleon processes over the measured phase space. The event rate that lies above the level defined by the sum of the templates provides a measure of the 2p2h-MEC process. The semi-inclusive 2p2h-MEC cross section over the entire analysis region obtained by this analysis is

$$(0.72 \pm 0.39) \times 10^{-39}\ \mathrm{cm}^2/\mathrm{GeV}/\mathrm{GeV/c}/\mathrm{nucleon}.$$

This rate represents $12.0 \pm 6.5\%$ of the observed CC inclusive cross section; its cross section ratio relative to CCQE scattering is estimated to be $44.3 \pm 23.9\%$. The distribution of the 2p2h-MEC signal in $|\vec{q}|$ and E_{avail} shows good agreement with a data-driven model for these processes however the estimated 2p2h-MEC rate lies above the estimates of two current phenomenological models. The distributions in $|\vec{q}|$ and E_{avail} also show modest tension with respect to distribution shape with those models. These results provide valuable input for future descriptions of 2p2h-MEC as is required by the long baseline neutrino oscillation experiments.

Bibliography

[1] P. Adamson, K. Anderson, and M. Andrews, *The NuMI Neutrino Beam*, Nucl. Instrum. Meth. **A806**, 279 (2016).

[2] T. Katori and M. Martini, *Neutrino nucleus cross sections for oscillation experiments*, J. Phys. G **45**, 1, 013001 (2018).

[3] J. Nieves, I. R. Simo, and M. V. Vacas, *Two particle two hole excitations in charged current quasielastic antineutrino nucleus scattering*, Phys. Lett. B **721**, 1-3, 90 (2013).

[4] J. Nieves, I. R. Simo, F. Sanchez, and M. J. Vicente Vacas, *2p2h Excitations, MEC, Nucleon Correlations and Other Sources of QE-like Events*, JPS Conf. Proc. **12**, 010002 (2016).

[5] 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).

[6] M. A. Acero, P. Adamson, G. Agam, L. Aliaga, T. Alion *et al.*, *Adjusting neutrino interaction models and evaluating uncertainties using NOvA near detector data*, The European Physical Journal C **80**, 12 (2020).

[7] O. Benhar, D. Day, and I. Sick, *Inclusive quasielastic electron-nucleus scattering*, Rev. Mod. Phys. **80**, 189 (2008).

[8] A. Gil, J. Nieves, and E. Oset, *Many-body approach to the inclusive (e, e') reaction from the quasielastic to the Δ excitation region*, Nuclear Physics A **627**, 4, 543 (1997).

[9] M. Martini, M. Ericson, G. Chanfray, and J. Marteau, *Unified approach for nucleon knock-out and coherent and incoherent pion production in neutrino interactions with nuclei*, Phys. Rev. C **80**, 065501 (2009).

[10] 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).

[11] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), *Measurement of the neutrino neutral-current elastic differential cross section on mineral oil at $E_\nu \sim 1$ GeV*, Phys. Rev. D **82**, 092005 (2010).

[12] M. Martini, M. Ericson, and G. Chanfray, *Neutrino quasielastic interaction and nuclear dynamics*, Phys. Rev. C **84**, 055502 (2011).

[13] J. Nieves, I. Ruiz Simo, and M. J. Vicente Vacas, *The nucleon axial mass and the MiniBooNE Quasielastic Neutrino-Nucleus Scattering problem*, Phys. Lett. B **707**, 72 (2012).

[14] A. Bodek, H. S. Budd, and M. E. Christy, *Neutrino quasielastic scattering on nuclear targets*, The European Physical Journal C **71**, 9, 1726 (2011).

[15] O. Lalakulich, K. Gallmeister, and U. Mosel, *Many-body interactions of neutrinos with nuclei: Observables*, Phys. Rev. C **86**, 014614 (2012).

[16] J. E. Amaro, M. B. Barbaro, J. A. Caballero, and T. W. Donnelly, *Meson-Exchange Currents and Quasielastic Antineutrino Cross Sections in the Superscaling Approximation*, Phys. Rev. Lett. **108**, 152501 (2012).

[17] G. D. Megias, T. W. Donnelly, O. Moreno, C. F. Williamson, J. A. Caballero *et al.*, *Meson-exchange currents and quasielastic predictions for charged-current neutrino-*^{12}C *scattering in the superscaling approach*, Phys. Rev. D **91**, 073004 (2015).

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

[19] C. Andreopoulos *et al.* (GENIE Collaboration), *The GENIE neutrino Monte Carlo generator*, Nucl. Instrum. Methods Phys. Res., Sect. A **614**, 1, 87 (2010).

[20] K. Gallmeister, U. Mosel, and J. Weil, *Neutrino-induced reactions on nuclei*, Phys. Rev. C **94**, 035502 (2016).

[21] R. Gran, J. Nieves, F. Sanchez, and M. J. V. Vacas, *Neutrino-nucleus quasi-elastic and 2p2h interactions up to 10 GeV*, Phys. Rev. D **88**, 113007 (2013).

[22] J. Nieves, I. R. Simo, and M. J. V. Vacas, *Inclusive charged-current neutrino-nucleus reactions*, Phys. Rev. C **83**, 045501 (2011).

[23] G. D. Megias, J. E. Amaro, M. B. Barbaro, J. A. Caballero, and T. W. Donnelly, *Inclusive electron scattering within the SuSAv2 meson-exchange current approach*, Phys. Rev. D **94**, 013012 (2016).

[24] G. D. Megias, J. E. Amaro, M. B. Barbaro, J. A. Caballero, T. W. Donnelly *et al.*, *Charged-current neutrino-nucleus reactions within the superscaling meson-exchange current approach*, Phys. Rev. D **94**, 093004 (2016).

[25] P. Rodrigues, J. Demgen, E. Miltenberger, L. Aliaga, O. Altinok *et al.*, *Identification of Nuclear Effects in Neutrino-Carbon Interactions at Low Three-Momentum Transfer*, Phys. Rev. Lett. **116**, 7 (2016).

[26] Gran, R. *et al.* (MINERvA Collaboration), *Anti-Neutrino Charged-Current Reactions on Scintillator with Low Momentum Transfer*, Phys. Rev. Lett. **120**, 221805 (2018).

[27] K. Abe, J. Amey, C. Andreopoulos, L. Anthony, M. Antonova *et al.*, *Characterization of nuclear effects in muon-neutrino scattering on hydrocarbon with a measurement of final-state

kinematics and correlations in charged-current pionless interactions at T2K, Phys. Rev. D **98**, 032003 (2018).

[28] B. Abi *et al.*, *Long-baseline neutrino oscillation physics potential of the DUNE experiment*, Eur. Phys. J. C **80**, 10, 978 (2020).

[29] Fermilab, Fermilab Accelerator Complex (2021).

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

[31] J. A. Vasel, A. Sheshukov, and A. Habig, in *Meeting of the APS Division of Particles and Fields* (2017).

[32] D. Phan, *A Search for Neutron-Antineutron Oscillationin the NOvA Experiment*, University of Texas Thesis (2020).

[33] K. Bays and J. Wolcott, *2018 cross-section tuning tech note*, NOVA-doc-27755 (2018).

[34] L. Aliaga Soplin, B. Behera, C. Johnson, S. kai Lin, M. Muether *et al.*, *Technical Note on the ND numu CC Inclusive Double-Differential Cross Section Measurement*, NOVA-doc-32688-v1 (2018).

[35] J. Paley, *Kalman Filter Track and Vertex Reconstruction with RecPack*, NOVA-doc-1062-v1 (2007).

[36] D. Torbunov, *NuMu Energy Estimator Technote for Prod4 MC*, NOVA-doc-27626-v2 (2018).

[37] B. Zamorano, *Cross-checks on Second Analysis energy estimator*, NOVA-doc-5645-v1 (2016).

[38] T. Olson, *2p2h In-Out test*, NOVA-doc-26193-v1 (2017).

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

[40] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*, Nucl. Instrum. Methods Phys. Res., Sect. A **362**, 2, 487 (1995).

[41] H. B. Prosper and L. Lyons, editors, *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, CERN, Geneva (2011).

[42] D. Pershey, NOvA Cross Section Package svn/trunk/CAFAna/XSec/CrossSectionAnalysis.h.

[43] Altinok, O. *et al.* (MINERvA Collaboration), *Measurement of ν_μ charged-current single π^0 production on hydrocarbon in the few-GeV region using MINERvA*, Phys. Rev. D **96**, 072003 (2017).

[44] T. Olson, *Determination of systematic uncertainties for available energy and three momentum transfer double differential cross-section measurement using muon neutrino interactions in the Near Detector*, NOVA Document 38676-v2 (2019).

[45] T. Katori, *Meson Exchange Current (MEC) Models in Neutrino Interaction Generators*, AIP Conf. Proc. **1663**, 030001 (2015).

[46] B. Behera, *CVN based numu CC inclusive*, NOVA-doc-24430-v1 (2017).

[47] N. Nayak, *Flux PCA on new beam focussing shifts*, NOVA-doc-34924-v1 (2018).

[48] M. Strait, *Tech note: muon energy scale systematic*, NOVA-doc020816-v21 (2019).

[49] Birks' Law, https://en.wikipedia.org/wiki/Birks'_law.

[50] A. Antoshkin, N. Anfimov, A. Sotnikov, and O. Samoylov, *Test bench for measurements of scintillators(NOvA, LAB-based etc.) properties*, NOvA-Doc-31369 (2019).

[51] J. Nieves, I. R. Simo, and M. J. V. Vacas, *Inclusive charged-current neutrino-nucleus reactions*, Phys. Rev. C **83**, 045501 (2011).

[52] R. Gran, J. Nieves, F. Sanchez, and M. J. V. Vacas, *Neutrino-nucleus quasi-elastic and 2p2h interactions up to 10 GeV*, Phys. Rev. D **88**, 113007 (2013).

[53] C. L. Smith, *Neutrino reactions at accelerator energies*, Physics Reports **3**, 5, 261 (1972).

[54] J. Nieves, R. Gran, I. Ruiz Simo, F. Sanchez, and M. J. Vicente Vacas, *Neutrino-nucleus CCQE-like scattering*, Nucl. Part. Phys. Proc. **273-275**, 1830 (2016).