Title

A Measurement of νe Appearance and νμ Disappearance Using 10 Years of Data from the NOvA Experiment

Author

Jargowsky, Benjamin

Publication Date

2024

UNIVERSITY OF CALIFORNIA,
IRVINE

A Measurement of ν_e Appearance and ν_μ Disappearance Using 10 Years of Data from the NOvA Experiment

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Physics

by

Benjamin Jargowsky

Dissertation Committee:
Professor Jianming Bian, Chair
Professor Henry Sobel
Professor Juan Pedro Ochoa-Ricoux

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

There are more people than I can list who I owe gratitude for the help and support I've received over the course of my graduate studies. I'll list only a few here.

First, I would like to thank my advisor, Jianming Bian, for his constant direction, and his dedication to his role as an academic mentor. I also want to thank other members of our group, starting with Wenjie Wu, for his guidance and all of the fun times in the lab. Next, I want to thank fellow grad students Alejandro Yankelevich and Yiwen Xiao, for their help and collaboration over the years. Thanks also to Nitish Nayak, who was a veteran graduate student at the beginning of my graduate studies, for answering all of my dumb questions and imparting his knowledge of NOvA to me.

Thank you to the other dissertation committee members, Hank Sobel and Pedro Ochoa-Ricoux, for taking the time to review this dissertation, as well as to other neutrino people at UCI, such as Michael Smy, for the informative weekly meetings, which serve as a useful reminder to a new grad student that there's more going on in neutrino physics than what's in your small corner of the world. Thanks also to Jeff Griskevich for managing our lab and ordering everything we've ever needed.

Thanks to everyone in NOvA who I've worked with and learned from over the years. Special thanks to everyone who helped me along the way with this analysis, such as the analysis conveners: Erika Catano-Mur, Liudmila Kolupaeva, Ryan Nichol, Artur Sztuc, and Zoya, as well as other very knowledgeable people in the group who have provided feedback and guidance for my work, Alex Himmel and Jeremy Wolcott.

Thanks to my parents, Michael and Lori, for the lifetime of support, and their backing of my educational career, nurturing my passion in physics. Thanks also for taking the time to proofread this dissertation. Thank you to my partner, Angelina, for her support, and her patience with my preoccupation with work during my graduate studies.

Lastly, I need to thank those who have helped keep the NOvA experiment running, specifically those in Fermilab Accelerator Division and NOvA Operations. Without these people's diligent labor, there would be no NOvA data for this analysis.

This work was supported by United States Department of Energy Grant DE-SC0009920.

VITA

Benjamin Jargowsky

EDUCATION

Doctor of Philosophy in Physics **2024**
University of California, Irvine *Irvine, California*

Bachelor of Science in Physics **2019**
University of California, Santa Barbara *Santa Barbara, California*

RESEARCH EXPERIENCE

Graduate Research Assistant **2020–2024**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2019–2021**
University of California, Irvine *Irvine, California*

ABSTRACT OF THE DISSERTATION

A Measurement of ν_e Appearance and ν_μ Disappearance Using 10 Years of Data from the NOvA Experiment

By

Benjamin Jargowsky

Doctor of Philosophy in Physics

University of California, Irvine, 2024

Professor Jianming Bian, Chair

NOvA is a long baseline neutrino experiment with an 810 km baseline, using the NuMI beam at Fermilab, and a functionally identical near and far detector operating at an angle 14 mrad off axis from the beam. NOvA jointly measures muon neutrino (and antineutrino) disappearance and electron neutrino (and antineutrino) appearance to make a measurement of $sin^2\theta_{23}$, δ_{CP}, and Δm^2_{32}, including its sign, the mass ordering.

This dissertation reports a new measurement from NOvA, using 10 years of data, with a total exposure of 26.6 $\times 10^{20}$ POT of neutrino beam and 12.5 $\times 10^{20}$ POT of antineutrino beam. This represents 95.6% more neutrino beam exposure since the last NOvA analysis. Aspects of the analysis are discussed in detail, including neutrino energy estimation, analysis systematic uncertainties, including the implementation of systematics new to the NOvA analysis, and the Bayesian fit infrastructure using Markov Chain Monte Carlo (MCMC).

The analysis yields the following credible intervals for the oscillation parameters assuming the normal ordering: $\delta_{CP} = 0.930^{+0.210}_{-0.290}\,\pi \,\cup\, 0.150^{+0.150}_{-0.110}\,\pi$, $\Delta m^2_{32} = 2.424^{+0.035}_{-0.040} \times 10^{-3}\mathrm{eV}^2$, and $sin^2\theta_{23} = 0.55^{+0.02}_{-0.06}$, with a 76% preference for normal ordering using a prior for $sin^2(2\theta_{13})$ using Daya Bay's measurement. If the Daya Bay $sin^2(2\theta_{13})$ vs Δm^2_{32} constraint is used as a prior instead, an 87% preference for normal ordering emerges.

Chapter 1

Introduction to Neutrinos

1.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics is the most complete theory of fundamental interactions. The Standard Model is a gauge quantum field theory, which includes six quarks and six leptons, as well as four gauge bosons which give rise to the strong, electromagnetic, and weak forces, and a Higgs boson, arising from the mechanism which gives some of these particles mass [121].

Quarks and leptons come in three "generations". Each generation has two members, with subsequent generations being identical in terms of interactions, only different in mass. For the quarks, the first generation has an up quark, with $\frac{2}{3}$ elementary electric charge, and a down quark, carrying $-\frac{1}{3}$ elementary charge. This electromagnetic interaction is brokered by the photon field. In addition to quarks being charged under this field, they also interact via the strong force, mediated by the gluon field, and the weak force, mediated by the W and Z fields. It's the strong force which leads quarks to form bound states: baryons, consisting of 3 quarks, such as protons and neutrons, and mesons, consisting of a quark and antiquark, such as pions.

For the leptons, the first generation consists of the electron, with electric charge -1, and an electron neutrino, with electric charge 0. The electron neutrino is coupled only to the weak force, and through the weak force, couples directly to the electron. The electron's higher generation versions are called the muon and the tau, and they couple to their own flavors of neutrino.

The Higgs mechanism gives the Higgs field a vacuum expectation value, which through spontaneous symmetry breaking, gives the quarks, the charged leptons, the W and Z bosons, and the Higgs boson mass. At the time the Standard Model was created, neutrinos were assumed massless, so the model did not include any terms to give neutrinos mass. Mass terms can be added, however, which in general allow the possibility of oscillations from one flavor state to another, which was the first experimental indication of neutrino mass [63].

A full treatment of the Standard model is beyond the scope of this dissertation; see one of the many references on this topic for more information [108, 98, 121].

1.2 Neutrino Masses

Though neutrinos are massless in the original Standard Model, mass terms can be added, which in general can mix the flavor states with mass eigenstates as a general linear combination. Any non-trivial mixing between flavor eigenstates and mass eigenstates will lead to neutrino flavor eigenstates changing from one to another over time. This can be seen looking at the rest frame of the neutrino, where the phases of the respective mass eigenstates will rotate in time as $e^{-im_i c^2 t/\hbar}$, becoming out of phase with one another. The various possible constructions of neutrino mass terms is discussed in the next section.

1.2.1 Neutrino Masses

In the following section, we'll follow the notation of Srednicki [108].

Dirac Masses

Any mass term which respects the gauge invariance of the standard model must include another field, in addition to the left handed Weyl fields of the Standard Model, ν_I, where I is the index for flavors, of which there are 3 (e, μ, τ). We will call this new field $\bar{\nu}_J$, which is a separate field, and not a charge conjugate of the field ν_I. We will for now assume there are also 3 of these new fields. We can add the trivial kinetic term for $\bar{\nu}_I$ as there is for ν_I. From there, we can then add a mass term by added a Yukawa coupling to the Higgs:

$$\mathcal{L}' = -\phi^{\dagger i}\ell_{iI}y_{IJ}\bar{\nu}_J + \text{h.c.} \tag{1.1}$$

where repeated indices are summed over, and ℓ_{iI} is the SU(2) doublet $\begin{pmatrix} \nu_I \\ e_I \end{pmatrix}$, e_I is a left handed Weyl field in the SU(2) doublet (becoming the left chiral part of the electron Dirac field), and $\phi^{\dagger i}$ is the Higgs doublet before spontaneous symmetry breaking.

After spontaneous symmetry breaking, the term becomes

$$\mathcal{L}' = -\frac{1}{\sqrt{2}}(v + H)\nu_I y_{IJ}\bar{\nu}_J + \text{h.c.} \tag{1.2}$$

where $\frac{1}{\sqrt{2}}v y_{IJ}$ is identified as the mass matrix of a Dirac field for the neutrino, $\begin{pmatrix} \nu_I \\ \bar{\nu}_I^{\dagger} \end{pmatrix}$ which in general is expected to mix the flavor and mass eigenstates. This term will break individual lepton number conservation, but conserve total lepton number.

Majorana Masses

Besides the Dirac mass term discussed in the last section, we can add a Majorana mass term for the new field of the form

$$\mathcal{L}'' = -\frac{1}{2}M_{IJ}\bar{\nu}_I\bar{\nu}_J + \text{h.c.} \tag{1.3}$$

. With this Majorana term in addition to the Dirac term, the behavior of the resultant eigenstates depends on the scale of the M_{IJ} matrix.

If the masses of the Majorana term are much larger than the mass eigenvalues of $\frac{1}{\sqrt{2}}vy_{IJ}$, than we can integrate out the heavier $\bar{\nu}_I$ fields, and we get an effective theory with Majorana neutrinos with mass matrix $\frac{1}{2}v^2(y^TM^{-1}y)_{IJ}$. Note the masses of the neutrinos than scale as the inverse of the M matrix, which could explain the tiny masses of neutrinos compared to other masses in the Standard Model. This is called the see-saw mechanism. This leaves us with very heavy "sterile" neutrino states, made of mostly the $\bar{\nu}_I$ states, which therefore barely interact with the charged leptons, and will be extremely difficult to detect directly.

If all or some of the Majorana term eigenstates of $\bar{\nu}_I$ are not much heavier than the Dirac masses, say n of them, then diagonalizing will give $3 + n$ light neutrinos, being various proportions of ν_I and $\bar{\nu}_I$. n of these will be mostly $\bar{\nu}_I$, therefore mostly sterile, and difficult to detect directly, yet of comparable mass scale to the active neutrinos. These are usually called light sterile neutrinos.

Note there's nothing guaranteeing that $\bar{\nu}_J$ has exactly 3 states, so in the situations outlined above, we can also have any arbitrary number of states $\bar{\nu}_J$, and therefore any number of sterile neutrino states.

1.3 Neutrino Oscillations

1.3.1 Neutrino Mixing

Assuming we have 3 mass eigenstates, we can generally parameterize the mixing matrix between mass and interaction eigenstates, which we'll call U. In the case of Majorana neutrinos, we have

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{-i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\eta_1} & 0 & 0 \\ 0 & e^{i\eta_2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.4}$$

where $c_{ij} \equiv cos\theta_{ij}$ and $s_{ij} \equiv sin\theta_{ij}$. The angles θ_{ij}, without loss of generality, can be defined to be in the first quadrant, with δ_{CP} and η_i being in the range [0, 2π].

Dirac neutrinos can be parameterized the same way, but the phases η_i, called the Majorana phases, can then be absorbed into the neutrino states. The matrix then becomes

$$U = \begin{pmatrix} c_{12}s_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta_{CP}} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta_{CP}} & c_{13}c_{23} \end{pmatrix} \tag{1.5}$$

. This is called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix.

Neutrino Oscillations

For a neutrino created in one flavor eigenstate, each mass eigenstate, j, has a time dependent phase $e^{-iE_jt/\hbar}$, where $E_j = \sqrt{p_j^2 + m_j^2}$, where the masses m_j are different between different mass eigenstates. Over a distance, L, we can work out

$$P_{\alpha\beta} = \delta_{\alpha\beta} - 4\sum_{i<j}^{n} Re[U_{\alpha i}U^*_{\beta i}U^*_{\alpha j}U_{\beta j}]sin^2 X_{ij} + 2\sum_{i<j}^{n} Im[U_{\alpha i}U^*_{\beta i}U^*_{\alpha j}U_{\beta j}]sin2X_{ij} \tag{1.6}$$

where n is the number of neutrinos, and

$$X_{ij} = \frac{\Delta m^2_{ij} L}{4E} \tag{1.7}$$

, with $\Delta m^2_{ij} \equiv m^2_i - m^2_j$ [121].

It should be noted that the oscillations aren't dependent on the absolute masses of the neutrinos themselves, but only the differences between the squares of the mass states. The oscillation scale, in terms of energy and oscillation length, becomes a function of $\frac{L}{E}$, with a scale factor $|\Delta m^2_{ij}|/4$. The mass mixing parameters are amplitudes for these oscillation terms.

1.3.2 Neutrino Interactions

To detect neutrinos, they must first interact in our detectors. Neutrinos are neutral, and do not have color charge, so they only interact through the weak interaction, through the W and Z boson fields.

There are various ways which neutrinos can interact with the particles in matter. First, neutrinos can interact with the electrons. Tree level diagrams of these interactions are shown in figure 1.1.

First, on the left, we see the interaction through the W field, which we call a charged current (CC) interaction, as there is an exchange of charge. Note this can only happen for ν_e since there are no heavier charged leptons in matter. The right side interacts through the Z field,

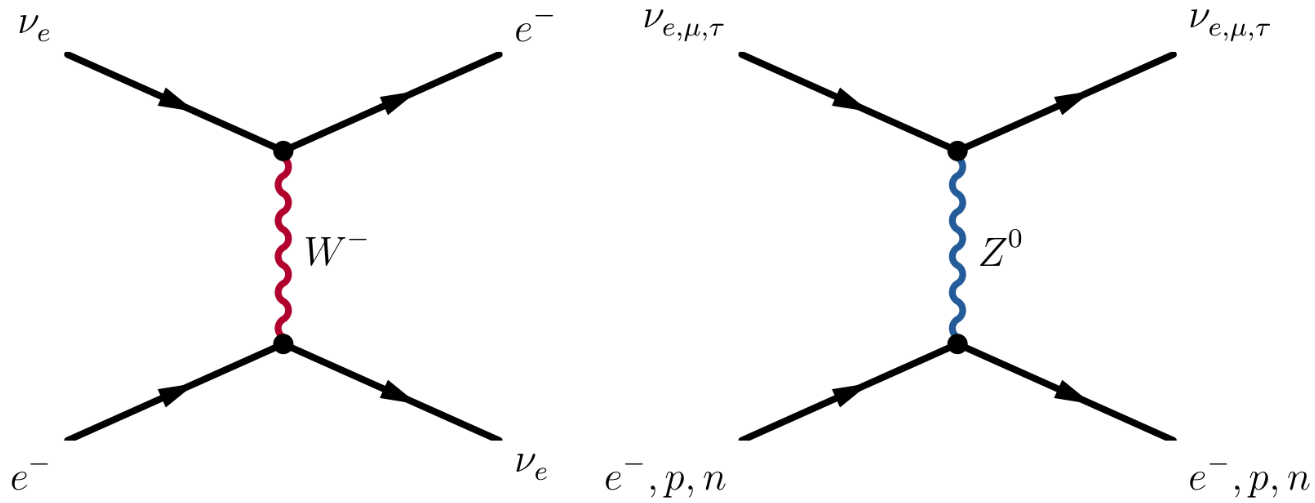


Figure 1.1: Tree level Feynman diagrams showing neutrino CC and NC interactions with matter particles, specifically electrons for the CC interaction. Image taken from reference [100].

and is called a neutral current (NC) interaction. This can happen for any flavor of neutrinos, and also for other target particles besides electrons.

For both of these interaction channels, we see that there is an outgoing neutrino. The probability of the neutrino interacting again in the detector is tiny, so effectively this is energy leaving the interaction that we cannot estimate. This will make energy estimation difficult. Also, for the NC events, we cannot tell which flavor of neutrino made this interaction. Lastly, the cross section for neutrino-electron interactions is very small, $\sim 1/2000$ of the total CC cross section. For these reasons, neutrino-electron scattering is not used directly in oscillation analyses, but can be used for other purposes, such as constraining the beam flux [119].

Besides interactions on the electrons, we can also consider the interactions on particles in the nuclei, such as in figure 1.2. Here, we also have CC and NC interactions, on the left and right respectively. Again, NC interactions have an outgoing neutrino, so we will not be able to identify the flavor, and it will be difficult to estimate energy. For CC interactions though, we go from a neutrino to a charged lepton, and specifically to the lepton that corresponds to the flavor of neutrino. For these, all the leptonic energy can then be deposited by the

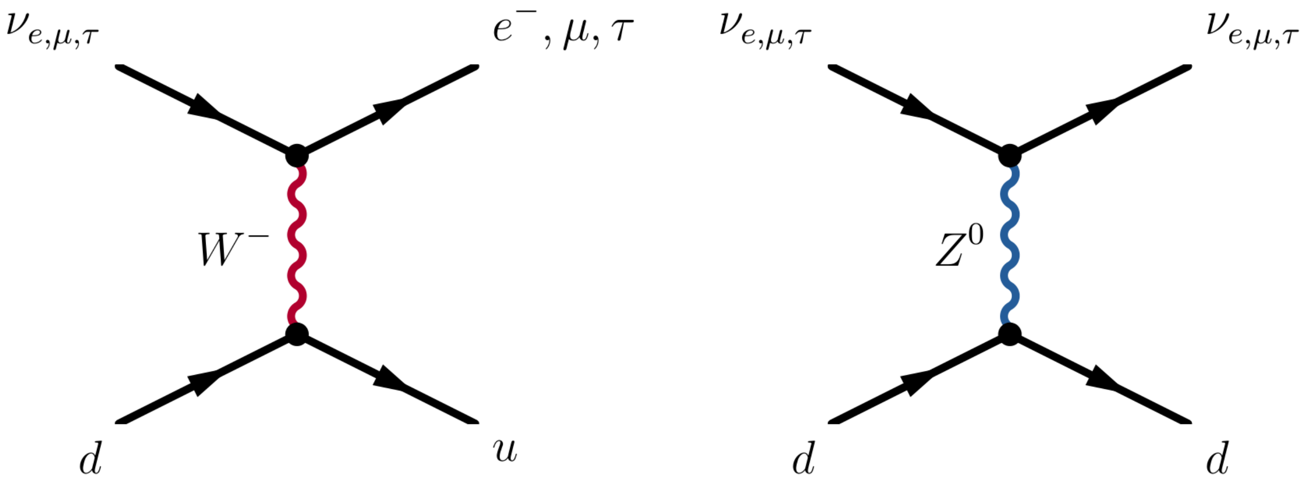


Figure 1.2: Tree level Feynman diagrams for CC and NC interactions for neutrinos with hadronic parts of matter. Image taken from reference [100].

charged lepton in our detector. The other part will be a hadronic component, where again most of the energy will be deposited in our detector. Also, if we can identify the lepton, we can identify the flavor of the neutrino that interacted.

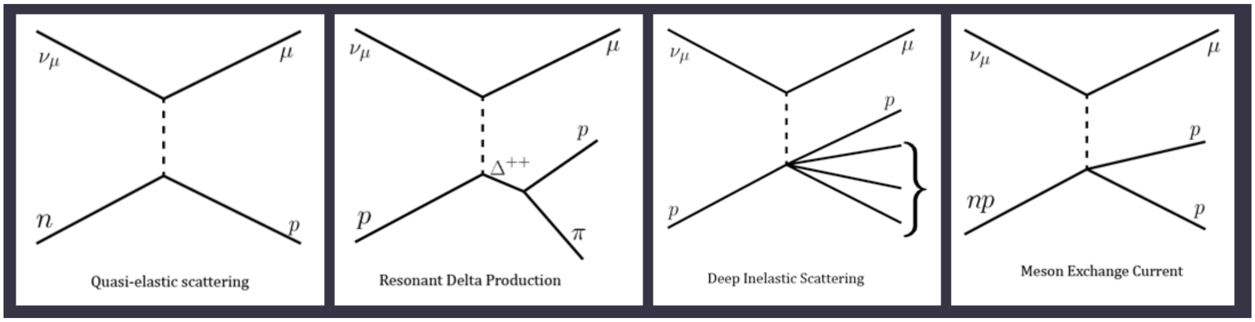


Figure 1.3: Illustrations for various neutrino – nuclei interactions: QE, Res, DIS, and MEC. Image taken from reference [106].

Therefore, these CC interactions in the nuclei are those we are interested in for our oscillation analysis, where the others are backgrounds. Neutrino-nucleus scattering is a very complicated process, however, and there are several different interaction types. A schematic of the common types we see at the energies relevant for NOvA is shown in figure 1.3. These are commonly abbreviated, from left to right, as QE, Res, DIS, and MEC. The simulated spectrum of true neutrino energies for NOvA is shown in figure 1.4, broken up in terms of these interaction types.

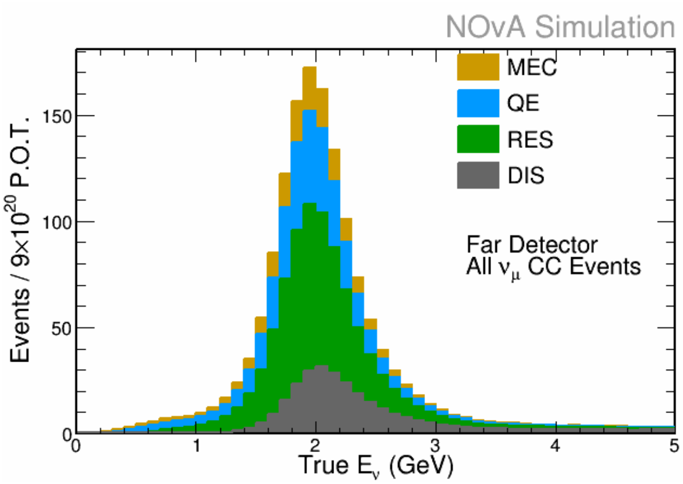


Figure 1.4: Simulated neutrino true energy spectrum of NOvA broken down by true interaction types.

1.3.3 Matter Effects

We have seen neutrino oscillation probabilities for oscillations in a vacuum, but in practice we have another complication to deal with, which is that neutrinos have some interaction with matter along the way. For coherent CC interactions with matter, only ν_e's have a non-zero term, as electrons are present in matter, but μ and τ's are not [87]. For ν_e's, we get a potential term for this interaction of $\sqrt{2}G_F n_e$, and for $\bar{\nu_e}$'s, we have an interaction of the opposite sign, $-\sqrt{2}G_F n_e$, where n_e is the electron density, and G_F is the Fermi constant. For coherent NC interactions, we can have an interactions with any flavor of neutrinos equally, which effectively can be taken to be an overall phase, so the effect will not be observable in oscillations. For a complete derivation, see reference [84].

The oscillations relevant for the NOvA experiment (the topic of this dissertation) are those from ν_μ to ν_μ or ν_e. The $\nu_\mu \to \nu_\mu$ oscillation probability becomes to first order [121]:

$$P_{\nu_\mu \to \nu_\mu} \approx 1 - cos^2\theta_{13} sin^2(2\theta_{23}) \frac{\Delta m^2_{32} L}{4E_\nu} + \mathcal{O}(\alpha, s^2_{13}) \tag{1.8}$$

For $\nu_\mu \rightarrow \nu_e$, now including the matter effects

$$P_{\nu_\mu \rightarrow \nu_e (\bar{\nu}_\mu \rightarrow \bar{\nu}_e)} \approx 4 sin^2\theta_{13} sin^2\theta_{23} \frac{sin^2\Delta(1-A)}{(1-A)^2} + \alpha^2 sin^2\theta_{12} cos^2\theta_{23} \frac{sin^2 A\Delta}{A^2} \tag{1.9}$$

$$+8\alpha J_{CP}^{max} cos(\Delta \pm \delta_{CP}) \frac{sin\Delta A}{A} \frac{sin\Delta(1-A)}{1-A} \tag{1.10}$$

with

$$\Delta \equiv \frac{\Delta m_{31}^2 L}{4E} \quad ; \quad A \equiv \frac{2EV}{\Delta m_{31}^2} \quad ; \quad J_{CP}^{max} = c_{12}\ s_{12}\ c_{23}\ s_{23}\ c_{13}^2\ s_{13}$$

and the ± is + for neutrinos, and − for antineutrinos. Note that, as expected, we get terms with different signs for electron neutrinos and antineutrinos.

Figure 1.5 shows an illustration of this matter effect on the oscillation from $\nu_\mu \rightarrow \nu_e$ for a scale of length and energy relevant for the NOvA experiment. Note the opposing directions of the matter effects for neutrinos and antineutrinos. Also notice these swap between the "normal hierarchy" and "inverted hierarchy", which is the sign of Δm_{32}^2. Positive is normal, and negative is inverted. This is called the "mass hierarchy", or "mass ordering", and in this dissertation, the two will be used interchangeably.

See figure 1.6 for an illustration of the different mass ordering scenarios. The sign of Δm_{12}^2 is already well established to be positive (see section 1.4 for a discussion). So differing signs of Δm_{32}^2 also mean a different sign of Δm_{31}^2. Therefore, different mass orderings give different signs of A, Δ, and α, which explains the differences shown in figure 1.5 for the different mass orderings.

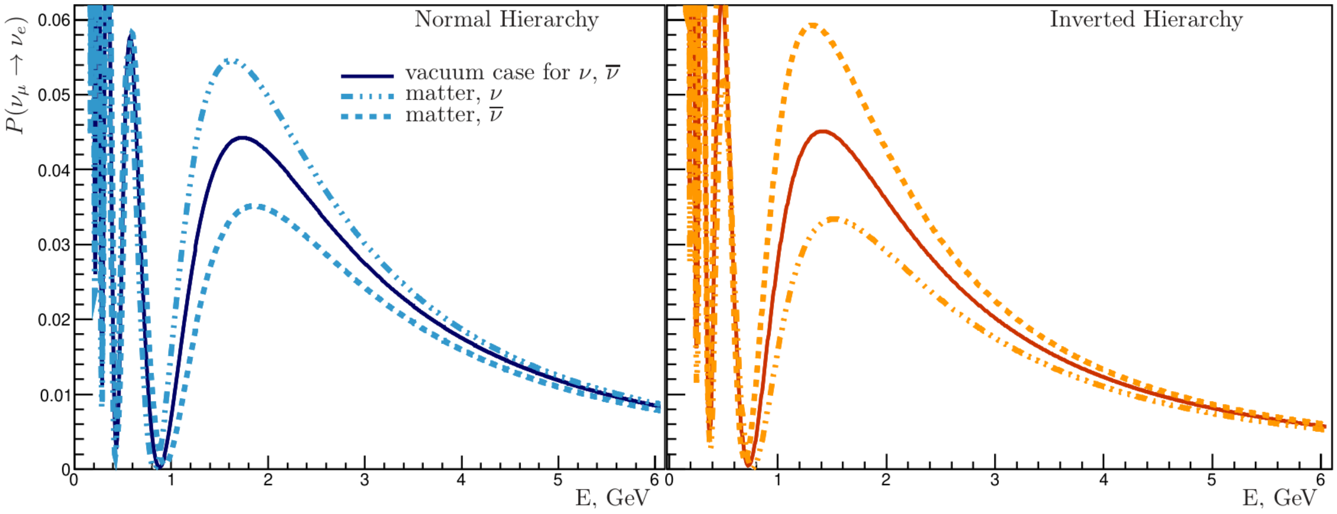


Figure 1.5: The matter effects on the $\nu_\mu \to \nu_e$ oscillation probability for neutrinos and antineutrinos, and for normal and inverted mass ordering. Image taken from reference [87].

1.4 Neutrino Physics Experiments

In the following section, a very brief outline of the history of different neutrino experiments will be laid out. Then, the status of measurements for various neutrino parameters will be discussed, as well as a discussion of what remains to be measured.

See the Particle Data Groups review [121] for a more thorough review of the history.

1.4.1 Solar Neutrinos

Energy is generated in the Sun's core through nuclear interactions. There are two main chains of reactions: the pp chain, and the CNO cycle. These reactions provide a natural source of pure ν_e events. The rates for these reactions, as well as their energy spectra, are given precisely by the model of nuclear reactions in the Sun, called the Standard Solar Model (SSM).

The first experiment to detect solar neutrinos was at Homestake starting in the late 1960s

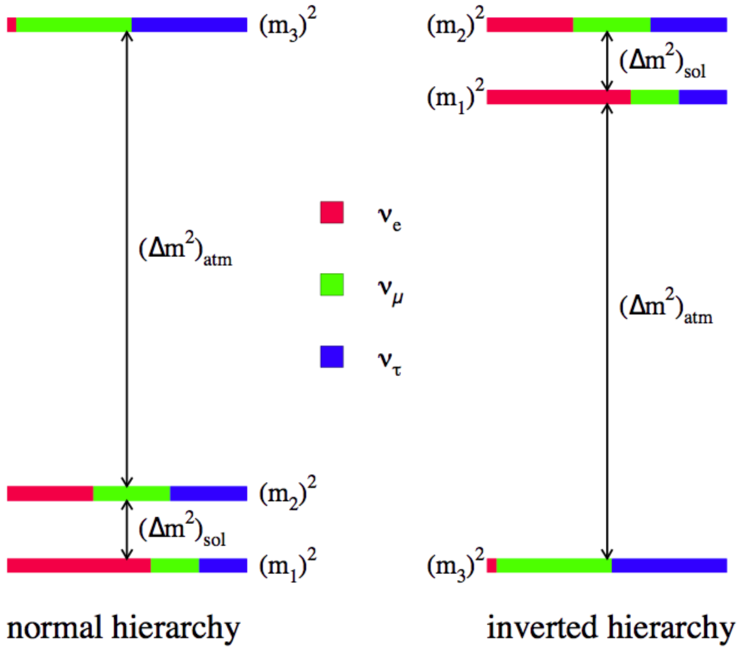


Figure 1.6: Illustration showing the two possible mass orderings between the 3 mass eigenstates: normal and inverted ordering. The ordering between 1 and 2 is known, but between 3 and the other two are not. Image taken from reference [39].

[53]. The detector utilized a radiochemical technique based on inverse beta decay:

$$\nu_e \; + \; ^{37}Cl \; \rightarrow \; ^{37}Ar \; + \; e^-$$

The experiment measured a rate significantly smaller than was predicted from SSM, which was called the solar neutrino problem [50].

Other solar neutrino experiments later probed other solar neutrino energies in an attempt to better understand the solar neutrino problem. Gallium based detectors could measure solar neutrino rates with a much lower energy threshold using the reaction

$$\nu_e \; + \; ^{71}Ga \; \rightarrow \; ^{71}Ge \; + \; e^-$$

. Several experiments using gallium based detectors were created, particularly SAGE [2] and GALLEX [71]. These gallium based detectors confirmed the measured rate of solar neutrinos was far lower than the expected rate. Other types of experiments that confirmed the lower rate were Kamiokande [62] and Super-Kamiokande [65] using water based detectors.

The problem was finally understood when the Canadian experiment SNO, which is a heavy water based detector experiment, made precise measurements of both the CC and NC rates, which showed part of the solar neutrino flux was composed of other neutrino flavors besides electron neutrinos [23]. This showed the observed ν_e solar rate was consistent with the presence of neutrino oscillations combined with matter effects from inside the sun.

1.4.2 Atmospheric Neutrinos

Another natural source of neutrinos comes from interactions of high energy cosmic rays from space with nucleons in Earth's atmosphere. The neutrinos are made when pions and kaons made in these interactions decay. The energy spectrum of these is particularly wide, ranging from around 100 MeV all the way to the TeV scale and beyond [121]. The first detection of atmospheric neutrinos was done by the Kolar Gold Field experiment [15] in India.

Some experiments measuring the ratio of fluxes between ν_μ and ν_e reported a ratio lower than predicted [74]. This was later explained by Super-K, which confirmed the deficit, and reported results consistent with neutrino oscillations [64]. A later analysis would rule out other proposed solutions, and measure the L/E dependence directly to be consistent with oscillations [28].

1.4.3 Accelerator Neutrinos

Besides the natural sources of neutrinos described in past sections, there are also human made sources of neutrinos. A neutrino beam for neutrino experiments can be made, using a proton beam on a target, creating mesons which later decay into neutrinos and other decay products. After the other decay products are stopped by matter, only the neutrinos remain. Magnetic horns are used to focus and sort the parent mesons, focusing the beam and choosing mesons with a specific charge, effectively choosing a neutrino or antineutrino beam. These experiments choose a particular distance for the far detector from the beam, as to probe the specific region of L/E of interest. In general, accelerator neutrino experiments can be broken up into long baseline and short baseline. Long baseline experiments are designed to probe the oscillations predicted from the 3 neutrino flavor oscillations. Short baseline experiments probe neutrino oscillations that could be visible at shorter baselines from sterile neutrinos.

Long Baseline Experiments

The first long baseline neutrino experiment was K2K, which used the KEK 12 GeV proton beam to aim neutrinos at the Super-K detector at a 250 km baseline [24]. K2K also used a set of near detectors close to the beamline to constrain the beam flux, as is common for long baseline experiments. K2K confirmed the muon neutrino disappearance first observed by atmospheric experiments.

The MINOS experiment used the NuMI beam at Fermilab, using a 120 GeV proton beam, with a 735 km baseline from the far detector [16]. The MINOS near and far detectors were iron scintillator based detectors with tracking and calorimetry capabilities.

The CNGS neutrino beamline at CERN has a 17 GeV beam. OPERA used this beam for a search of tau neutrino appearance [19], the first to measure this oscillation channel. OPERA didn't use a near detector, as the beam should have no significant tau neutrino background. ICARUS also used the CNGS beam, with a liquid argon time projection chamber (LArTPC) [104].

T2K uses the J-PARC synchrotron facility, off axis from the Super-K detector [3]. T2K made the first measurement of $\nu_\mu \rightarrow \nu_e$ oscillation.

NOvA (the topic of this dissertation) uses the NuMI beam in an off axis configuration from a far detector 810 km away. NOvA's near and far detectors use segmented liquid scintillator technology. NOvA also measures the $\nu_\mu \to \nu_e$ oscillation channel in additional to the muon neutrino disappearance channel. For more background on NOvA, see chapter 2.

Future long baseline accelerator experiments are DUNE [8], which will use LArTPC technology and an upgraded beam at Fermilab, and Hyper-Kamiokande [4], which will be a large water Cherenkov detector, and use the J-PARC beam.

Short Baseline Experiments

There are also several short baseline experiments that have looked for oscillations from sterile neutrino states. The LSND experiment at Los Alamos found an excess of $\bar{\nu}_e$ events [21]. The KARMEN experiment at Rutherford Appleton Laboratory was unable verify this excess [27].

The MiniBooNE experiment at Fermilab searched for $\bar{\nu}_e$ appearance in the same parameter spaces as LSND using the BSB beam using an 8 GeV proton beam. MiniBooNE's detector used mineral oil and had a 541 m baseline. MiniBooNE reported excesses in ν_e and $\bar{\nu}_e$ event counts [22].

A program of short baseline experiments at Fermilab have been created to probe this excess further, using several detectors at different baselines, unlike MiniBooNE and LSDN, which had only one detector. One of these experiments, MicroBooNE, has failed to verify the MiniBooNE excess [9].

1.4.4 Reactor Neutrinos

Another man made source of neutrinos is from nuclear reactors, which are high power sources of $\bar{\nu}_e$'s. Calculating the expected energy dependent flux of neutrinos from reactors is inherently difficult, with contributions from several isotopes.

KamLAND is a liquid scintillator detector in the Kamiokande cavern in Japan, with an 180 km baseline from the reactor being measured. KamLAND was the first experiment to measure $\bar{\nu}_e$ disappearance from a nuclear reactor [60].

Since then, several experiments were devised to make precision measurements of oscillations from reactor sources at shorter baselines: Double Chooz in France, Daya Bay in China, and RENO in Korea. Double Chooz uses a gadolinium-doped liquid scintillator detector about 1050 m from two reactors, and a near detector [7]. Daya Bay has four far detectors and four near detectors, with the near detectors broken up into two different near detector halls [25]. Daya Bay's far detector is 1648m from six reactors. RENO has an identical near and far detector, with the far detector 1383 m from six reactors [33]. These experiments all released their first results in 2012, all reporting measurements of $\bar{\nu}_e$ disappearance.

1.4.5 Measurements of Oscillation Parameters

With the broad range of neutrino physics experiments discussed previously, many parameters of the PMNS matrix, equation 1.5, are well constrained. The Majorana phases of the Majorana mixing parameterization, equation 1.4, are also in theory measurable in the case of neutrinos being Majorana, but in oscillation probabilities, the terms including those are highly suppressed, so they are extremely difficult to measure.

Different classes of experiments measure different oscillation channels, at different baselines, and are therefore sensitive to different sets of parameters. Solar neutrino experiments currently have made the best measurements of θ_{12}, although they also have sensitivity to Δm^2_{21} and θ_{13}.

Reactor experiments with longer baselines, specifically KamLAND, have the most precise measurement of Δm^2_{21}, but also have sensitivity to θ_{12} and θ_{13}. The reactor experiments with shorter baseline, specifically Double Chooz, Daya Bay, and RENO, have the best measurement of the angle θ_{13}, and also have sensitivity to $|\Delta m^2_{32}|$.

Long baseline accelerator experiments looking at $\nu_\mu \rightarrow \nu_\mu$ (and $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$) are most sensitive to $|\Delta m^2_{32}|$, and θ_{23}. Measuring the channel $\nu_e \rightarrow \nu_e$ (and $\bar{\nu}_e \rightarrow \bar{\nu}_e$) gives the best sensitivity to δ_{CP}, as well as some sensitivity to θ_{13} and θ_{23}.

Lastly, atmospheric neutrino experiments see a mix of different flavors, originating as ν_μ and ν_e (and $\bar{\nu}_\mu$ and $\bar{\nu}_e$), as well as different oscillation lengths. For this reason, atmospheric neutrino experiments have sensitivity to a range of oscillation parameters: θ_{23}, $|\Delta m^2_{32}|$, θ_{13}, and δ_{CP}.

Of these oscillation parameters, the parameters with the least precise measurements are δ_{CP} and θ_{23}, as well as the mass ordering, which is the sign of Δm^2_{32}. δ_{CP} is of particular importance for theoretical reasons, as it's the CP violating phase and significant CP violation in the lepton sector is a possible explanation of the matter-antimatter asymmetry in the universe [66]. The θ_{23} octant isn't yet known decidedly, i.e. whether $sin^2\theta_{23}$ is greater than or equal to 0.5. The mass ordering is one of the basic parameters of the oscillations, yet difficult to measure, but long baseline accelerator experiments can resolve it by making use of differences that it causes to matter effects, see figure 1.5. Reactor experiments also have sensitivity to the mass ordering. JUNO, for example, will have good sensitivity to the mass ordering by measuring fine interference patterns in the reactor oscillation spectrum [10].

Chapter 2

The NOvA Experiment

NOvA (NuMI Off-axis ν_e Appearance) is a currently running long baseline neutrino experiment. It uses the NuMI neutrino beam at Fermilab, a near detector (ND) at Fermilab, and a far detector (FD) 810 km away in Ash River, Minnesota [31]. The main goal of NOvA is to make measurements of oscillation parameters, specifically $sin\theta_{23}$, Δm^2_{32}, and δ_{CP}. The NuMI beam is off axis from the NOvA detectors, to concentrate the beam flux around 2 GeV, the region NOvA is most sensitive to the oscillation parameters, see figure 2.1.

A central aspect of NOvA's design is a functionally identical near and far detector. NOvA uses the near detector data / Monte Carlo ratio as a data-driven correction to the FD spectrum. This helps constrain systematic uncertainties, such as beam flux and neutrino interaction models.

2.1 The NuMI Beam

NOvA's neutrino source is the Neutrinos at the Main Injector (NuMI) beam at Fermilab. NuMI uses the Fermilab Main Injector proton accelerator, which was previously used as part

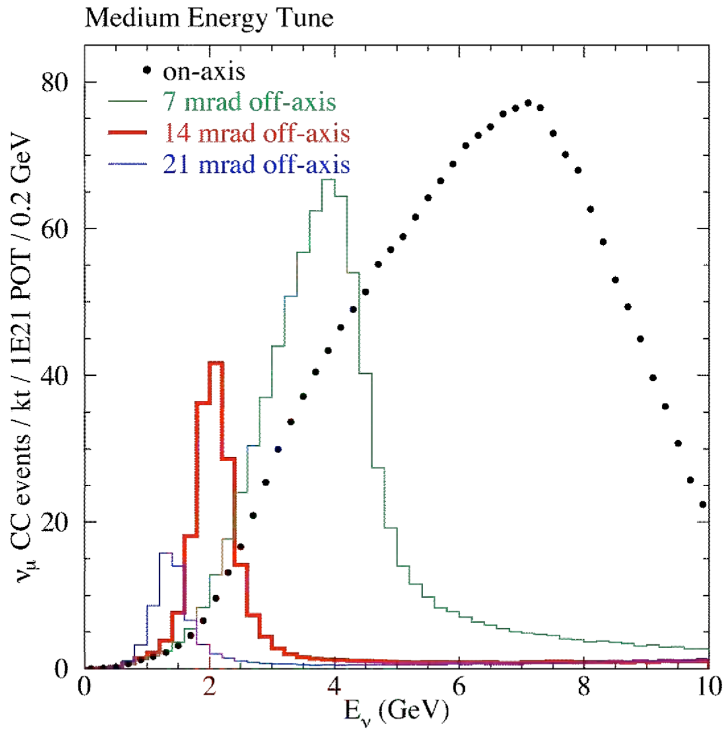


Figure 2.1: The neutrino energy flux NOvA would see at various angles to the beam, taken from the NOvA TDR [31]

of the Tevatron proton collider [17]. This 120 GeV proton beam is the basis of the neutrino beam.

A schematic of the NuMI beam is shown in figure 2.2. This proton beam is first directed at a graphite target. The protons collide with the graphite nuclei to create charged mesons, primarily charged pions and kaons.

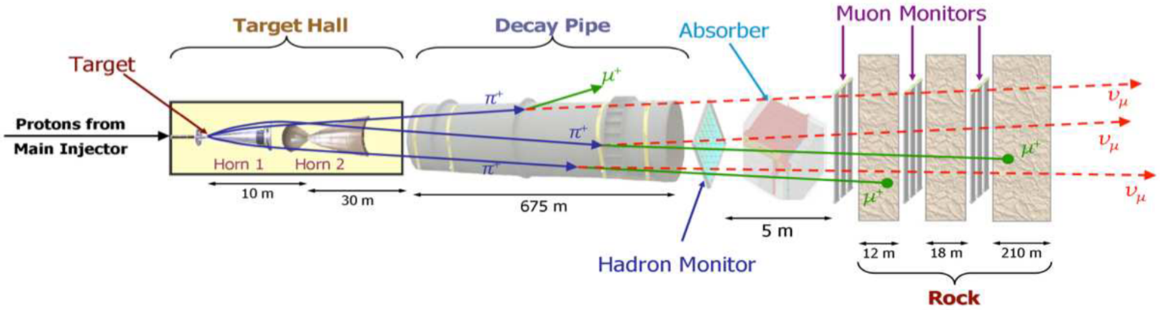


Figure 2.2: An illustration on the components of the NuMI beam. Image taken from reference [17].

Next, magnetic horns focus the charged mesons which will be the parent particles of the neutrinos. The focusing horns also can choose between keeping positive or negative mesons. The configuration that keeps the π^+'s, is called Forward Horn Current (FHC). The π^+'s decay primarily through the decay $\pi^+ \rightarrow \mu^+ + \nu_\mu$. Therefore, the FHC configuration will give a mostly muon neutrino beam. The horn configuration that keeps the negative pions will therefore give us a mostly muon antineutrino beam, called Reverse Horn Current (RHC). The beam isn't 100% pure, either in flavor or neutrino/antineutrino makeup. In either beam mode, there are decay modes that create neutrinos of the wrong sign, which will be called wrong sign background in the analysis. There are also some ν_e's besides the mostly ν_μ's, called the intrinsic Beam Electron Neutrinos (BEN), which will be a background to the ν_e's appearing through oscillations at the far detector.

Next, the selected and focused mesons are sent through the decay pipe, which is a length where the muons are allowed to decay in a vacuum. The decays create the neutrinos that are NOvA's signal neutrinos. Antimuons (and muons for RHC) are created in the decay pipe as well. A large absorber block, made of aluminum, steel and concrete, is just after the decay pipe, meant to absorb any hadrons that remain. From there, the muons will lose energy when traveling into the rock underground.

We have no way to know the exact number of neutrinos being creating at NuMI, but what we can measure is the number of protons in our beam being sent to the target. We will therefore measure our exposure with Protons on Target (POT), which should be proportional to the amount of neutrinos created.

2.2 NOvA's Detectors

2.2.1 Detector Technology

NOvA's detectors are composed of stacked PVC cells filled with liquid scintillator [31]. The liquid scintillator is made up of 4.1 % pseudocumene which is the active scintillant, with the rest being mostly mineral oil. Charged particles will created scintillation light peaked at 360 - 390 nm in the pseudocumene. PPO and bis-MSB are added to shift the peak to 400-450 nm. A wavelength shifting fiber optic cable in the cell then absorbs light in the 400-450 nm range (blue), and shifts it to 490 - 550 nm (green).

The signal from the wavelength shifting fiber is read out by avalanche photodiodes (APDs) [86]. The signals from the APDs are amplified by a low noise pre-amp on the MASDA chip [120] made by Fermilab. The amplified signals are then multiplexed by a custom ASIC designed for NOvA. The front end electronics trigger continuously, without an external trigger. The time stamps are compared to the NuMI beam spill times in the Data Acquisition System (DAQ) to determine if they're in the beam spill.

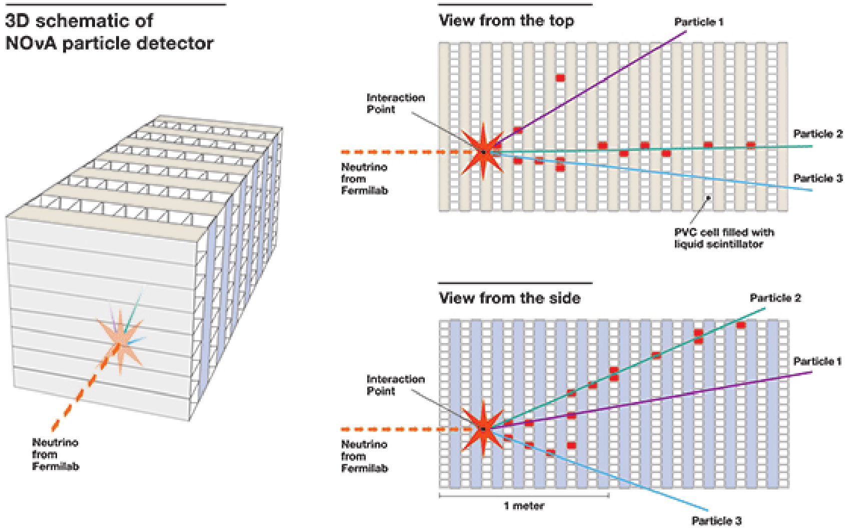


Figure 2.3: An illustration of the two views of an event seen inside NOvA's detectors.

2.2.2 Near Detector (ND)

The length of the cells alternate between horizontal and vertical, giving two different views of each event in the detector, see figure 2.3. The Near Detector (ND) at Fermilab, and the Far Detector (FD) in Ash River, Minnesota are functionally identical, but different sizes. The ND is 64 cells wide and 96 cells high, with 186 of these planes. There are 10 additional planes on the end, with planes of steel in between, to help tag muons exiting the active region of the detector, called the muon catcher. In total, the ND is 2.9 m wide, 4.2 m tall, and 14.3 m long, with a 222 ton active volume.

The ND is located in the MINOS shaft, under the MINOS hall at Fermilab. A new cavern was added to the shaft for the ND, 14 mrad off axis from the NuMI beamline.

2.2.3 Far Detector (FD)

The FD has a 15 kton active volume. It is a larger, structured arrangement of the cell construction used for the ND, in all being 15.6 m wide, 15.6 m high, and 78 m long, with 385,000 cells in total. The FD is located in Ash River, Minnesota, at an 810 baseline, above ground.

Like the ND, it is 14 mrad off axis from the NuMI beam. The off axis detector shifts the energy spectrum of the neutrinos from the beam, as shown in figure 2.1. The breakdown of the flux and neutrino energy of neutrinos coming from parent charged pions at various angles is shown in figure 2.4.

Unlike the ND, the FD is above ground, so it will be constantly showered with the spectrum of cosmic rays, mostly secondary cosmic muons. The FD timing resolution therefore is required to be large enough to reject cosmics with great efficiency to be able to pick out beam neutrinos.

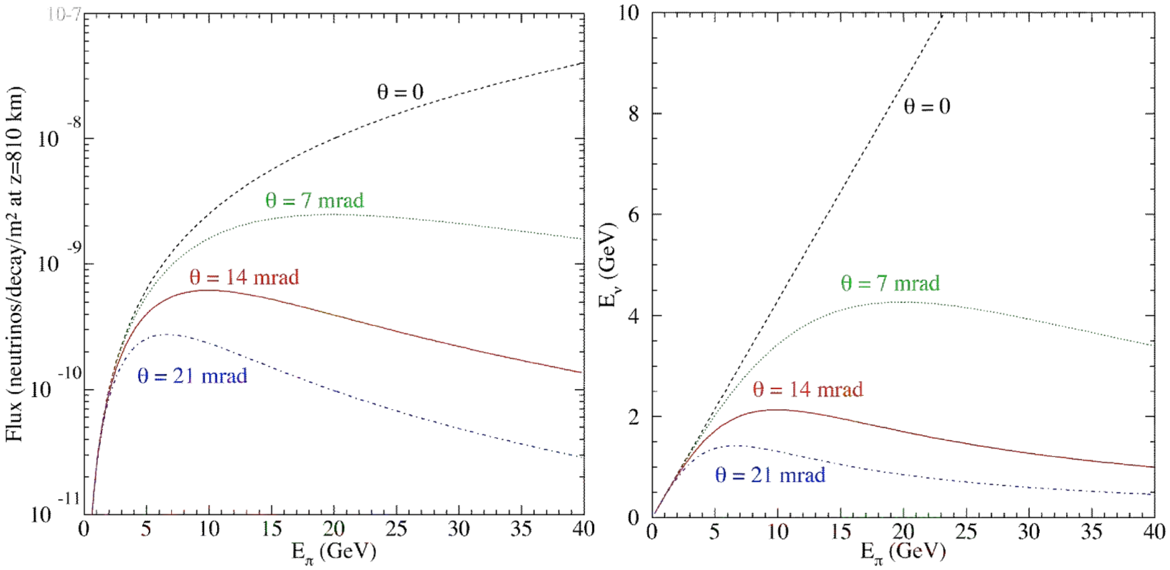


Figure 2.4: The flux and neutrino energy at various angles from the NuMI beam given the energy of the parent pion. Taken from the NOvA TDR [31].

2.2.4 Calibration

Cosmic muons are used for calibration in NOvA's detectors. Photons produced along the length of NOvA's detector cells are attenuated by different factors at different lengths along the cell by the wavelength shifting fiber. The first step of the calibration is to correct for this attenuation using through going muons, as shown in figure 2.5 for the FD. "W" is the length along the given cell. This is called "attenuation calibration".

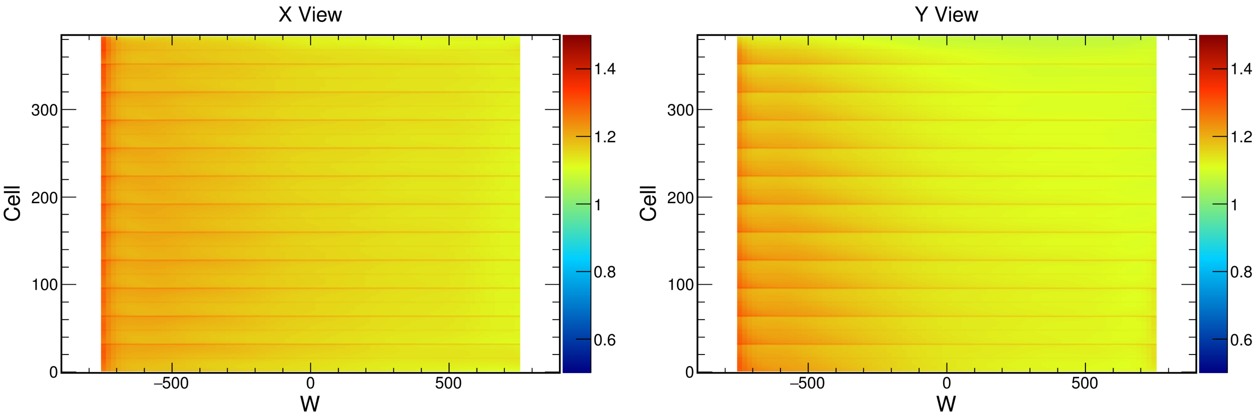


Figure 2.5: The attenuation calibration factors for different cells in the FD at different distances from the middle of the cell, W.

Figure 2.6 shows the reconstructed energy by true energy using simulation at different distances along the cell to evaluate the performance of the "attenuation calibration".

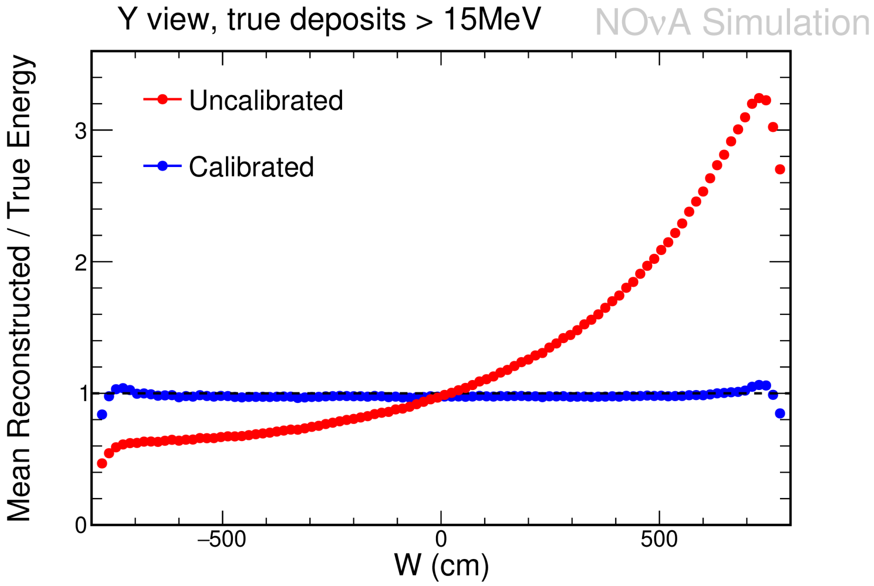


Figure 2.6: The hit reconstructed energy, calibrated and uncalibrated, by the true hit energy at different distances alone the cell, W, for simulated muons.

Now with the effects of the calibration along the length of the cell corrected, absolute calibration is done to translate the measured charge read out into a calibration measurement of energy. This step is called "absolute" calibration. Stopping muons are used to correct for this energy scale. Figure 2.7 shows the dE/dx of hits for these stopping muons as a function of the distance from the end of the track. Hits from the distance of 100-200 from the track end are selected to use for the absolute calibration. This is a flat region which follows the Bethe-Block distribution well.

Figure 2.8 shows distribution of the hits in the above selected region for the stopping muons. Figure 2.8a shows the true dE/dx from simulation. Figure 2.8b shows the measured uncorrected hit dE/dx for data and simulation. And figure 2.8c shows using the ideal distribution (a) and uncorrected values (b) to correct the hit dE/dx, showing a more aligned distribution between data and simulation.

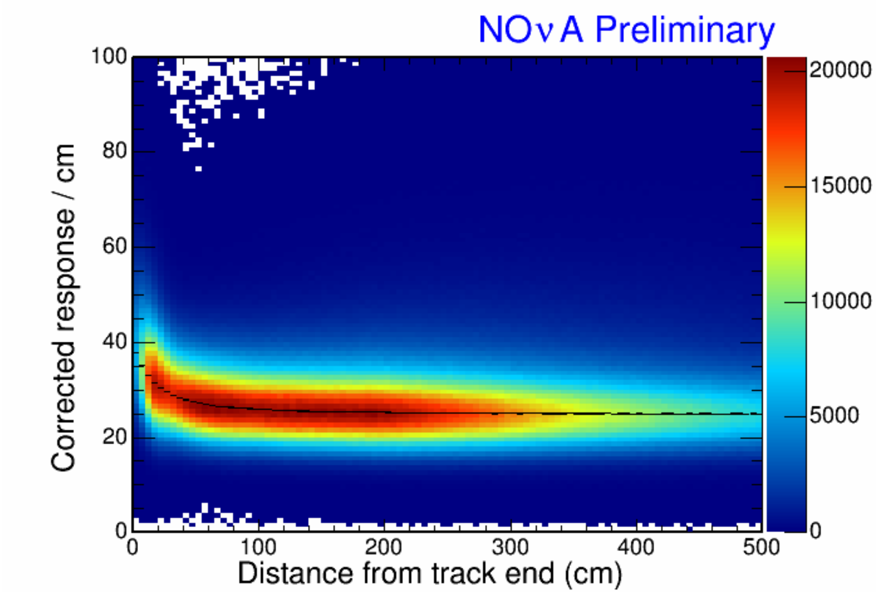


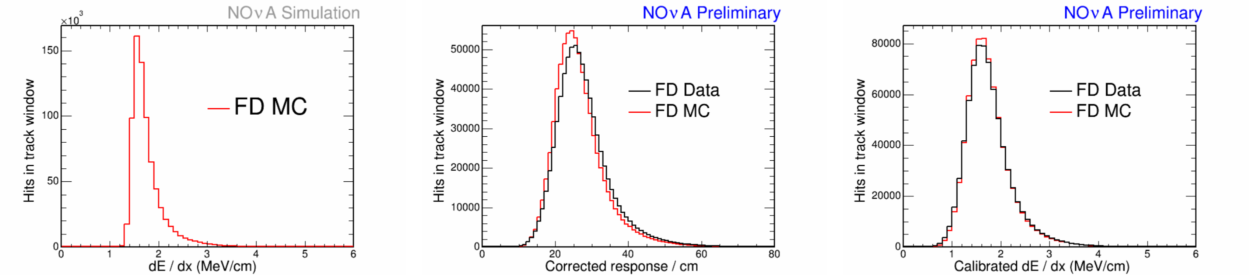
Figure 2.7: The dE/dx of each hit for stopping muons plotted against the distance of the hit from the end of the track.

2.3 Simulation

Simulating particles in the detector is a basic part of any physics analysis. To know which physics parameters the data is consistent with, you must know what the data should look like for the a set of parameters in range of allowable parameters. For this reason, we simulate neutrino events in our detectors, and send them through the whole reconstruction chain to see what these look like after going through the same process the data will. The process of simulating individual samples to get the expected output is called Monte Carlo (MC).

The first stage of simulation is simulating the NuMI beam [29]. The beam is fully simulated from the target to the magnetic horns, to downstream production with Geant4 [20]. The spectra for different neutrino flavors come directly from these Geant4 simulations.

From there, with our neutrino spectra transported to the positions of the near and far detectors, we simulate the neutrino interactions in the detector with GENIE neutrino event generator [26]. We don't simulate cosmic rays for usual samples used for the analysis,



(a) (b) (c)

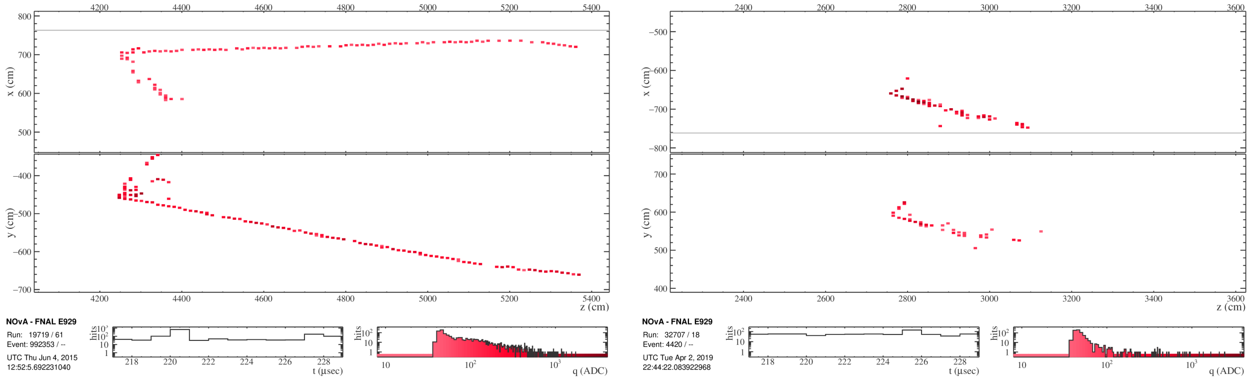
Figure 2.8: Distributions of hits from 100 - 200 cm from the end of the stopping muon track. (a) shows simulated muon true dE/dx for these hits. (b) Shows the data and MC reconstructed dE/dx per hit. (c) shows the calibrated dE/dx for data and MC, showing better consistency.

as we have a large sample from cosmic only periods, so these cosmics are overlaid over the simulation [52]. For samples we need simulated cosmic rays, such as calibration, we use CRY [70]. From there, the charged particles created from the interaction generator are passed to Geant4, which propagates the particles through the detector geometry, and simulates their energy deposits.

Geant4 could simulate the light transport through the cell and wavelength shifting fiber, but this is extremely computations expense. Instead, a deterministic parameterized model is developed to translate the energy deposited by Geant4 into readouts by the APD.

2.4 Reconstruction

NOvA has a chain of reconstruction software which algorithmically identifies features of events that occur in NOvA's detectors, separated in time, which can be used in downstream analyses. The reconstruction is done in NOvA's software framework, NOvASoft, and written on top of the ART framework. [88]. The raw data is saved in files, grouped by the detector, "run" and "subrun" of data taking. The simulated MC events are identical in format to the actual data files at this stage, as to be as identical as possible when treated by reconstruction algorithms.



(a) (b)

Figure 2.9: Event display for candidate ν_μ event (left) and ν_e event (right), with two views for each from the two detector views.

The following objects are reconstructed from the data and MC in reconstruction:

- Raw Hits – The raw detector readout indicating light seen in a cell. These simply including the ADC counts of the detector, the plane and cell, and the time.

- "CalHits" – First reconstructed object created by ART, an object representing a hit in NOvA's reconstruction framework.

- Slices – Collection of hits which should represent a single physics event, such as a cosmic muon or a neutrino event, including all the charged daughter particles. Our current slicer algorithm is called "TDSlicer", which has been fine tuned to avoid issues seen with previous slicers, such as issues caused by ND event pileup [97]. Slices are the basic reconstruction objects used for all other downstream reconstruction algorithms. From here, hits located in slices are considered signal, when all others are considered noise.

- Kalman tracks – Charged particles that don't produce showers will create long tracks in out detector. This is particularly true for muons created from ν_μ CC events, see the ν_μ CC candidate event display with the long muon track in figure 2.9a. We have a

Kalman filter based tracking algorithm [83] that works well for these long tracks. This is done separately for each view at first, then combined later. This algorithm will play a large role in the identification and energy reconstruction of the ν_μ CC events to be used in the oscillation analysis

- Vertex – A reconstructed vertex is needed for further reconstruction steps for more sophisticated feature identification needed for analysis of ν_e CC events. The first step is to map the hits into a Hough space with a Hough transform [77], which identifies associated pairs of hits possibly coming from the same particle. Then, the elastic arms algorithm [95] identifies vertex candidates given the output from the Hough transform. Vertex candidates are ranked to find the optimal vertex, with a number of edge cases corrected for.

- Prongs – With the vertex, the next step is to identify "prongs", which should correspond to a single daughter particle of a neutrino event, such as a pion, electron, or proton. These are used to help identify features of the ν_e CC events, such as the electron and individual hadronic daughter prongs. See the fuzzy electron shower and hadronic component in the ν_e CC candidate event display shown in figure 2.9b. Prong reconstruction uses the Fuzzy K means algorithm [57]. This is done first individually in each plane, and then the 2D prongs are combined to 3D prongs. A flow chart and illustration of each step is shown in figure 2.10

The MC production of simulation and reconstruction used for the 2024 analysis was production 5.1, which has only a few simulation and reconstruction changes from production 5, used for the 2020 analysis [42].

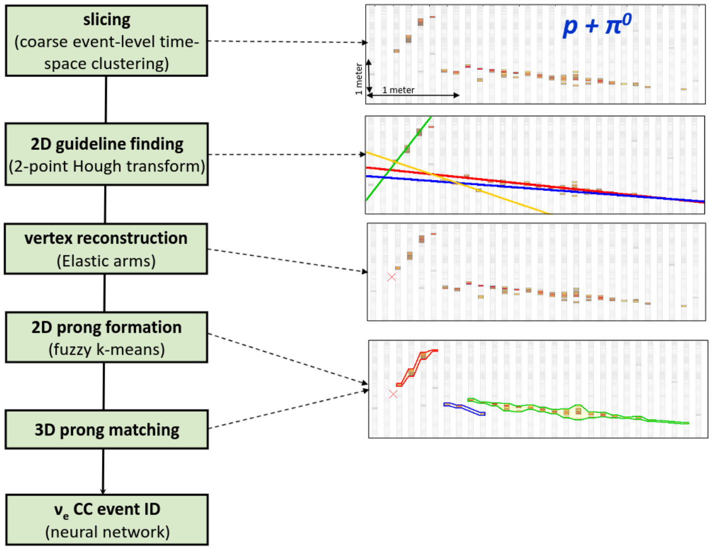
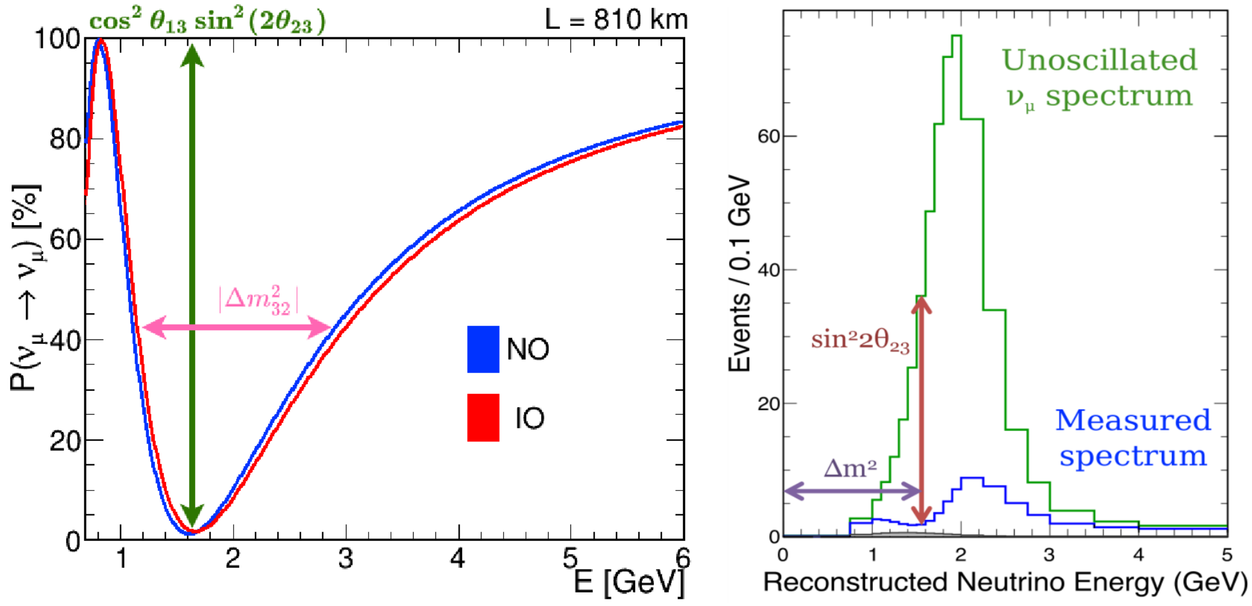


Figure 2.10: NOvA reconstruction flowchart for the prong reconstruction chain, showing an example event for each. The last step is the higher level event CVN, discussed in section 4.1.3. Image taken from reference [32].

2.5 NOvA's Physics Goals

NOvA's main physics goals are measuring θ_{23}, δ_{CP}, and Δm^2_{32}, including its sign, the mass ordering. It measures two oscillation modes to measure this, muon neutrino disappearance, $\nu_\mu \to \nu_\mu$ (and $\bar{\nu}_\mu \to \bar{\nu}_\mu$), as well as electron neutrino appearance, $\nu_\mu \to \nu_e$ (and $\bar{\nu}_\mu \to \bar{\nu}_e$).

Muon neutrino disappearance is most sensitive to $|\Delta m^2_{32}|$, as well θ_{23}. See figure 2.11 for an illustration of the qualitative effects of these parameters. Electron neutrino appearance is most sensitive to δ_{CP}, but also θ_{23}, the mass ordering, and θ_{13} (which will be much better constrained by reactor experiments, though). Figure 2.12 shows the expected rates for ν_e vs $\bar{\nu}_e$ event counts for different values of δ_{CP} (the ellipses), θ_{23} with example upper octant (UO) and lower octant (LO) values, and the mass orderings. This also illustrates the need to have data for both neutrinos and antineutrinos, which NOvA gets from running in FHC and RHC beam modes. See figure 1.5 for how the matter effects increase the sensitivity to the mass ordering.



(a) (b)

Figure 2.11: (a) $\nu_\mu \rightarrow \nu_\mu$ oscillation probability by energy for normal and inverted ordering, illustrating the width and amplitude related to the oscillation parameters. (b) The resulting ν_μ energy spectrum with and without oscillations in the FD

2.5.1 Outline of Oscillation Analysis

The rough outline of how NOvA's oscillation analysis works is as follows:

- Run experiment to get data, both with FHC and RHC beam.
- Simulate MC for ν_μ and ν_e events, for the FHC and RHC beam.
- Run reconstruction on data and MC.
- Estimate energy of events, as neutrino oscillations are a function of energy (see chapter 3).
- Send events through selection for ν_μ and ν_e samples, choosing the events that look like ν_μ or ν_e events (see section 4.1).

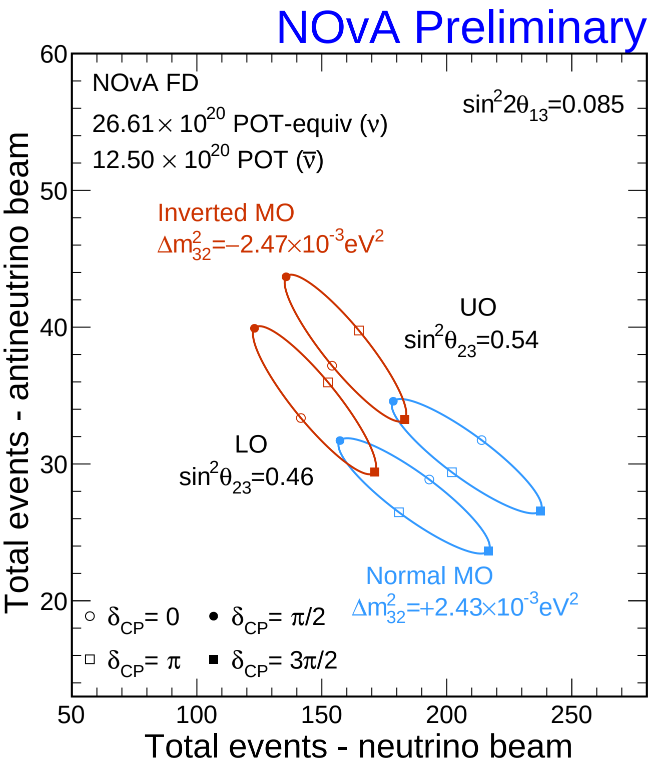


Figure 2.12: $\nu_e - \bar{\nu}_e$ event count rates against varying values of δ_{CP} in the normal and inverted ordering, and upper and lower octant

- Use the selected MC events to make "predictions", which are the expected values of ν_μ and ν_e energy spectra at the FD, given a set of oscillation parameters.

- Include in these the systematic uncertainties of our detectors and analysis, and leave them as constrained free parameters in the predictions (see section 4.3).

- Fit the oscillation parameters and systematics to the data, using the predictions as expected rates for a set of oscillation parameters, finding the oscillation parameters most consistent with the data (see chapter 5).

Chapter 3

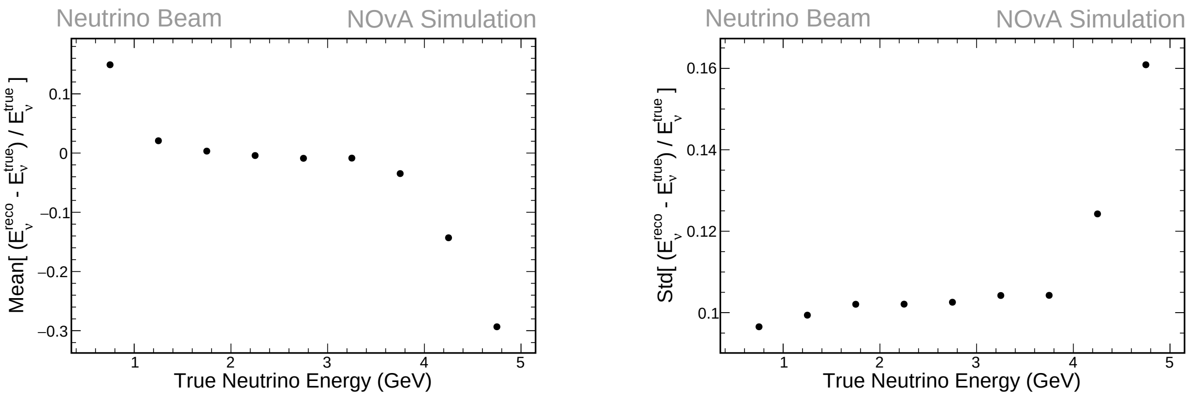
Neutrino Energy Estimation

3.1 Energy Estimation

Neutrino oscillations are inherently dependent on the neutrino energy. Therefore, neutrino energy estimation is central to the inference of neutrino oscillation parameters. For NOvA's oscillation analysis, we are most interested in resolving the region between 1 – 3 GeV.

After selecting ν_e CC and ν_μ CC event samples, energy estimation is performed individually, with two different methods used for each sample. In this analysis, NOvA used energy estimation methods similar to those used in recent analyses, but significant consideration was given to switching to a method based on deep learning algorithms, as outlined in section 3.3. These traditional and deep learning methods considered are discussed in section 3.2.

When discussing energy estimation performance, we will frequently discuss "energy resolution", is $(recoE - trueE)/trueE$, which in general should be a sharp peak around 0. The narrower the peak, the better the performance, and a shift off of 0 indicates a bias.



(a) (b)

Figure 3.1: The mean and standard deviations of the resolution of the calorimetric ν_e energy estimation.

3.2 Developed Techniques

3.2.1 ν_e Energy

The Calorimetric Method

The traditionally used ν_e energy estimation uses calorimetric energy of the electromagnetic and hadronic components, and corrects them individually. To identify particle clusters as electromagnetic or hadronic in nature, NOvA has a particle CNN-based identification method [101].

NOvA uses quadratic correction factors to then estimate the overall ν_e energy:

$$E_\nu = AE_{EM} + BE_{had} + CE_{EM}^2 + DE_{had}^2$$

The quadratic correction factors help keep the performance stable in the energy region from 1 – 3 GeV, see figure 3.1. This figure shows the mean and standard deviation of the resolution, broken up into regions of true neutrino energy. In general, this distribution should be narrow

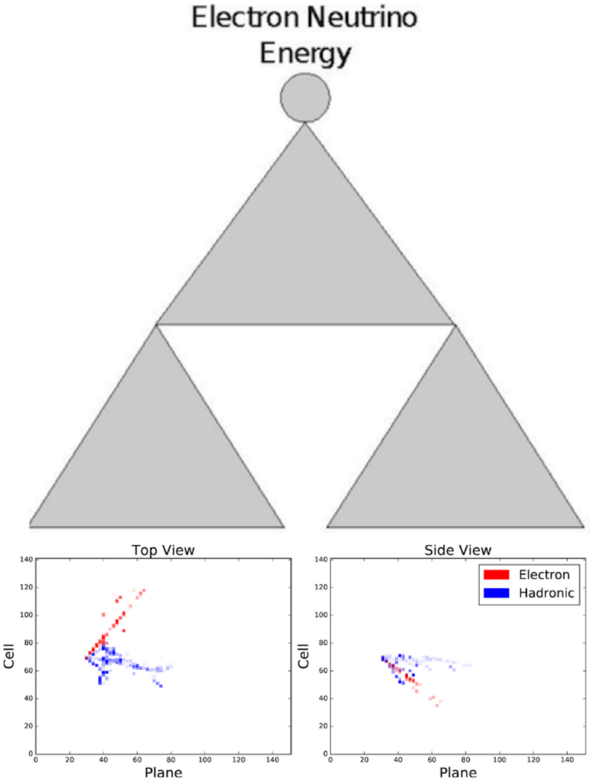


Figure 3.2: An illustration of the energy regression CNN which takes the two images of the event from each view as input, and output an estimate of neutrino energy.

and centered around 0, so the mean should be close to 0, and the standard deviation as low as possible. Notice how in the figure, the mean and standard deviation have good and flat performance from 1 – 3 GeV.

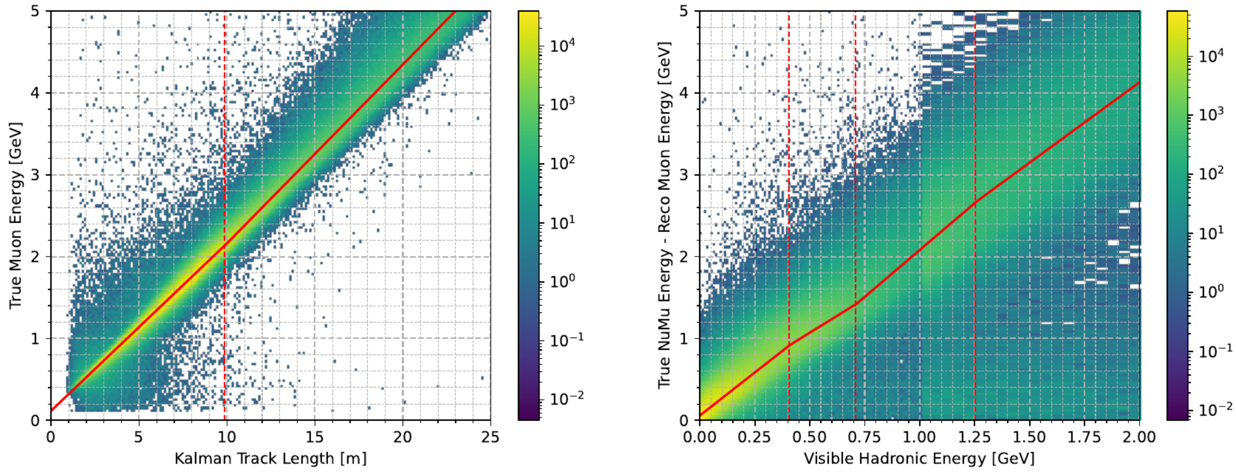
The CNN-Based Method

In addition to the traditional calorimetric based method, a method using a convolutional neural network (CNN) was also developed, which is a neural network specialized at taking images as inputs (see section A.2) [37]. This CNN takes two images of each neutrino event, one for each detector view, then spits out an estimate of electron neutrino energy, as illustrated in figure 3.2. We use a ResNet CNN architecture [72].

3.2.2 ν_μ Energy

The Spline-Based Method

The traditionally used method for ν_μ uses linear splines to estimate energy [111]. This method used separate spline fits for muon and hadronic parts individually, then added them together to get an estimate for the total neutrino energy. See figure 3.3 for an example. Muon energy is estimated from a fit to Kalman track length. Hadronic energy is estimated from a fit to visible deposited hadronic energy. Note that the ND has two different spline fits for the muon: one for the active region, and one from the muon catcher.

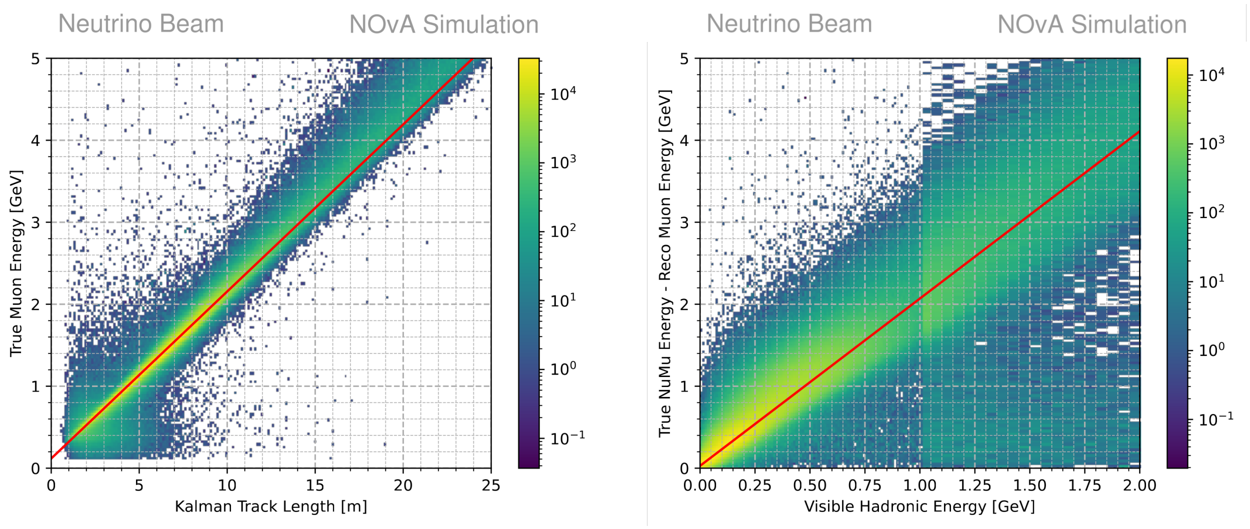


(a) Spline Fit to Muon Kalman Track Length

(b) Spline Fit to Deposited Hadronic Energy

Figure 3.3: Muon and hadronic energy spline fits, to the muon Kalman track length and the visible hadronic energy respectively.

In this analysis, it was found using splines for energy estimation can result in some jagged features of the hadronic energy spectrum in particularly, which happen in places where the slope of the line changes suddenly, when going from one line to the next in the spline. For this analysis, it was decided to use a single line instead of a series of lines (a trivial spline), the resulting fits shown in figure 3.4.



(a) (b)

Figure 3.4: Linear fits to the muon and hadronic energy, without multiple splines as shown into Figure 3.3, which was the set used for the final analysis

The LSTM Energy Estimator

An energy estimator that uses both slice level and prong level inputs in an LSTM neural network [75] was originally proposed in 2017 [103]. Extensive development and studies have been done since the 2020 NOvA analysis. One key feature of this new energy estimator is that it was trained to predict both the muon and the muon neutrino energy.

One of the challenges of working with prong-level variables is that the number of prongs per slice is not fixed. This means that the energy estimator needs to accept a variable number of inputs, this is why the LSTM (Long Short-Term Memory) neural network architecture was chosen as it can naturally work with varying input lengths. The final LSTM energy estimator architecture is shown in figure 3.5.

The LSTM energy estimator takes a list of reconstructed quantities as input, both event-level quantities, and information about individual reconstructed 2D and 3D prongs. The complete list of input variables is shown in Table 3.1.

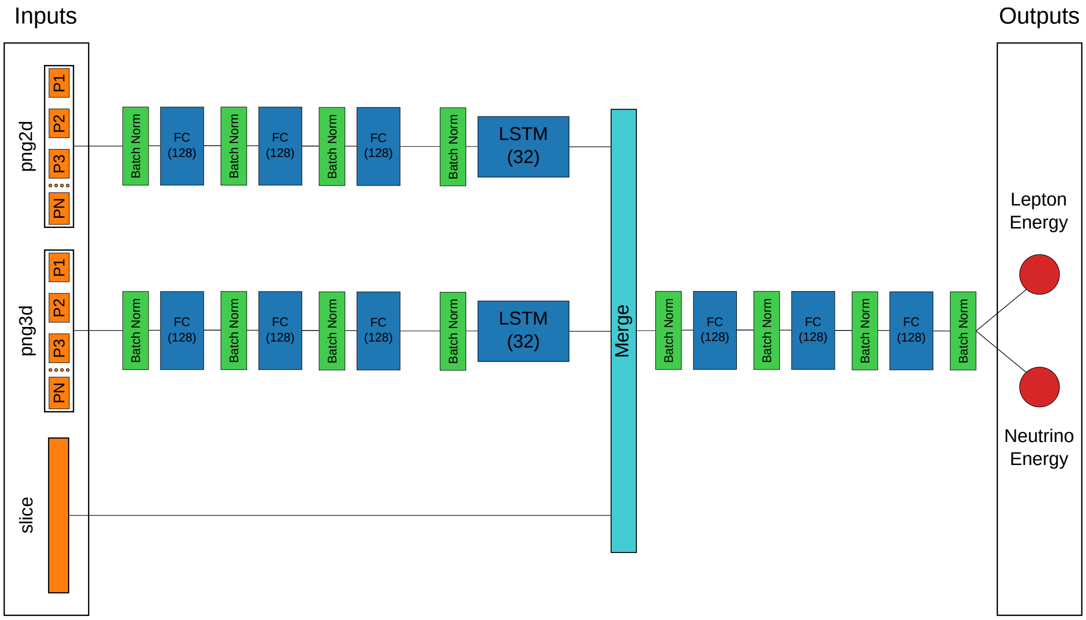


Figure 3.5: A schematic of the LSTM architecture, showing the slice and prong level inputs.

3D Prong Variables	2D Prong Variables	Slice Variables
reco energy	cal energy	cal energy
momentum (x,y,z)	direction (x,y,z)	timing info
overlapping energy	length	gain info
cal energy	# hits total and per plane	# hits
particle CVN	# planes	cal energy not from prongs
direction (x,y,z)	start (x,y,z)	
length	weighted cal energy	
# hits total and per plane		
# planes		
start (x,y,z)		
weighted cal energy		

Table 3.1: The input variables of the LSTM energy estimator. The estimator takes 2D and 3D prong variables in addition to slice-level variables and flags for specific modes of detector operation (e.g. low gain and coarse timing).

A number of considerations had to be made to make sure what the LSTM learned could generalize from the training sample well:

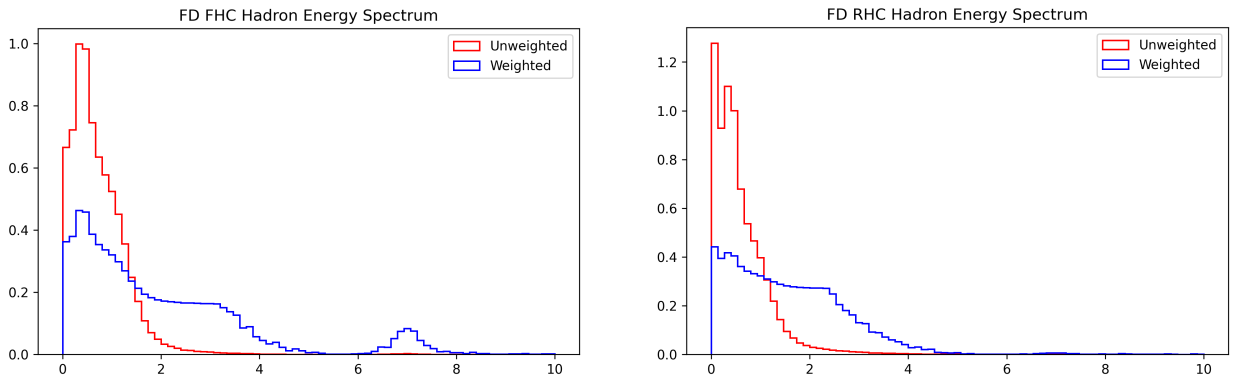
- Applying preselection only to avoid sharp features in the reconstructed-true energy mapping at the 5 GeV cutoff.
- Weighting the energy spectrum, so it looks flat during training, and the LSTM doesn't become biased toward the most probable value
- To reduce sensitivity to calibration systematic uncertainties, noise was added to the calorimetric energies during the training process, where the noise for each event was uniformly picked from $\{-20\%, 0, +20\%\}$.

There are many more details about the development and training of the LSTM energy estimator in the technical note, reference [112].

There have been a number of robustness studies on the performance and susceptibility to systematic uncertainties of the LSTM energy estimator [118]. Since prong length is one of the key inputs it was studied directly by scaling the individual prong lengths up or down by 5% and seeing how the output energy changed. From this study, it was found that the LSTM is only directly sensitive to the muon prong length, where the ±5% shifts in muon prong length resulted in ±0.8% shifts in neutrino energy, all of the other prong length shifts resulted in negligible changes in the neutrino energy.

The CNN-based method

There is also a CNN based energy estimation method adapted for ν_μ energy as there is in the ν_e case. The hadronic energy is traditionally the bottleneck of ν_μ energy estimation, and the long muon track can sometimes leave the pixelmap, so this CNN estimates the hadronic



(a) True Hadronic Energy (GeV)

(b) True Hadronic Energy (GeV)

Figure 3.6: Unweighted and weighted true hadronic energy spectrum for FHC (left) and RHC (right) samples used for CNN training

energy component instead of the neutrino energy. This hadronic energy estimate can then be added to the spline-based muon energy EE described above to be an estimate of the neutrino energy.

The architecture used is based on GoogLeNet [110], adapted to take 2 input images, one for each view, each 80 planes by 100 cells [79], same inputs as the ν_e CNN [37]. A ResNet based architecture was also trained with nearly identical results, but the GoogLetNet based architecture was chosen for a speedup in evaluation speed. FHC and RHC models were trained separately. The CNNs are trained using Keras with Tensorflow version 1.12 [1]. The models were trained using GPU resources at the Wilson Cluster at Fermilab. The training optimizer used was Adam [85]. The loss function used is mean loss error, intended to give less weight to outliers.

The datasets used for training were nonswap miniproduction 5 FD FHC and RHC HDF5 files, available on the Wilson Cluster. For cuts, true ν_μ CC events were selected, and analysis ν_μ selection was applied. After cuts, there were ~800,000 events for training, and ~90,000 for evaluation. The performance was tested on ~280,000 events. A weighting scheme was devised to make the network focus less on the peak in the energy spectrum. Trying to make the spectrum totally flat proved unsuccessful, with the network's performance becoming worst in the most populated region from 0-2 GeV. A less aggressive scheme that was settled on is shown in figure 3.6. Weights were implemented in training using the sample weights functionality in Keras.

The LSTM showed the strength of training with calibration robustness augmentation, so for future NOvA analyses, the CNN is being retrained with a similar calibration augmentation scheme. For this CNN training, the scale of the pixelmap is scaled by a value sampled from a normal distribution with mean 1 and standard deviation 0.1 (10 %). Another future improvement for the CNN is that the target will be the neutrino energy itself, instead of the hadronic energy, as biases can arise from differences in the muon and hadronic energy. The CNN is then given the muon track length as an extra input to make up for the scenario where the muon track is long enough to leave the pixelmap. The CNN is then weighted to make the neutrino energy spectrum flat. Early results show these updates have promising effects on performance [80].

3.3 Energy Estimator Studies for 2024 Analysis

3.3.1 ν_e Energy

For 2024, there were two methods considered for use in the analysis for ν_e energy reconstruction: the traditionally used calorimetric method, as well as the CNN. The systematics are expected to be comparable, but the performance of the CNN is better overall on the nominal MC, as shown in figure 3.7 for FHC.

It is unclear how much an overall improvement in performance will translate into better sensitivity for the ν_e sample, however, as the binning of the ν_e is fairly low, see section 4.1.4. To see the improvements we can get from this CNN based method, we studied the effects

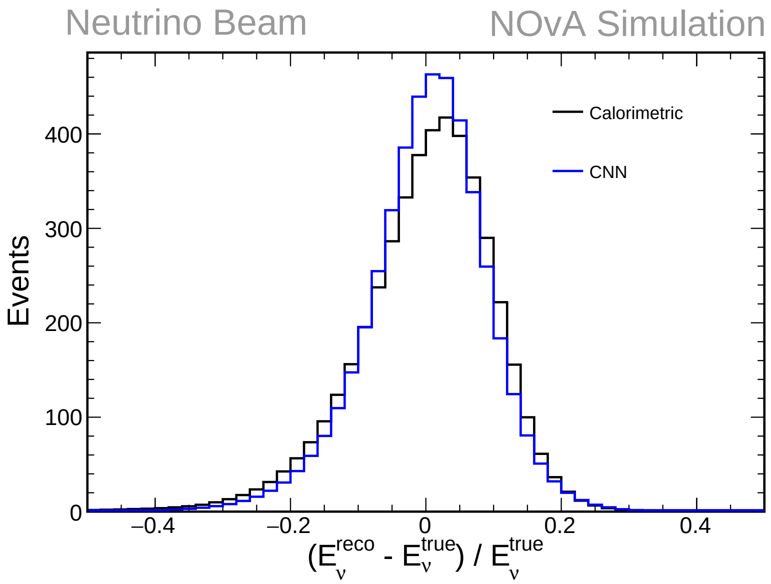


Figure 3.7: The ν_e energy resolution for the calorimetric and CNN method, for FHC

on the final analysis sensitivities using fake data. Figures 3.8 and 3.9 show the effects of using the CNN over the calorimetric method traditionally used with statistics only, where every other part of the analysis is the same as the 2020 analysis [12]. The figures show an improvement with the CNN, but only a very marginal one.

Given how little the improvements in performance help our overall sensitivity to the oscillation parameters, it was chosen to stay with the calorimetric method for this analysis. The calorimetric method is a conservative choice, as its performance and dependence on systematic uncertainties is well understood.

3.3.2 ν_μ Energy

Comparing ν_μ EE performance - energy resolution

In the following section, we'll review the many studies performed to choose the ν_μ EE method to use for this analysis, specifically of overall performance on nominal MC, the robustness

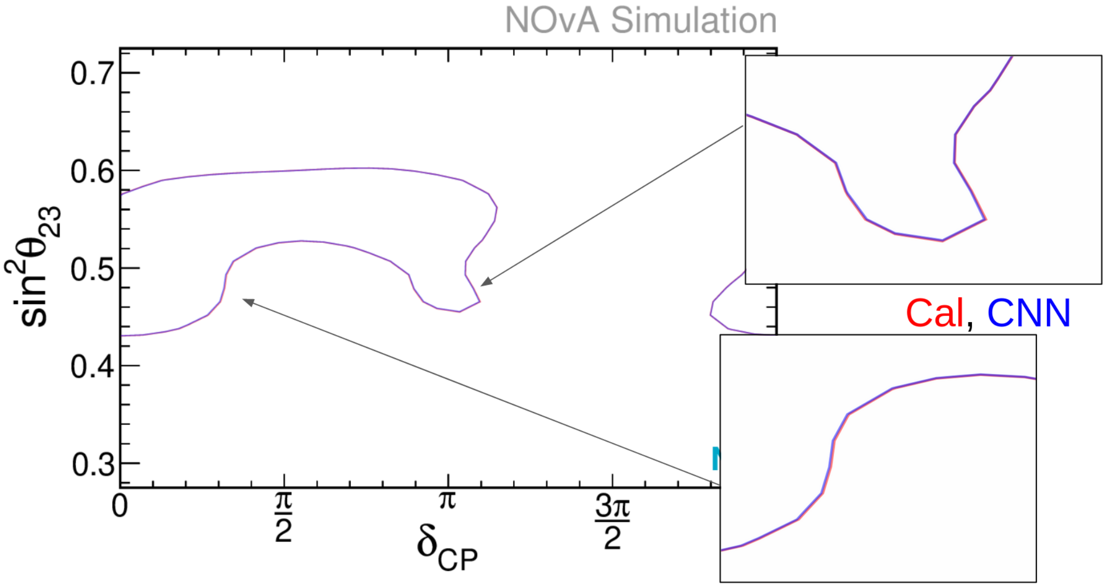


Figure 3.8: Statistics only contours for δ_{CP} vs $sin^2\theta_{23}$ comparing the sensitivity of the calorimetric and CNN based methods, NO

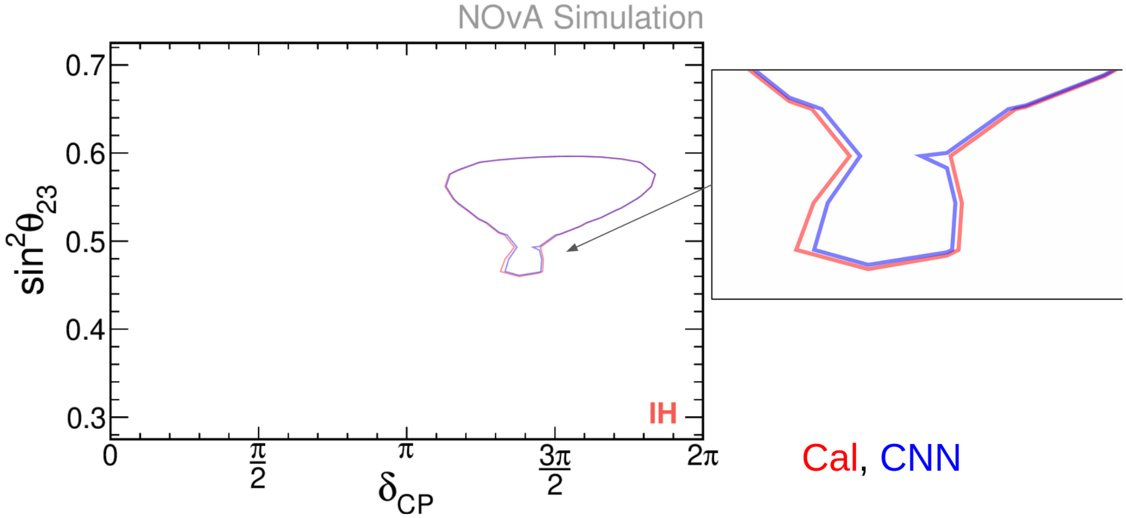


Figure 3.9: Statistics only contours for δ_{CP} vs $sin^2\theta_{23}$ comparing the sensitivity of the calorimetric and CNN based methods, IO

to systematic uncertainties, and the corresponding analysis sensitivities [82].

Figure 3.10 and figure 3.13 show the energy spectrum and overall resolution ($(recoE - trueE)/trueE$) of the EEs on the nominal FD samples, for FHC and RHC, respectively. Figure 3.11 and figure 3.14 show the mean and standard deviation of the resolution broken up by 5 bins of true neutrino energy. The overall mean and standard deviation for each are listed in table 3.2.

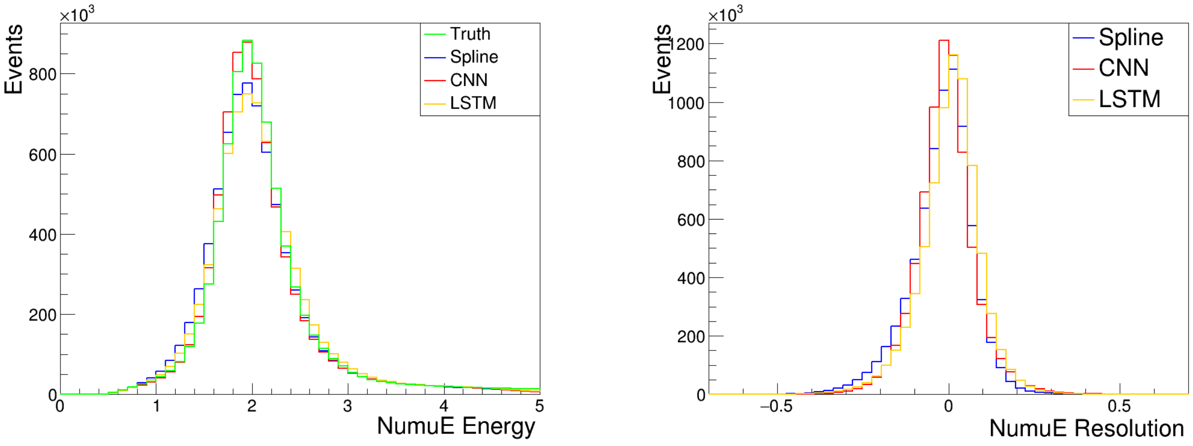
The cut applied for the following performance plots is 2020 analysis cut, with the cuts on energy removed, which is the same cut used for the spline training. The weights used are the production 5.1 cross-section weights, and the standard flux weights.

In terms of overall performance, both deep-learning methods see an improvement over the traditional, spline-based method. From best to worst for resolution the order is the CNN, the LSTM, and then the spline, for both FHC and RHC. For bias, the CNN is the best for FHC and the LSTM for RHC.

Looking at this performance broken up into bins of true neutrino energy for FHC (figure 3.11a), we can see the energy dependence of the mean follows a similar trend for all three EEs, starting positive and becoming negative at higher energies. RHC (figure 3.14a) follows a similar trend, except for the first bin of the CNN which is near 0.

For the standard deviation, we also see similar trends for FHC and RHC (figures 3.11b and 3.14b). For the lowest bin, the LSTM has the lowest, then the spline, then CNN. From the two bins from 1-3 GeV, the CNN has the lowest, followed by the LSTM, then spline. From 3-4 GeV, the LSTM does the best, followed by the CNN, then spline. For the highest bin, the CNN and LSTM are roughly comparable.

Energy resolutions are also shown for the muon and hadronic energy individually (figures 3.12 and 3.15). The spline-based method has spline EEs for both muon and hadronic energy



(a) Energy Spectrum (b) Energy Resolution

Figure 3.10: FHC ν_μ energy Spectra and Resolution

FHC Total Performance			
	Spline	LSTM	CNN
Mean	-0.0198	0.0051	-0.0071
Std	0.0948	0.0901	0.0875

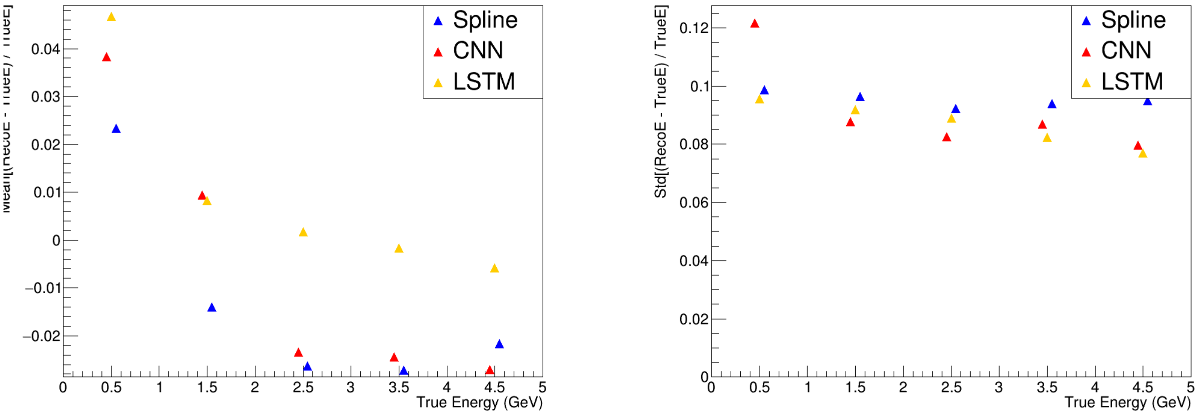
RHC Total Performance			
	Spline	LSTM	CNN
Mean	-0.0115	0.0097	-0.0086
Std	0.0802	0.0784	0.0747

Table 3.2: ν_μ Energy Estimation Performance Statistics

individually, while the CNN has an output only for the hadronic energy, and uses the muon spline EE. Note that's why there's no "CNN" for the muon energy resolutions. The LSTM outputs estimates for the muon and neutrino energy, individually. The hadronic energy for the LSTM is defined as the neutrino energy minus the muon energy.

The energy spectra for the 3 methods are show in figures 3.16 - 3.21 broken up by hadronic energy fraction quartile. Data and MC are shown together. Noteworthy features are the weird distributions for the LSTM, particularly in the first quartiles. The most worrisome of these is figure 3.20a. Note the jagged V shape right at the peak around 2 GeV.

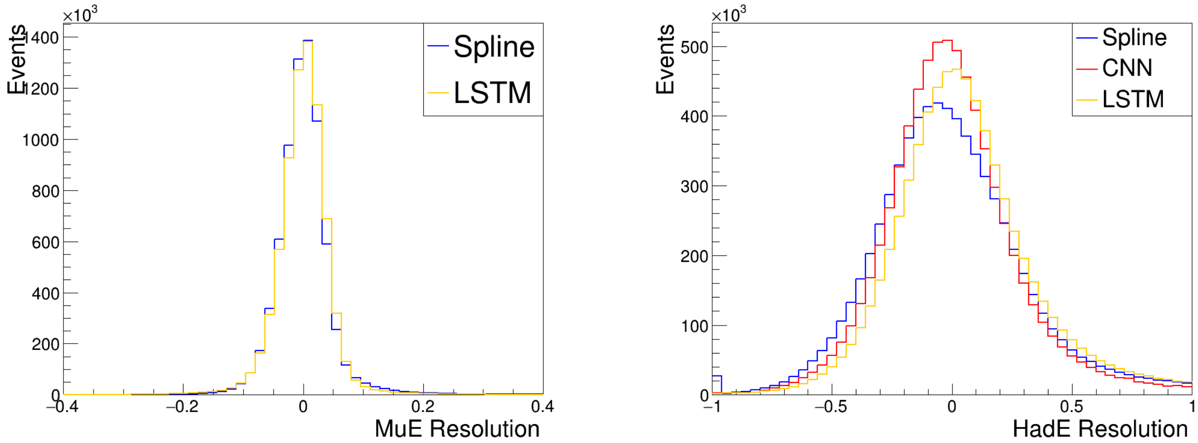
The CNN also has nonphysical features, particularly in the fourth quartile. Figure 3.21d shows a much narrower distribution in the fourth quartile than the other three. These features are also present for FHC. The narrowing of energy at higher hadronic energy fraction is not the expected behavior from calorimetric considerations and may be a bias from the



(a) Mean of resolution by true energy

(b) Std of resolution by true energy

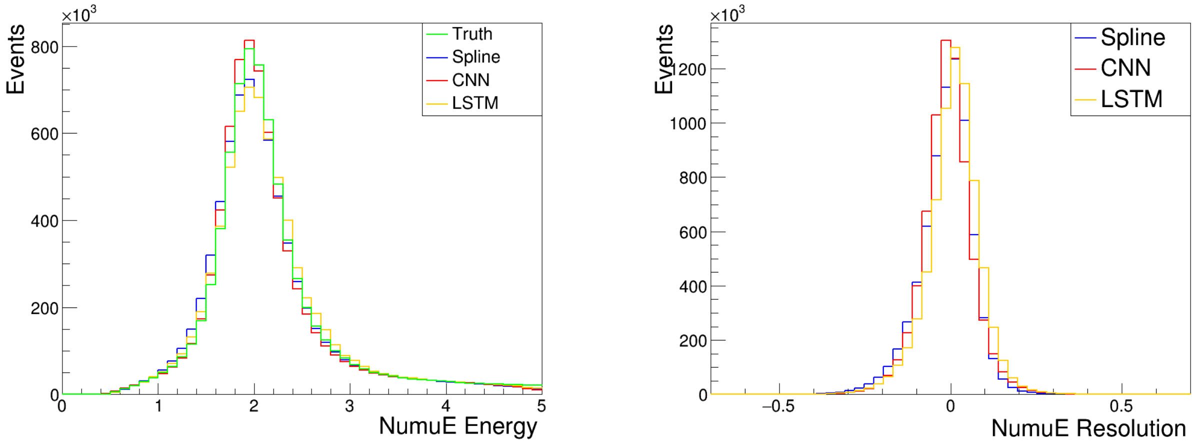
Figure 3.11: FHC ν_μ energy performance binned by true Energy



(a) FHC ν_μ Muon Energy Resolution

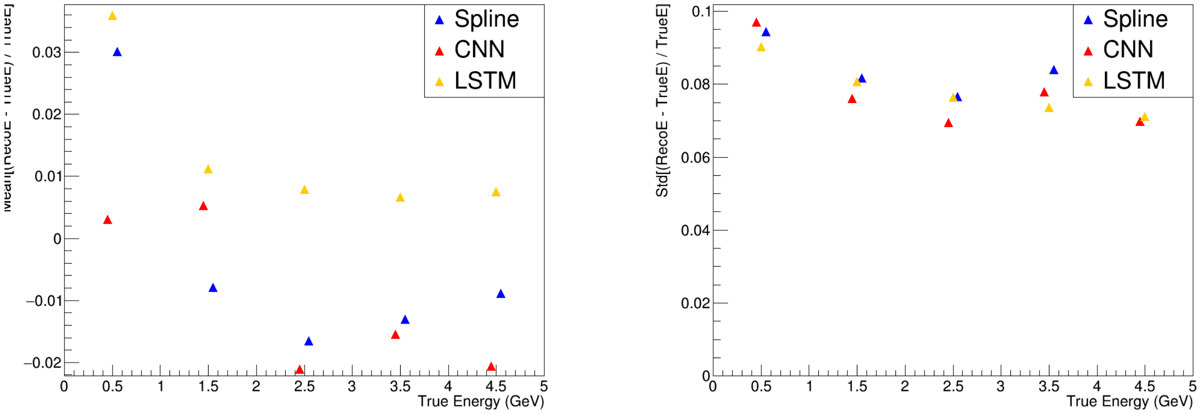
(b) FHC ν_μ Hadronic Energy Resolution

Figure 3.12: FHC ν_μ Muon and Hadronic Energy Resolution



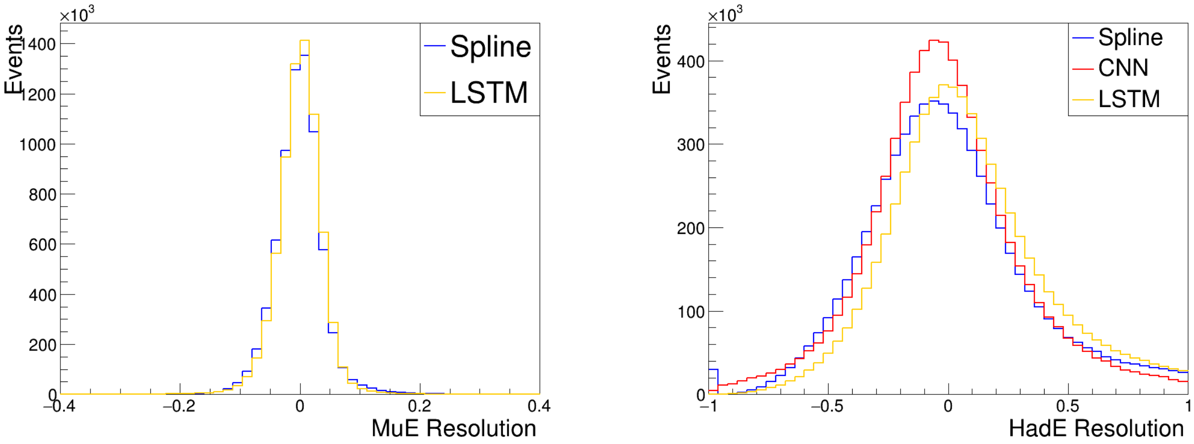
(a) Energy Spectrum (b) Energy Resolution

Figure 3.13: RHC ν_μ energy Spectra and Resolution



(a) Mean of resolution by true energy (b) Std of Resolution By True Energy

Figure 3.14: RHC ν_μ energy performance binned by true energy



(a) RHC ν_μ Muon Energy Resolution (b) RHC ν_μ Hadronic Energy Resolution

Figure 3.15: RHC ν_μ Muon and Hadronic Energy Resolution

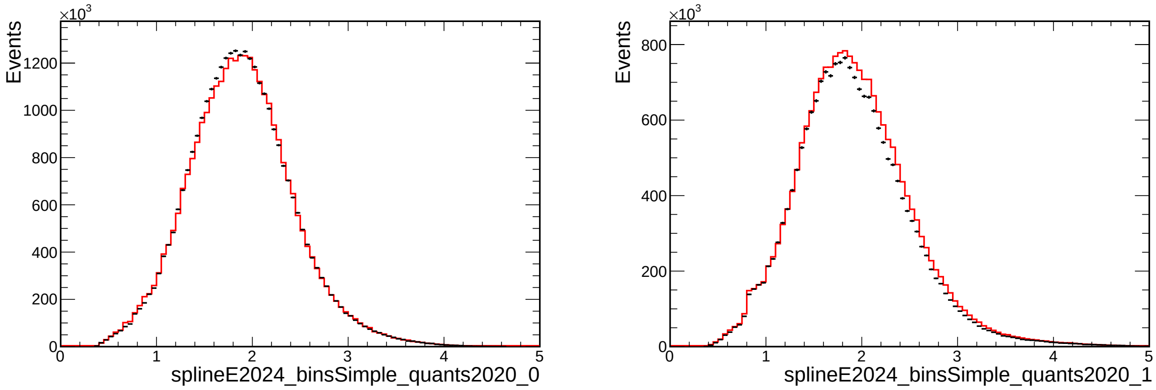
training of the algorithm.

Comparing ν_μ EE performance - systematics

The effects of systematics were also explored for the three EEs. The shifts in resolution from detector systematics for calibration (figures 3.22 and 3.23), Cherenkov (figures 3.28 and 3.29), and light levels (figures 3.26 and 3.27) are shown. For the largest of these, calibration, shifts in resolution statistics, shown as error bars, are broken up by ranges of true energy for mean and standard deviation (figures 3.24 and 3.25 for FHC and RHC, respectively).

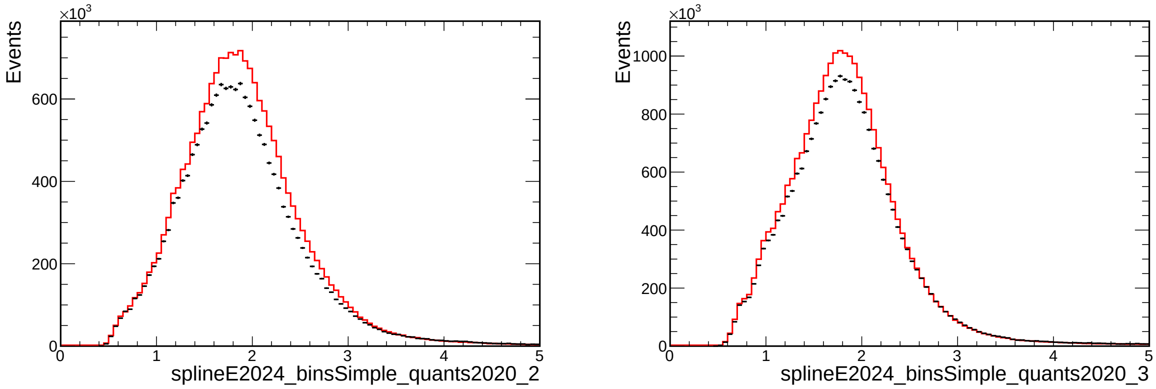
In addition to these detector systematic uncertainties, cross-section systematic uncertainties were tested together using a multiverse with 1000 universes. The corresponding shifts from these are shown in figure 3.30 for FHC and figure 3.31 for RHC.

The shifts in statistics from these detector systematic uncertainties as well as the shifts from the cross-section systematic uncertainties are summarized in Table 3.3 for FHC and Table 3.4 for RHC. In general, the LSTM sees the smallest shift in bias for calibration for both FHC and RHC. It also has a smaller shift in bias for light levels in FHC, where the size of



(a) Hadronic Fraction Quartile 1

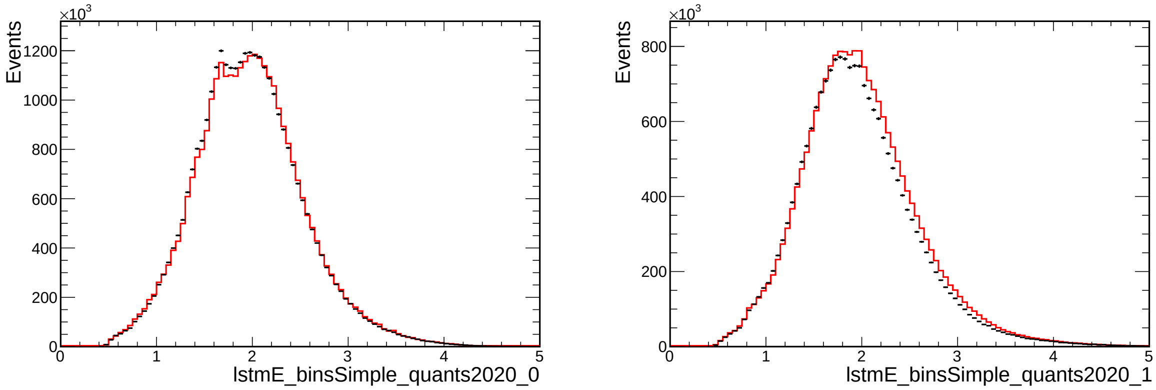
(b) Hadronic Fraction Quartile 2



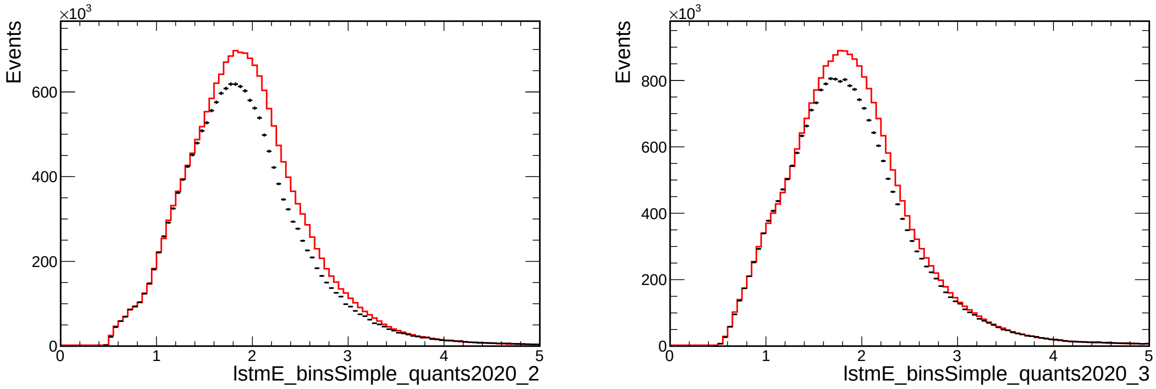
(c) Hadronic Fraction Quartile 3

(d) Hadronic Fraction Quartile 4

Figure 3.16: Spline ND FHC ν_μ Energy Spectra Broken Up By Hadronic Energy Fraction Quartiles

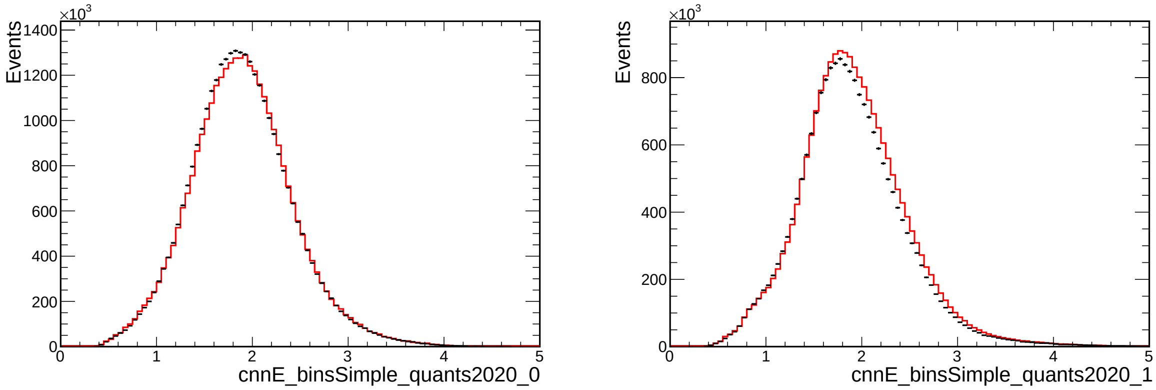


(a) Hadronic Fraction Quartile 1 (b) Hadronic Fraction Quartile 2



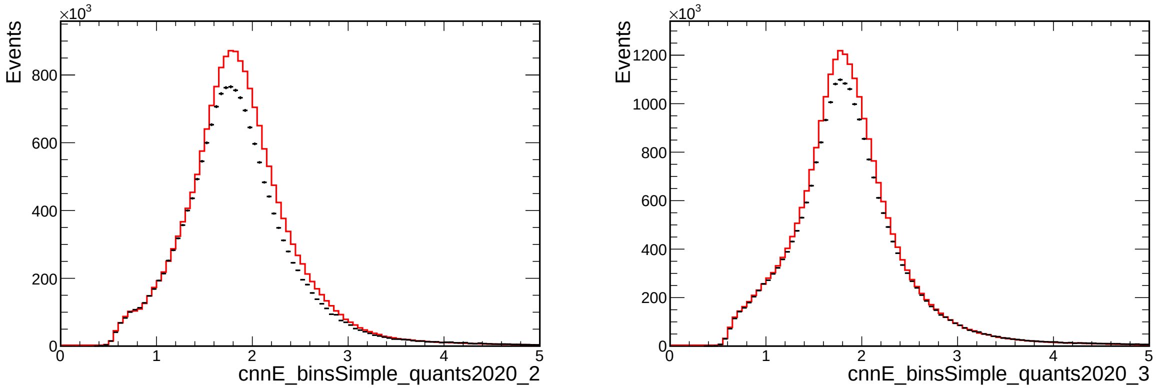
(c) Hadronic Fraction Quartile 3 (d) Hadronic Fraction Quartile 4

Figure 3.17: LSTM ND FHC ν_μ Energy Spectra Broken Up By Hadronic Energy Fraction Quartiles



(a) Hadronic Fraction Quartile 1

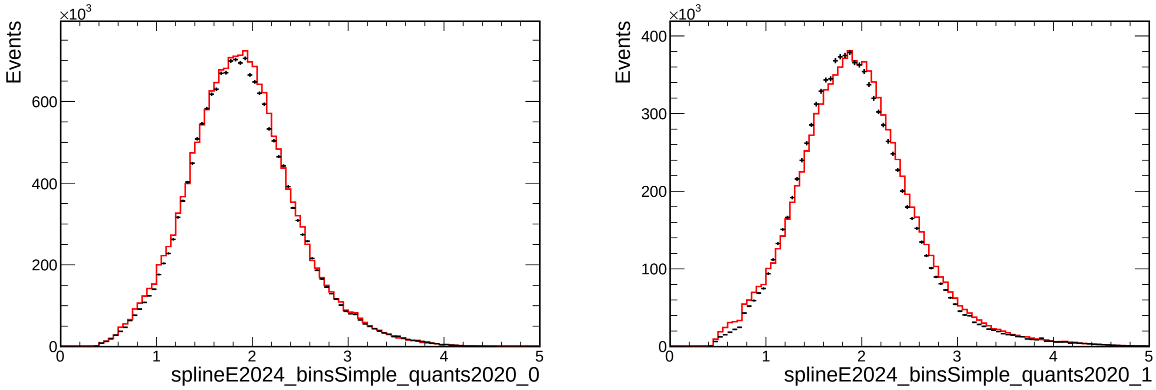
(b) Hadronic Fraction Quartile 2



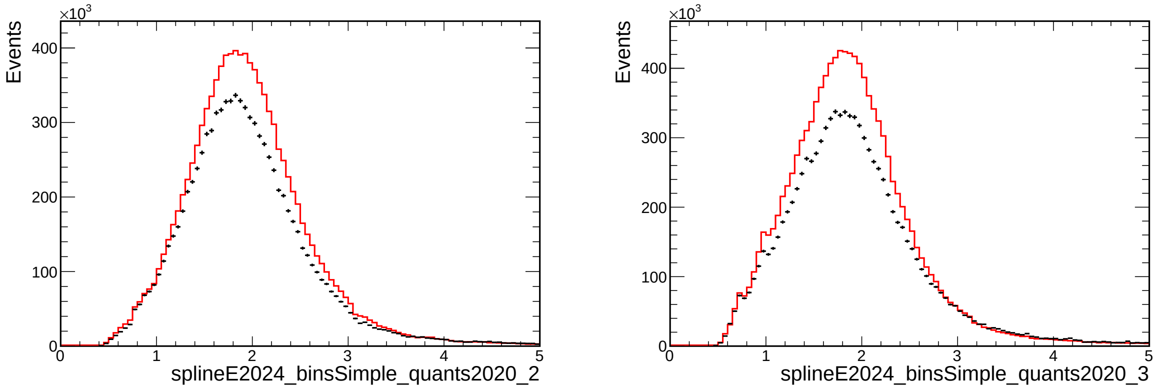
(c) Hadronic Fraction Quartile 3

(d) Hadronic Fraction Quartile 4

Figure 3.18: CNN ND FHC ν_μ Energy Spectra Broken Up By Hadronic Energy Fraction Quartiles

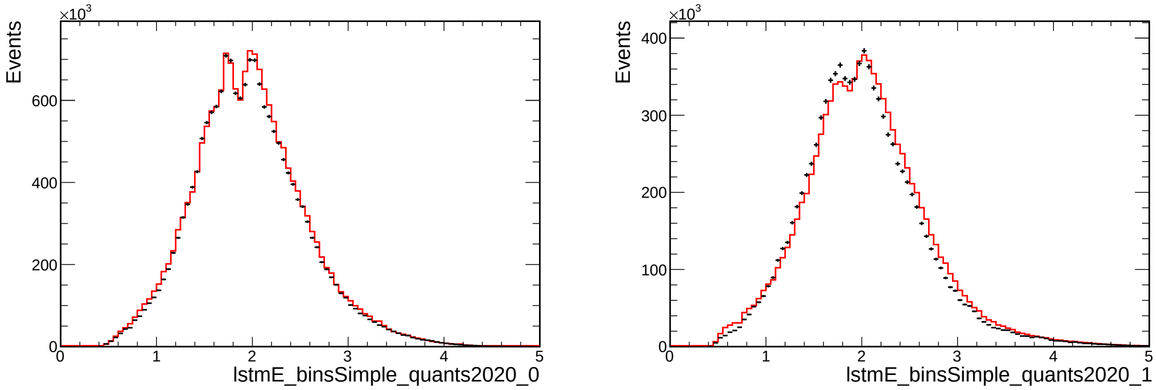


(a) Hadronic Fraction Quartile 1 (b) Hadronic Fraction Quartile 2

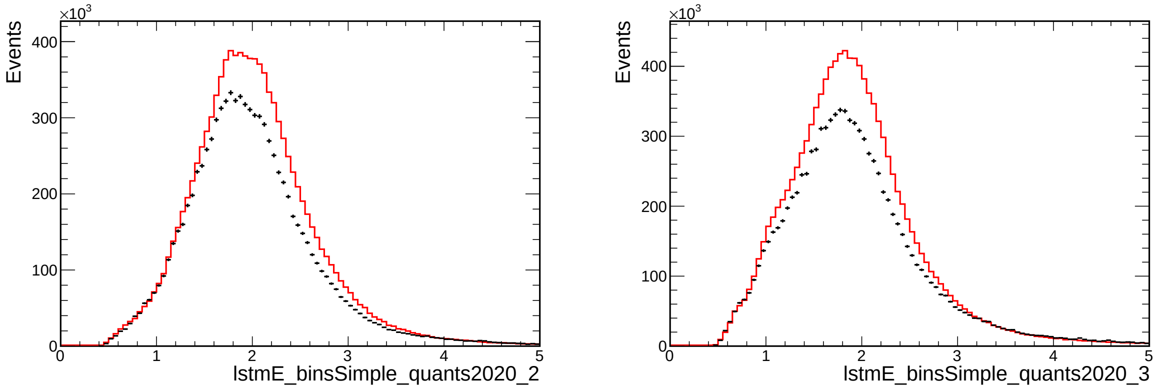


(c) Hadronic Fraction Quartile 3 (d) Hadronic Fraction Quartile 4

Figure 3.19: Spline ND RHC ν_μ Energy Spectra Broken Up By Hadronic Energy Fraction Quartiles

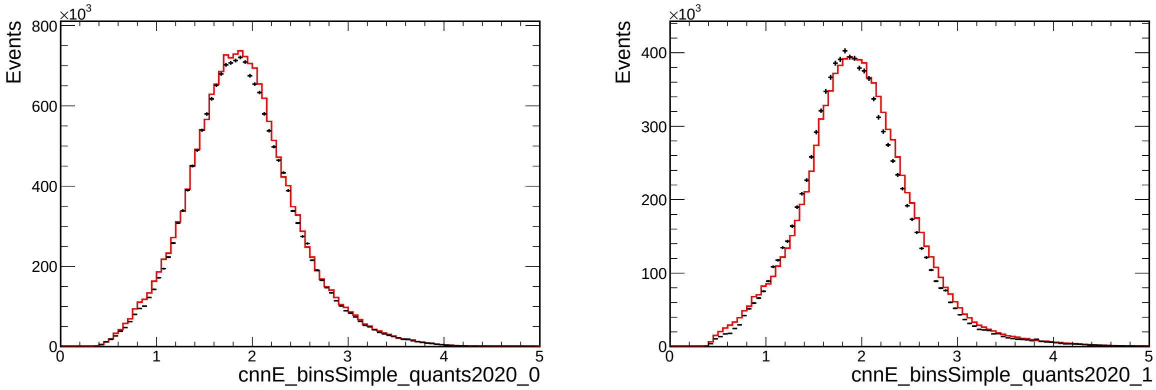


(a) Hadronic Fraction Quartile 1 (b) Hadronic Fraction Quartile 2

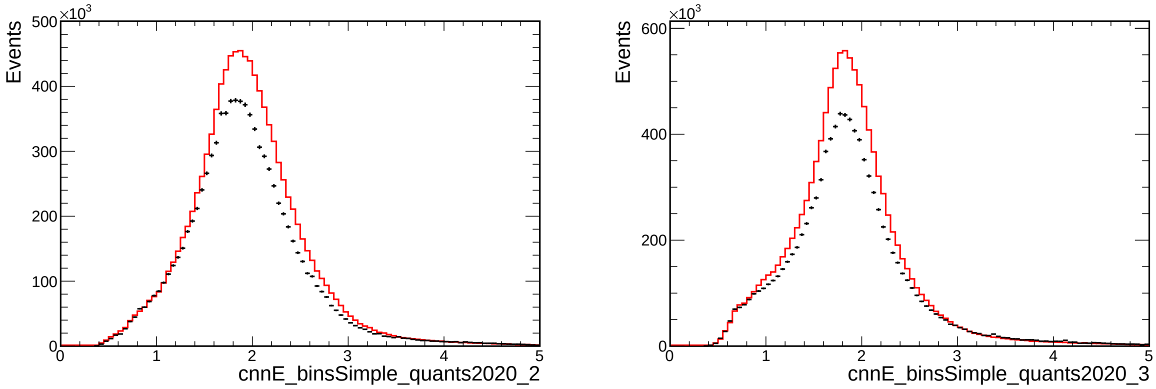


(c) Hadronic Fraction Quartile 3 (d) Hadronic Fraction Quartile 4

Figure 3.20: LSTM ND RHC ν_μ Energy Spectra Broken Up By Hadronic Energy Fraction Quartiles



(a) Hadronic Fraction Quartile 1 (b) Hadronic Fraction Quartile 2



(c) Hadronic Fraction Quartile 3 (d) Hadronic Fraction Quartile 4

Figure 3.21: CNN ND RHC ν_μ Energy Spectra Broken Up By Hadronic Energy Fraction Quartiles

FHC Mean Sensitivities			
	Spline	LSTM	CNN
Calibration	0.0336	0.0080	0.0302
Light Levels	0.0338	0.0082	0.0304
Cherenkov	0.0014	0.0006	0.0017
XSec Systs	0.0031	0.0020	0.0002

FHC Std Sensitivities			
	Spline	LSTM	CNN
Calibration	0.0013	0.0023	0.0036
Light Levels	0.0015	0.0019	0.0038
Cherenkov	0.0001	0.0001	0.0002
XSec Systs	0.0022	0.0022	0.0024

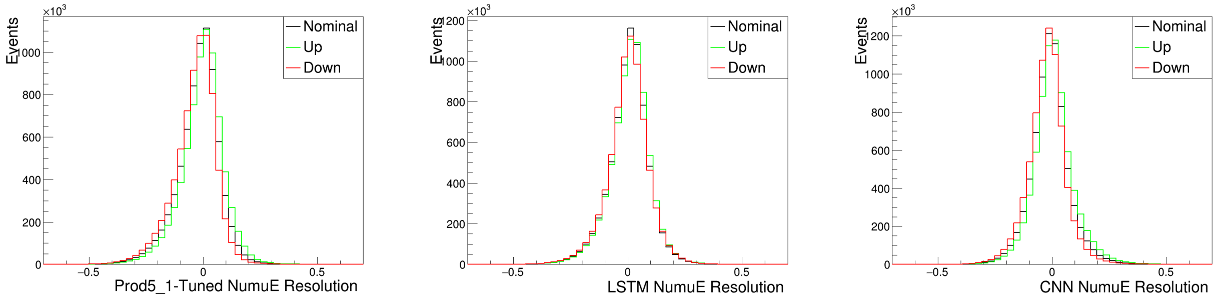
Table 3.3: ν_μ Energy Estimation Sensitivity for mean and standard deviation to systematic uncertainties for FHC

shifts is comparable to the calibration systematic.

Also, considered is the performance of these on the MENATE samples, which is an alternative neutron model [59]. The shifts in overall resolution distributions, comparing the nominal and MENATE samples, are shown for FHC in figure 3.32 and for RHC in figure 3.33. The shifts in the overall statistics, MENATE - nominal, are tabulated in 3.5. In general, there is an improvement in standard deviation for the MENATE sample compared to the nominal sample for FHC, and a small degradation in standard deviation for RHC. As for mean, the LSTM has the largest shift in mean (by magnitude), followed by the CNN, then spline. This is true for both FHC and RHC.

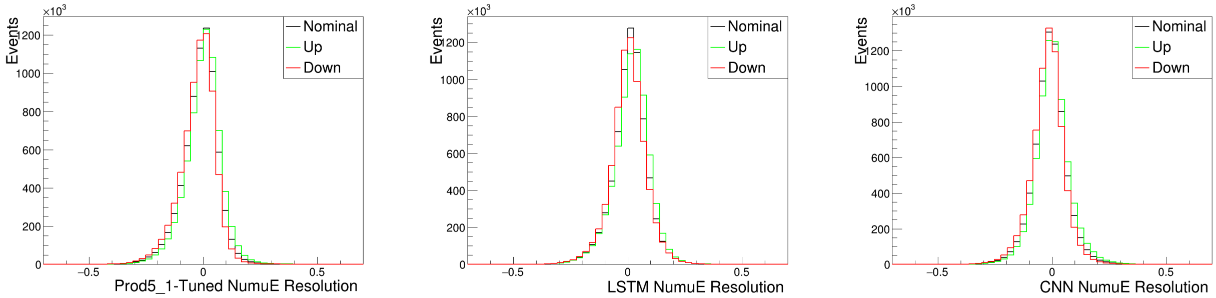
Lastly, the Geant4Reweight reweighting was checked for the 3 EEs. The technical note for this method is available as reference [109], which implements the Geant4Reweight [40] scheme to NOvA. The FHC and RHC shifts are available in figures 3.34 and 3.35, respectively. Compared to the other systematics considered, these have very small effects. Overall, the sizes are roughly comparable among the 3, with the spline having the largest for `kProtonTotSyst`, which is the largest of the list available. Also, the π^- shifts are larger for FHC, and π^+ for RHC, which serves as a sanity check that the Geant4Reweight method is giving physical results.

More about these systematic uncertainties themselves, as well as the others not discussed here, can be found in section 4.3.



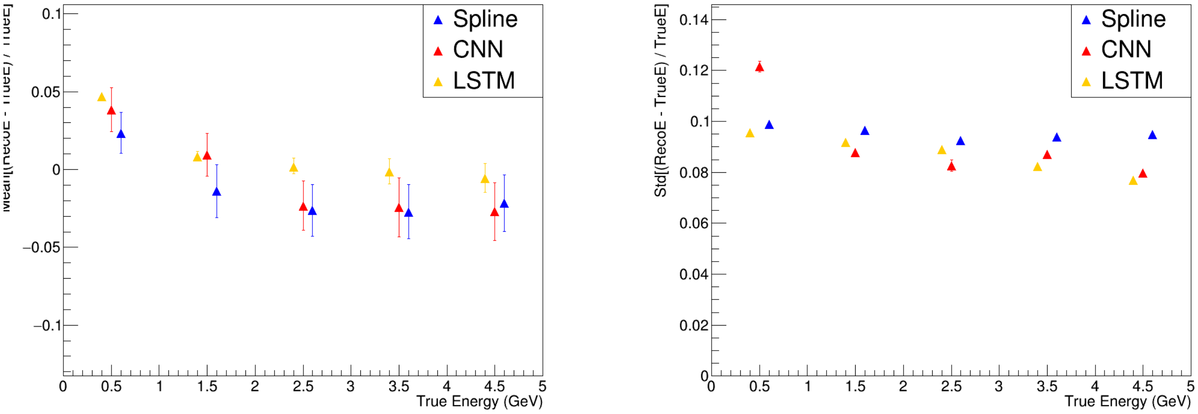
(a) Spline (b) LSTM (c) CNN

Figure 3.22: FHC ν_μ energy performance with calibration systematic shifts



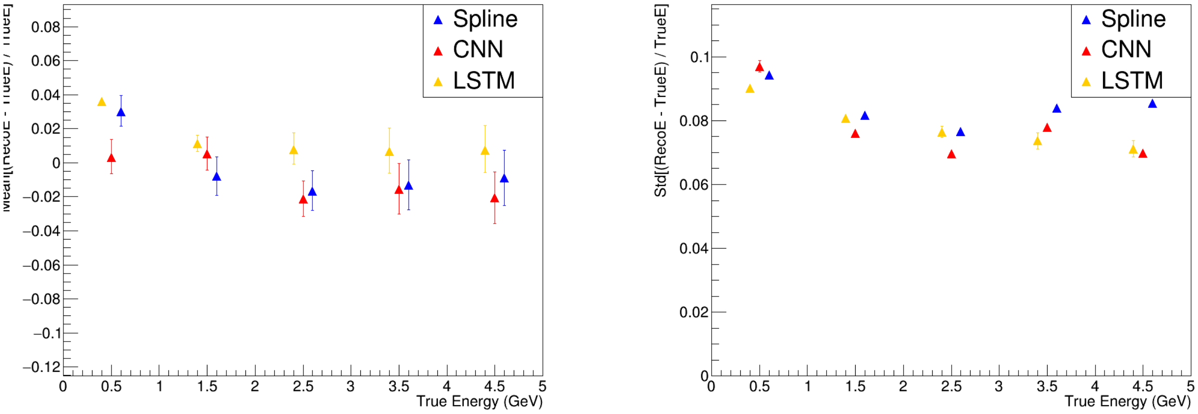
(a) Spline (b) LSTM (c) CNN

Figure 3.23: RHC ν_μ energy performance with calibration systematic shifts



(a) Mean of Resolution By True Energy (b) Std of Resolution By True Energy

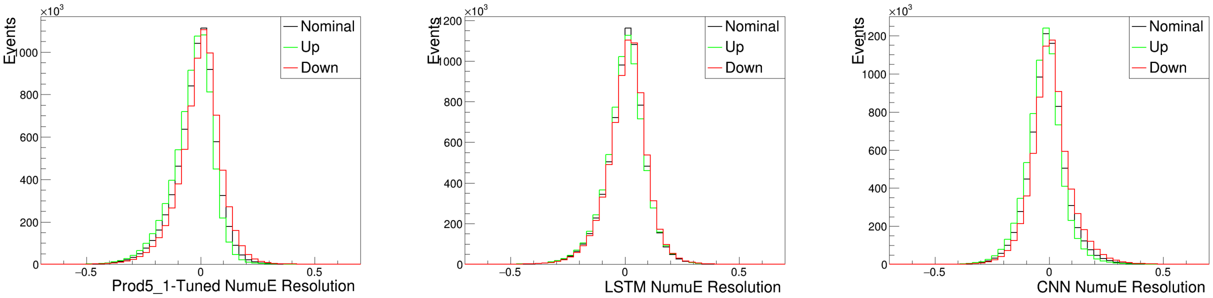
Figure 3.24: FHC ν_μ energy performance Binned By true energy, where error bars are calibration systematics



(a) Mean of Resolution By True Energy

(b) Std of Resolution By True Energy

Figure 3.25: RHC ν_μ energy performance binned by true energy, where error bars are calibration systematics

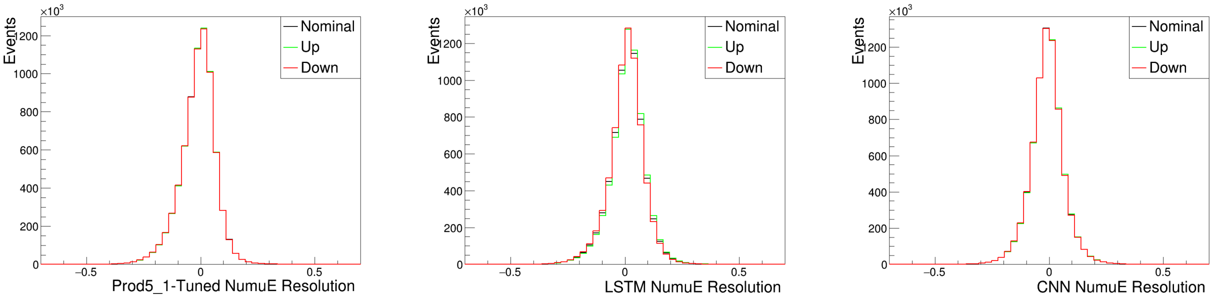


(a) Spline

(b) LSTM

(c) CNN

Figure 3.26: FHC ν_μ energy performance with light level systematic shifts

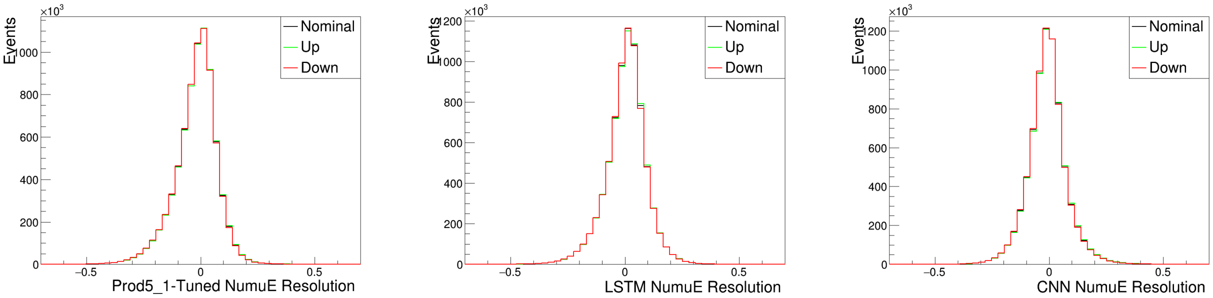


(a) Spline

(b) LSTM

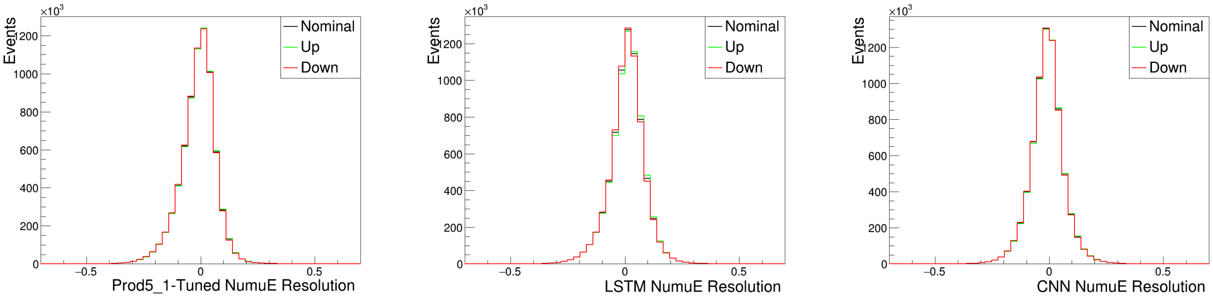
(c) CNN

Figure 3.27: RHC ν_μ energy performance with light level systematic shifts



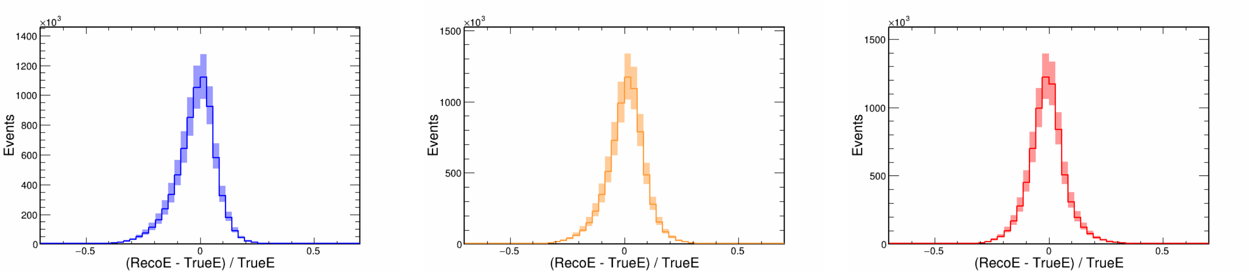
(a) Spline (b) LSTM (c) CNN

Figure 3.28: FHC ν_μ energy performance with cherenkov systematic shifts



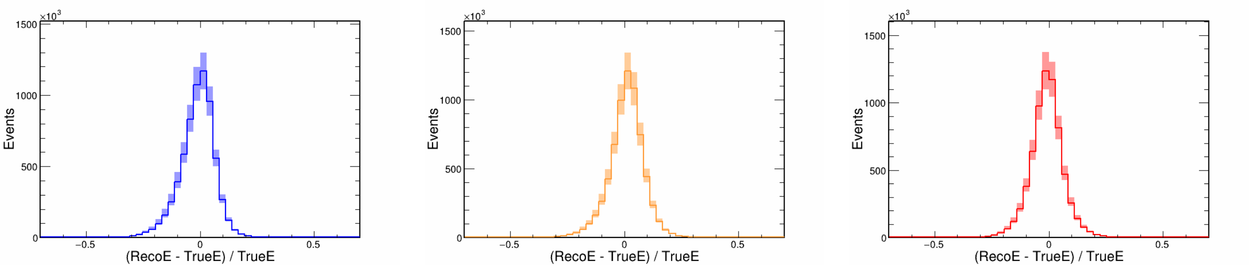
(a) Spline (b) LSTM (c) CNN

Figure 3.29: RHC ν_μ energy performance with cherenkov systematic shifts



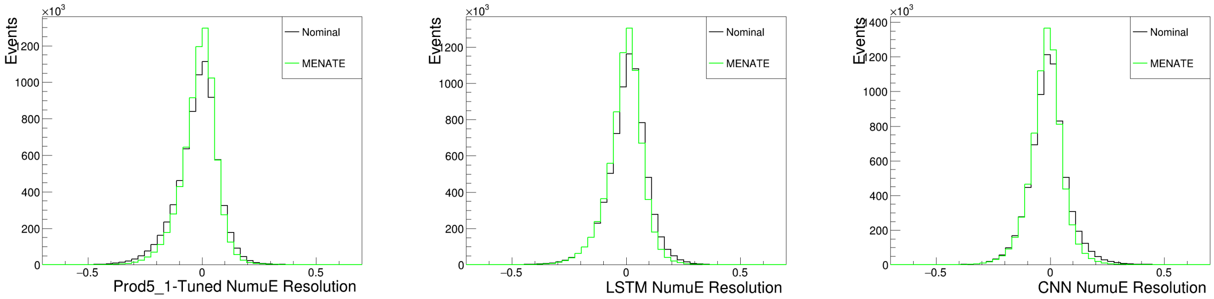
(a) Spline (b) LSTM (c) CNN

Figure 3.30: FHC ν_μ energy performance with cherenkov systematic shifts



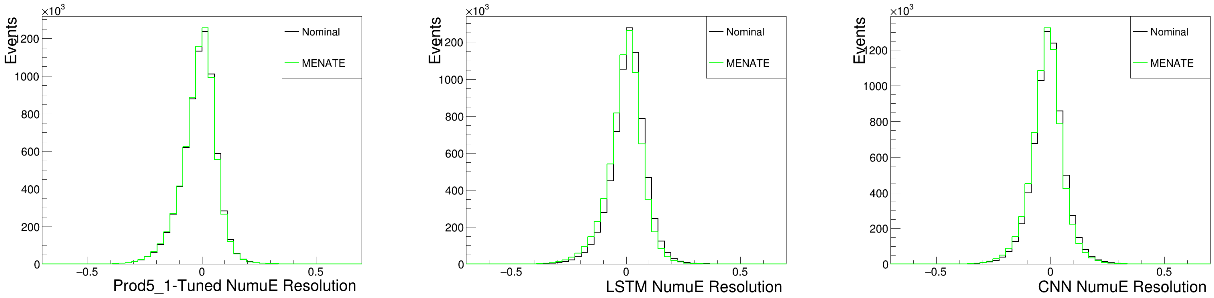
(a) Spline (b) LSTM (c) CNN

Figure 3.31: RHC ν_μ energy performance with XSec systematic shifts



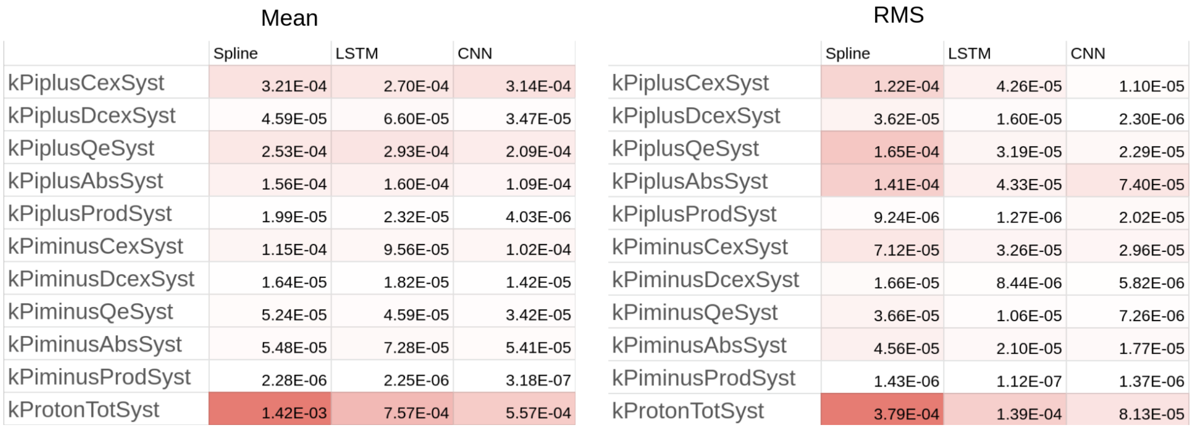
(a) Spline (b) LSTM (c) CNN

Figure 3.32: FHC ν_μ energy MENATE samples compared to nominal samples



(a) Spline (b) LSTM (c) CNN

Figure 3.33: RHC ν_μ energy MENATE samples compared to nominal samples



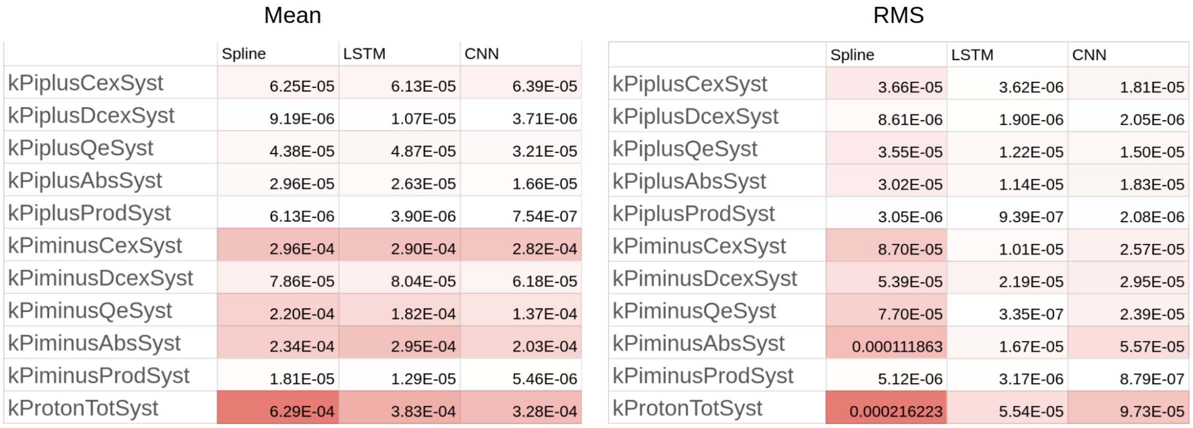
Mean

	Spline	LSTM	CNN
kPiplusCexSyst	3.21E-04	2.70E-04	3.14E-04
kPiplusDcexSyst	4.59E-05	6.60E-05	3.47E-05
kPiplusQeSyst	2.53E-04	2.93E-04	2.09E-04
kPiplusAbsSyst	1.56E-04	1.60E-04	1.09E-04
kPiplusProdSyst	1.99E-05	2.32E-05	4.03E-06
kPiminusCexSyst	1.15E-04	9.56E-05	1.02E-04
kPiminusDcexSyst	1.64E-05	1.82E-05	1.42E-05
kPiminusQeSyst	5.24E-05	4.59E-05	3.42E-05
kPiminusAbsSyst	5.48E-05	7.28E-05	5.41E-05
kPiminusProdSyst	2.28E-06	2.25E-06	3.18E-07
kProtonTotSyst	1.42E-03	7.57E-04	5.57E-04

RMS

	Spline	LSTM	CNN
kPiplusCexSyst	1.22E-04	4.26E-05	1.10E-05
kPiplusDcexSyst	3.62E-05	1.60E-05	2.30E-06
kPiplusQeSyst	1.65E-04	3.19E-05	2.29E-05
kPiplusAbsSyst	1.41E-04	4.33E-05	7.40E-05
kPiplusProdSyst	9.24E-06	1.27E-06	2.02E-05
kPiminusCexSyst	7.12E-05	3.26E-05	2.96E-05
kPiminusDcexSyst	1.66E-05	8.44E-06	5.82E-06
kPiminusQeSyst	3.66E-05	1.06E-05	7.26E-06
kPiminusAbsSyst	4.56E-05	2.10E-05	1.77E-05
kPiminusProdSyst	1.43E-06	1.12E-07	1.37E-06
kProtonTotSyst	3.79E-04	1.39E-04	8.13E-05

Figure 3.34: FHC ν_μ energy G4Rwgt shifts



Mean

	Spline	LSTM	CNN
kPiplusCexSyst	6.25E-05	6.13E-05	6.39E-05
kPiplusDcexSyst	9.19E-06	1.07E-05	3.71E-06
kPiplusQeSyst	4.38E-05	4.87E-05	3.21E-05
kPiplusAbsSyst	2.96E-05	2.63E-05	1.66E-05
kPiplusProdSyst	6.13E-06	3.90E-06	7.54E-07
kPiminusCexSyst	2.96E-04	2.90E-04	2.82E-04
kPiminusDcexSyst	7.86E-05	8.04E-05	6.18E-05
kPiminusQeSyst	2.20E-04	1.82E-04	1.37E-04
kPiminusAbsSyst	2.34E-04	2.95E-04	2.03E-04
kPiminusProdSyst	1.81E-05	1.29E-05	5.46E-06
kProtonTotSyst	6.29E-04	3.83E-04	3.28E-04

RMS

	Spline	LSTM	CNN
kPiplusCexSyst	3.66E-05	3.62E-06	1.81E-05
kPiplusDcexSyst	8.61E-06	1.90E-06	2.05E-06
kPiplusQeSyst	3.55E-05	1.22E-05	1.50E-05
kPiplusAbsSyst	3.02E-05	1.14E-05	1.83E-05
kPiplusProdSyst	3.05E-06	9.39E-07	2.08E-06
kPiminusCexSyst	8.70E-05	1.01E-05	2.57E-05
kPiminusDcexSyst	5.39E-05	2.19E-05	2.95E-05
kPiminusQeSyst	7.70E-05	3.35E-07	2.39E-05
kPiminusAbsSyst	0.000111863	1.67E-05	5.57E-05
kPiminusProdSyst	5.12E-06	3.17E-06	8.79E-07
kProtonTotSyst	0.000216223	5.54E-05	9.73E-05

Figure 3.35: RHC ν_μ energy G4Rwgt shifts

RHC Mean Sensitivities

	Spline	LSTM	CNN
Calibration	0.0236	0.0150	0.0211
Light Levels	0.0004	0.0056	0.0008
Cherenkov	0.0013	0.0019	0.0013
XSec Systs	0.0028	0.0015	0.0010

RHC Std Sensitivities

	Spline	LSTM	CNN
Calibration	0.0008	0.0026	0.0023
Light Levels	0.0002	0.0000	0.0001
Cherenkov	0.0002	0.0002	0.0002
XSec Systs	0.0025	0.0024	0.0022

Table 3.4: ν_μ Energy Estimation Sensitivity for mean and standard deviation to systematic uncertainties for RHC

FHC MENATE Shifts (MENATE - Nominal)

	Spline	LSTM	CNN
Mean	0.0064	-0.0099	-0.0097
Std	-0.0144	-0.0095	-0.0127

RHC MENATE Shifts (MENATE - Nominal)

	Spline	LSTM	CNN
Mean	-0.0018	-0.0145	-0.0083
Std	0.0002	0.0023	0.0001

Table 3.5: ν_μ Energy Estimation Sensitivity for mean and standard deviation to neutron model

Comparing ν_μ EE performance - sensitivities

Lastly, the sensitivities for $sin^2\theta_{23}$ vs Δm^2_{32} are made with fake data using the 2020 best fit point, shown in figure 3.36 for 3 σ. They are made using the same analysis features as the 2020 NOvA analysis, with the modifications made where necessary to use these new EEs, as well as the new production 5.1 MC. The predictions were made with all the 2020 systematics, except a few that are difficult to run with time constraints or implement to the deep learning methods, such as flux, small GENIE knob PCA, muon energy scale, and angle acceptance systematics.

The 3 EEs are well ordered by figure 3.36, so that from best to worst, they are LSTM, CNN, Spline. The dotted lines show the stats-only contours for comparisons (the LSTM and CNN dotted are practically overlapping). This makes it clear that the main improvements from the LSTM come from its robustness to the systematics.

Another Asimov point tested is what's called point C in the list of Asimov points to test for the 2024 analysis[46], contours shown in figure 3.37. This point is

$$\begin{aligned}\Delta m^2_{32} &= -2.45 \times 10^{-3} eV^2 \\ sin^2\theta_{23} &= 0.50 \\ \delta_{CP} &= 1.5\pi\end{aligned}$$

chosen as the point preferred by global fitters.

Conclusion

The three ν_μ energy estimators (EEs) available in Production 5.1 were compared, and considered for use in the 2024 3-Flavor analysis. The two deep-learning methods have an improved

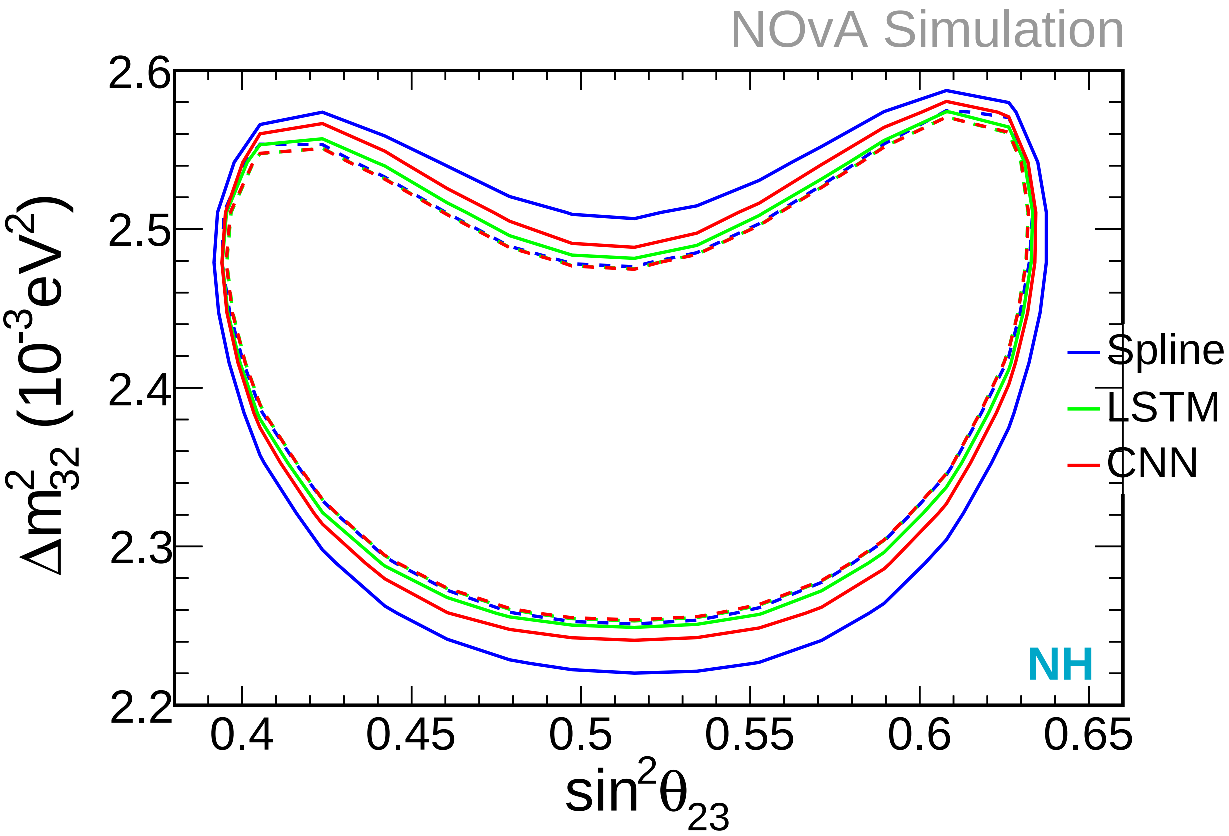


Figure 3.36: Comparison of contours using 2020 best fit point for each of the 3 ν_μ EEs, at 3 sigma. Solid lines are with systematics, dotted lines are stats-only

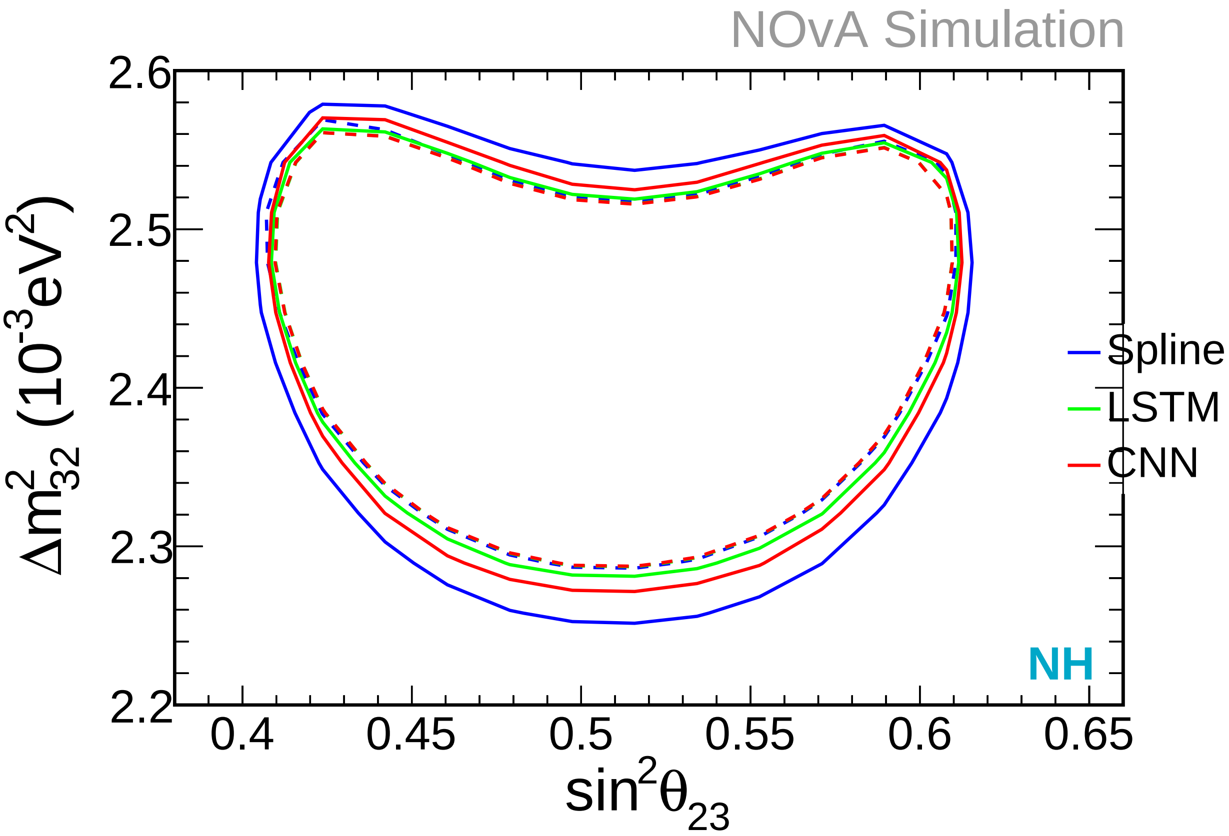


Figure 3.37: Comparison of contours using Asimov point C for each of the 3 ν_μ EEs, at 3 sigma. Solid lines are with systematics, dotted lines are stats-only

performance over the traditional spline-based method which has been used in all prior analyses. The LSTM was trained very specifically for robustness to the calibration systematic and thus has a very small shift in bias when considering the up and down calibration systematic, which is the largest of the systematic uncertainties. This overall good performance and learned robustness to the calibration scale give the LSTM the best sensitivities when looking at the 2020 best-fit point.

Based on the sensitivities, it would have seemed like the LSTM is the best choice for the 2024 analysis. It was trained to have a good balance of overall performance, as well as robustness to calibration uncertainty, which is the largest systematic for the EEs. This gives it good overall stats-only performance, as well as leading to the smallest pull from the systematic uncertainties for the final contours.

However, after looking at the ND energy spectra broken up by hadronic energy quartiles, the nonphysical jaggedness of the first quartile for the LSTM is quite worrying. Furthermore, the extremely narrow fourth quartiles of CNN do not seem to follow the principles of calorimetry.

For this reason, it seems the spline is the most conservative choice, given the unexplained features of the deep learning based methods. The Spline with one line only, shown in figure 3.4, was the ν_μ energy estimator decided on for the 2024 analysis.

Chapter 4

NOvA Oscillation Analysis Elements

The 2024 NOvA oscillation analysis was the first NOvA analysis since 2020 [12] with new data. Figure 4.1 shows the beam exposure accumulation of NOvA. The data included in this analysis has 26.6 $\times 10^{20}$ POT of neutrino beam exposure and 12.5 $\times 10^{20}$ POT of antineutrino beam [58]. This POT was the result of 10 years of NOvA data taking, and 95.6% more neutrino beam exposure since the last NOvA analysis.

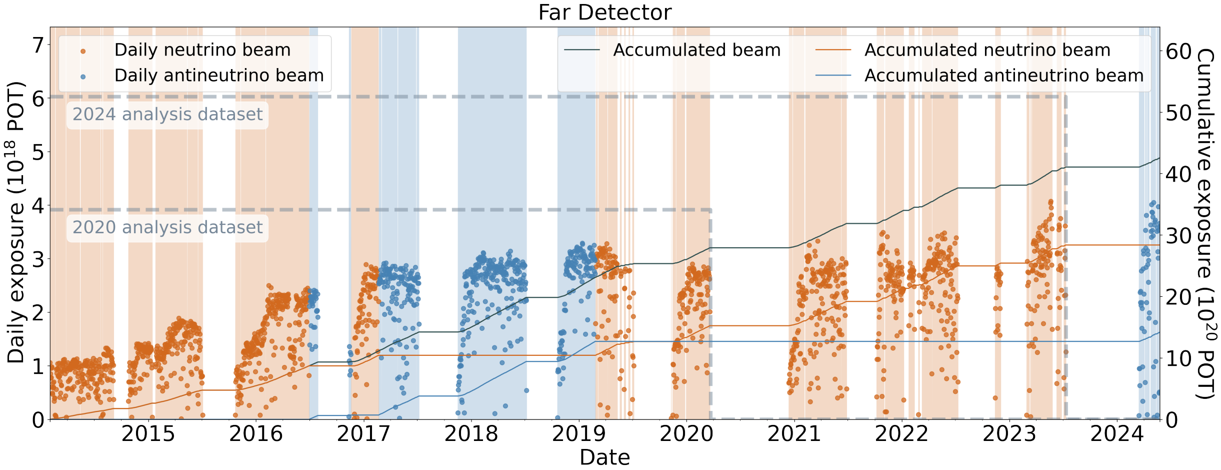


Figure 4.1: POT accumulation of NOvA over the running of the experiment

Many of the elements of the analysis remain the same as the 2020 analysis, with a few significant improvements [47]. First, besides more beam exposure, we had several improvements in simulation and reconstruction.

Three new processes have been added in the simulation: diffractive scattering, NC meson exchange current, and CC quasi-elastic lambda [42]. There was also significant changes in the light modeling, with an 80% increase in the amount of simulated Cherenkov light, and changes to attenuation parameters. In the ND, there was a 15% decrease in the light level scale in both views, where in the FD, there was a 20% increase in the X view and no change in the Y. Reconstruction didn't change as much, but still saw several improvements in track reconstruction. The GENIE cross section tuning saw several improvements as well [36].

4.1 Selection and Binning

4.1.1 Quality and Containment Cuts

Quality

The basic quality cuts provide a sanity check that the slice is something we would want to include in an analysis, with some basic reconstruction having had been completed successfully. In short, these cuts remove slices with too many or too few hits, too much energy, or slices where no tracks or prongs were reconstructed. The cuts are different for ν_μ and ν_e events, and they have been retuned for this analysis [44].

For ν_μ event reconstruction quality cuts, we require a reconstructed muon track with non-zero energy, a successful reconstruction of a 3D Kalman track, more than 20 hits overall in the slice, hits in more than 4 detector planes, and basic cosmic cut. For ν_μ selection quality, the only cut applied is that the reconstructed energy is below 5 GeV.

For ν_e event reconstruction quality cuts, we have a cut to remove slices that go through too many planes, a requirement that a vertex was reconstructed, and that at least one prong was reconstructed. And for selection quality, the cuts are slightly different between the ND and FD, but for the ND and FD core samples, there is high and low threshold on number hits in the slice, the reconstructed energy, and the longest prong. For the FD peripheral, we have a looser lower bound on the energy than in the Core, and a max threshold on electromagnetic energy. For a new low energy ν_e sample we included in this analysis, we have similar quality cuts than in the core sample, but with lower energy bounds.

Containment

The containment cuts check that all the charged particles from a slice are completely contained in the detector. This is needed to perform decent energy reconstruction, so except for special samples, we will want everything to be contained. The values of the cuts were retuned for the 2024 analysis [43]. The cuts are different for ν_μ and ν_e events, and for ND and FD. In short, there is a threshold for distances of vertex, showers, and tracks to be from the edge of the detector, both in terms of 3D coordinates and in number of detector planes.

4.1.2 Cosmic Rejection

NOvA's FD is above ground, without any significant shielding from cosmics, so good cosmic rejection is a basic requirement to do an oscillation analysis. NOvA uses two different cosmic rejection mechanisms: a traditionally used method, the "Cosmic Rejection Veto" which uses tuned cuts on direction and position information of tracks to remove cosmic rays, and a CNN based method, the "Cosmic CNN" which removes chunks of time that contain only cosmics early on in the reconstruction process.

The Cosmic CNN is a method to remove the bulk of cosmics early in the reconstruction process, to avoid the disk space required to save reconstruction information about the extreme amount of slices that are clearly cosmics [73]. This mechanism breaks up the full trigger window into 32 windows of time, where the beam spills by design are included entirely in one of these windows. The CNN then looks at the raw pixelmap of the entire detector for each of these 32 windows, where each pixel is made binary to 0 for no energy deposition, and 1 for a deposition, to remove dependence on calibration. The CNN makes a decision for each window to decide if it is only cosmics, or includes a neutrino. For each slice in the analysis, we then put a cut on its corresponding window's cosmic score, in addition to the slice level Cosmic Rejection Veto based on direction and position information.

In addition to these cosmic rejection mechanisms, each analysis sample have a separately tuned boosted decision tree (BDT) tuned to remove any remaining cosmics from the specific analysis sample.

4.1.3 Particle Identification

For particle identification (PID) in our selection, besides the BDTs for cosmic rejection mentioned in 4.1.2, NOvA also uses a CNN trained to identify neutrinos, called EventCVN [30]. This CNN is trained to take slices as input, and output different "scores" to indicate the belief that it is in one of several categories of neutrino, or a cosmic.

Besides EventCVN, there is another method called ReMId, which is used to identify Kalman tracks as muon like by using a k-Nearest-Neighbor (kNN) algorithm and dE/dx, scattering, track length, and non-hadronic plane fraction [99].

4.1.4 Analysis Samples

ν_μ Samples

The ν_μ event samples to measure ν_μ disappearance remain mostly unchanged since the 2020 analysis, predictions for this sample shown with and without oscillations in figure 4.2. The binning has been re-optimized, now having 22 bins from 0 to 5 GeV, which is a ∼25% increase in bins from the 2020 analysis [105].

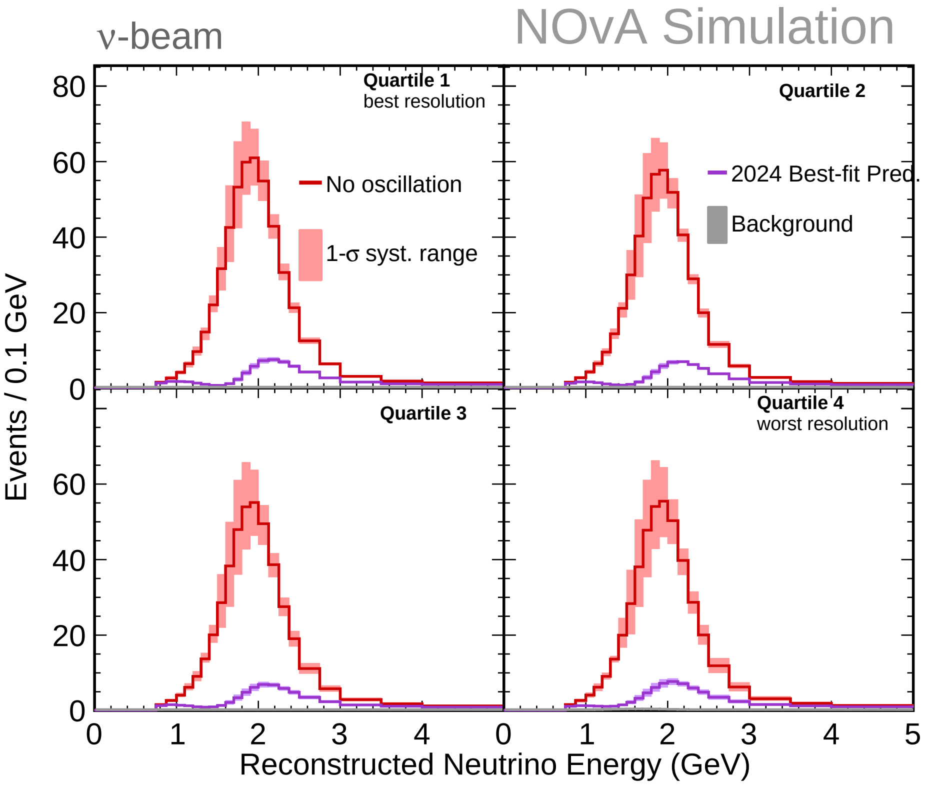


Figure 4.2: Extrapolated ν_μ predicted sample broken up by hadronic energy fraction, FHC

As in the 2020 analysis, the sample is broken up into 4 quartiles of hadronic energy fraction, $E_{had}^{Reco}/E_{\nu}^{Reco}$. The breakup is done on a bin-by-bin level, shown in figure 4.3.

The selection for the ν_μ was re-optimized for the 2024 analysis [43]. The ν_μ sample applies the standard quality and containment cuts, discussed in 4.1.1, the the cosmic rejection veto

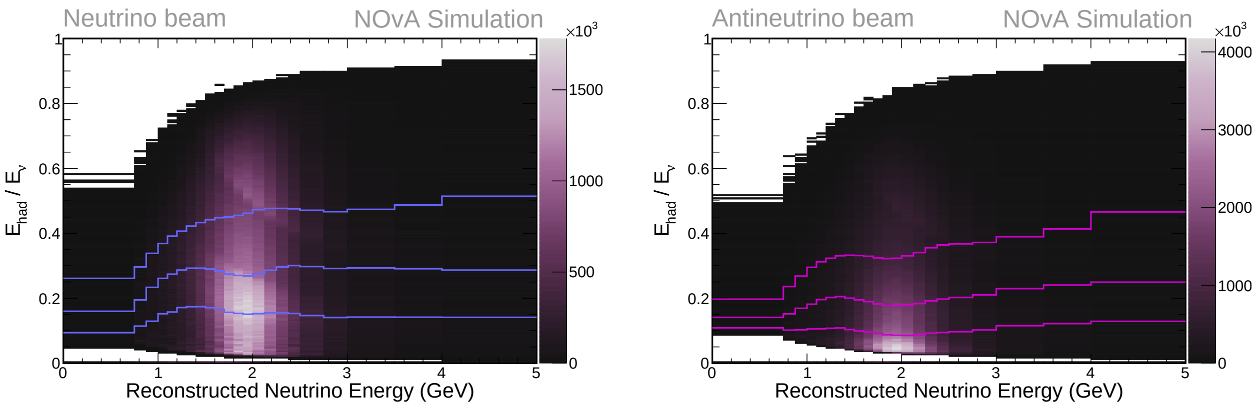


Figure 4.3: Hadronic energy fraction quantiles for ν_μ events, for FHC and RHC, taken from [43].

and CNN, a ν_μ cosmic rejection BDT, a cut on EventCVN ν_μ score, and on ReMId.

ν_e Samples

The ν_e event samples to measure ν_e appearance include the same samples as the 2020 analysis, with a somewhat modified definition of the "peripheral" sample, and a new low energy sample. Predictions for these samples are shown with this analysis best fit point in figure 4.4. The binning for the "Core" sample shown in the center, with high and low PID subsamples, are kept the same as the 2020 analysis, with 6 bins from 1 to 4 GeV, each 0.5 GeV wide. There is then one additional bin, the "Peripheral" bin, shown right in the figure, which includes some events that failed containment. Lastly, there is the new low energy sample shown left in the figure, with two bins from 0.5 to 1.5 GeV, which was only included for the FHC sample in this analysis.

A rough overview of the ν_e sample cutflow is shown in figure 4.5. All ν_e events go through some basic quality cuts. When events fail the containment cut, they have a chance to go into the peripheral sample. This is different than the definition from 2020 analysis, where the peripheral sample was composed of events that fail any number of core selection cuts, not just the containment.

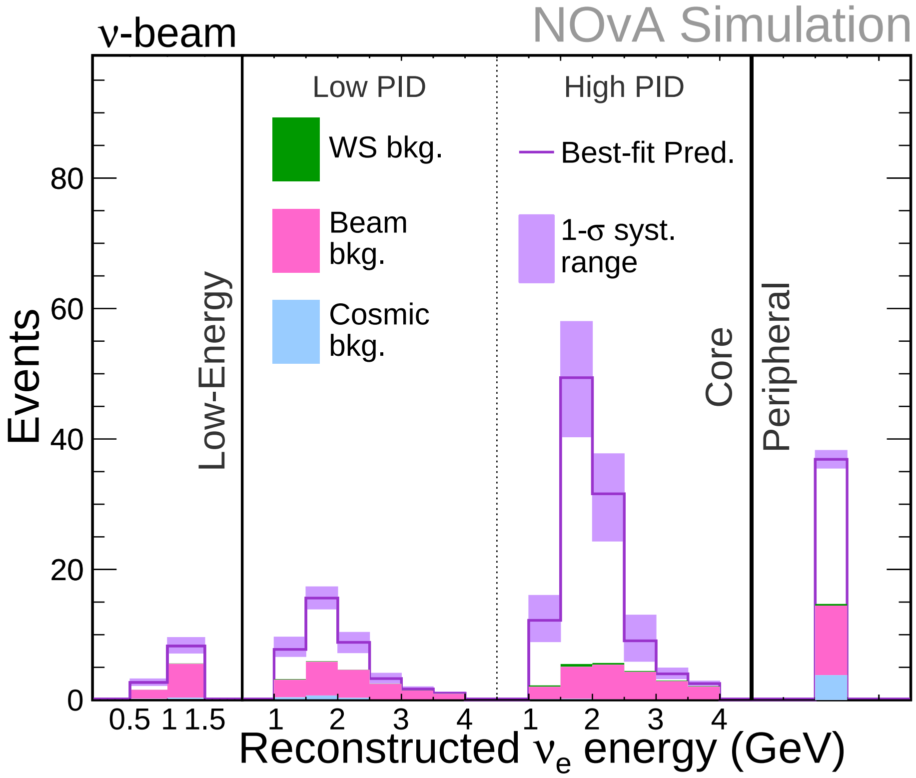


Figure 4.4: Extrapolated ν_e predicted sample, FHC

From here, events are selected for the peripheral bin that pass a set of EventCVN ν_e and cosmic score cuts, as well as the peripheral BDT. An electromagnetic energy cut is also applied, which was shown to be effective in removing intrinsic beam ν_e events. They are all put into the one peripheral bin, since without containment, energy estimation is not reliable, but there is still information in the event count.

For the events that do pass the containment cuts, the Core sample cuts are applied, applying a cut on EventCVN score as well as the Core BDT for cosmic rejection. If they fail this Core selection cut, they again have a chance to survive in an alternative sample, the low energy sample (in the FHC sample only). For this case, there is a cut on another specially tuned BDT for the low energy sample, as well as different quality cuts, which include a cut to choose events in this low energy region. Through this BDT, we again regain events not clean enough to end up in the Core sample.

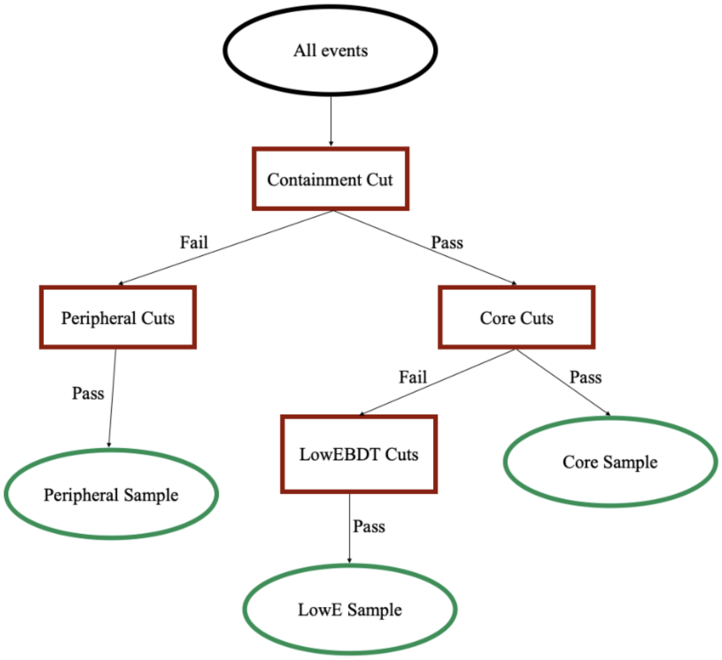


Figure 4.5: ν_e cutflow

The events that do pass the core cut are split into samples of high and low PID based on their EventCVN score. These cuts, as well as all the other ν_e selection cuts, have been re-tuned for the 2024 analysis.

4.2 Near to Far Extrapolation

A distinct feature of the NOvA oscillation analysis, as in recent analyses, is the near detector to far detector extrapolation technique which uses the near detector data to apply a data-driven correction to the predicted far detector energy spectrum. The extrapolation for this analysis remains mostly unchanged from the 2020 analysis [92]. The technique uses the data / MC energy bin ratios with ND data and MC, and uses the relationship of the true and reconstructed energy in MC to transfer these ratios as corrections to the FD spectrum by using the predicted oscillation probabilities for a proposed set of oscillation parameters.

Different signal and background samples are treated individually.

4.2.1 $\nu_\mu \to \nu_\mu$

An illustration of extrapolation for $\nu_\mu \to \nu_\mu$ is shown in figure 4.6. For this sample, the predicted ν_μ signal in the FD can be extrapolated directory from the ND ν_μ sample. During extrapolation, we further divide each of the hadronic energy fraction samples into a sub-sample of 3 $|\vec{p}_T^{\mu}|$ quantiles for each bin of hadronic energy fraction, i.e. the magnitude of the muon momentum transverse to the beam direction. These $|\vec{p}_T^{\mu}|$ sub-bins are recombined immediately after extrapolation, before fitting the samples. The p_T^{μ} extrapolation was found to reduce the effects of most systematic uncertainties, especially cross section systematics.

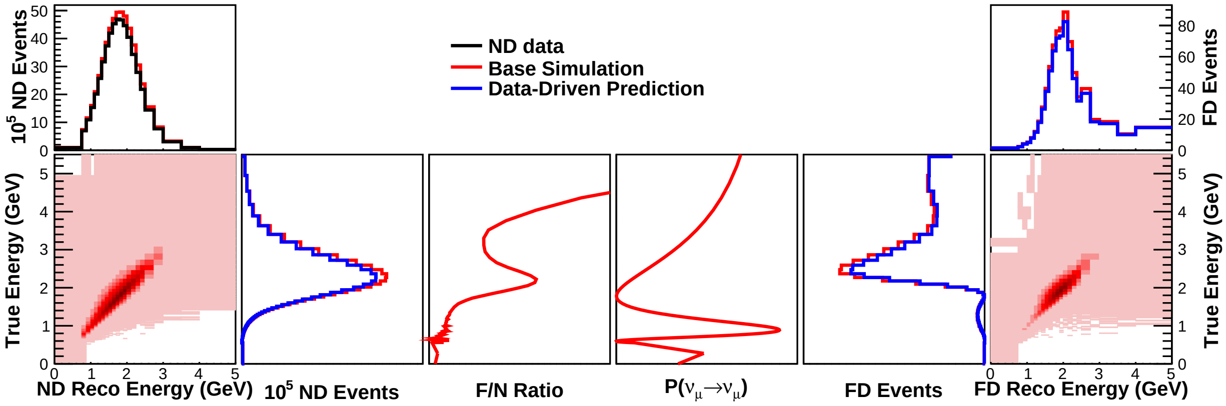


Figure 4.6: Illustration of near to far extrapolation for the $\nu_\mu \to \nu_\mu$ oscillation signal

4.2.2 ν_μ Background

Backgrounds from neutrinos for ν_μ sample will be small, and these numbers are taken directly from MC predictions. For the cosmic background, for both the ν_μ and ν_e background, we use the dedicated cosmic sample, and scale the integral to be the same as the NuMI beam window sideband.

4.2.3 $\nu_\mu \rightarrow \nu_e$

There is no appeared ν_e signal in the ND, by definition, so all the ν_e signal in the FD come from ν_μ in the ND. So, we use the same method as $\nu_\mu \rightarrow \nu_\mu$ method, where we use the $\nu_\mu \rightarrow \nu_e$ oscillation probability to correct the predicted ν_e FD sample with the data / MC ratios from the ν_μ ND sample. The p_T^μ samples are also used for this $\nu_\mu \rightarrow \nu_e$ extrapolation.

4.2.4 ν_e Background

To constrain the FD ν_e background, we use the ND ν_e sample, which are all backgrounds by definition in the FD. We break up the MC into true components of beam ν_e CC, ν_μ CC and NC. We use different methods to correct the ratios of these component for the FHC and RHC samples. In FHC, we use "Combo" decomposition, where we constrain each of these components from different samples. For the intrinsic beam ν_e's, we use details about the parent mesons from the NuMI, and the ν_μ sample.

Most of the intrinsic beam ν_e's come from the following decays:

$$\pi^+ \rightarrow \nu_\mu + \mu^+$$
$$\mu^+ \rightarrow \bar{\nu_\mu} + e^+ + \nu_e$$

and

$$K^+ \rightarrow \pi^0 + e^+ + \nu e$$
$$K_L^0 \rightarrow \pi^- + e^+ + \nu e$$

. The pion decays dominate the ν_μ energy events from 0 - 5 GeV, so we can use the ND ν_μ sample to constrain the expected intrinsic ν_e beam background, which is dominated by the pion production at lower energies, then use the uncontained ν_μ sample candidates to constrain the rate of the kaon decays. From there, the ν_μ CC to NC ratio is constrained using the Michel electrons in bins where the Michel rates are significant.

For RHC, we use "proportional" decomposition, where rely more on the MC for the ratios of the true backgrounds, and scale them to the observed ND ν_e sample.

4.3 Systematic Uncertainties

Overall, the systematics are mostly the same as those in the 2020 NOvA analysis, being retuned in some parameters, with a few newly implemented systematics.

4.3.1 Detector Systematics

We have several systematics that describe uncertainties in detector response and calibration that have been retuned with the new simulation [45].

The largest of these is the calibration energy scale, which we assume to be an uncertainty of 5% overall on the calibration scale. We include both an absolute calibration scale uncertainty, where the FD and ND are assumed to be calibrated identically, and a relative calibration uncertainty, where the FD and ND are assumed to be anticorrelated in calibration. We also have an uncertainty on the calibration shape, which takes into accounts differences in calibration in different parts of the detector. This was modeled assuming a linear difference in shape over the detector, and then constrained by using data/MC differences.

We also have uncertainties on the Cherenkov light collection efficiency. The Cherenkov collection efficiency is derived from light level tuning, and was estimated to be on the scale of 6.2% with the current simulation. In addition to Cherenkov, we have a systematic on the light level, tuned in a similar way as Cherenkov, and determined to be 5%.

We include a detector aging systematic, which allows for a simple linear drift downward in light level over time.

Lastly, we have a muon energy scale systematic, essentially constraining the expected length for a track of a muon of known energy. There are separate overall uncertainties for the ND and FD, as well as correlated and uncorrelated, and a separate factor for the muon catcher.

4.3.2 Cross Section Systematics

GENIE is the model we use for the generation of neutrino interactions [26]. GENIE has tunable parameters, which represent cross section parameters for the many different interaction processes. We tune the central value of these knobs to be consistent with the data in our ND. We then have systematic uncertainties in terms of these GENIE knobs, of which there are 78 in total. By judging the size of the effects of each of the individual knobs on our final sensitivities, we break up this list into "large" and "small", where we leave the large knobs as their own parameters in the fit, and throw the remaining into a principal component analysis (PCA), described in the next section.

GENIE Principal Component Analysis (PCA)

The small GENIE knob PCA is repeated with the 2024 small GENIE knobs, in a analogous fashion to the version done for the 2020 analysis [94], with similar results, but slightly faster convergence [81].

Principle component analysis (PCA) is a way of linearizing the correlated effects of a list of many variables. We use PCAs for NOvA systematic uncertainties with a large number of

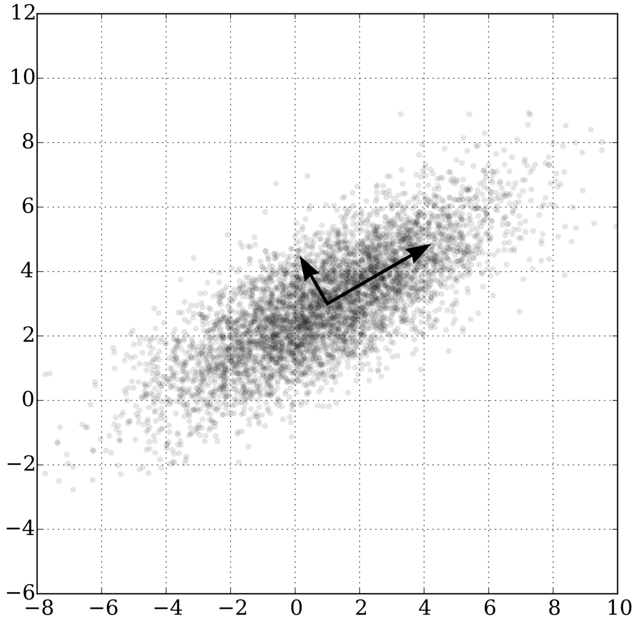


Figure 4.7: Illustration of principal component analysis (PCA), for a 2D Gaussian correlated in the two directions. Image taken from reference [115].

correlated, small model parameters. PCAs can pick out the largest emergent effects of all of the small knobs together, allowing us to reduce the dimensionality of the parameter space and still retain most of the effects of the full model. The recipe to make a PCA is to find the covariance matrix of all of the model parameters, then find the eigenvectors and eigenvalues of the covariance matrix. The eigenvectors are the principal components (PCs), where the eigenvalues are the effective sizes for each PC. An illustration of a PCA for a 2D correlated Gaussian is shown in figure 4.7.

For our purpose, we make a "multiverse" of the systematic parameters for the analysis energy spectra samples, i.e. we make many different possible spectra with random throws of all the systematics in the list. We then find the bin by bin covariance matrix, and use this to find the eigenvectors (the PCs) and eigenvectors. We rank the PCs by their eigenvalue, choosing as few as are needed to capture the effects of the original multiverse. We scale up the chosen

Coverage	2020 #	2024 #
90%	13	10
95%	19	16
99%	34	31
99.9%	66	60
99.99%	97	93

Table 4.1: Coverage of GENIE PCA by number of PCs

PCs by some amount to account for the full effect of the underlying systematic.

For the GENIE PCA, 1000 genie universes are generated using the “small” genie knobs, i.e. the knobs not treated specifically in the fit. This list was the same as last analysis, with updated knobs changed to their new version. The full list is 48 long. This list is made by judging from the size of the shifts on the predictions from each of the GENIE knobs. Roughly, a threshold of around $\chi^2 \geq 0.005$ chooses the “large” knobs. However, the final list was chosen to be mostly consistent with the 2020 analysis list, since there was no large changes between Production 5 and 5.1. So, the newly added knobs were kept in the “large” list, regardless of size, to have their impacts on the analysis more apparent.

The effects on the true energy spectrum on samples of true neutrino flavors are quantified, and the bin by bin covariance matrix is found. The eigenvectors of this covariance matrix are the principal components (PCs), and can be ranked in size by their eigenvalue.

The 1000 universe samples, along with their central value, and the ratio of the universes to their central value, are shown in figure 4.8. Their bin by bin covariance matrix is shown in figure 4.9. As with the 2020 analysis, the ND samples are suppressed so convergence focuses on the FD samples, but the ND components are still carried through to the end. The first 4 principle components of this covariance matrix, ranked by eigenvalue, are shown in figure 4.12. The eigenvalue by PC number is shown in figure 4.10.

The coverage of the total variance by number of PCs is shown in figure 4.11, and numbers

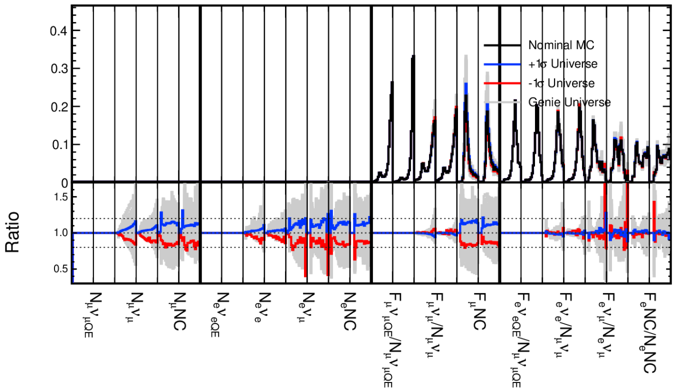


Figure 4.8: Effects of small GENIE knobs on PCA samples

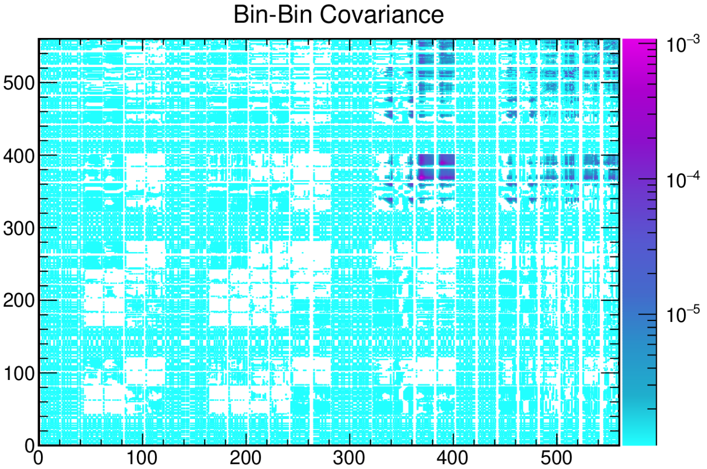


Figure 4.9: Bin by bin covariance of GENIE PCA samples

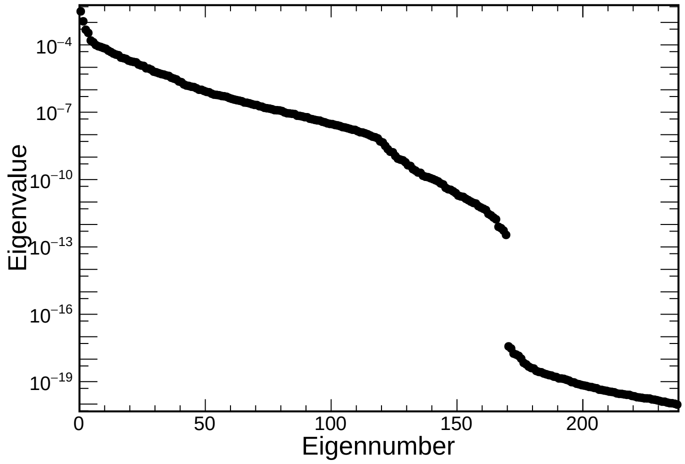


Figure 4.10: Ranked eigenvalues of GENIE PCs

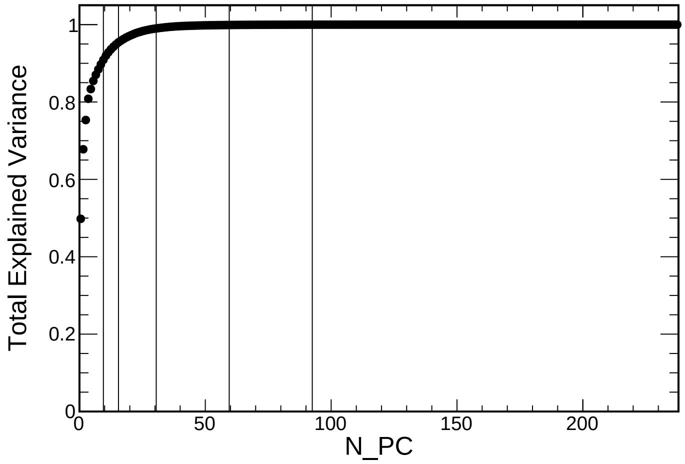
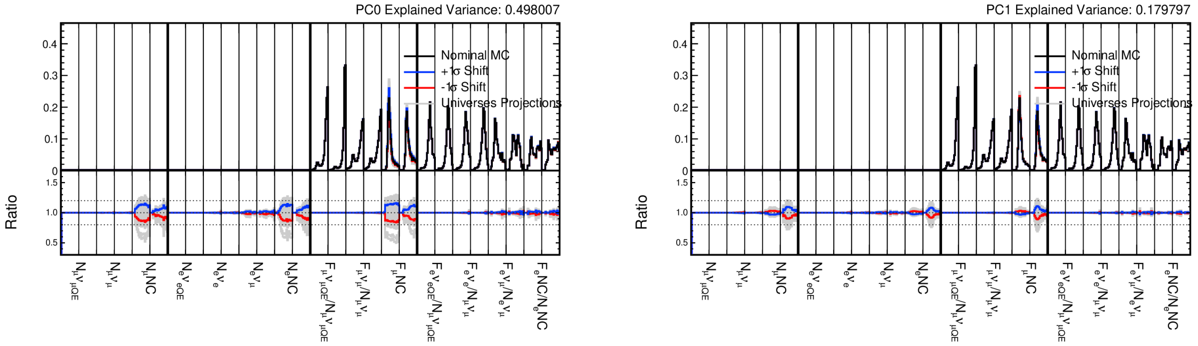


Figure 4.11: Coverage of total variance by largest GENIE PCs

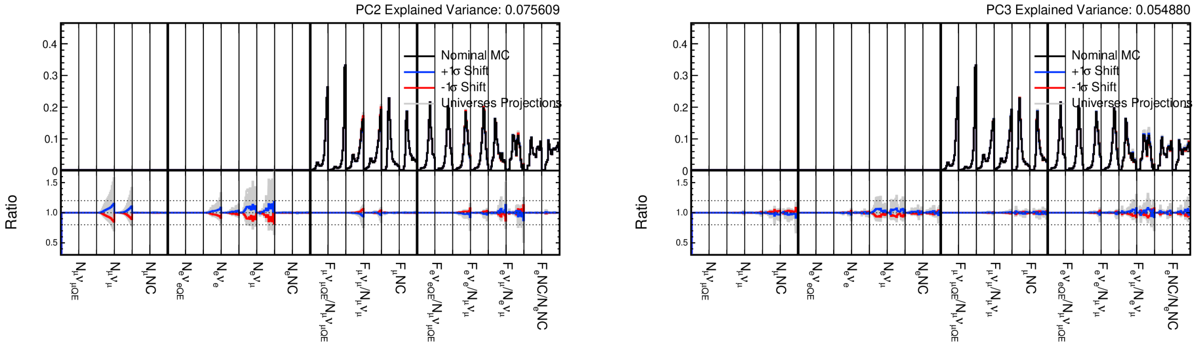
required for certain thresholds listed in Table 4.1. The 2024 PCA in general seems to converge slightly faster than in 2020. The PCs look roughly comparable to the corresponding 2020 PCs, seemingly capturing the same effects. The choice of number of PCs and amount of upscaling is motivated by looking at the effect on sensitivities, in the next section.

As a standard check to see if the PCA correctly captures the effects of the underlying multiverse distribution, we compare the effects of the PCA and multiverse shifts on sensitivities. To do this, we first choose fake data at some oscillation point, here the 2020 best fit point.



(a) PC0

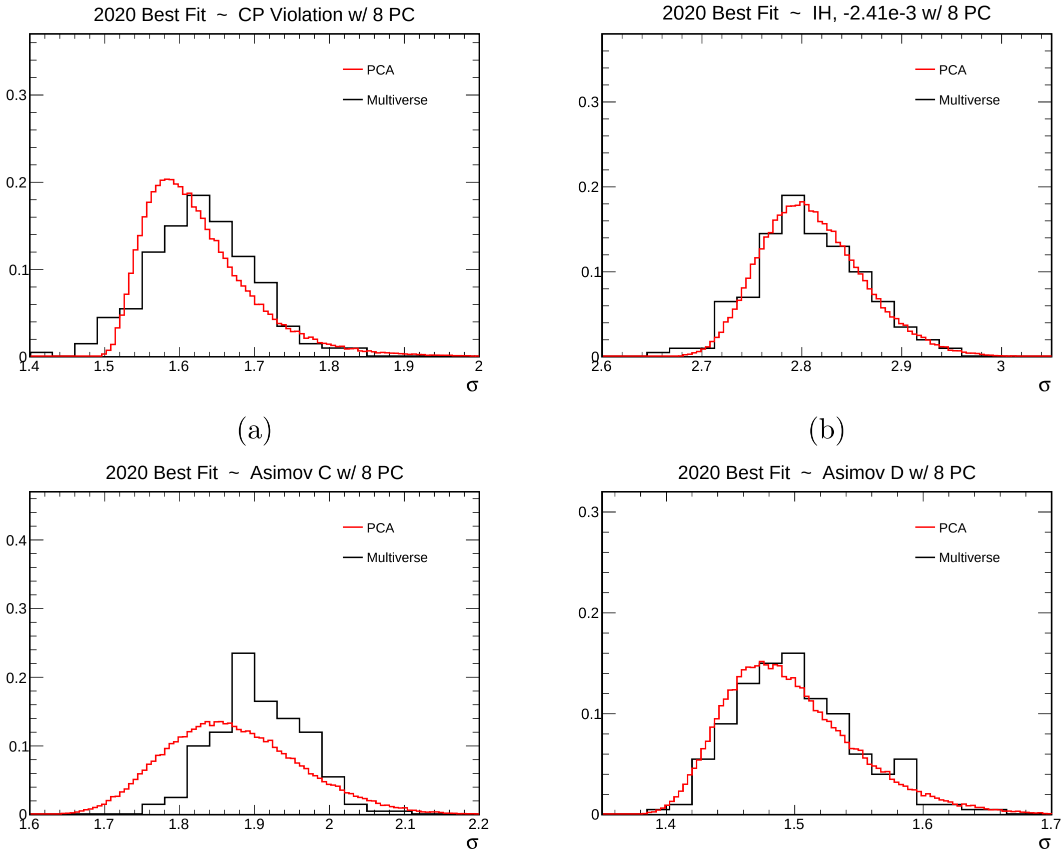
(b) PC1



(c) PC2

(d) PC3

Figure 4.12: First 4 GENIE PCs



(a) (b)

(c) (d)

Figure 4.13: Sensitivity comparisons for GENIE PCA and the underlying multiverse, for 4 oscillation points

Then we look at the predictions at another oscillation point. We fluctuate the systematics at this second point, either running a multiverse, or by fluctuating the PCA knobs by drawing values from a normal distribution. We can scale up and down the PCA effects by multiplying a number by the number drawn from the normal distribution. This is often needed to match the PCA effects to the multiverse. We then take the χ^2 between the fake data and this second, systematically shifted point. We compare the distribution of the square root of χ^2.

Figure 4.13 shows sensitivity comparisons with 8 PCs and 35% shift up. Adding more PCs doesn't further improve agreement, and reducing them degrades it, so this is the number of PCs and scaling used for the analysis.

4.3.3 Beam Flux and Focusing Systematics

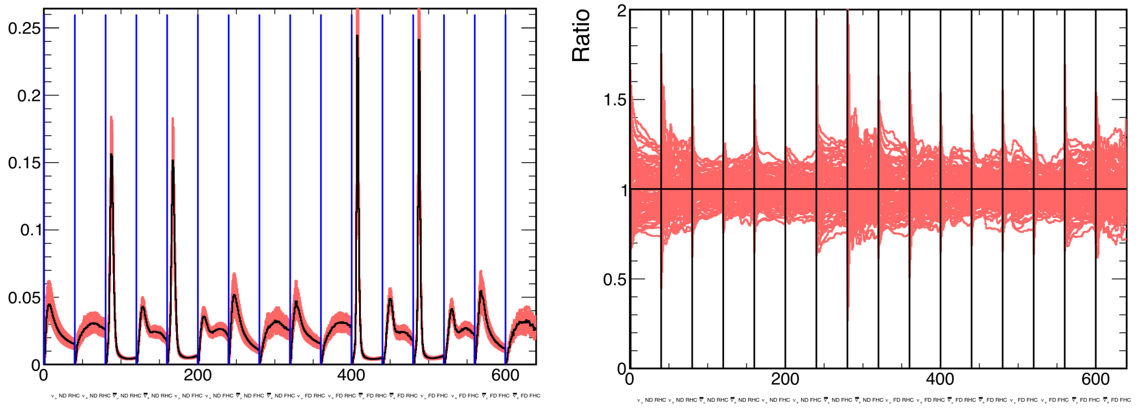
Other systematic uncertainties with a large number of small parameters are the beam flux and the beam focusing systematics, capturing effects of the NuMI beam modeling. These systematics were implemented as one combined PCA in 2018 [93], and this scheme was again used in 2020.

The flux and beam focusing PCA was also re-done for the production 5.1 samples used for this 2024 analysis [81]. The results work out similarly, and we use the same 5 PCs and 25% upscaling.

The procedure is unchanged from the 2018 analysis. The 100 flux universes are used with 20 beam focusing universes for each, creating 2000 flux and beam focusing universes. The effects on the true energy spectrum on samples of true neutrino flavors are quantified, and the bin by bin covariance matrix is found. The eigenvectors of this covariance matrix are the principal components (PCs), and can be ranked in size by their eigenvalue.

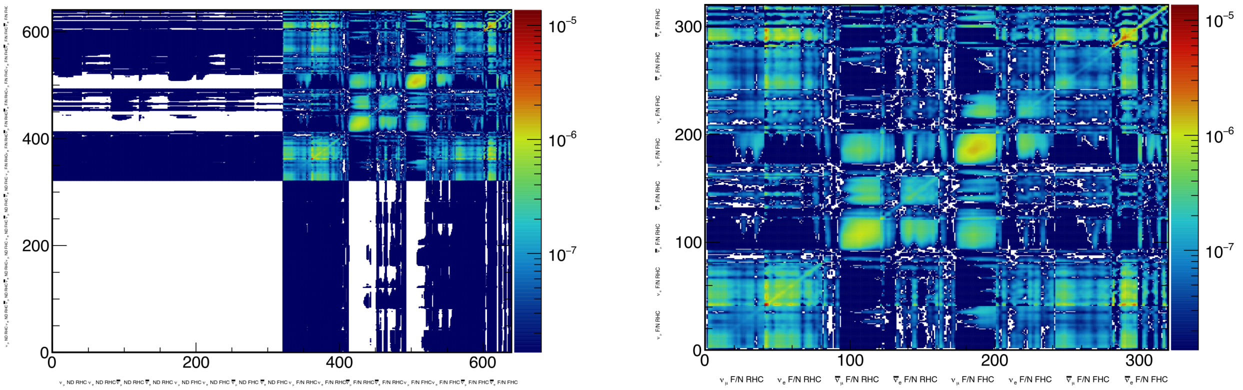
The 2000 universe samples, along with their central value, and the ratio of the universes to their central value, are shown in figure 4.14. Their bin by bin covariance matrix is shown in figure 4.15. The ND samples are suppressed so convergence focuses on the FD samples, but the ND components are still carried through to the end. The first 3 principle components of this covariance matrix, ranked by eigenvalue, are shown in figures 4.17 - 4.19. The eigenvalue by PC number is shown in figure 4.16.

As a standard check to see if the PCA correctly captures the effects of the underlying multiverse distribution, we compare the effects of the PCA and multiverse shifts on sensitivities, just as was done for the GENIE PCA in the last section. This procedure showed the same choice of PCs and scaling used in 2018 and 2020 are still reasonable.



(a) (b)

Figure 4.14: Flux PCA samples along with the multiverse for flux and beam focusing in red and the central value in black (a), and the ratio of the universes with their central value



(a) Bin by bin covariance of all PCA samples (b) Bin by bin covariance of F/N block

Figure 4.15: Bin by bin covariance of flux PCA samples

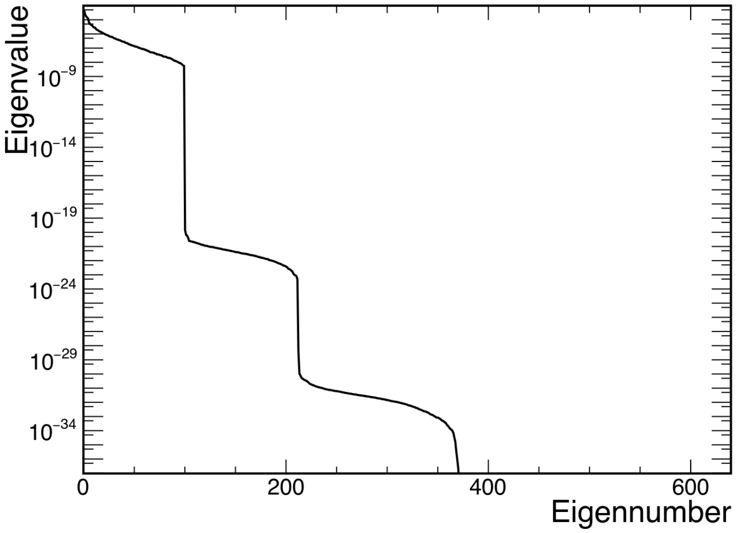


Figure 4.16: Eigenvalue by number of Flux PC

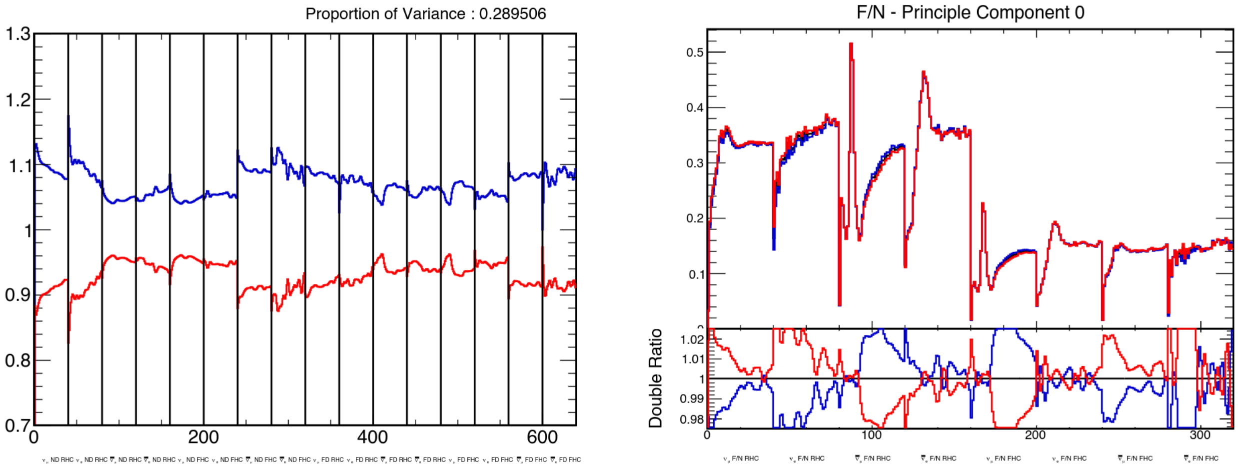


Figure 4.17: Flux PC 0

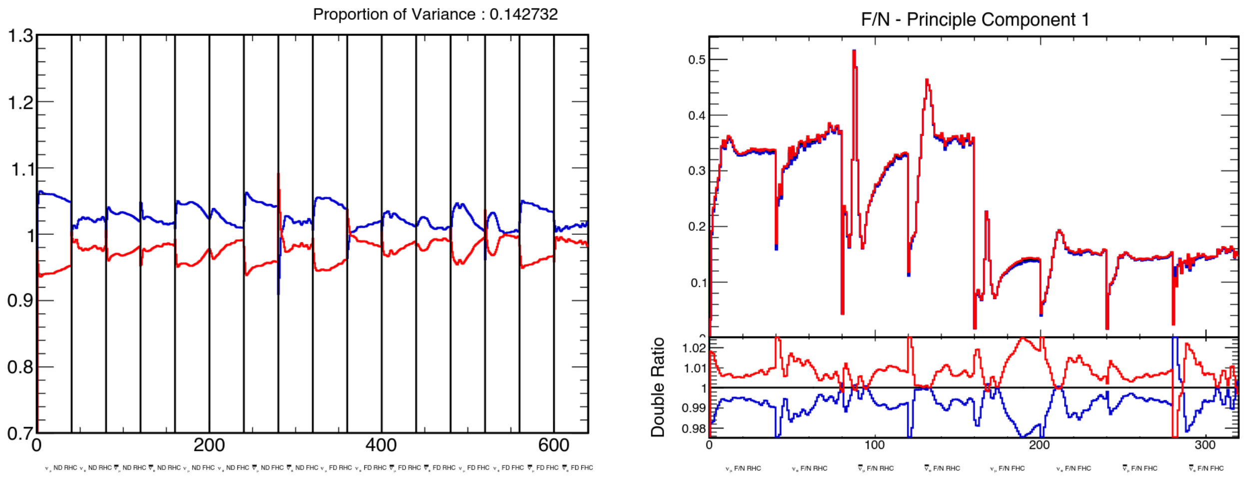


Figure 4.18: Flux PC 1

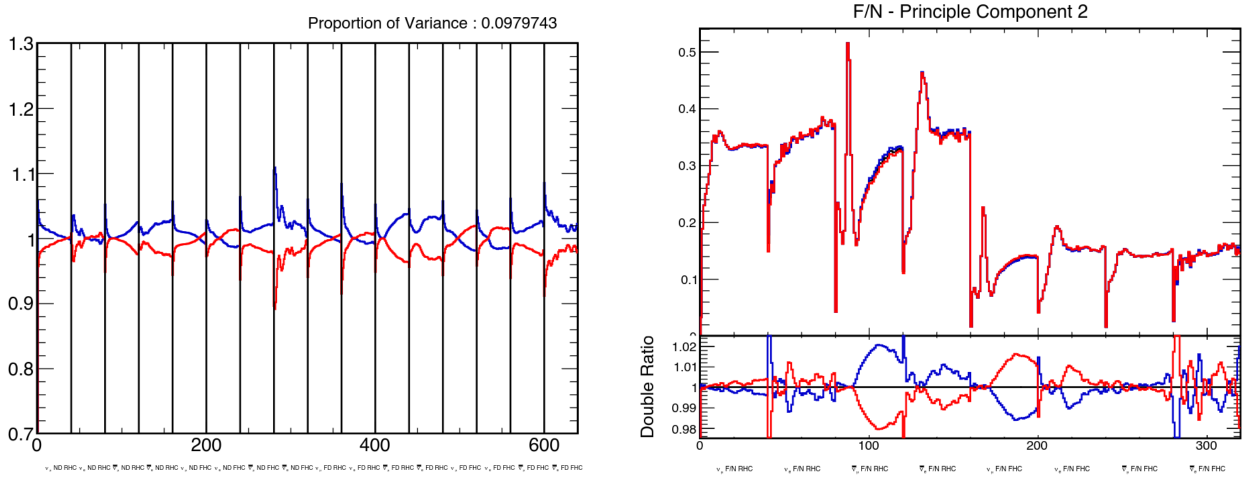


Figure 4.19: Flux PC 2

4.3.4 Geant4Reweight Systematic

Geant4Reweight [40] is a powerful event reweighting framework which allows one to quantify changes to Geant4 hadron modeling without the need to fully re-simulate. It assigns a weight to an event after modifying the cross section for a given interaction type. For the interaction steps, i, of a particle, each of length, L_i, the weight will be

$$w = \frac{\sigma'_N}{\sigma_N}\left(\frac{e^{-\sum_{i=1}^{N-1}\sigma'_i L_i}}{e^{-\sum_{i=1}^{N-1}\sigma_i L_i}}\right)$$

where the cross sections σ' indicate the new chosen cross section, and σ_N is the cross section of the last step of the process, either survival or interaction. It treats inelastic interaction types separately: absorption, quasi-elastic, charge exchange, double charge exchange, and production.

Geant4Reweight was implemented into NOvA after the 2020 analysis [109]. Practically speaking, it introduces 11 systematic knobs. There are 5 knobs for π^-, 5 for π^+, 1 for protons, and one neutron knob, the effects of which weren't fully understood, so it was excluded.

The effects of these are implemented into the 3 flavor analysis as a PCA, following the usual procedure, such as for the GENIE PCA. The main difference between this and the others is that Geant4Reweight shouldn't affect true energy. So instead of the PCA samples being samples of true energy as for the flux and Genie PCAs, we here use the analysis samples of reconstructed energy.

The Geant4Reweight universe samples, along with their central value, and the ratio of the universes to their central value, are shown in figure 4.20. Their bin by bin covariance matrix is shown in figure 4.21. A number of different multiverse sizes were tried. The first 3 principle components of this covariance matrix, ranked by eigenvalue, are shown in figures 4.24 - 4.26.

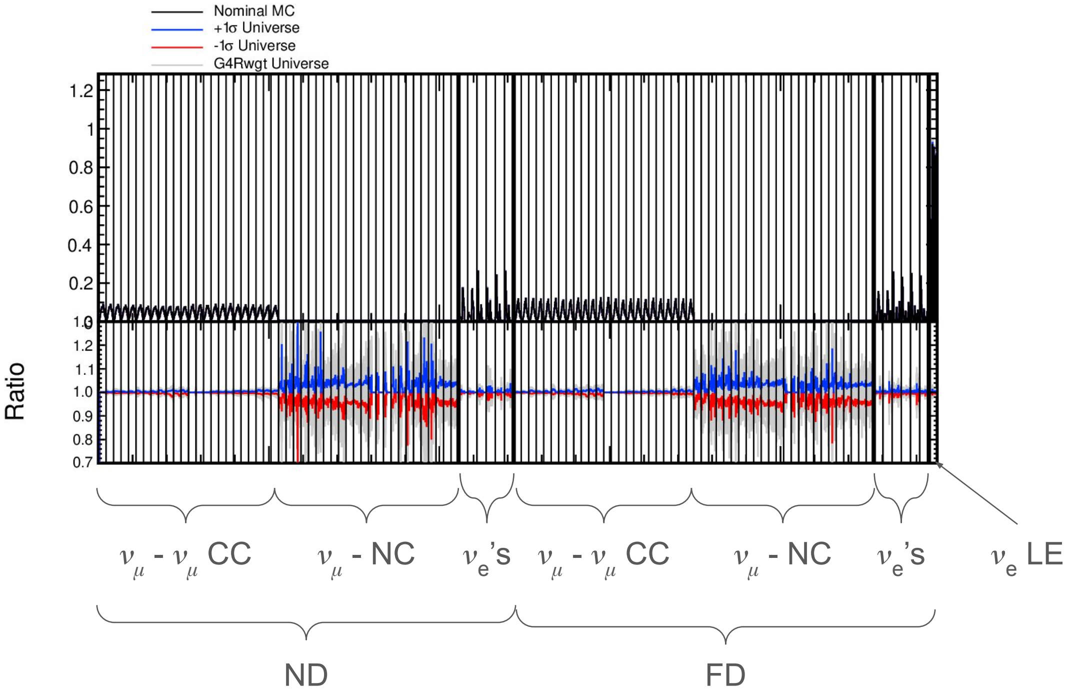


Figure 4.20: Geant4Reweight PCA Samples

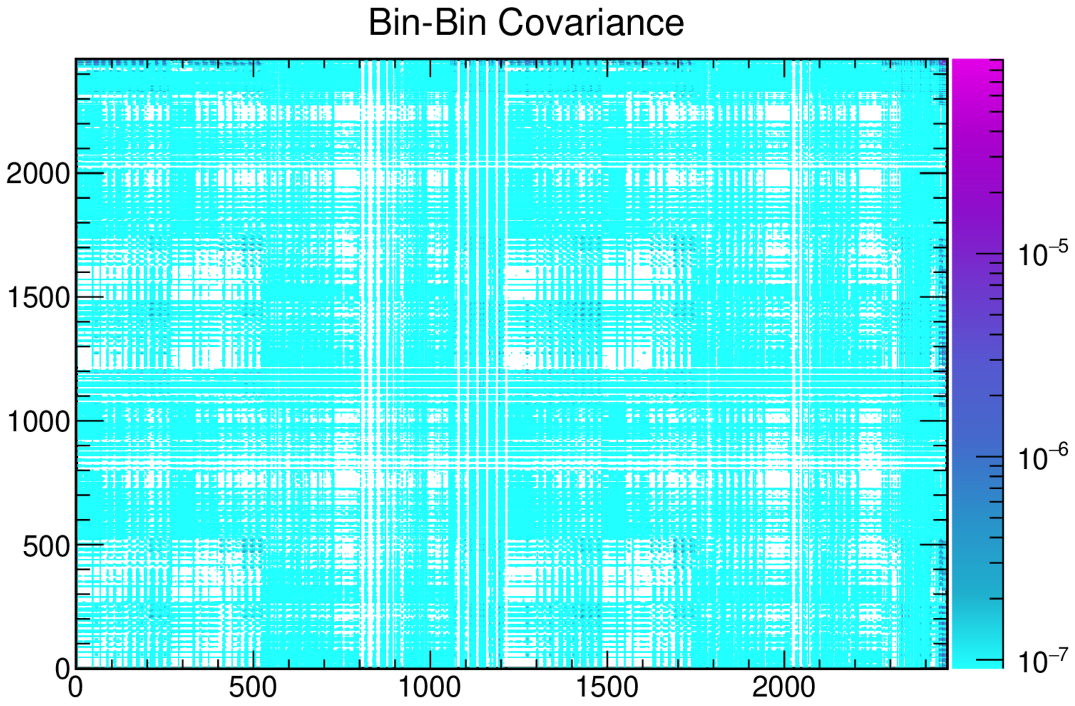


Figure 4.21: Bin by bin Geant4Reweight PCA Sample covariance

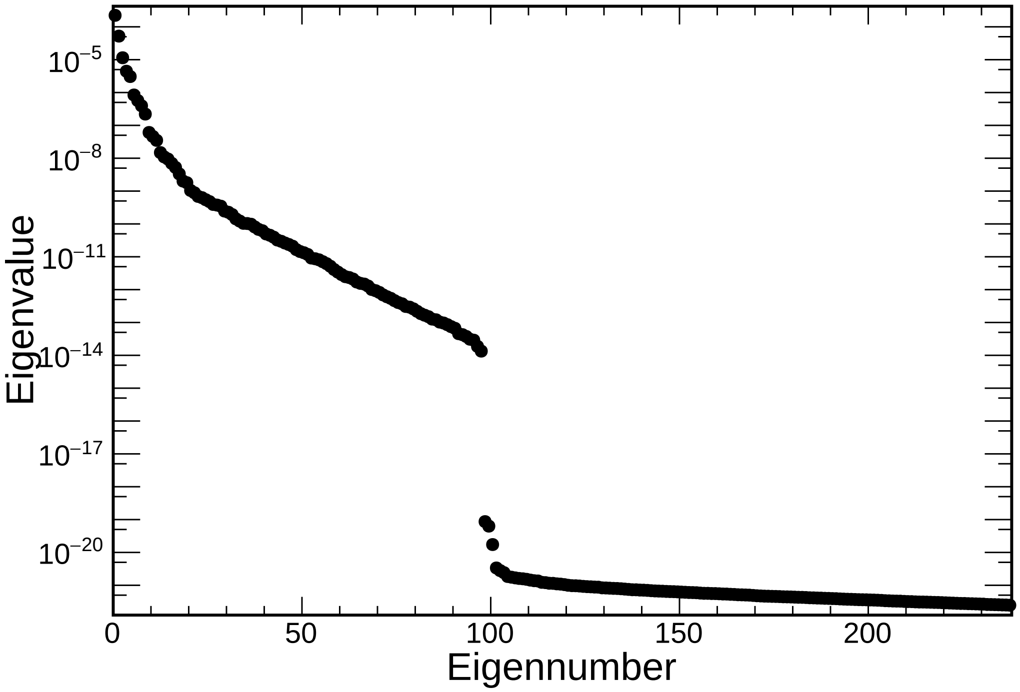


Figure 4.22: Ranked eigenvalues of Geant4Reweight PCA

The eigenvalue by PC number is shown in figure 4.22.

The coverage of the total variance by number of PCs is shown in figure 4.23, and numbers required for certain thresholds listed in Table 4.2. The choice of final number of PCs and upscaling factor required is chosen looking at sensitivities, see the next section.

As for the other two systematics implemented as a PCA, we check the number of required PCAs and the scaling we need by looking at the effects of the PCA on the sensitivities,

Coverage	#
90%	3
95%	4
99%	6
99.9%	9
99.99%	15

Table 4.2: Coverage of Geant4Reweight PCA by number of PCs

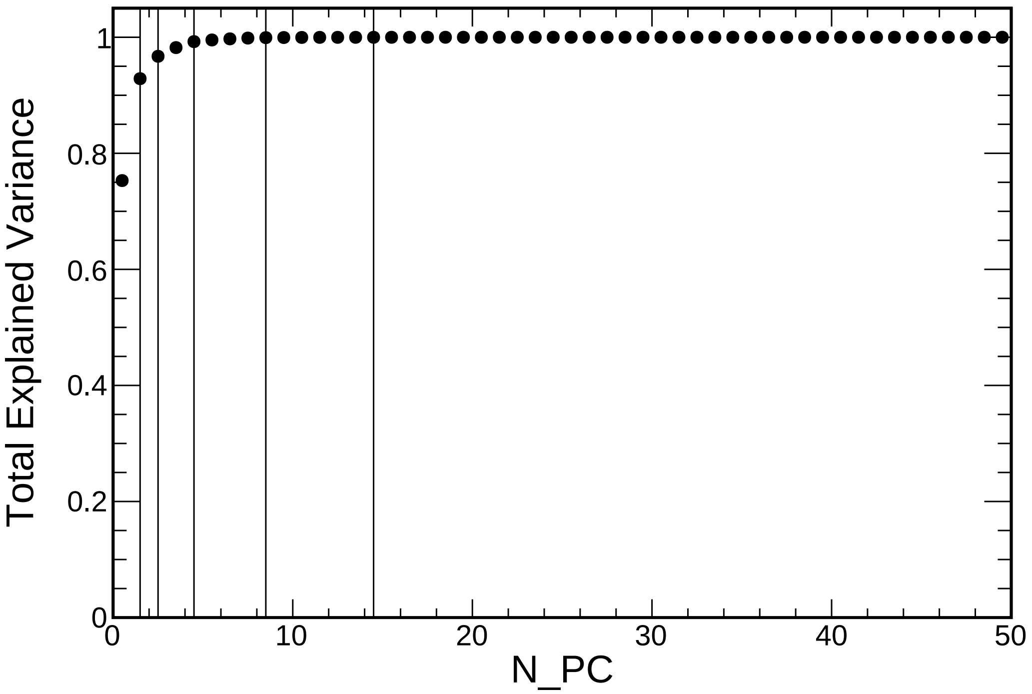


Figure 4.23: Coverage of covariance by largest PCs of Geant4Reweight PCA

and compared that to the underlying multiverse. This procedure showed we need to use 3 PCs. 100% upscaling seems to be needed to appropriately match the underlying multiverse distributions.

4.3.5 Other Systematics

NOvA also has a systematic on the cosmic scale, coming from the statistical uncertainty of the cosmic sample used to make predictions. The cosmic statistics used were significantly increased for the 2024 analysis, so the effect of the cosmic scale systematic has decreased.

A new systematic on neutron modeling was added, specifically implemented as an alternative neutron carbon interaction model, MENTATE [59, 102]. Geant4 is normally used for neutron simulation in NOvA, and this systematic quantifies the effect of using the alternative MENATE model in its place.

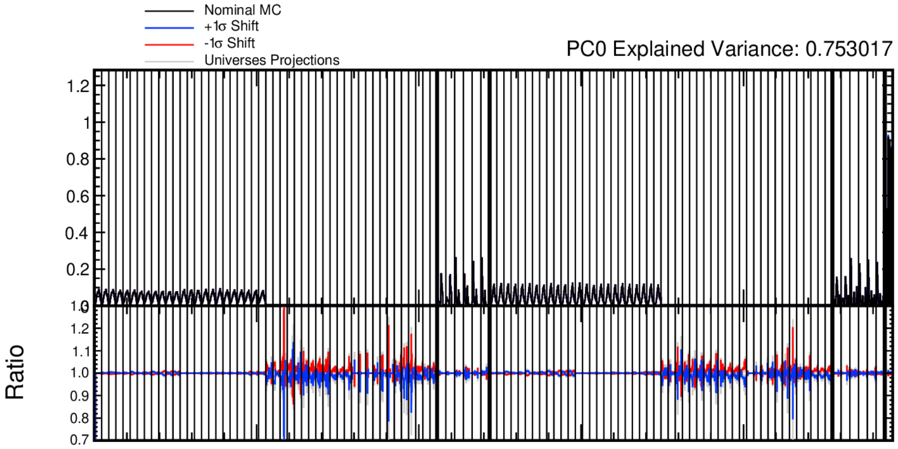


Figure 4.24: Geant4Reweight PCA PC 0

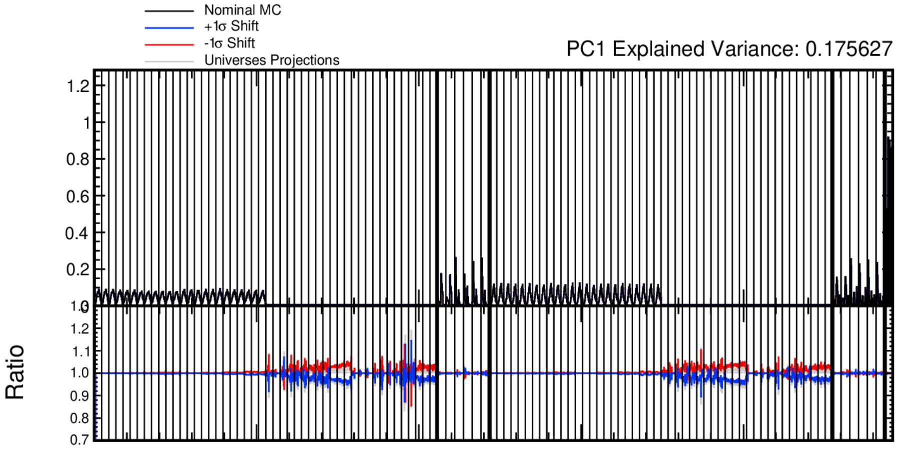


Figure 4.25: Geant4Reweight PCA PC 1

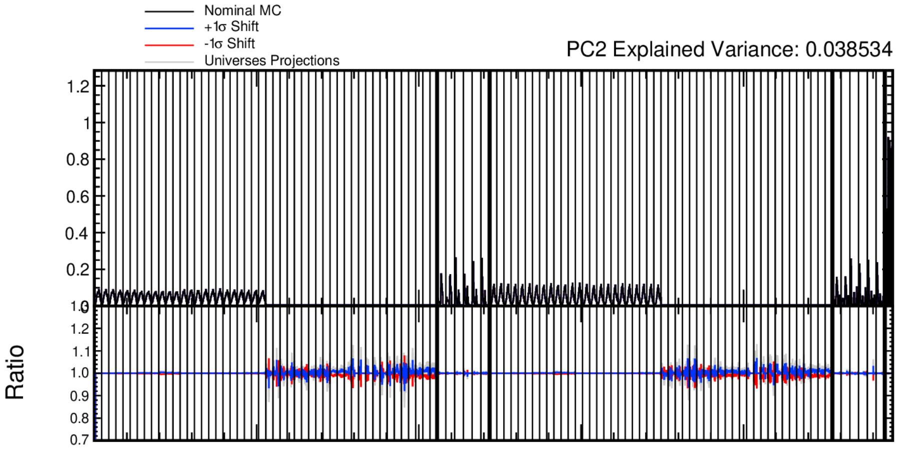


Figure 4.26: Geant4Reweight PCA PC 2

There are several other systematics in addition to those delved into above, such as normalization systematics, Michel tagging, FD rock systematics, and matter density systematics. The normalization systematic is the composite of multiple small systematics effects, such as uncertainty in POT, differences in detector masses, and pileup of events in the ND.

Chapter 5

Fit to Oscillation Parameters

5.1 The Likelihood Function

The log likelihood function for a set of neutrino oscillation parameters, $\vec{\theta}$, and a set of M systematic parameters, $\vec{\phi}$, is given by equation 5.1 [13]. The likelihood function is, in general, the probability of getting the measured data assuming some set of parameters of the underlying model. The first term is the Poisson log likelihood, for N bins of energy, applying Stirling's approximation. $e_i(\vec{\theta}, \vec{\phi})$ is the predicted energy bin value for bin i with oscillation and systematic values $\vec{\theta}$ and $\vec{\phi}$. o_i is the observed number of events in bin i. The second term is the systematic pull term, coming from the normal log likelihood, with each systematic term ϕ_k, normalized by the standard deviation of that pull term, σ_k.

$$-2\,\mathrm{ln}\mathcal{L}(\vec{\theta}, \vec{\phi}) = -2\sum_{i=1}^{N}\left[e_i(\vec{\theta}, \vec{\phi}) - o_i + o_i \mathrm{ln}\frac{o_i}{e_i(\vec{\theta}, \vec{\phi})}\right] + \sum_{k=1}^{M}\frac{\phi_k^2}{\sigma_k^2} \tag{5.1}$$

For the 2024 NOvA analysis, there are 4 oscillation parameters, included in the fit: $sin^2\theta_{23}$,

$sin^2 2\theta_{13}$, Δm^2_{32}, and δ_{CP}. The number of systematic parameters, M, is 68.

Constraints for unmeasured oscillation parameters from solar and reactor neutrino experiments are fixed to the following values [121]:

$$\Delta m^2_{21} = 7.53 \times 10^{-5} \text{eV}^2$$
$$sin^2\theta_{12} = 0.307$$

5.2 Frequentist Parameter Inference

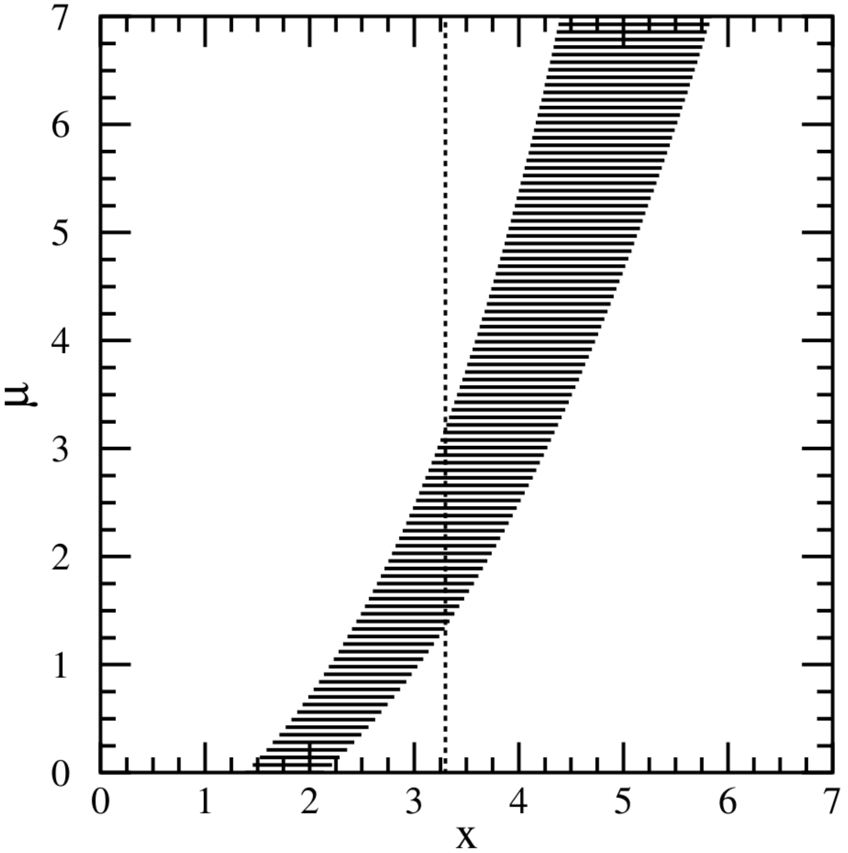


Figure 5.1: Confidence belt for a distribution with underlying parameter μ and estimator x. Image taken from Feldman-Cousins paper [61]

.

In Frequentist statistics, we are interested in making confidence intervals for parameters of interest using the data we measure. The general procedure will be outlined for a distribution of one parameter for simplicity, but the method is easily generalized to any number of parameters.

We consider a distribution $P(x|\mu)$. Here, x is a measured statistic, and μ is an underlying parameter we want to infer. If we want to make a confidence interval of "confidence" α, for each value of μ, we make an interval from some chosen c_1 to c_2 such that

$$\alpha = \int_{c_1}^{c_2} P(x|\mu)dx \tag{5.2}$$

. See a visual representation of these intervals together in figure 5.1. To get the confidence interval given a measurement x, you then take the interval of μ values contained within the belts intersecting x, as also shown in figure 5.1.

This is the general procedure, but the choice of c_1 and c_2 for each μ has to be made. The choice is in many situations obvious, such as for confidence intervals for the mean of a Gaussian distribution when the standard deviation is known. Clearly here, for any mean μ, we can choose c_1 and c_2 on either side of μ with equal distance from this mean. Or, if we want a upper or lower limit, we can bring either c_1 or c_2 to $\pm\infty$ and let the other side float to find the appropriate α.

The band also doesn't need to be continuous. We can frame the problem of these belts to any ordering principle which orders which values of x to add to the band first for a given value of μ, adding more values of x until we reach α.

5.2.1 The Feldman-Cousins Method

Choosing an ordering for confidence belts for the situation where your interval is near a boundary in values is non-trivial. These include situations such as the rate parameter of the Poisson distribution, or the standard deviation of the Gaussian, near 0. You may be tempted to choose a central interval around the mean for values of the data where the parameter doesn't include zero, then switch to an upper limit in the case it does. This method doesn't work properly though, as can be seen by drawing the confidence belts for this situation [61].

A unified method that will stitch these two situations together is often called the "Feldman-Cousins" method. In this method, for a value of the underlying model μ, we choose which x values to include in band by ordering them with the following metric:

$$R = \frac{P(x|\mu)}{P(x|\hat{\mu})} \tag{5.3}$$

where $\hat{\mu}$ is the value of μ that maximizes $P(x|\mu)$ for the given x.

5.2.2 Profiling

For making confidence regions of our Frequentist parameters of interest, we need a way of dealing with nuisance parameters. For our uses, nuisance parameters include all parameters we are not currently making confidence regions for, including all the systematic parameters, as well as the other oscillation parameters. When creating the confidence regions for 2 oscillation parameters at a time, the other floating 70 parameters must be dealt with.

For this, we use the technique of "profiling" over the nuisance parameters [13]. This means we choose values for the nuisance parameters where the likelihood is maximum. For every point of parameter space being fit to, $\vec{\theta}$, we maximize the nuisance parameters, $\vec{\phi}$:

$$\hat{\hat{\vec{\phi}}} = argmax_{\vec{\phi}} \mathcal{L}(\vec{\theta}, \vec{\phi}) \tag{5.4}$$

where note here $\vec{\theta}$ and $\vec{\phi}$ are the fit and nuisance parameters, not the oscillation and systematic parameters as before.

The Feldman-Cousins ordering now becomes

$$R = \frac{\mathcal{L}(\vec{\theta}, \hat{\hat{\vec{\phi}}})}{\mathcal{L}(\hat{\vec{\theta}}, \hat{\vec{\phi}})} \tag{5.5}$$

where it's important to note the bottom is the maximum over all parameter space, where in the top, $\vec{\phi}$ is maximized for each $\vec{\theta}$ individually.

5.2.3 Computational Considerations

Wilks' theorem [116] can be used as an approximation to compute the log likelihood values at different fit parameter points, as it shows the likelihood ratio asymptotically becomes a χ^2 distribution, allowing one to calculate thresholds for different confidence values at fit parameter points quickly. The conditions for Wilks' theorem to be accurate aren't met for our fit though, since we have relatively small sample sizes and boundaries in our fit parameters.

For this reason, using Wilks' theorem can give a quick approximate Frequentist fit, but a more accurate fit requires simulating large numbers of pseudo-experiments to get the more accurate distribution of confidence thresholds. Running the full Feldman-Cousins corrections with profiling for a fit to NOvA's oscillation parameters is therefore a very computationally expensive procedure, which is run on high performance computing (HPC) resources at the

National Energy Research Scientific Computing Center (NERSC) [107].

5.3 Bayesian Parameter Inference

In Bayesian statistics, we infer the parameters of our model (here called ‘θ’, which can be multidimensional) using data (‘x’, also possibly multidimensional) using Bayes’ rule:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta} \tag{5.6}$$

. Here, $p(x|\theta)$ is our model, and $p(\theta|x)$ is our “posterior”, which is the probability density of the parameters of the model, given the data we observed. We also need to specify $p(\theta)$, called the “prior”, which represents our knowledge of the model before the data. We can put information we already know about the model in the prior, or use an “uninformative” prior, which can be a uniform in the parameter you’re trying to measure. In general, for a prior with a non-zero density at all allowable points of the parameter space, the choice of prior won’t matter in the high statistics limit.

Note the bottom term is independent of θ, and serves as a normalization factor. Many texts therefore drop the bottom factor, and write Equation 5.6 as:

$$p(\theta|x) \sim p(x|\theta)p(\theta) \tag{5.7}$$

.

Also note the radically different interpretation than Frequentist statistics. Here in Bayesian statistics, we can talk freely about the probability that the parameters are in some range, and also transform the posterior to give the posterior of a related parameter.

Another large difference in Bayesian statistics is the treatment of nuisance parameters. Instead of profiling as is common in Frequentist statistics, Bayesian statistics marginalizes out nuisance parameters from the posterior distribution. If θ is our parameter of interest, and λ is a nuisance parameter, so the joint posterior is $p(\theta, \lambda|x)$, then we marginalize over the parameter λ by integrating over it:

$$p(\theta|x) = \int p(\theta, \lambda|x) d\lambda \tag{5.8}$$

5.3.1 Bayesian Hypothesis Testing

Bayesian hypothesis testing differs significantly from that of Frequentist statistics. Say we have two possible models, H_0 and H_1. The probability of a model could be given in the standard way as in Equation 5.6, where θ is instead here discrete, and the integral on the bottom becomes a sum. The issue with this is that the probability of a model given the data, such as $P(H_0|x)$, will be dependent on the priors. Also, there may be many different models, and we only want to compare two, H_0 and H_1.

So instead, it often useful to find the "Bayes Factor" [68], here called BF_{10}:

$$BF_{10} = \frac{P(x|H_1)}{P(x|H_0)} \tag{5.9}$$

. Note this will be the same as the ratios of their posterior probabilities assuming they have the same prior. This represents effectively the ratio of the evidence from the data for H_1 compared to H_0.

For hypotheses with models with ranges of continuous parameters, the Bayes factor becomes

$$BF_{10} = \frac{\int p(\theta_1|H_1) p(x|\theta_1, H_1) d\theta_1}{\int p(\theta_0|H_0) p(x|\theta_0, H_0) d\theta_0} \tag{5.10}$$

. For the results of the NOvA oscillation analysis, we can express the mass ordering and octant preference with a posterior probability assuming the priors of the two are the same, as well as the corresponding Bayes factor.

5.3.2 Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) is a method for inferring the posterior distribution of Bayesian parameters of interest. A Markov Chain is a series of objects where the next object of the chain is chosen from a probability distribution based on the last object, making a "chain" of sampled points. There are a number of MCMC algorithms that efficiently sample from any probability distribution based on a prescribed set of rules for choosing the next point based on the last point. By using MCMC to sample a large number of points from a Bayesian posterior, we can recreate its shape.

To sample from our oscillation parameter space, we sample from the joint posterior of all oscillation and systematic parameters simultaneously. To marginalized over the parameters not included in a particular result being shown, you simply drop the parameters you're not interested in from the chains. They're implicitly marginalized over this way, as the frequency of that parameter's occurrence in its respective phase space will weight the chain appropriately.

MCMC chain length must be sufficiently long so the chain has ample time to explore the entire phase space of the target distribution. The start point of the chain isn't important, as the chain should "fall into" the distribution over time. The beginning values of the chain may be biased though, before it has "found" the distribution, so it's common to remove some segment of the beginning of the chain, usually called "burn-in".

Credible Intervals

To make intervals like confidence intervals from Frequentist statistics, called "credible" intervals in Bayesian statistics, you bin the values sampled from the posterior. You then choose the appropriate number of bins necessary to reach the desired level of confidence, i.e., you add bins to the credible interval until the desired percent of probability of the posterior is contained.

There are numerous schemes for deciding the ordering of which bins to add first. We use the method of adding bins by largest to smallest value, which will in general give the smallest region. This method doesn't guarantee a contiguous region, as is apparent by looking at the results from later sections, but you can choose a scheme that does, such as choosing a central region, if a single contiguous region is needed or wanted.

Importance Sampling

Importance sampling is a powerful method that allows samples drawn from one distribution to be weighted to reflect another distribution. For our purposes in MCMC, this means that MCMC samples drawn sampled from a posterior using one prior can be individually reweighted to represent the posterior with another choice of prior. Therefore, we can create MCMC chains with one prior, and reweight them after, to give the distribution for a different prior.

If we drew samples with an original prior in a variable x, $p(x)$, and we want to reweight to represent a posterior with a different prior $p'(x)$, then we weight each of the draws of a sample, i, with value, x_i, with weight

$$\frac{p'(x_i)}{p(x_i)} \tag{5.11}$$

.

5.3.3 NOvA's MCMC Implementations

NOvA has implemented two MCMC algorithms: Metropolis and Hamiltonian Monte Carlo (HMC) [14]. The Metropolis-based sampler was implemented from scratch, and named ARIA. The HMC sampler uses an external library, called Stan, and has implemented an interface to it into NOvASoft.

ARIA

The Metropolis algorithm is a remarkably simple yet powerful algorithm. To sample a new point based on a previous point, we first choose a candidate point from a symmetric proposal distribution, the most common choice being the multivariate Gaussian. This proposed point is "accepted" with probability $Min[\frac{P(x')}{P(x)}, 1]$. Otherwise, the previous point is kept as the new point. An illustration of this process for a 1 dimensional Gaussian distribution is shown in figure 5.2.

NOvA's implementation is called ARIA, named after Arianna W. Rosenbluth, who wrote the first computer implementation of the method [91]. ARIA was originally developed for use in the NOvA-T2K joint fit [113].

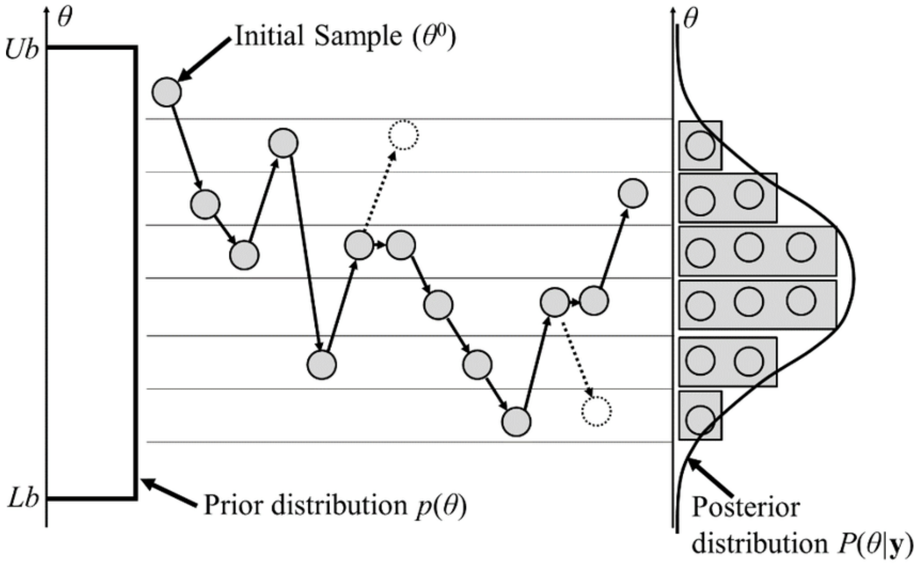


Figure 5.2: An illustration of the Metropolis algorithm. Image taken from external reference [89].

Stan

Stan [41] is a framework that implements the No U-Turn Sample (NUTS) [76], which is an adapted version of Hamiltonian Monte Carlo (HMC) [56]. HMC is a sophisticated algorithm that chooses a next point by simulating a particle through a potential defined as a function of the target probability distribution. It simulates the particle "falling" through the distribution. NUTS adapts this, simulating the particle moving both forward and backward in time, and adds a mechanism where it stops when the particle makes a "U-Turn", which is when it stops and starts moving in the opposite direction. It then chooses a point randomly along this path it has drawn out. The point is then accepted or rejected, in a similar fashion to Metropolis.

This complicated algorithm has a distinct advantage over Metropolis, in that adjacent points in the chain are significantly less correlated between one another, and the algorithm can explore the parameter space far more efficiently.

Mass Ordering Switching

The mass ordering is a more complicated consideration than other continuous parameters for MCMC. Typically, MCMC will have a difficult time getting from the "island" of one mass ordering to the other, so an MCMC chain started in one will stay in that mass ordering for the whole chain. For ARIA, with the Metropolis algorithm, this is simply solved by introducing a binary mass ordering parameter, where when choosing new proposed parameters, you also have a chance of proposing a mass ordering switch, where then the proposed point has a chance to switch ordering by the normal rules of the Metropolis algorithm.

For Hamiltonian Monte Carlo which Stan uses, discrete parameters are more awkward to handle, as you can't differentiate along this parameter. For this, we do what is often done with discrete parameters in HMC, which is to use another MCMC algorithm to handle this parameter [38]. Here we use the Metropolis algorithm we use for ARIA for the mass ordering, where after the HMC is done for other parameters, we propose a mass ordering flip, and then accept or reject it based on the normal rules of the Metropolis algorithm.

Computational Considerations

MCMC methods has distinct computational advantages over the Frequentist fit. Since many chains can be run independently to get large statistics, we can get sufficiently large MCMC samples from running many chains in parallel on FermiGrid [49]. Large enough samples can be generated in 1-2 days, as opposed to the full Feldman-Cousins corrected Frequentist fit, which takes ∼weeks to run on NERSC.

Also, the Frequentist confidence intervals must be run separately for each parameter space, such as δ_{CP} vs $sin^2\theta_{23}$ and $sin^2\theta_{23}$ vs Δm^2_{32}. In contrast, MCMC only has to be run once, then the chains can be marginalized over any set of parameters after, getting posteriors in any parameter space you want without needing to re-run MCMC.

Debugging MCMC can be more complicated than Frequentist fits, though, as it's a stochastic process, unlike Frequentist fits where each step can be checked in a more straightforward way. For this reason, we have a number of numerical diagnostics that validate MCMC is operating as expected, as well as the two independent MCMC implementations which validate against each other. Validation is the topic of section 5.3.4.

5.3.4 Validation of NOvA's Samplers

Diagnostics

MCMC is a very powerful method, but can be difficult to diagnose and validate compared to Frequentist methods of fitting. For this reason, many methods have been developed to validate MCMC methods. The most simple of these is simply to look at the "traces" of MCMC chains, which is simply plotting the value of one parameter from a chain against the step of the chain. This will help show you whether or not you have removed enough burn-in, and if you haven't you may see the beginning of the trace will look visibly different from the rest.

An example of traces from ARIA for oscillation parameters is show in figure 5.3 Notice the beginning of the traces are distinctly different from later aspects of it, which can help motivate the choice of how much burn in you need. For ARIA, we settled on a burn-in of 100,000 for this 2024 analysis.

Another standard diagnostics is the autocorrelation, r_k, defined as

$$r_k = \frac{\sum_{i=1}^{n-k}(\theta_i - \overline{\theta})(\theta_{i+k} - \overline{\theta})}{\sum_{i=1}^{n}(\theta_i - \overline{\theta})^2} \tag{5.12}$$

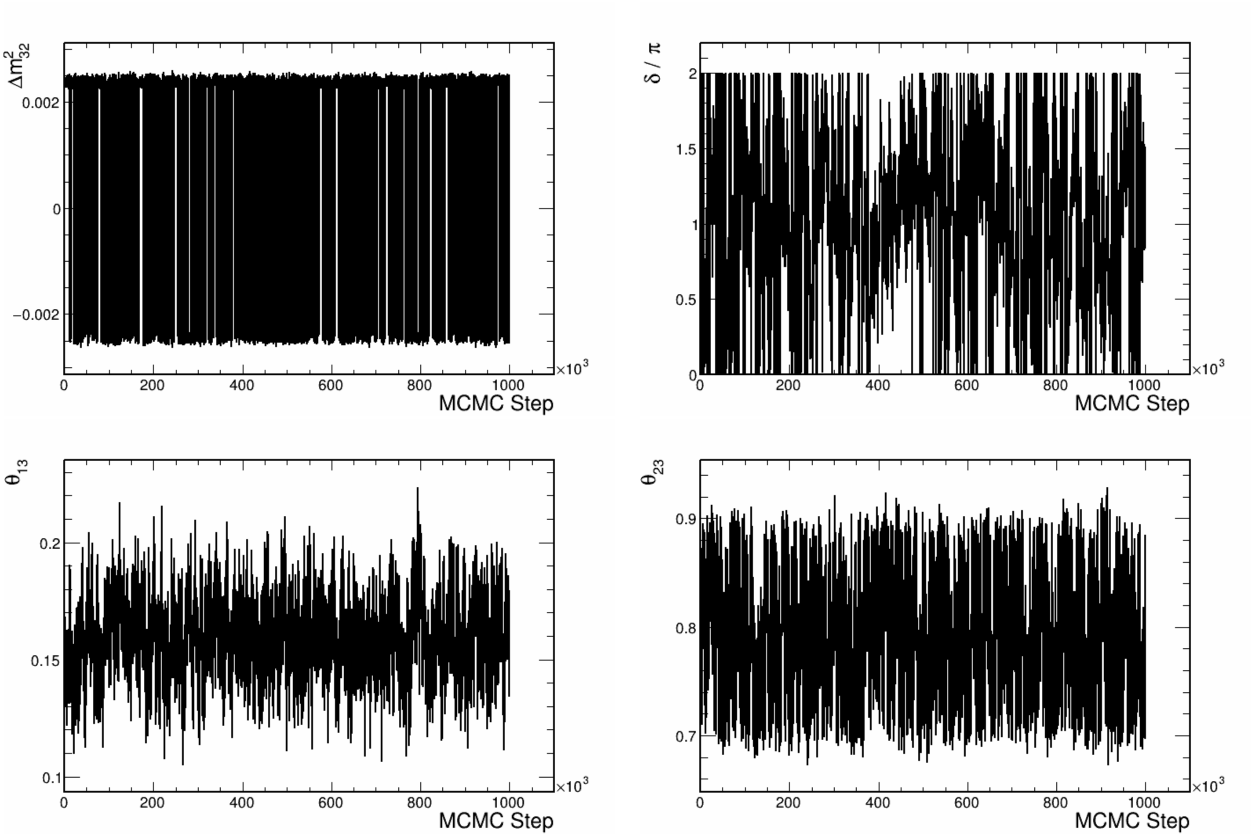


Figure 5.3: Traces of oscillation parameters from MCMC sampler ARIA

can roughly be thought of as the correlation between elements of the chain with the elements k later [38]. An example of this for the oscillation parameters from ARIA can be seen in figure 5.4. Notice how r_k starts around 1, then drops to 0 at higher lag.

Lastly, you can look at $\hat{R}$ to monitor the convergence of the chains. $\hat{R}$ will drop to 1 as chain length goes to infinity, and will tell you if the chain length chosen is long enough [68]. The $\hat{R}$ should be below 1.05, and ideally mostly under 1.01. $\hat{R}$ for each of 500 Stan chains is tabulated and shown in figure 5.5 Note around 10 of the 500 chains have a slightly elevated $\hat{R}$ for Δm^2_{32}.

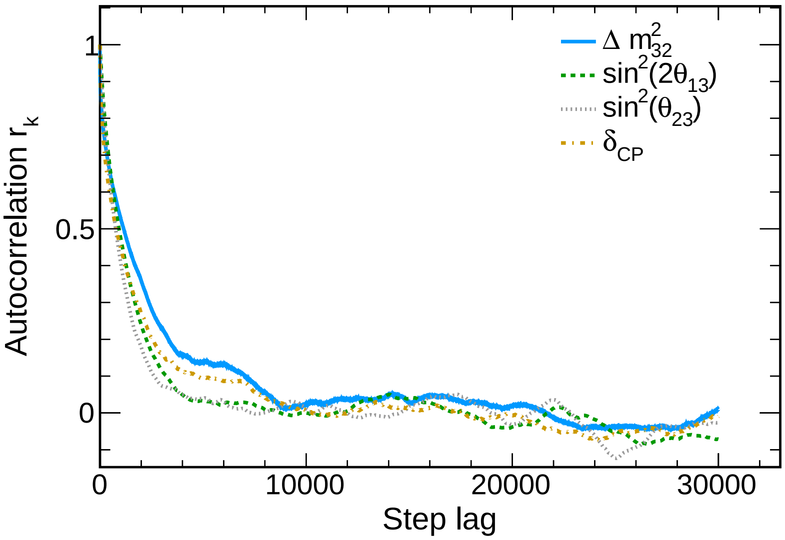


Figure 5.4: Autocorrelation of oscillation parameters from MCMC sampler ARIA

Stan-ARIA Comparisons

Another powerful validation method is provided by our two MCMC implementations. The samplers are implemented entirely separately, with one being external to NOvA altogether, so agreement between them strongly backs up their validity.

Figure 5.6 shows Stan - ARIA comparisons for Asimov fake data at the 2020 best fit point. The comparisons show great agreement.

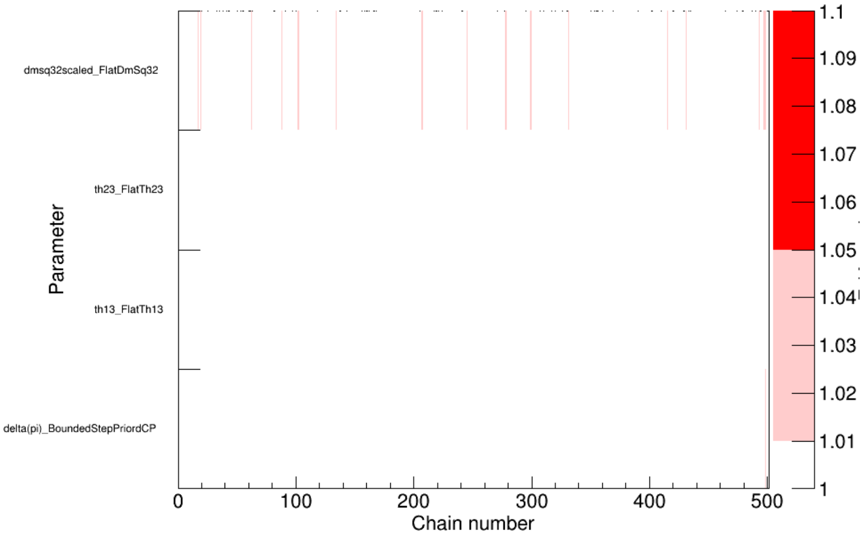
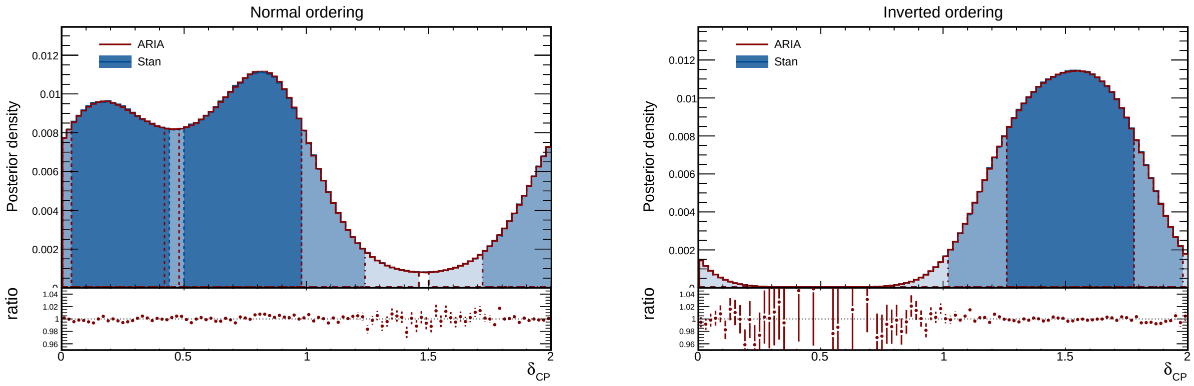
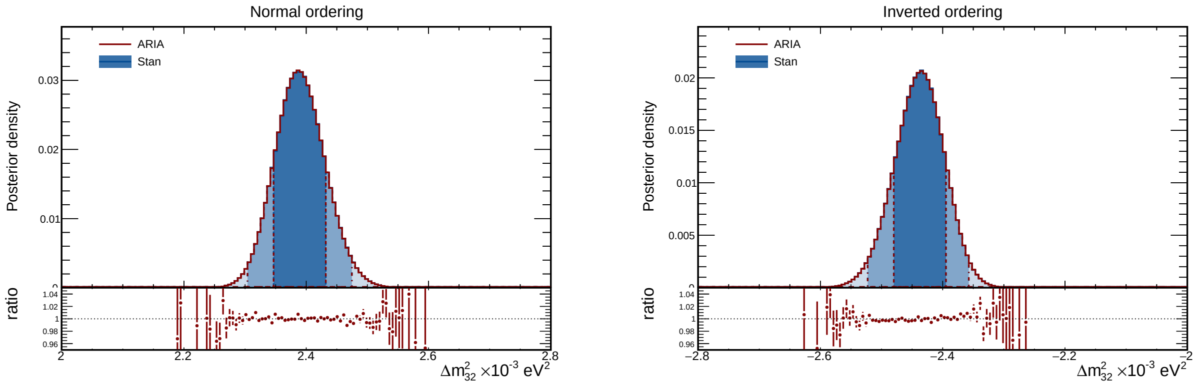


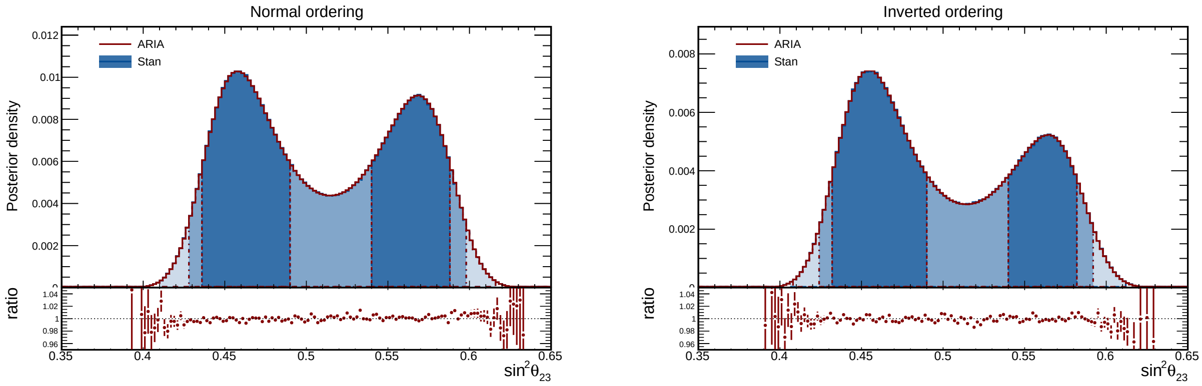
Figure 5.5: $\hat{R}$ for 500 Stan chains from a fake data fit, for each oscillation parameter. The left most bin represents $\hat{R}$ of all chains together, which the other bins are the per chain $\hat{R}$.



(a) δ_{CP} NO (b) δ_{CP} IO



(c) Δm^2_{32} NO (d) Δm^2_{32} IO



(e) $sin^2\theta_{23}$ NO (f) $sin^2\theta_{23}$ IO

Figure 5.6: ARIA and Stan comparisons for Asimov fake data at the 2020 NOvA best fit point

Asimov Point	$\Delta m^2_{32}(\times 10^{-3}\text{eV}^2)$	$sin^2\theta_{23}$	δ_{CP}	Description
A	2.41	0.57	0.82π	2020 NOvA Best Fit
B	2.51	0.53	1.50π	Near T2K's 2020 Best Fit
C	-2.45	0.50	1.5π	Near Global Fitter Preference
D	2.40	0.46	1.0π	LO and CP Conserving
E	-2.44	0.46	0.5π	Far From NOvA's Preferred Region

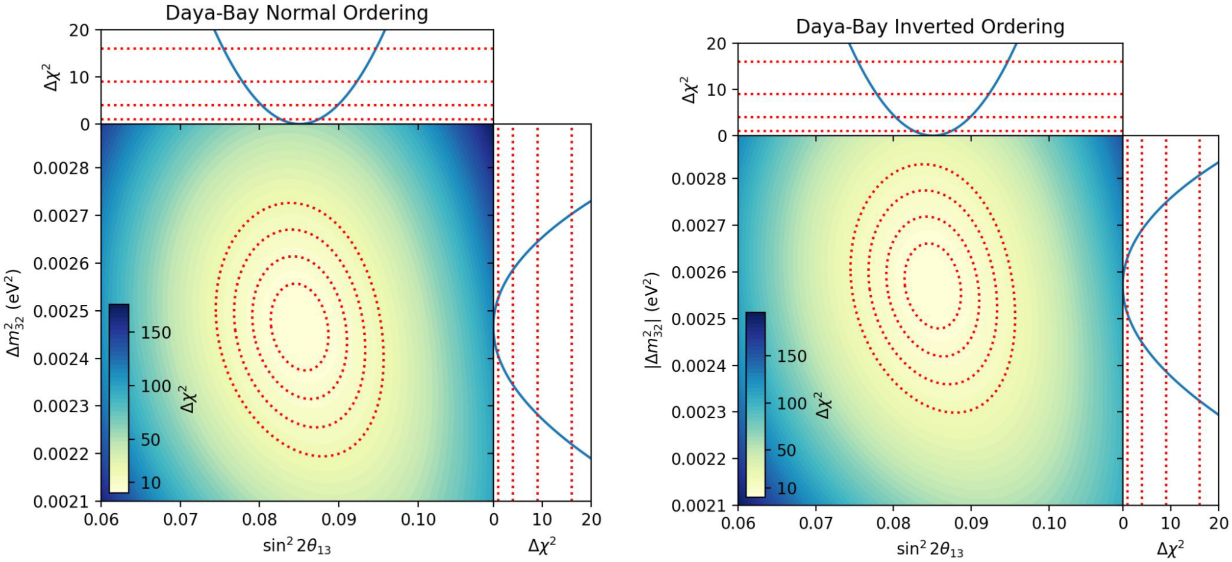
Table 5.1: Standard "Asimov Points" chosen to test NOvA fitters for this analysis

5.3.5 Reactor Constraints

Traditionally, both in the Frequentist analyses and the first Bayesian analysis, NOvA introduces a constraint on $sin^2(2\theta_{13})$ from the latest high-precision reactor experiment. For the 2024 NOvA result, we use Daya Bay's latest measurement [25]. For the Bayesian analysis, this is implemented as a normal prior on $sin^2(2\theta_{13})$ using their measurement. We call this the "1D" reactor constraint (or Daya Bay constraint).

Besides a measurement of $sin^2(2\theta_{13})$ alone, Daya Bay also reports their results in $sin^2(2\theta_{13})$ vs Δm^2_{32}, which are reported separately for normal and inverted ordering hypotheses, shown in figure 5.7. Daya Bay is not sensitive to the mass ordering itself, that is the sign of Δm^2_{32}, but we can use the information in this 2D reactor constraint in each mass ordering to further constrain the mass ordering in NOvA. This is implemented as an alternative prior on $sin^2(2\theta_{13})$ vs Δm^2_{32}. We call this the "2D" reactor constraint (or Daya Bay constraint).

With importance sampling, it is easy for us to sample with a prior flat in $sin^2(2\theta_{13})$, then reweight the events later to get the results for the 1D and 2D reactor constraint without needing to resample



(a) (b)

Figure 5.7: $sin^2(2\theta_{13})$ vs Δm^2_{32} contours from Daya Bay, in normal (a) and inverted (b) orderings, taken from reference [48]

5.3.6 Fake Data Fits

For the 2024 NOvA analysis, we used a set of standard test oscillation parameter point to test fitters [46]. Their names and parameters are listed in Table 5.1, as well as a description justifying its choice.

For these fake data studies, we will use “Asimov” fake data, in which the fake data is the average fake data you get for a choice of oscillation parameters. That is, the bin values of energy spectra are the expectation value of the oscillation parameter, and don’t include the Poisson fluctuations in each bin. Studies with the Poisson fluctuations in each bin can also be done, but note there are many different Poisson fluctuated fake datasets for the same oscillation point, where the Asimov fake data used here is their average value.

The reactor and solar parameter values will be kept constant between different points:

$$\Delta m^2_{21} = 7.53 \times 10^{-5}\text{eV}^2$$

$$sin^2\theta_{12} = 0.307$$

$$sin^2\theta_{13} = 2.20 \times 10^{-2}$$

The Asimov fake data fits here are with 12.5M Stan samples, which isn't as high statistics as you'd like for the full data fits and may be jagged in places, but it's more than enough to make conclusions about the points of interest.

Asimov A – No reactor constraint

The NOvA 2020 best fit point, Asimov A, is the most common Asimov point to look at as it gives the best guess as what our results may look like. With no reactor constraint, the prior in $sin^2(2\theta_{13})$ is flat, and this shows us how well NOvA can measure these parameters with NOvA data only.

Various 1D and 2D posterior densities are shown in figures 5.8 - 5.14. Notice in 5.8 how NO and IO prefer differing regions of δ_{CP}, see figure 2.12.

For 5.9, it can be seen that the lower octant of θ_{23} is preferred. This is true only with the reactor constraint, and the reason why can be seen in figure 5.12. The Daya Bay θ_{13} measurement is shown overlapping the 2D posterior, and it becomes clear the upper octant is more consistent with the reactor measurement. For Δm^2_{32}, NO is preferred but only slightly.

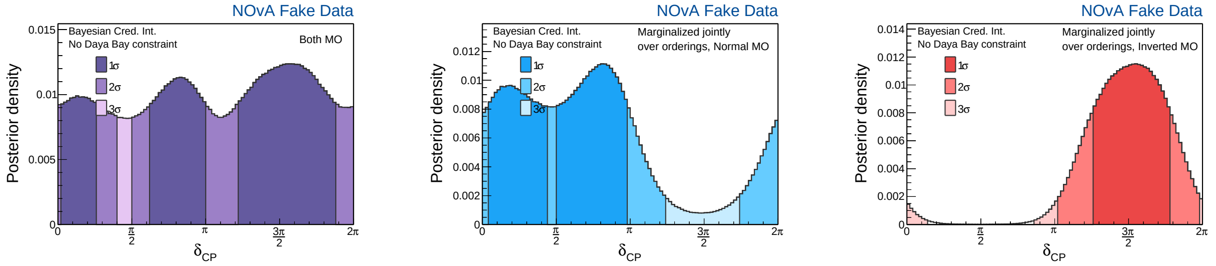


Figure 5.8: δ_{CP} with no reactor constraint applied, with fake data drawn from Asimov A

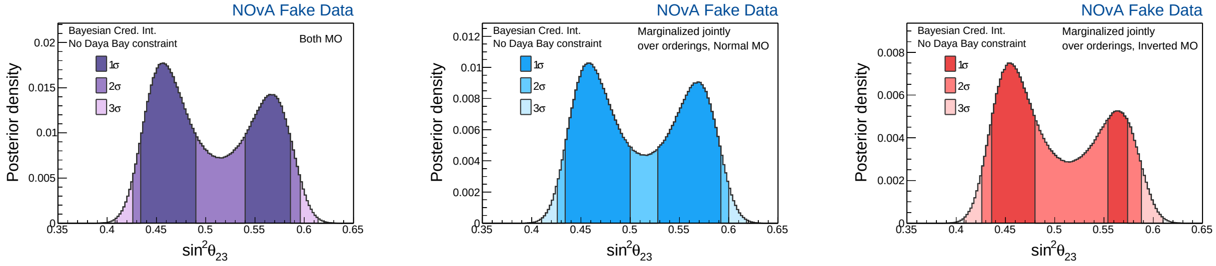


Figure 5.9: $sin^2\theta_{23}$ with no reactor constraint applied, with fake data drawn from Asimov A

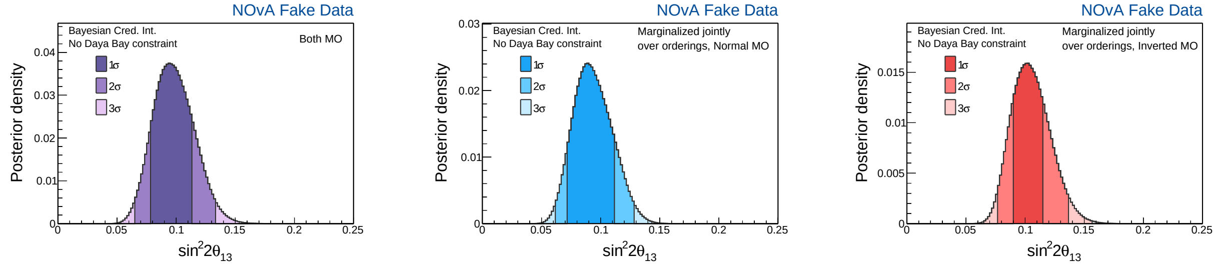


Figure 5.10: $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov A

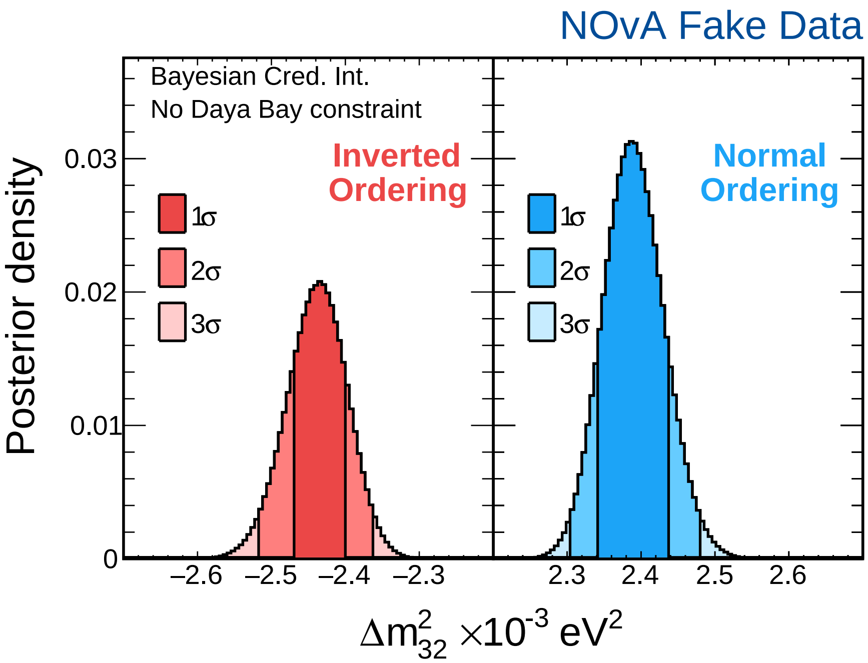


Figure 5.11: Δm^2_{32} with no reactor constraint applied, with fake data drawn from Asimov A

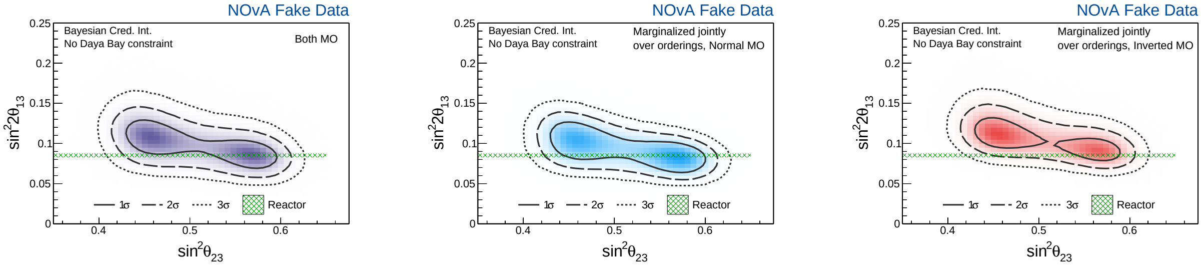


Figure 5.12: $sin^2\theta_{23}$ vs $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov A

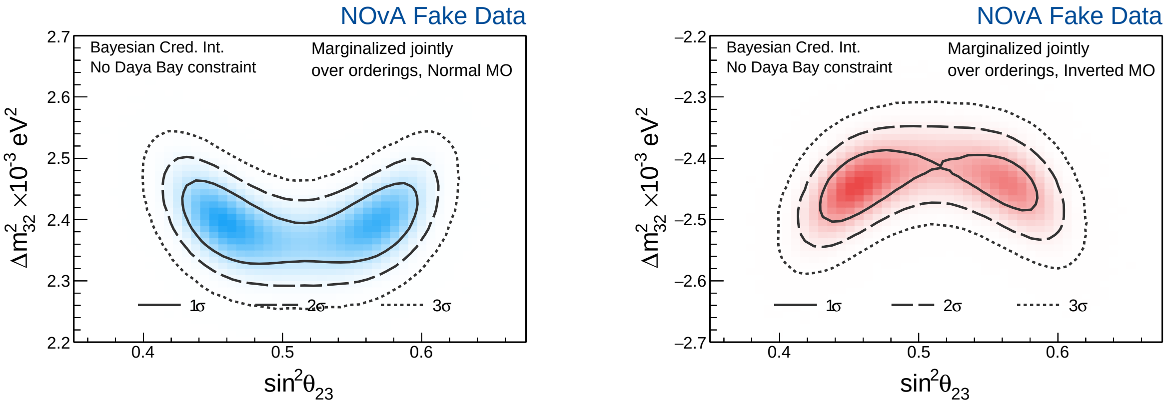


Figure 5.13: $sin^2\theta_{23}$ vs Δm^2_{32} with no reactor constraint applied, with fake data drawn from Asimov A

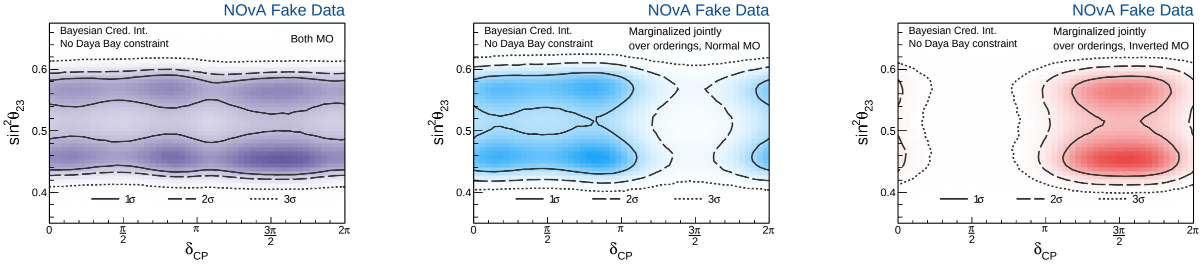


Figure 5.14: δ_{CP} vs $sin^2\theta_{23}$ with no reactor constraint applied, with fake data drawn from Asimov A

Asimov A – 1D reactor constraint

Figures 5.15 - 5.19 show various posteriors with the traditional 1D reactor constraint applied as a prior. Figure 5.15 shows δ_{CP} doesn't change significantly after applying this constraint. θ_{23} on the other hand does see a significant shift towards upper octant, as discussed in the no reactor constraint section. Δm^2_{32} also sees a significant shift towards normal ordering.

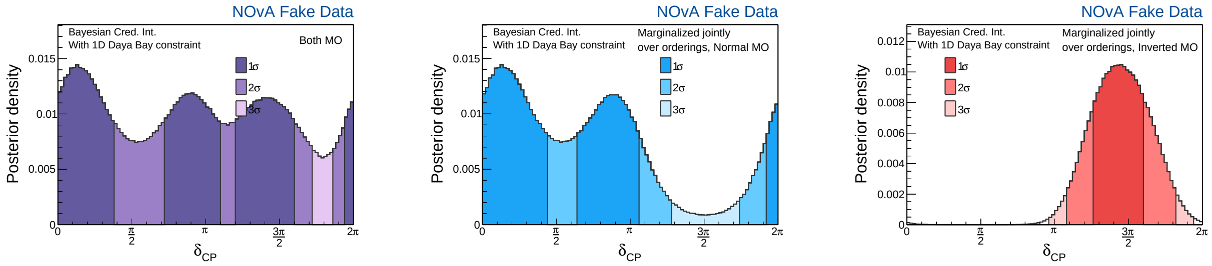


Figure 5.15: δ_{CP} with the 1D reactor constraint applied, with fake data drawn from Asimov A

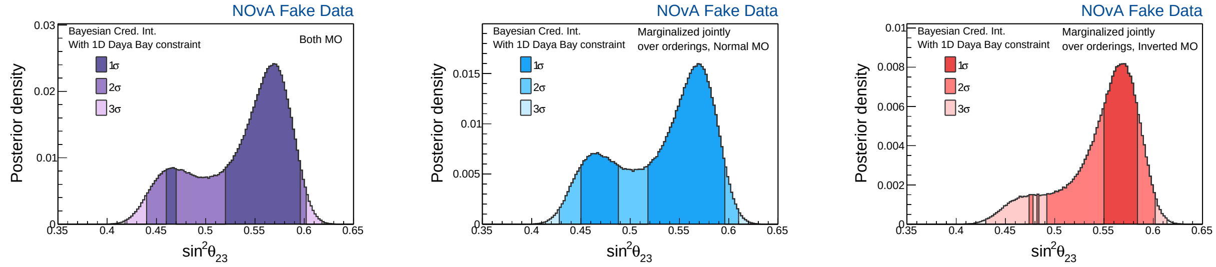


Figure 5.16: $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov A

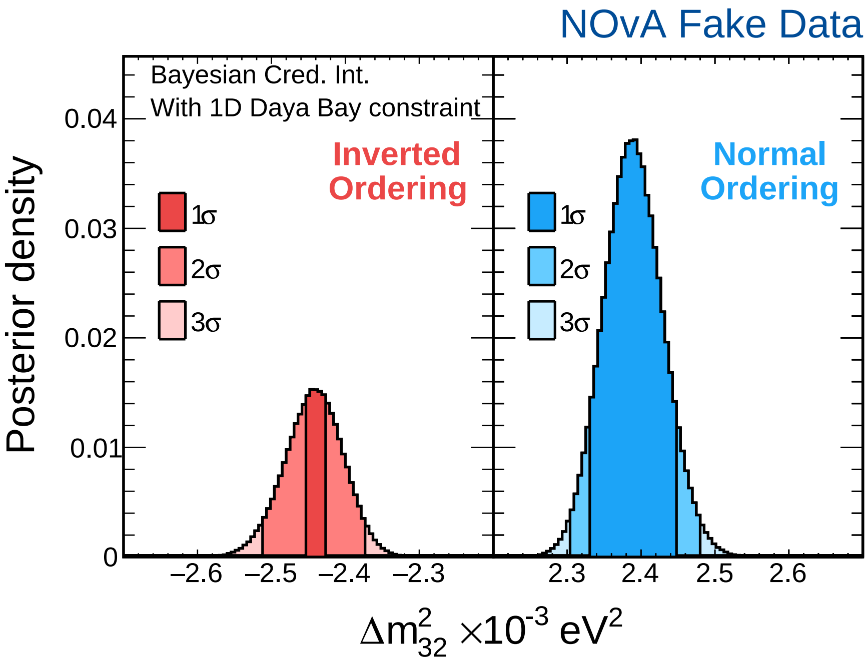


Figure 5.17: Δm^2_{32} with the 1D reactor constraint applied, with fake data drawn from Asimov A

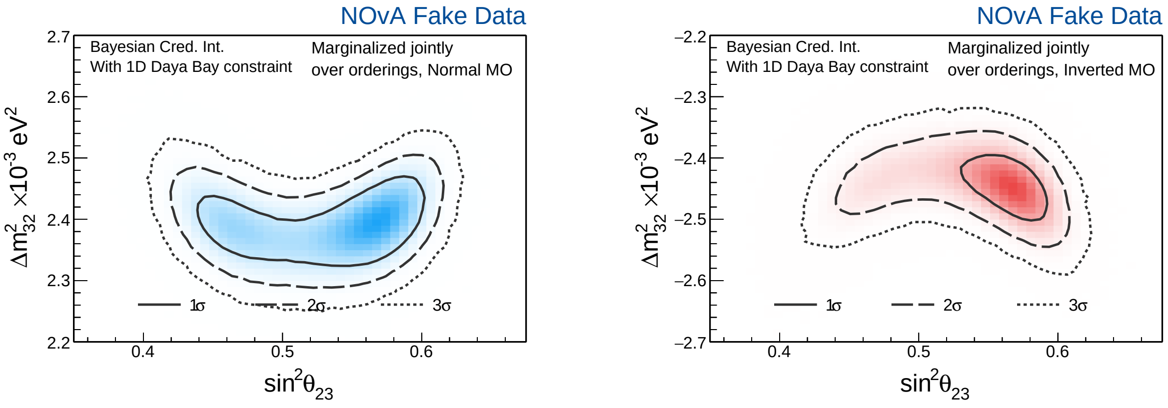


Figure 5.18: $sin^2\theta_{23}$ vs Δm^2_{32} with the 1D reactor constraint applied, with fake data drawn from Asimov A

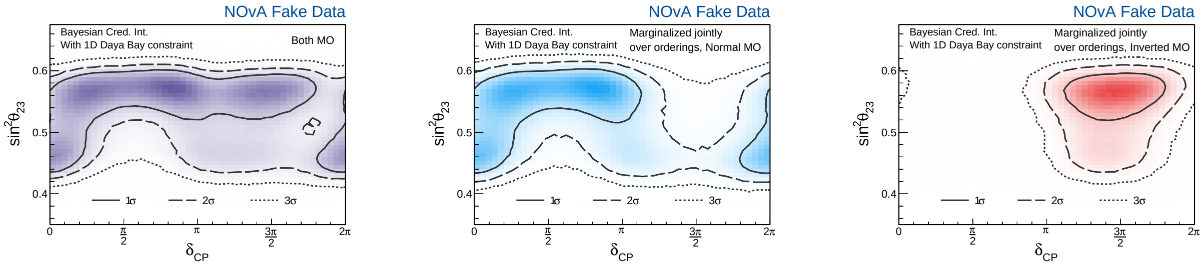


Figure 5.19: δ_{CP} vs $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov A

Asimov A – 2D reactor constraint

Figures 5.20 - 5.24 show Asimov A fake data fits, this time with the 2D reactor constraint. The IO model has even more tension with Daya Bay in this full phase space, pushing even further towards NO. The octant on the other hand becomes less decisive, with a small move back towards lower octant when compared with the 1D reactor constraint.

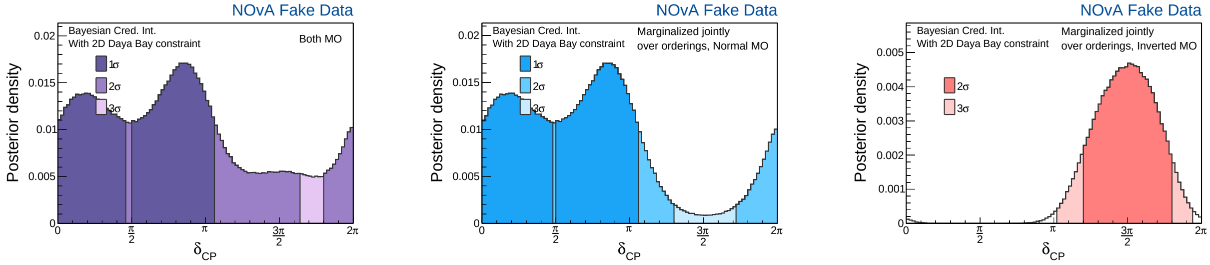


Figure 5.20: δ_{CP} with the 2D reactor constraint applied, with fake data drawn from Asimov A

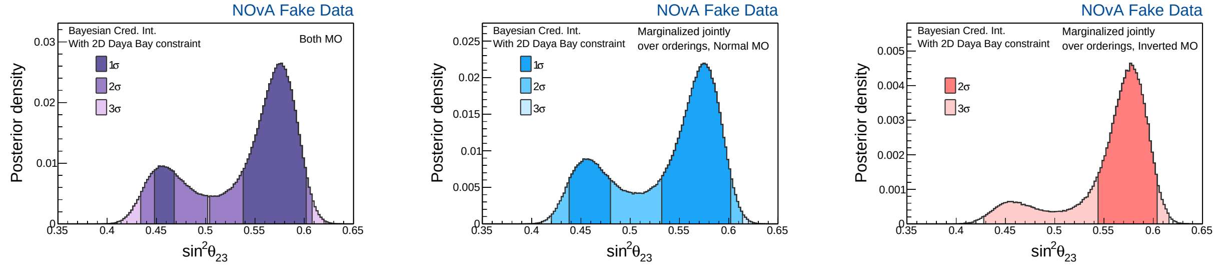


Figure 5.21: $sin^2\theta_{23}$ with the 2D reactor constraint applied, with fake data drawn from Asimov A

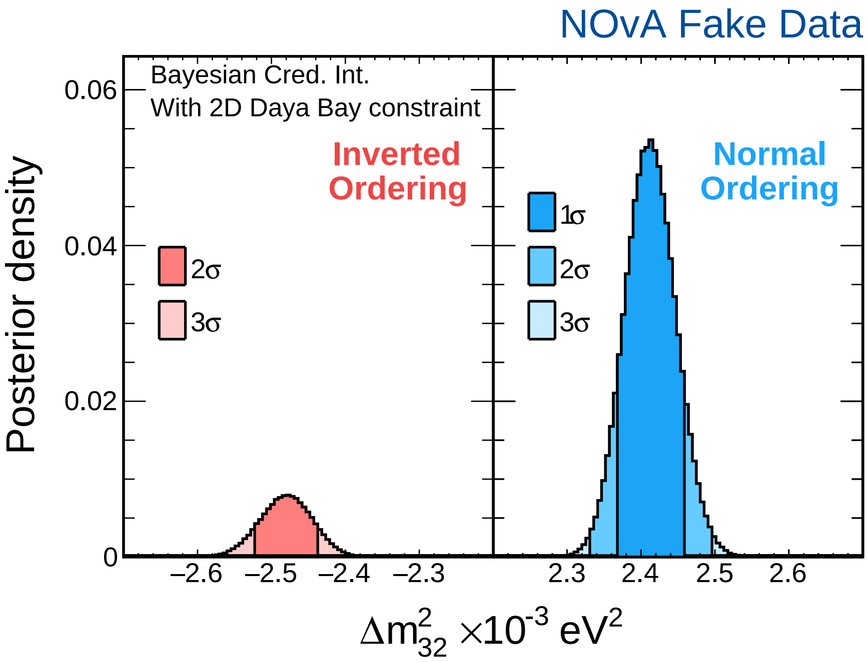


Figure 5.22: Δm^2_{32} with the 2D reactor constraint applied, with fake data drawn from Asimov A

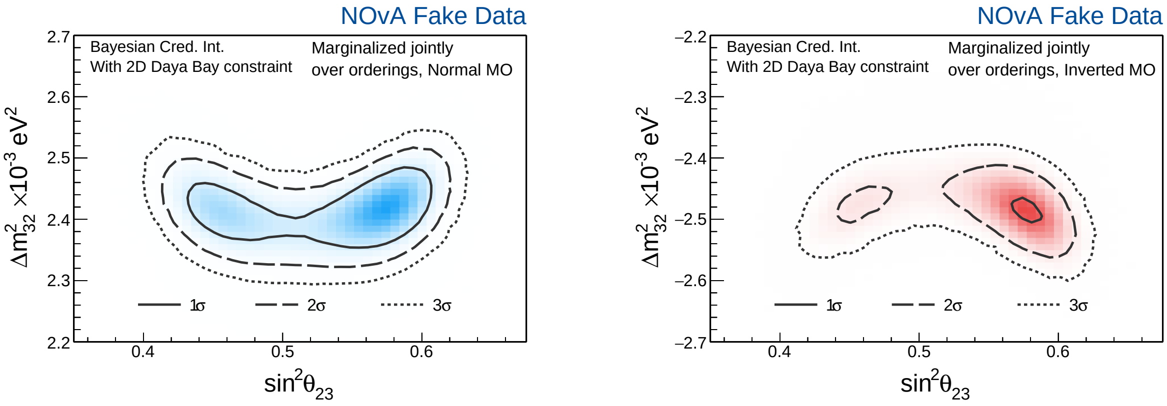


Figure 5.23: $sin^2\theta_{23}$ vs Δm^2_{32} with the 2D reactor constraint applied, with fake data drawn from Asimov A

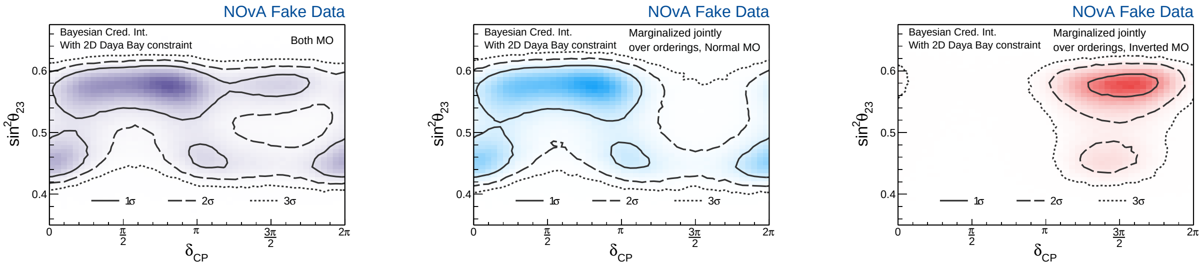


Figure 5.24: δ_{CP} vs $sin^2\theta_{23}$ with the 2D reactor constraint applied, with fake data drawn from Asimov A

Asimov B

Asimov B is fairly close to Asimov A, except Δm^2_{32} is somewhat higher, $sin^2\theta_{23}$ is slightly lower towards max mixing, and the largest difference is δ_{CP} which is all the way at $\frac{3\pi}{2}$, a region disfavored by NOvA for the NO model, but the best fit point for T2K.

The fake data fits for Asimov B are shown in figures 5.25 - 5.31. The most striking detail is the order preference, where NO is now highly preferred. Many IO plots look quite jagged do to lack of statistics. With 2D reactor constraint here, we see a NO preference of 99.98 %, over 3 σ.

Besides the large order preference, the same region of δ_{CP} is preferred now in both NO and IO as seen in figure 5.25. A fairly strong upper octant preference is shown in figure 5.26. A normal ordering preference is shown before the reactor constraint, but figure 5.29 shows why the reactor constraint further strengthens the NO preference, as the reactor θ_{13} value ends up far outside the IO posterior.

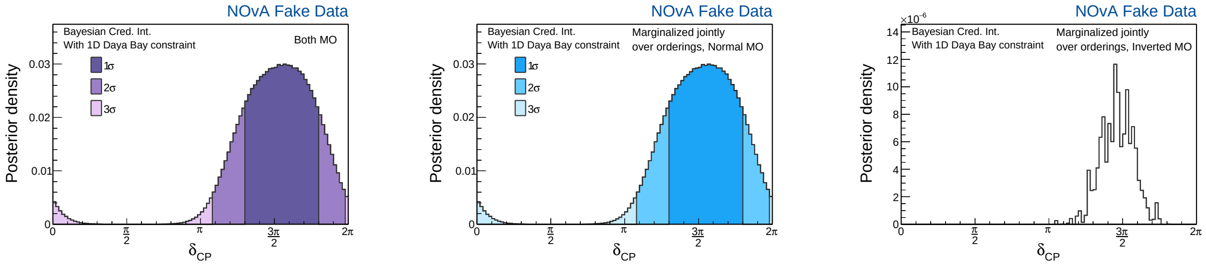


Figure 5.25: δ_{CP} with the 1D reactor constraint applied, with fake data drawn from Asimov B

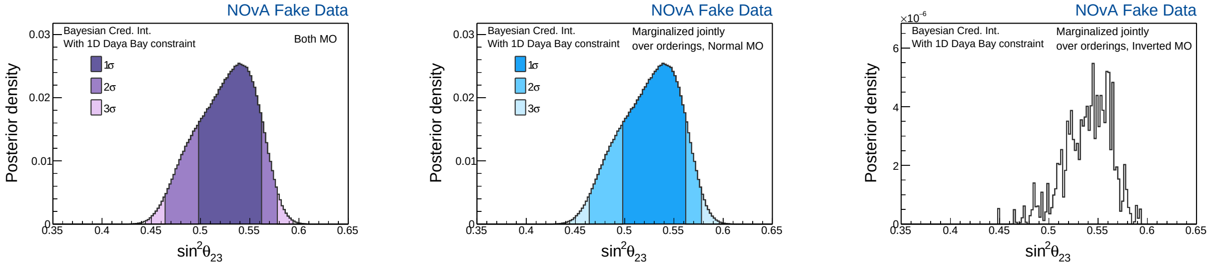


Figure 5.26: $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov B

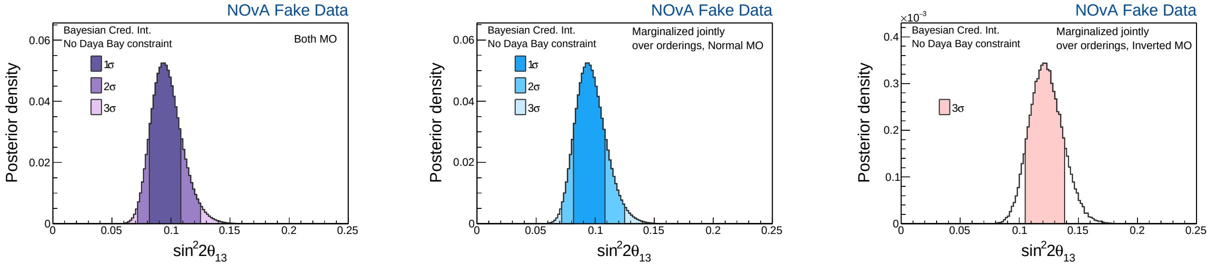


Figure 5.27: $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov B

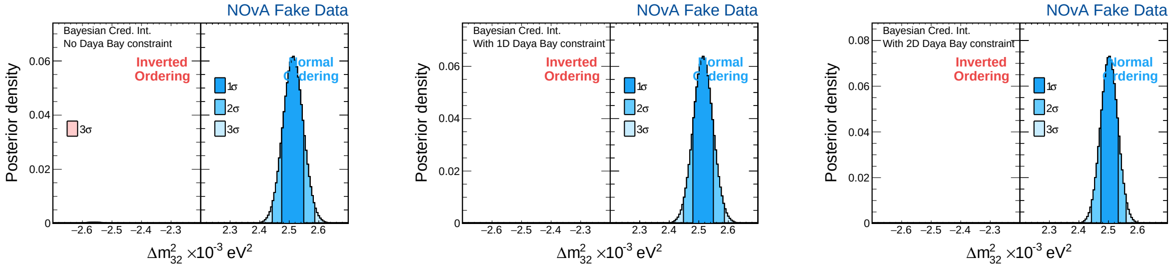


Figure 5.28: Δm^2_{32} with the different reactor constraints applied, with fake data drawn from Asimov B

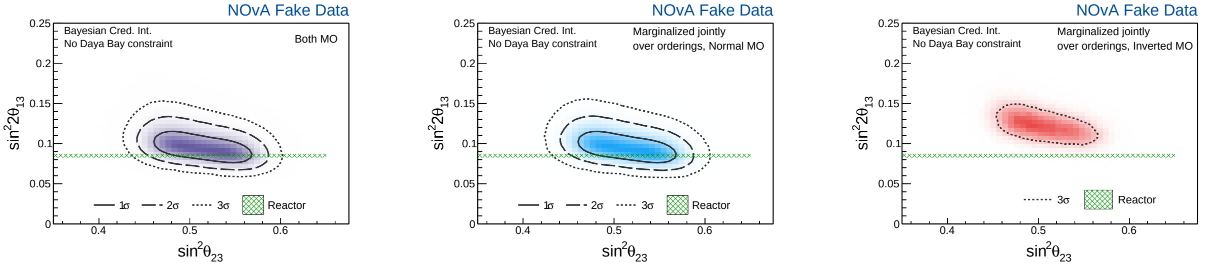


Figure 5.29: $sin^2\theta_{23}$ vs $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov B

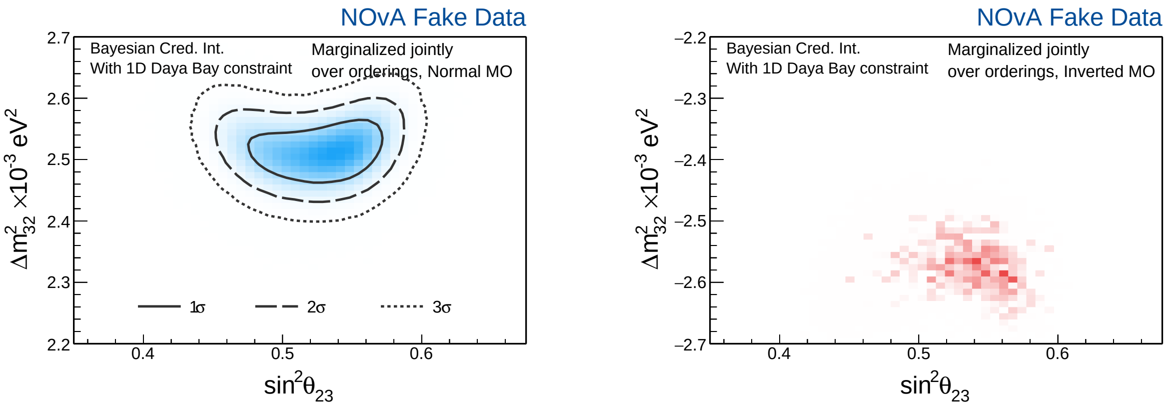


Figure 5.30: $sin^2\theta_{23}$ vs Δm^2_{32} with the 1D reactor constraint applied, with fake data drawn from Asimov B

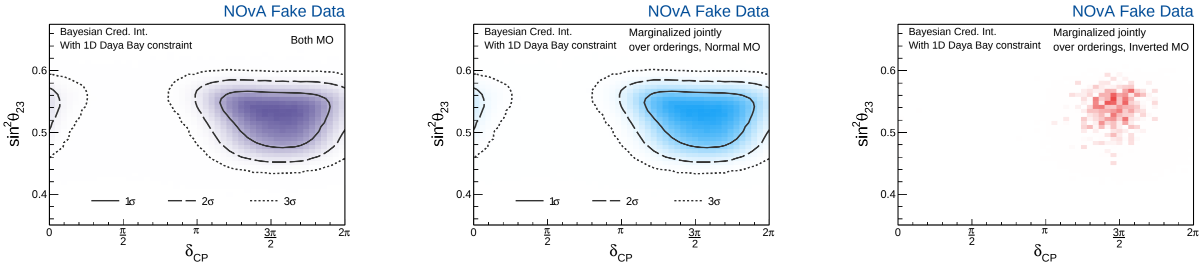


Figure 5.31: δ_{CP} vs $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov B

Asimov C

Figures 5.32 - 5.38 show fake data fits at Asimov C, near the global fitter preferred region in the IO, with δ_{CP} being $\frac{3\pi}{2}$. θ_{23} is at the max mixing point.

Here, δ_{CP} in figure 5.32 looks like Asimov A, where NO and IO prefer opposite regions, because in NO and IO, values of δ_{CP} prefer opposite sides of ν_e - $\bar{\nu_e}$ count space, see figure 2.12. Also, mass ordering is poorly differentiated, as can be seen in figure 5.35. NO/IO are in comparable proportion for no reactor constraint and the 1D reactor constraint, and NO is actually preferred more strongly with the 2D reactor constraint. The reason for this unexpected feature is because the assumed (negative) value of Δm^2_{32} in Asimov C is in tension with the preferred value of Δm^2_{32} in Daya Bay's $sin^2(2\theta_{13})$ vs Δm^2_{32} surface in the IO.

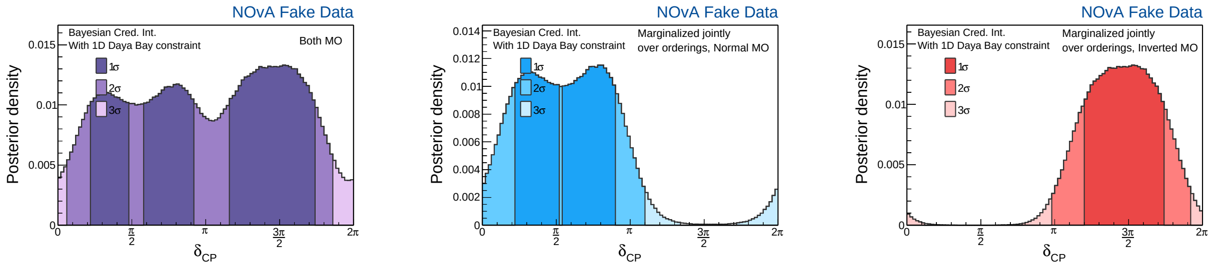


Figure 5.32: δ_{CP} with the 1D reactor constraint applied, with fake data drawn from Asimov C

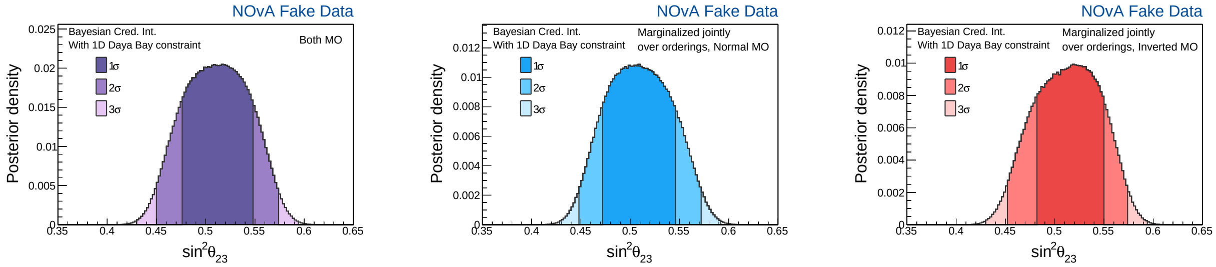


Figure 5.33: $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov C

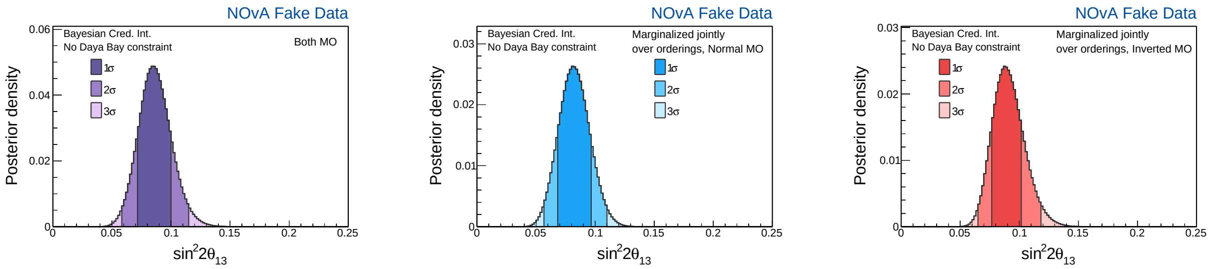


Figure 5.34: $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov C

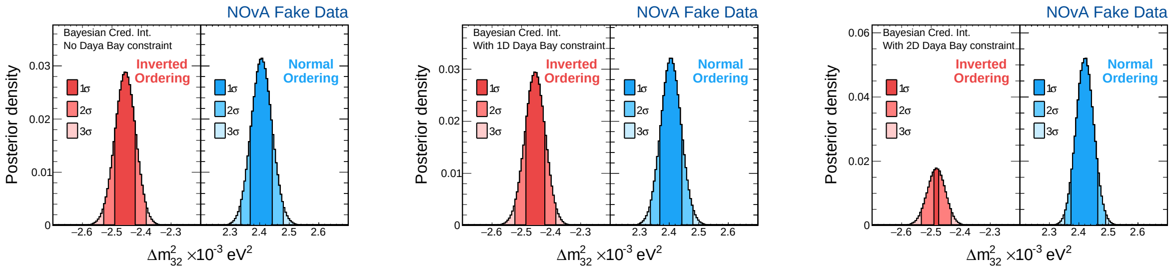


Figure 5.35: Δm^2_{32} with the different reactor constraints applied, with fake data drawn from Asimov C

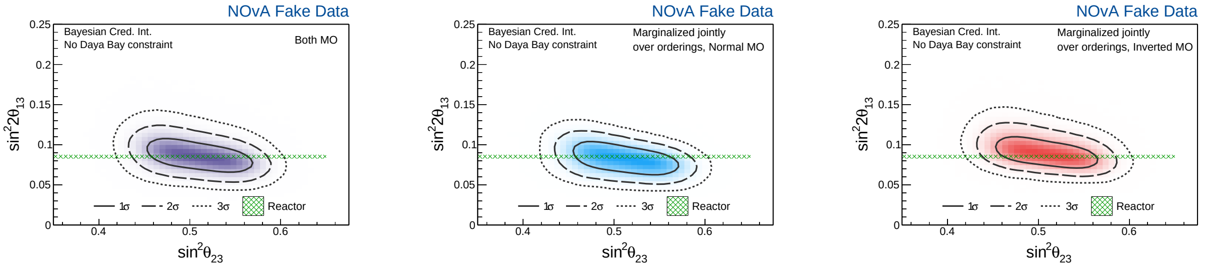


Figure 5.36: $sin^2\theta_{23}$ vs $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov C

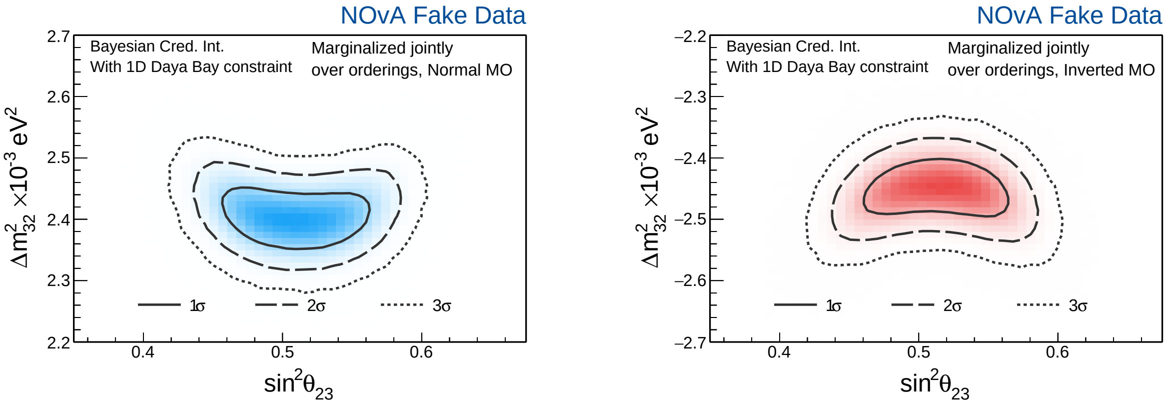


Figure 5.37: $sin^2\theta_{23}$ vs Δm^2_{32} with the 1D reactor constraint applied, with fake data drawn from Asimov C

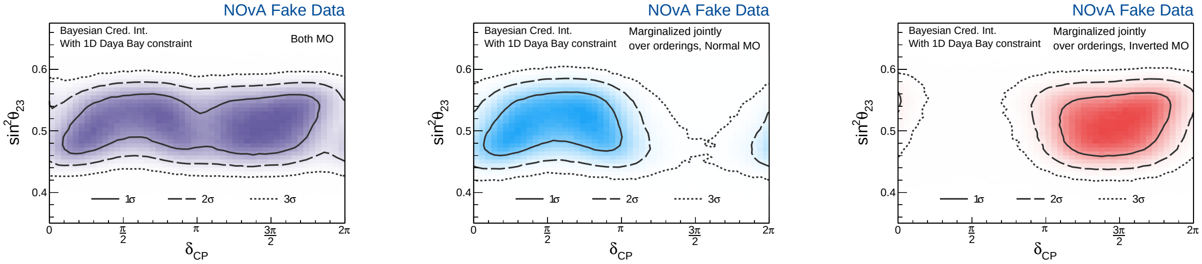


Figure 5.38: δ_{CP} vs $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov C

Asimov D

Asimov D fake data fits are in figures 5.39 - 5.45. Asimov D is a point in the lower octant and a CP conserving value of δ_{CP}.

δ_{CP} is similar in preferred region to Asimov A, with separate preferred spaces for δ_{CP} in NO and IO assumptions. The posterior peaks around 1.0π in NO, the "correct" Asimov point. In figure 5.40, you see θ_{23} prefers the lower octant in both NO and IO

For figure 5.42, we see the mass ordering prefers the NO with no reactor constraint, which gets slightly weaker actually with the 1D constraint, but adding the Δm^2_{32} information as well in the 2D reactor constraint pushes things to a fairly strong NO preference.

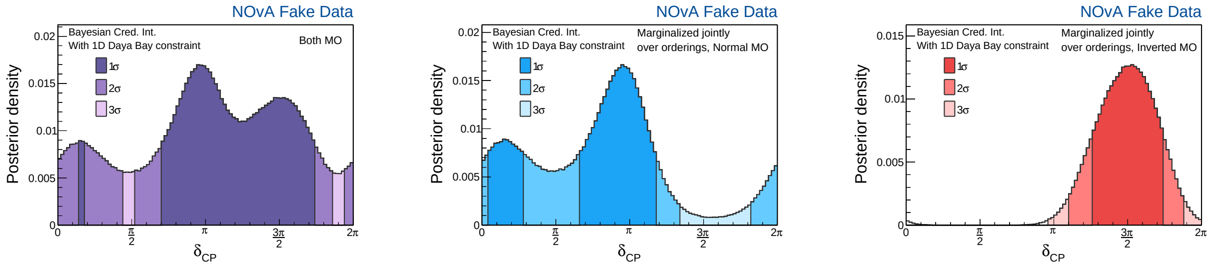


Figure 5.39: δ_{CP} with the 1D reactor constraint applied, with fake data drawn from Asimov D

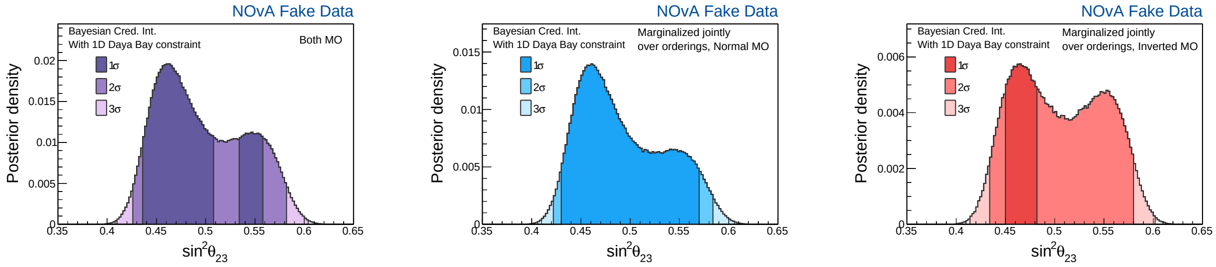


Figure 5.40: $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov D

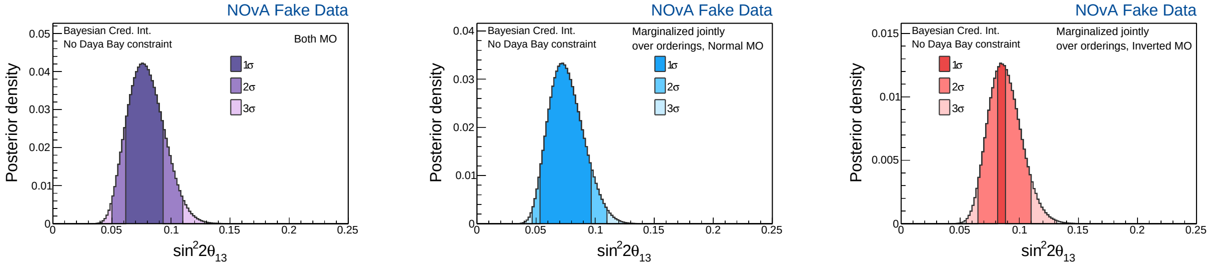


Figure 5.41: $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov D

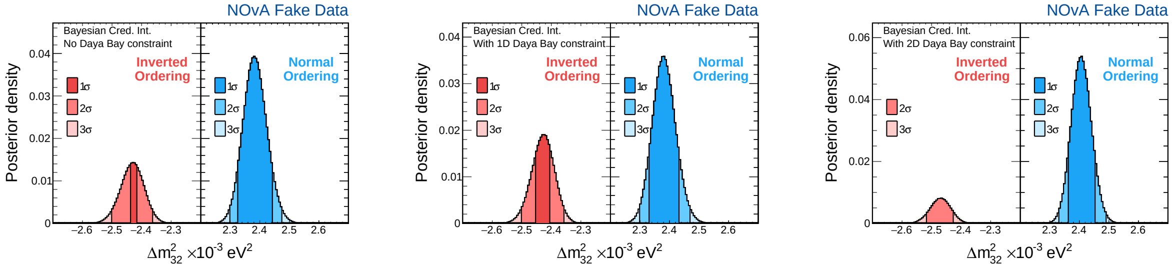


Figure 5.42: Δm^2_{32} with the different reactor constraints applied, with fake data drawn from Asimov D

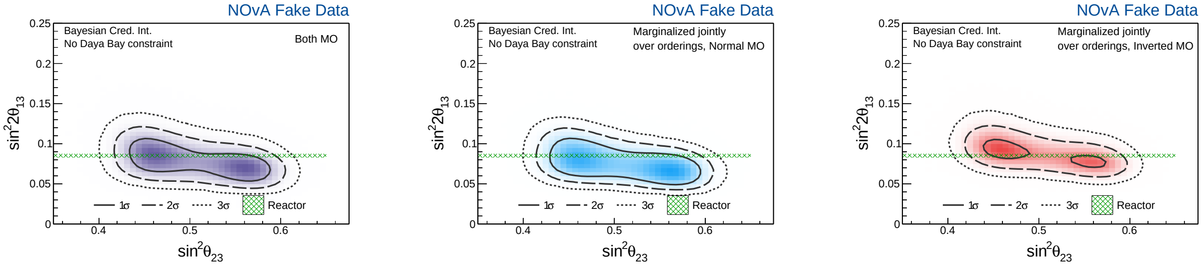


Figure 5.43: $sin^2\theta_{23}$ vs $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov D

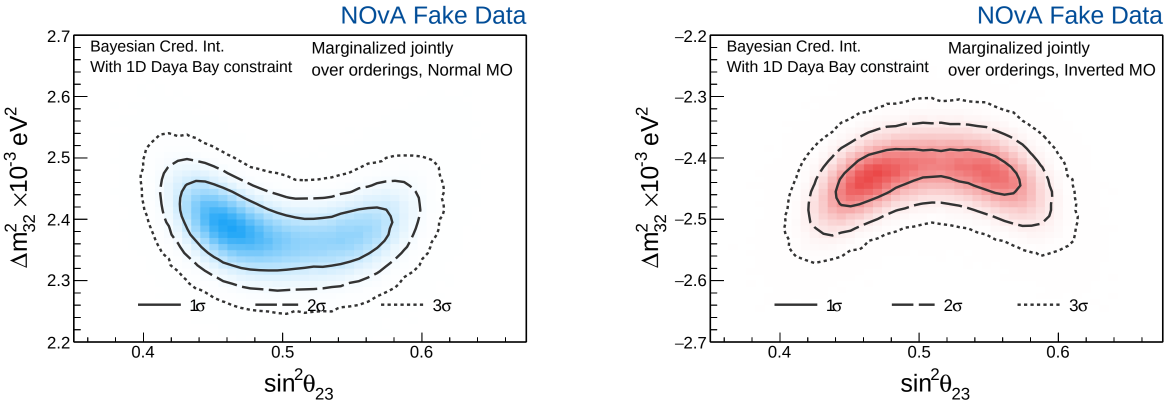


Figure 5.44: $sin^2\theta_{23}$ vs Δm^2_{32} with the 1D reactor constraint applied, with fake data drawn from Asimov D

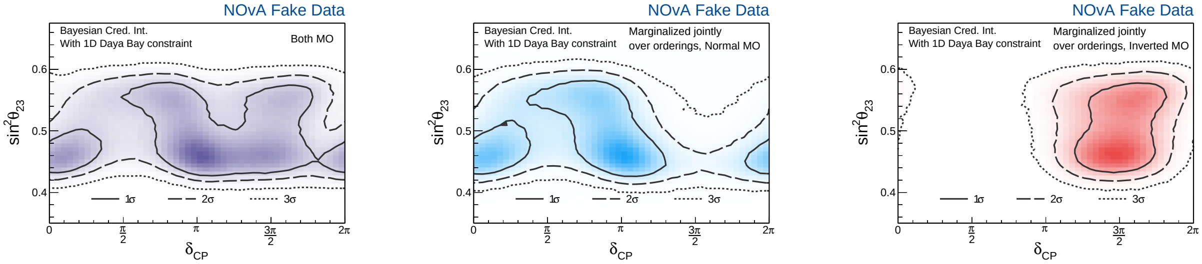


Figure 5.45: δ_{CP} vs $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov D

Asimov E

The last Asimov point we consider, E, is far from NOvA's preferred region, being in the IO, lower octant, with δ_{CP} being $\frac{\pi}{2}$ (which is an unfavored region for IO normally), shown in figures 5.46 - 5.52. We see this point, like B, has low statistics in one of the mass orderings, this time in NO, with a strong IO preference. Figure 5.49 shows this IO preference, which strengthens with the the 1D reactor constraint, the reason for which is shown by figure 5.50. Here the 1D Daya Bay band is far from the preferred region NO region. The preference shrinks with the 2D Daya Bay constraint, where Daya Bay's preferred Δm^2_{32} point is fairly far from the Asimov point.

For δ_{CP}, figure 5.46 shows both NO and IO plots have posteriors that prefer this "correct" δ_{CP} region. Figure 5.47 shows a relatively strong preference for the lower octant in both mass orderings, near the Asimov point.

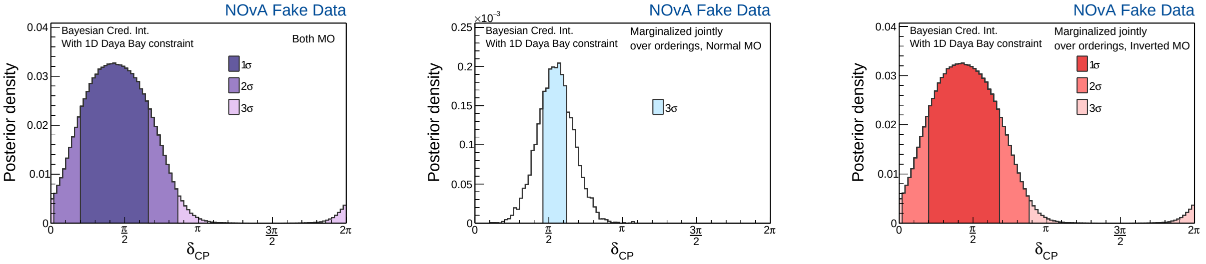


Figure 5.46: δ_{CP} with the 1D reactor constraint applied, with fake data drawn from Asimov E

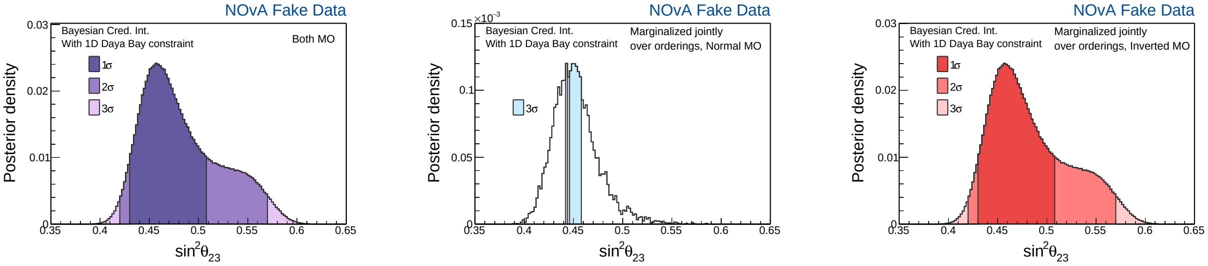


Figure 5.47: $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov E

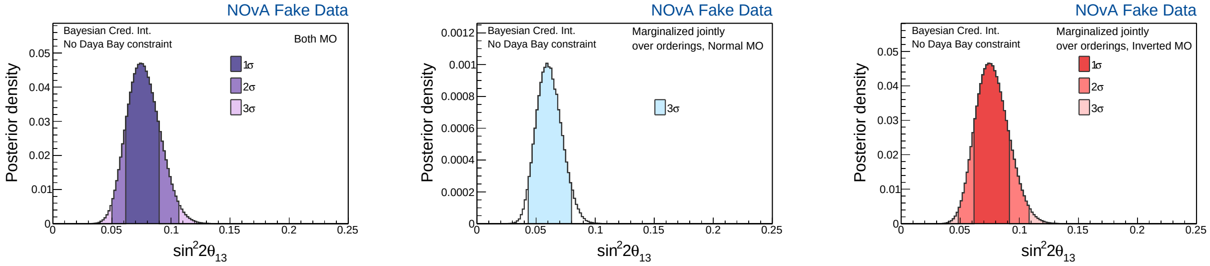


Figure 5.48: $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov E

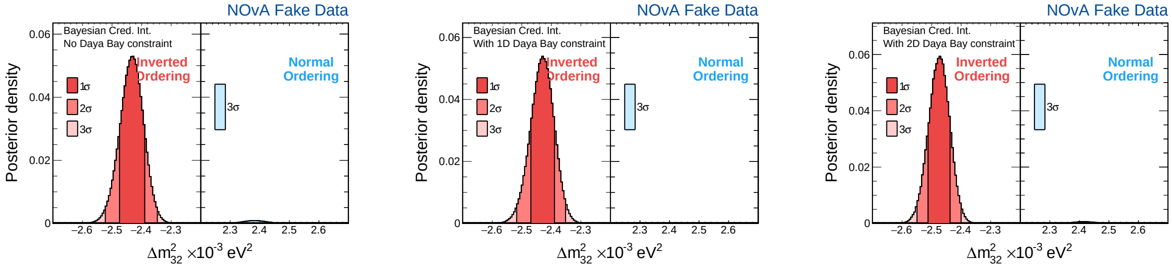


Figure 5.49: Δm^2_{32} with the different reactor constraints applied, with fake data drawn from Asimov E

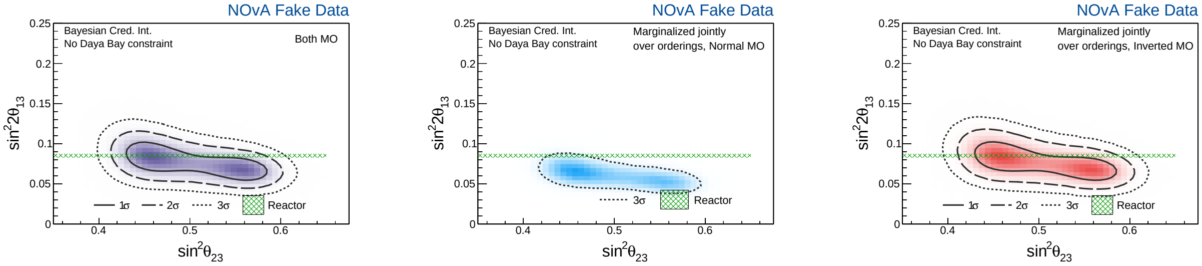


Figure 5.50: $sin^2\theta_{23}$ vs $sin^2 2\theta_{13}$ with no reactor constraint applied, with fake data drawn from Asimov E

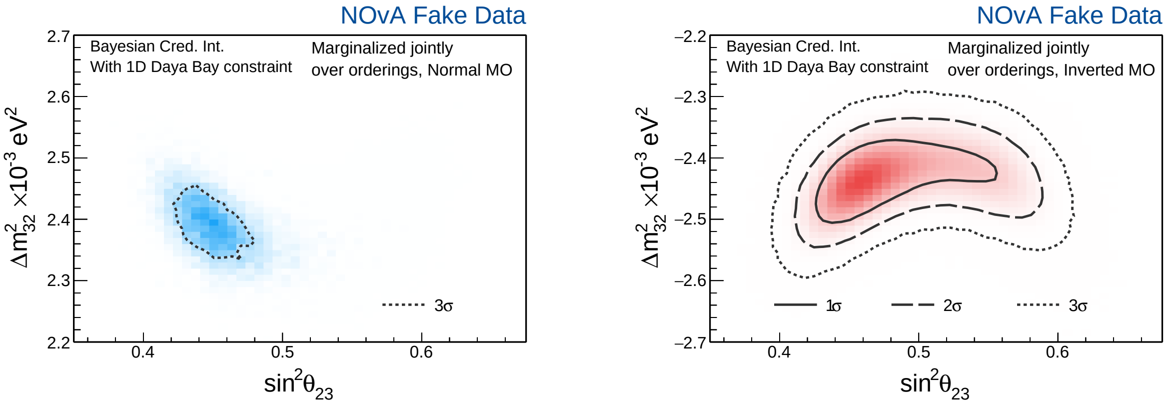


Figure 5.51: $sin^2\theta_{23}$ vs Δm^2_{32} with the 1D reactor constraint applied, with fake data drawn from Asimov E

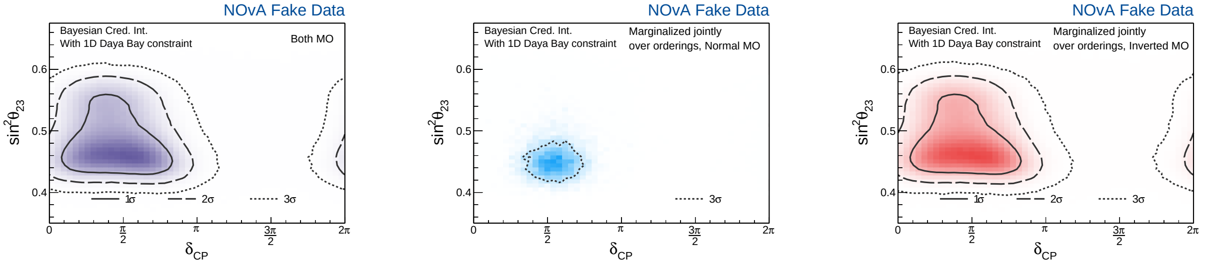


Figure 5.52: δ_{CP} vs $sin^2\theta_{23}$ with the 1D reactor constraint applied, with fake data drawn from Asimov E

Chapter 6

2024 NOvA Results

6.1 Event Counts and Spectra

The 2024 analysis results were first presented at Neutrino 2024 [117]. Table 6.1 shows the observed event counts for all samples, compared to the prediction numbers at the Frequentist best fit oscillation point. For the ν_μ and $\bar{\nu}_\mu$ samples, the predicted rates of background are low, with the largest other channel being the wrong sign channel. For the ν_e and $\bar{\nu}_e$ samples, the backgrounds are larger, coming from the beam ν_e's, NC, and cosmics, in order of largest to smallest. Lastly, for the new low E ν_e sample, it's predicted to be dominated by NC backgrounds.

Figure 6.1 shows the spectrum of selected ν_μ CC events for FHC and RHC and figure 6.2 shows these same spectra broken up by hadronic energy fraction quartile. Figure 6.3 shows the spectrum of selected ν_e CC events for FHC and RHC, for the core, peripheral, and low energy samples (FHC only).

	Neutrino Beam			Anti-neutrino Beam	
	ν_μ	ν_e	LowE ν_e	$\bar{\nu}_\mu$	$\bar{\nu}_e$
$\nu_\mu \rightarrow \nu_\mu$	372.7	4.0	0.3	24.7	0.2
$\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$	24.7	0.1	0.0	71.8	0.2
$\nu_\mu \rightarrow \nu_e$	0.4	121.6	2.9	0.0	2.1
$\bar{\nu}_\mu \rightarrow \bar{\nu}_e$	0.0	1.7	0.1	0.0	18.2
Beam $\nu_e + \bar{\nu}_e$	0.1	26.3	0.8	0.0	6.5
NC	5.5	16.3	5.0	0.8	2.0
Cosmic	4.6	5.7	0.5	0.7	1.1
Others	1.5	0.8	0.1	0.2	0.1
Signal	398.2	121.6	2.9	96.7	18.2
Background	11.3	54.9	6.8	1.7	12.2
Best fit	409.5	176.5	9.7	98.4	30.4
Observed	384.0	169.0	12.0	106.0	32.0

Table 6.1: The observed and predicted $\nu_\mu(\bar{\nu}_\mu)$ disappearance and $\nu_e(\bar{\nu}_e)$ appearance event counts at the Far Detector

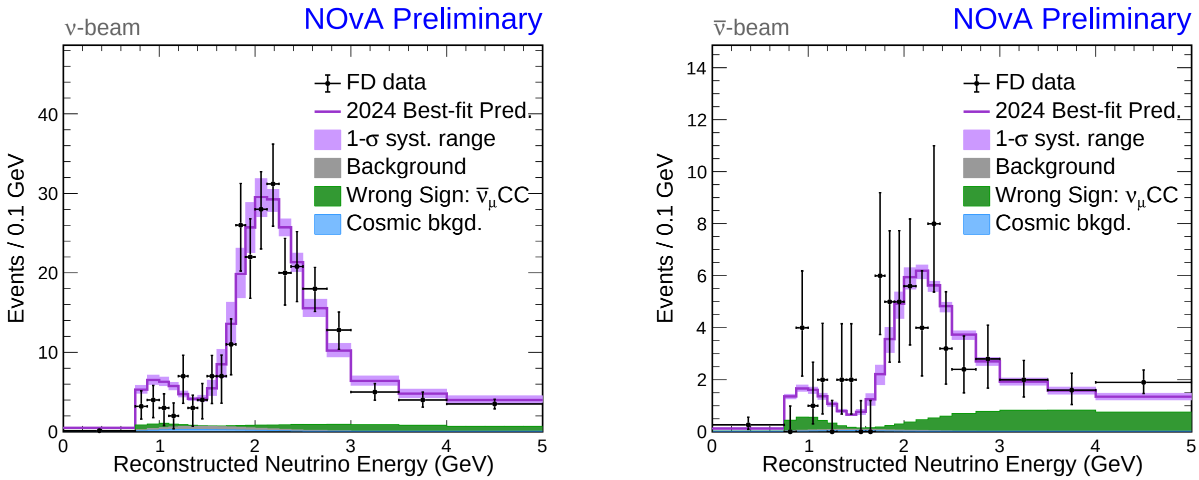


Figure 6.1: ν_μ sample predictions shown at analysis best fit point, overlaid with data, with hadronic energy fraction quartiles combined, FHC (left) and RHC (right)

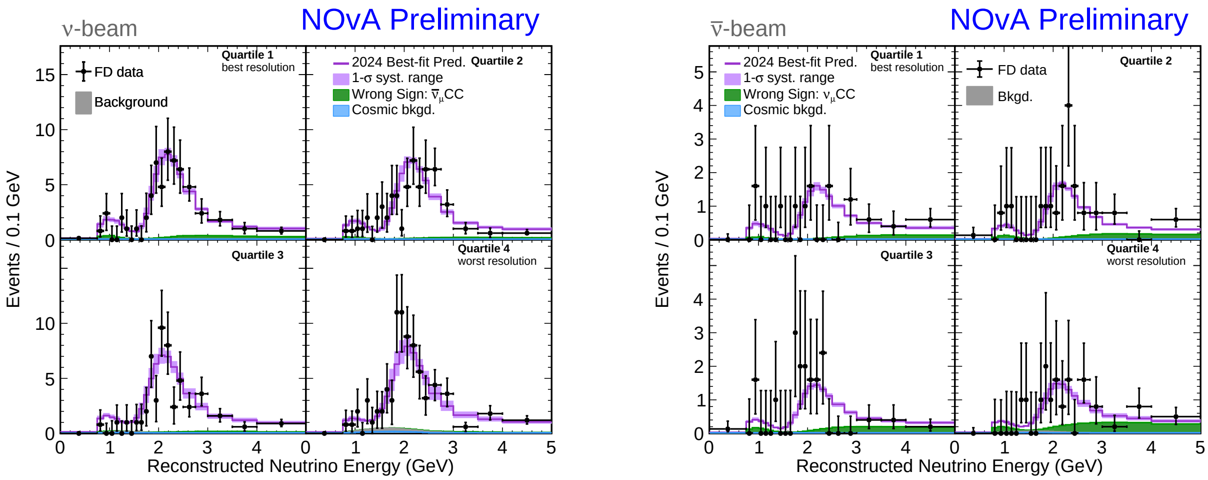


Figure 6.2: ν_μ sample predictions shown at analysis best fit point, overlaid with data, FHC (left) and RHC (right)

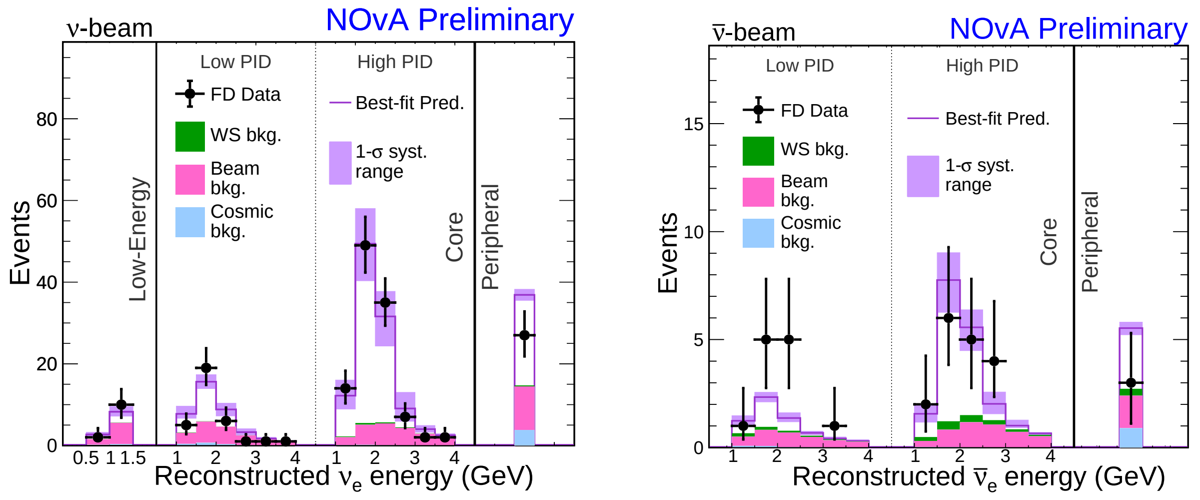


Figure 6.3: ν_e sample predictions shown at analysis best fit point, FHC (left) and RHC (right). Includes low and high PID core samples in the middle, the peripheral bin on the right, and for FHC only, the low energy sample on the left.

6.2 Frequentist Results

Figure 6.4 shows the best fit points of the 2024 NOvA Frequentist fit. Note the best fit point is again in the upper octant (UO) and normal ordering (NO). The δ_{CP} best fit is at 0.875π, close to the NOvA 2020 point. See section 6.2.2 for an illustration of different δ_{CP} values compared to our ν_e event rates.

Also shown are the preferences for NO with both the 1D and 2D reactor constraint. We see with the normal 1D reactor constraint, we have a mild preference at 1.36 σ for the NO. The preference slightly strengthens if we use the 2D reactor constraint instead to 1.57 σ.

Parameter	**Best-fit**	**Normal Ordering Preference** (σ)	
$\sin^2(\theta_{23})$	$0.546^{+0.032}_{-0.075}$	W/ 1D Daya Bay constraint	p-value 0.1731 $1.36\,\sigma$
$\Delta m^2_{32}\left(10^{-3}\,\text{eV}^2\right)$	$2.433^{+0.035}_{-0.036}$		
$\delta_{\text{CP}}\,(\pi)$	0.875	W/ 2D Daya Bay constraint	p-value 0.1158 $1.57\,\sigma$

Figure 6.4: Frequentist best fit values for oscillation parameters, and preferences for NO, with 1D and 2D Daya Bay constraint

Figure 6.5 shows the Feldman-Cousins corrected $sin^2\theta_{23}$ vs Δm^2_{32} contours for the NO, with the standard 1D reactor constraint. The surface is mostly in the same region as the 2020 analysis, but tighter in both dimensions. Figure 6.6 shows the 90% surface compared to other experiments. It falls right in line with other experiments surfaces, with no obvious tensions.

At the time of writing, the Feldman-Cousin corrected $sin^2\theta_{23}$ vs Δm^2_{32} IO contours, and the δ_{CP} vs $sin^2\theta_{23}$ contours were not finished running on NERSC, so only the Bayesian results are available for these (see section 6.3).

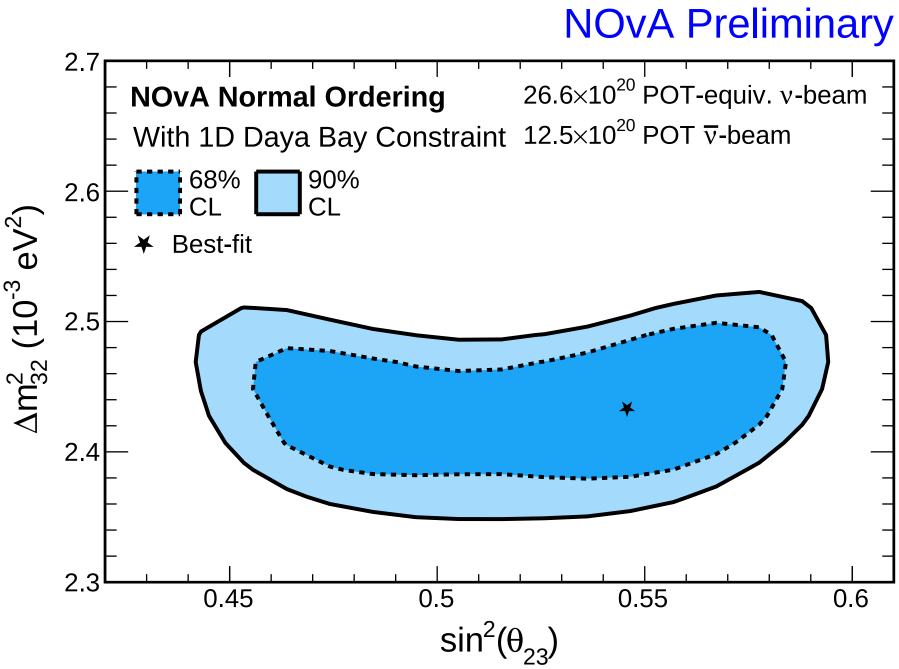


Figure 6.5: NO $sin^2\theta_{23}$ vs Δm^2_{32} Frequentist contours

6.2.1 Effects of Systematics

Figure 6.7 shows the uncertainties of various groups of systematics on the 3 oscillation parameters of interest at the Frequentist best fit point. Like previous analyses, detector calibration remains NOvA's largest uncertainty. For $sin^2\theta_{23}$, the detector calibration error bar is much larger than the others, with the uncertainty pulling mostly left, towards the LO. Statistical uncertainty dominates uncertainties here. For Δm^2_{32}, the error bars are mostly symmetric, with detector calibration not as dominant compared to other systematic groups. Systematics have a larger relative effect compared to statistical uncertainty here. Lastly, for δ_{CP}, the error bars on the systematic groups pull mostly left, closer to max CP violation.

6.2.2 Bi-Event Plot

Figure 6.8 shows the expected ν_e vs $\bar{\nu}_e$ event counts for different values of δ_{CP}, shown as ellipses, the upper and lower octant best fits, and the NO and IO best fits. NOvA remains in

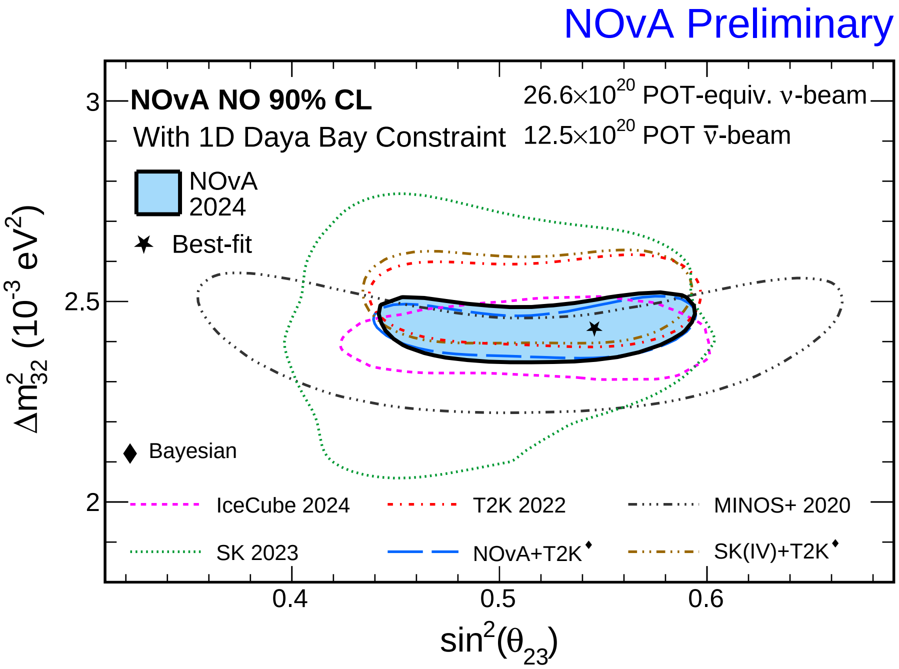


Figure 6.6: NO $sin^2\theta_{23}$ vs Δm^2_{32} Frequentist contours compared to other experiments [51, 114, 5, 113, 18, 6]

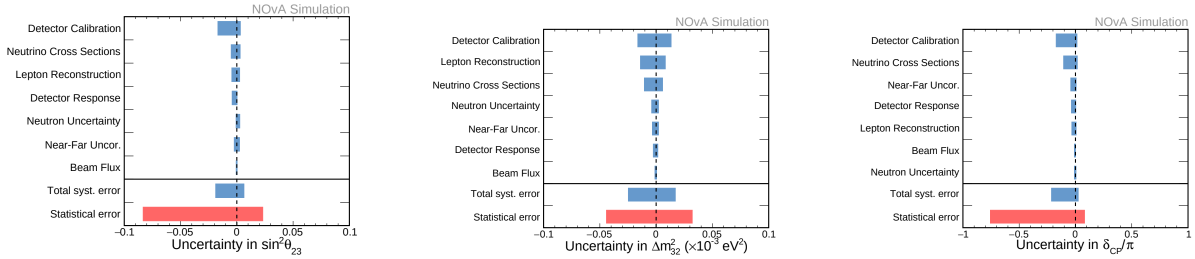


Figure 6.7: Uncertainties of oscillation parameters at best fit point in the Frequentist fit

a degenerate region, consistent with both octants and mass orderings, with a slight preference for the NO.

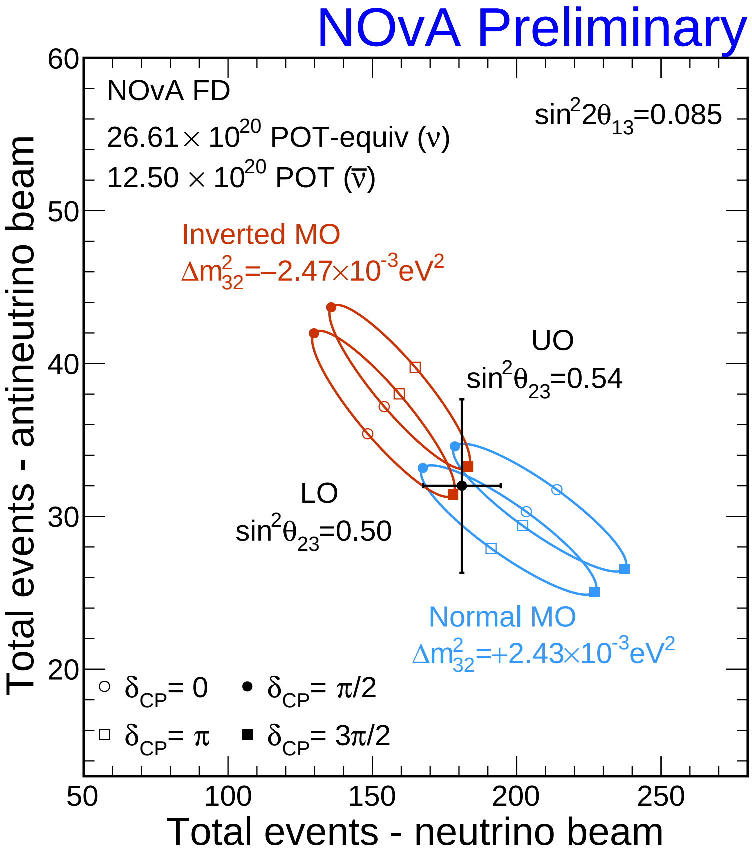


Figure 6.8: ν_e vs $\bar{\nu}_e$ event counts, compared to expected value for different values of δ_{CP}, upper and lower octant best fit, and normal and inverted ordering

	No Constraint		1D Constraint		2D Constraint	
	Prob	BF	Prob	BF	Prob	BF
Upper Octant Preference	57%	1.3	69%	2.2	67%	2.0
Normal Ordering Preference	69%	2.2	76%	3.2	87%	6.8

Table 6.2: Probabilities and corresponding Bayes' factors for upper octant and normal ordering preference with various Daya Bay constraints

6.3 Bayesian Results

The probabilities and corresponding Bayes factors for ordering and octant preference from Bayesian MCMC sampling are shown in Table 6.2. For octant preference, NOvA shows a slight preference for the upper octant, which is strongest with the 1D reactor constraint, and gets slightly weaker with the 2D reactor constraint. These results show a noteworthy preference for the normal ordering, which is strongest with the 2D reactor constraint, at 87%.

6.3.1 Oscillation Parameters – No reactor constraint

NOvA-only fits for oscillation parameters are shown in Figures 6.9 - 6.12.

The NOvA only measurement of $sin^2(2\theta_{13})$ is shown in figure 6.9. The Daya Bay value is at 0.0851 ± 0.0024, more consistent with the NO than the IO, which can be seen as the reason for the shift towards NO preference with the 1D reactor constraint shown in Table 6.2.

Figure 6.10 shows $sin^2\theta_{23}$ with no Daya Bay constraint. Note although there is still an overall preference for $sin^2\theta_{23} > 0.5$, the highest posterior density point is in the lower octant. This is not the case with reactor constraints applied, and figure 6.11 illustrates why, as the Daya Bay value is more consistent with the upper octant in $sin^2\theta_{23}$ vs $sin^2(2\theta_{23})$ space.

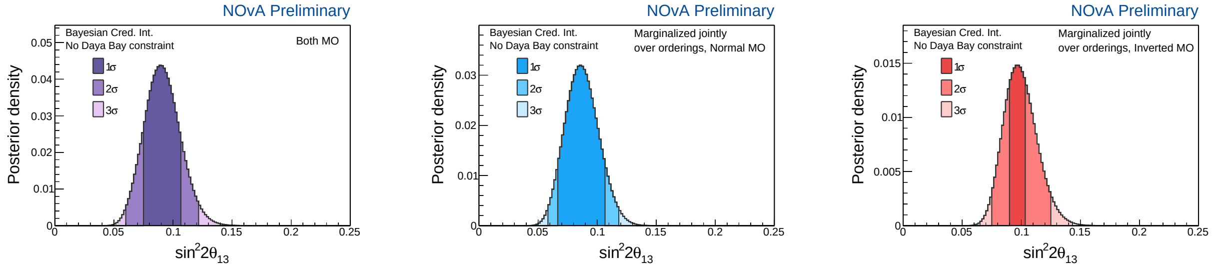


Figure 6.9: $sin^2 2\theta_{13}$ with no reactor constraint applied

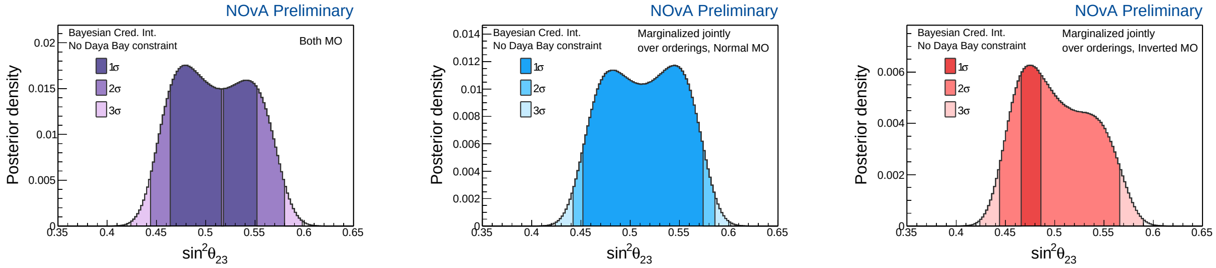


Figure 6.10: $sin^2\theta_{23}$ with no reactor constraint applied

Lastly, figure 6.12 shows the NOvA-only $sin^2(2\theta_{13})$ vs Δm^2_{32} posterior against Daya-Bay's 2D contours used for the 2D reactor constraint. It shows that not only does IO have a slight tension with Daya Bay in $sin^2(2\theta_{13})$, but also in the Δm^2_{32} direction, which gives the 2D reactor constrain the largest mass ordering preference.

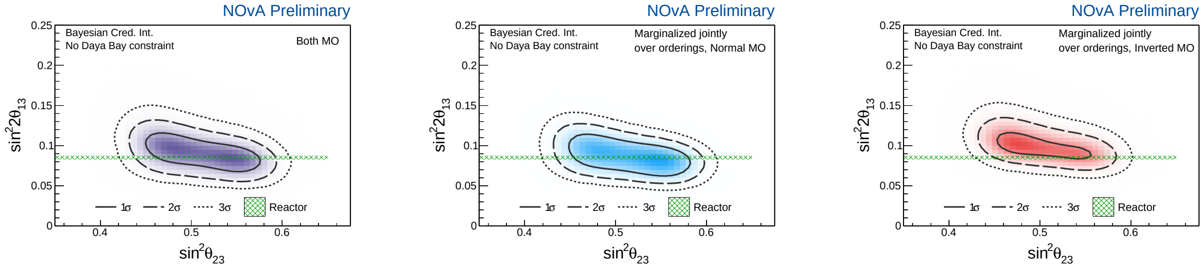


Figure 6.11: $sin^2\theta_{23}$ vs $sin^2 2\theta_{13}$ with no reactor constraint applied

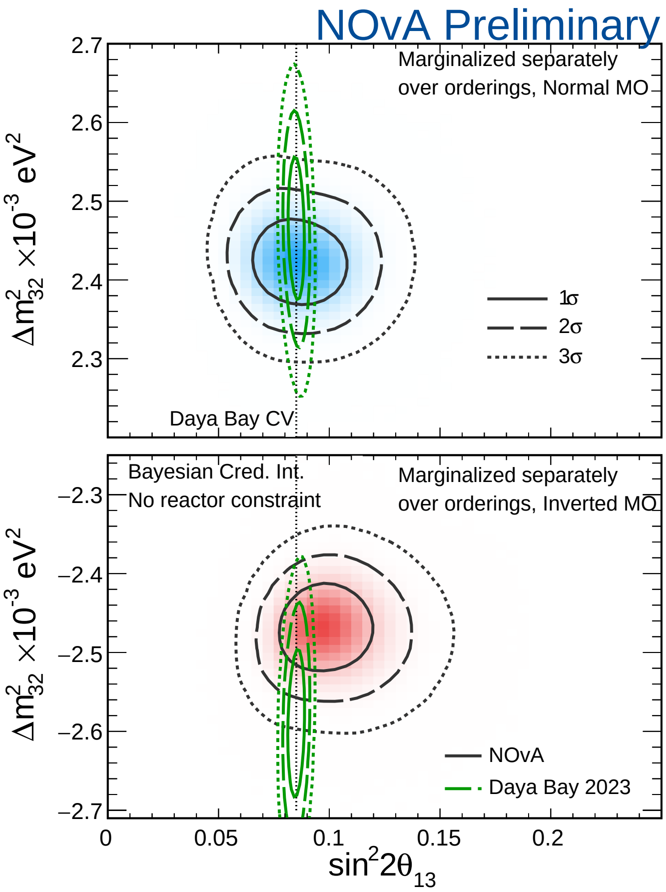


Figure 6.12: $sin^2 2\theta_{13}$ vs Δm^2_{32} posterior, for NO and IO, compared to Daya Bay contours used for the 2D reactor constraint [25].

	HPD		
	Both MO	Normal MO	Inverted MO
δ_{CP}	0.93π	0.93π	1.49π
$sin^2\theta_{23}$	0.54	0.55	0.54
$\Delta m^2_{32}(\times 10^{-3}\ \mathrm{eV}^2)$	–	2.42	-2.47

Table 6.3: Highest Posterior Density (HPD) points for oscillation parameters with the 1D Daya Bay constraint applied

6.3.2 Oscillation Parameters – 1D reactor constraint

The fits for the oscillation parameters applying the 1D Daya Bay constraint are shown in figures 6.13 - 6.18. The 1D reactor constraint is traditionally the constraint NOvA results have been presented with in past analyses. The highest posterior density (HPD) points are listed in Table 6.3.

Figure 6.13 shows δ_{CP}, broken up into different orderings with their respective credible intervals. As in the past, NOvA is in a degenerate region, supporting NO and IO models with different regions of δ_{CP}. The NO density has two peaks, one near π and the other near 0, the two CP conserving values of δ_{CP}. IO peaks near $\frac{3\pi}{2}$, being inconsistent with CP conserving points at more than 2 and 3 σ, assuming IO.

Figure 6.14 shows $sin^2\theta_{23}$ for different ordering breakdowns. Note these results are consistent with max mixing, $sin^2\theta_{23} = 0.5$, within 1 σ.

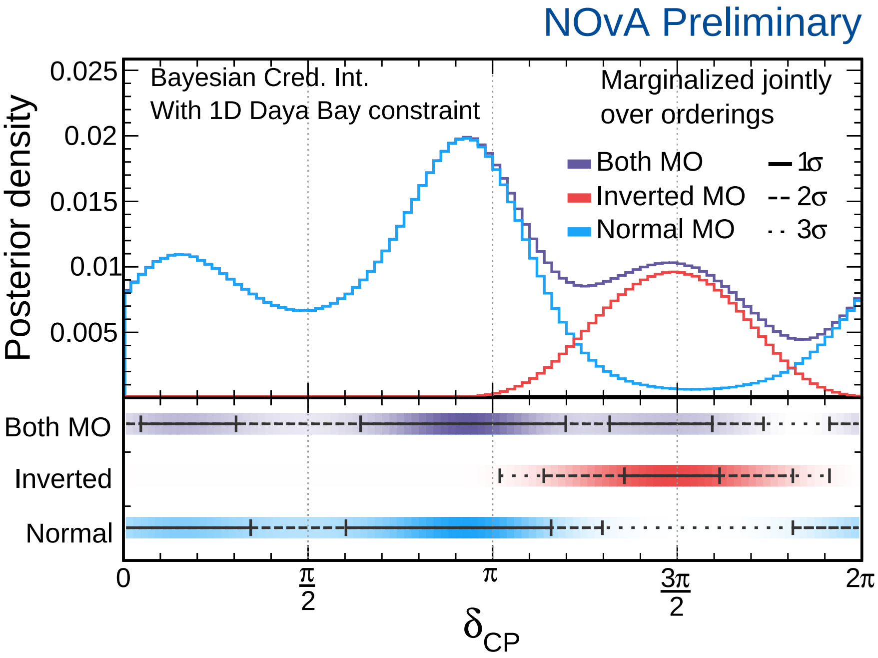


Figure 6.13: δ_{CP} with the 1D reactor constraint applied

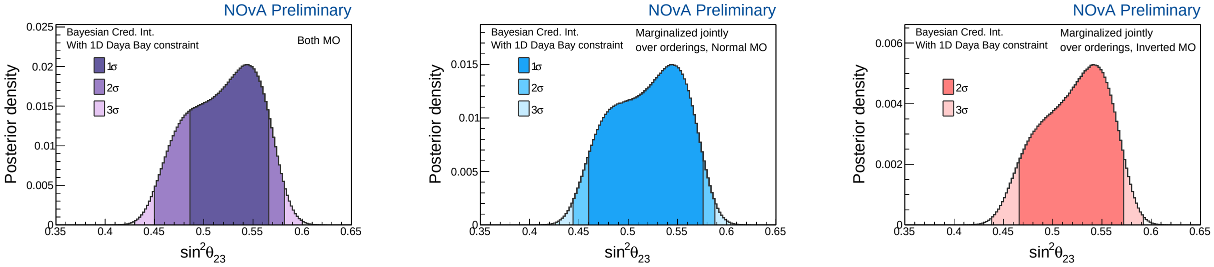


Figure 6.14: $sin^2\theta_{23}$ with the 1D reactor constraint applied

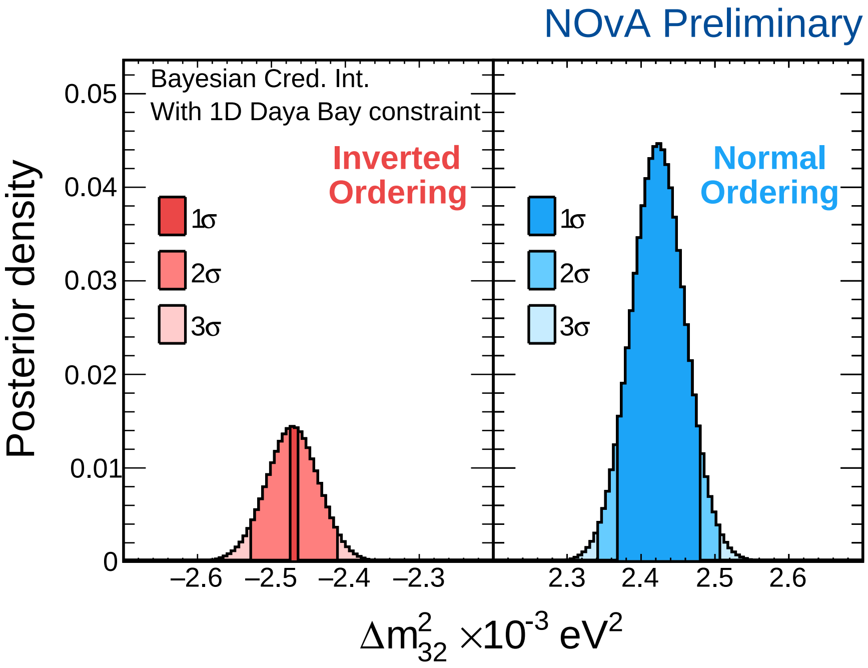


Figure 6.15: Δm^2_{32} with the 1D reactor constraint applied

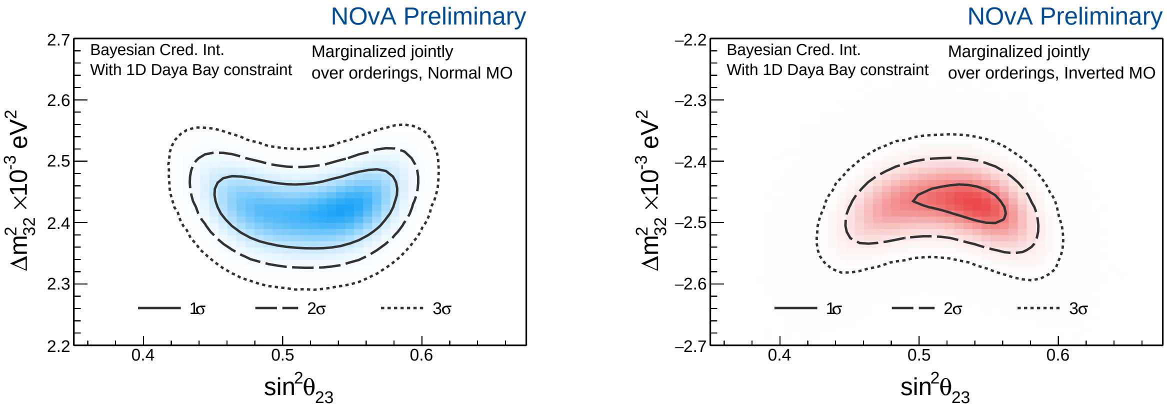


Figure 6.16: $sin^2\theta_{23}$ vs Δm^2_{32} with the 1D reactor constraint applied

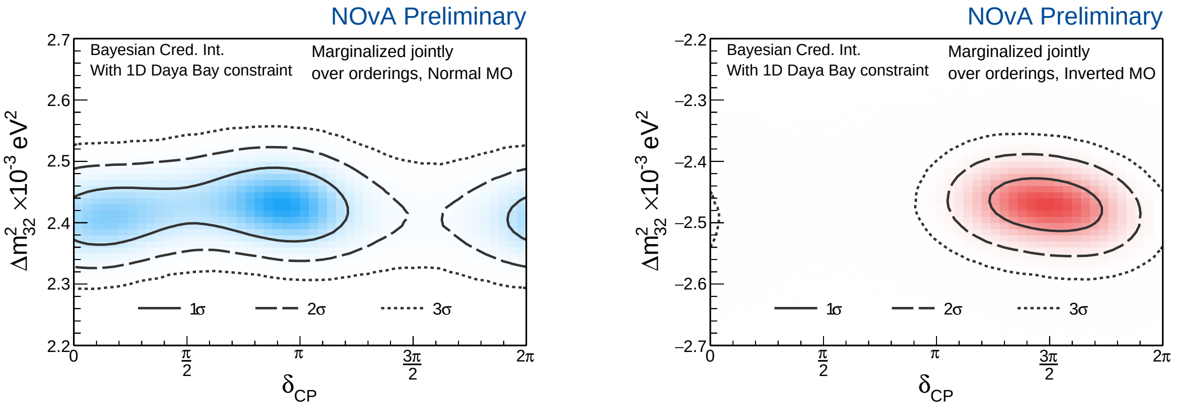


Figure 6.17: δ_{CP} vs Δm^2_{32} with the 1D reactor constraint applied

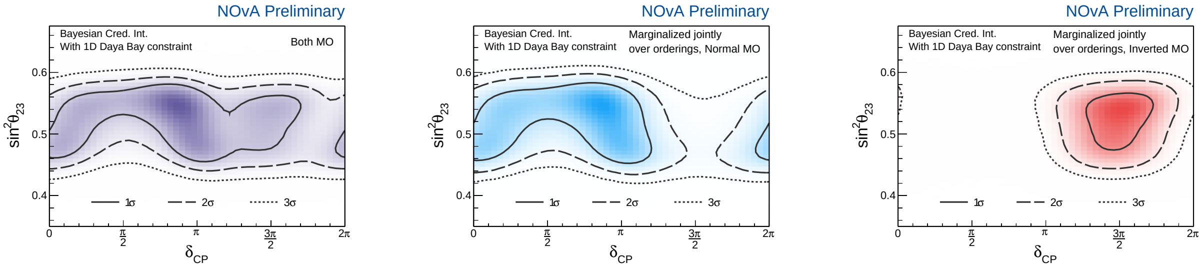


Figure 6.18: δ_{CP} vs $sin^2\theta_{23}$ with the 1D reactor constraint applied

6.3.3 Oscillation Parameters – 2D reactor constraint

Figures 6.19 - 6.22 show posteriors with the 2D reactor constraint in $sin^2(2\theta_{13})$ vs Δm^2_{32} applied. Figure 6.19 shows $sin^2\theta_{23}$ posteriors, showing that with the 2D Daya Bay constraint, $sin^2\theta_{23} = 0.5$ is now just outside the both orderings 1 σ credible interval.

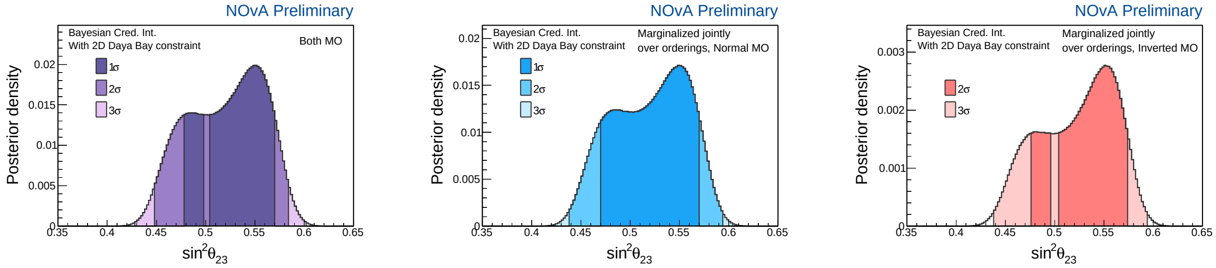


Figure 6.19: $sin^2\theta_{23}$ with the 2D reactor constraint applied

Figure 6.20 shows the most noteworthy aspect of the 2D constraint: a stronger preference for NO, at 87% preference.

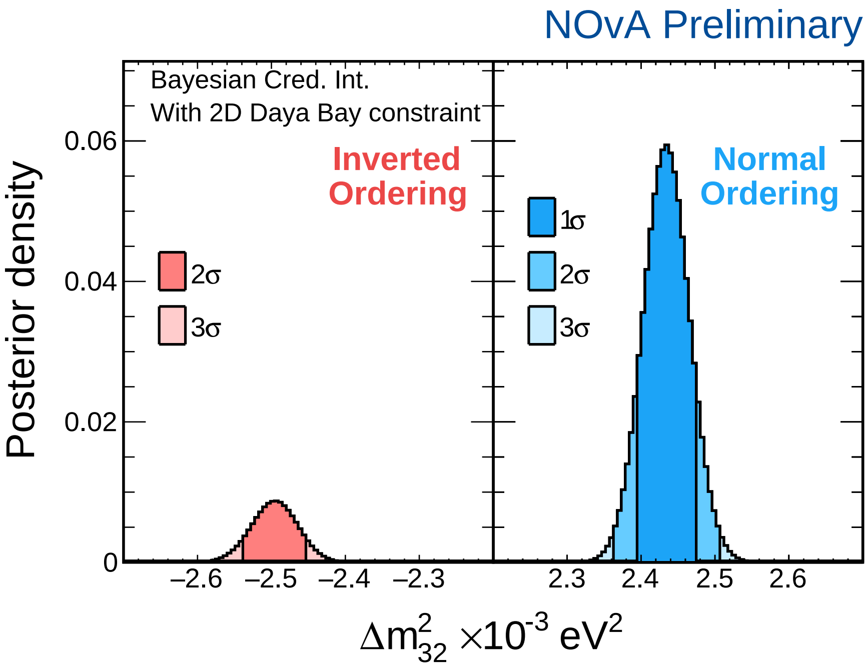


Figure 6.20: Δm^2_{32} with the 2D reactor constraint applied

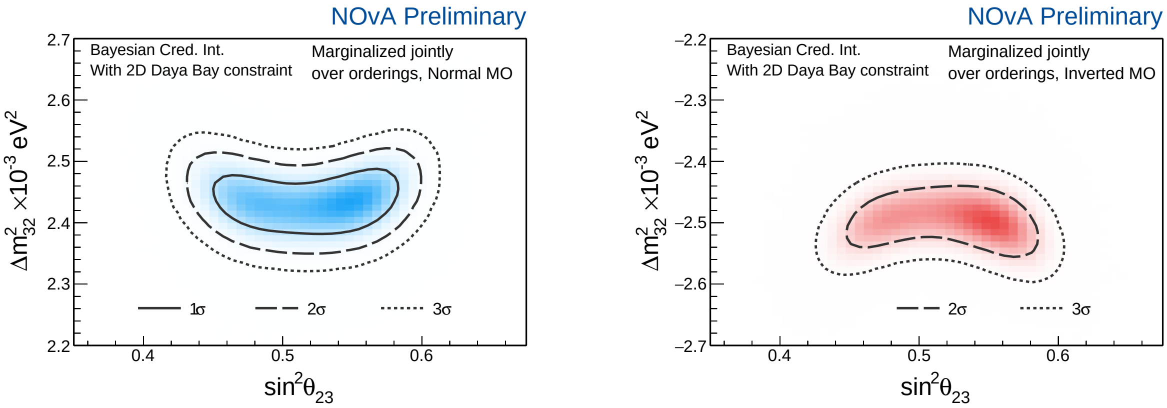


Figure 6.21: $sin^2\theta_{23}$ vs Δm^2_{32} with the 2D reactor constraint applied

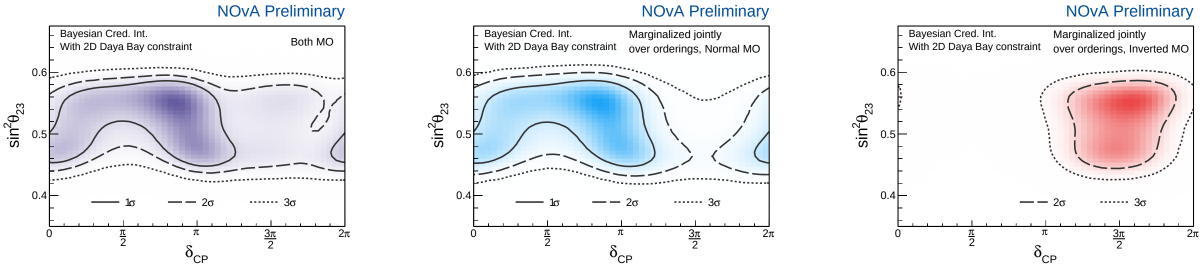


Figure 6.22: δ_{CP} vs $sin^2\theta_{23}$ with the 2D reactor constraint applied

6.3.4 Systematic Systematic Posteriors

MCMC will generate chains in the full 72 oscillation parameter – systematic space, which can be marginalized over any set of parameters to isolate the parameters of interest. The parameters of interest are usually combinations of oscillation parameters such as in the previous sections, but we can also look at the systematic parameters to see how the data constrains them. The prior distribution for the systematics are all normal distributions, with mean 0 and standard deviation 1. The distributions of the systematic posteriors, marginalized over all other parameters, are shown in figure 6.23.

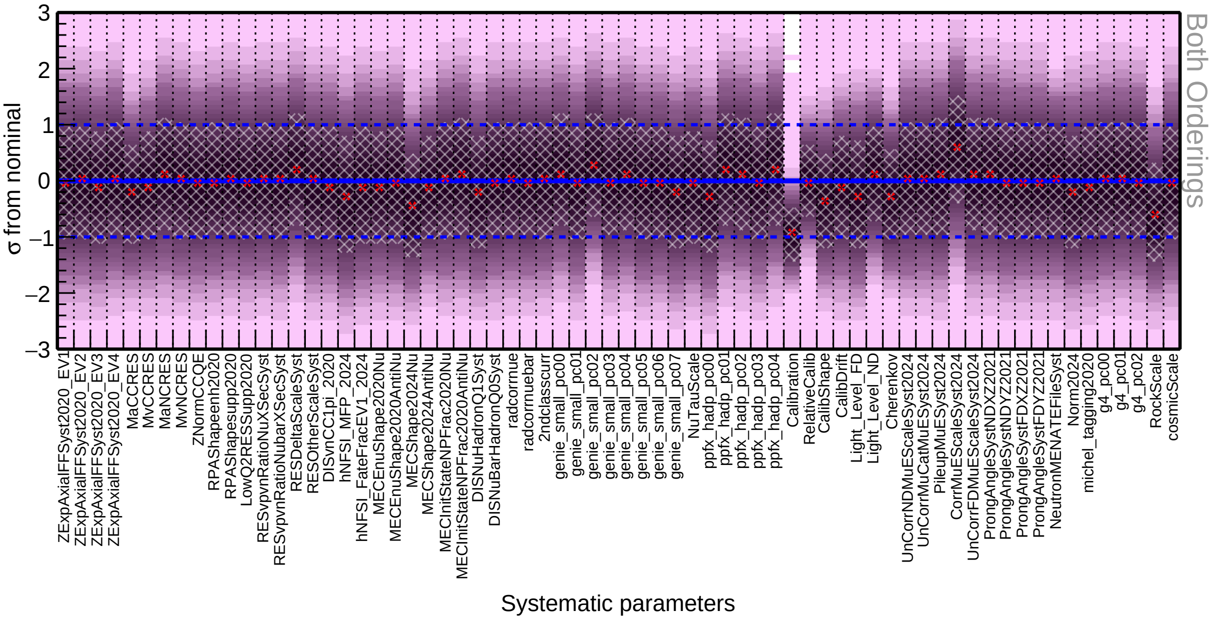


Figure 6.23: Posteriors of systematic uncertainties individually, marginalized over all other parameters. Red cross is the HPD point, cross shaded region is 1 σ interval, and blue lines are at 0 and ±1, the HPD and 1 σ interval for the prior

Most systematic posteriors are very similar to the prior, with a few exceptions. "Calibration" for example sees a strong pull down, with the HPD being near -1 σ. "CorrMuEScaleSyst2024", the muon energy scale correlated between the near and far detectors, has a pull up, with the HPD being around 0.5 σ. "RockScale", a conservative systematic on the expected number of muons from neutrinos interacting in the rock for the FD, sees a pull down with an HPD around -0.5 σ.

6.3.5 Jarlskog Invariant

The Jarlskog invariant is a metric that quantifies CP violation that's independent of the parameterization of the neutrino mass mixing matrix [54]. For 3 flavor neutrino oscillations, it's defined as:

$$J = \cos(\theta_{12})\cos^2(\theta_{13})\cos(\theta_{23})\sin(\theta_{12})\sin(\theta_{13})\sin(\theta_{23})\sin(\delta_{CP}) \tag{6.1}$$

$J = 0$ indicates CP conservation, where values farther from 0 indicate larger CP violation.

MCMC offers an easy way to evaluate the posterior of the Jarlskog invariant since the sampled points of the oscillation parameter posteriors can be manipulated with ease. The priors for terms with θ_{12} use the global average of solar neutrino experiments for a prior [121]. We consider two different priors for δ_{CP}: a prior flat in sinδ_{CP} since J itself is in terms of this, and also flat in δ_{CP} for theoretical reasons [55].

This Jarlskog invariant with these two priors for different mass ordering breakdowns is shown in figure 6.24 with the 1D reactor constraint. In general, NOvA doesn't reject $J = 0$ with any significance. If one assumes NO, positive values of J are preferred, but $J = 0$ is in the 1 or 2 σ credible interval depending on the prior. If one assumes IO, negative values are more strongly preferred, with $J = 0$ being strongly disfavored.

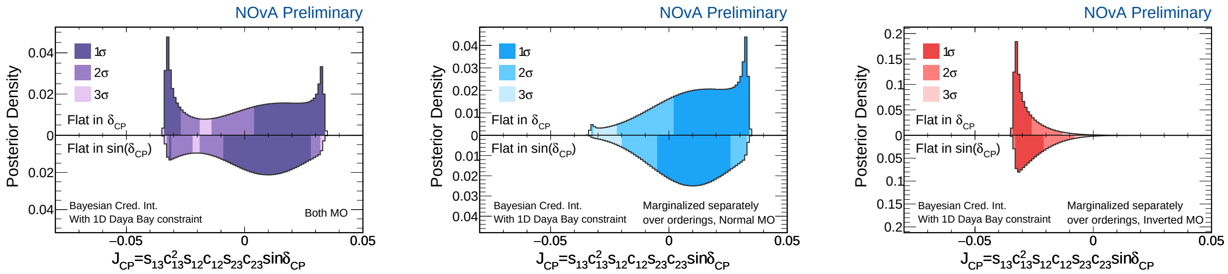


Figure 6.24: Jarlskog Invariant with the 1D reactor constraint applied

Dataset	P-Value
ν_μ	0.68
$\bar{\nu}_\mu$	0.44
ν_e	0.64
ν_e LE	0.81
$\bar{\nu}_e$	0.20

Table 6.4: Posterior predictive p-values for individual samples, showing consistency of data fit model

6.3.6 Posterior P-Values and Ranges

It's natural to wonder how consistent the data is with the 3 flavor model, as large inconsistencies could point to new physics. With Bayesian statistics, this is can be done looking at posterior p-values and posterior ranges. Posterior p-values use the posterior distributions of the models parameters (including oscillation parameters and systematics) to measure how likely the sampled data is if the experiment was completed many times, analogous to the Frequentist concept of p-value.

The posterior p-values are shown in Table 6.4 for the 5 individual analysis samples. In general, they show great consistency with the data.

We can also look at the spectrum against the posterior ranges. Posterior ranges effectively show the expected average value for each energy bin, given the fit posteriors of the model parameters. The don't include the Poisson errors on the bin, which is the difference between posterior ranges, and posterior predictive distributions.

The posterior ranges for the spectra of ν_μ samples is shown in Figure 6.25, and the ν_e samples in figure 6.26. Visibly, these show reasonable agreement between the data points and the posterior ranges.

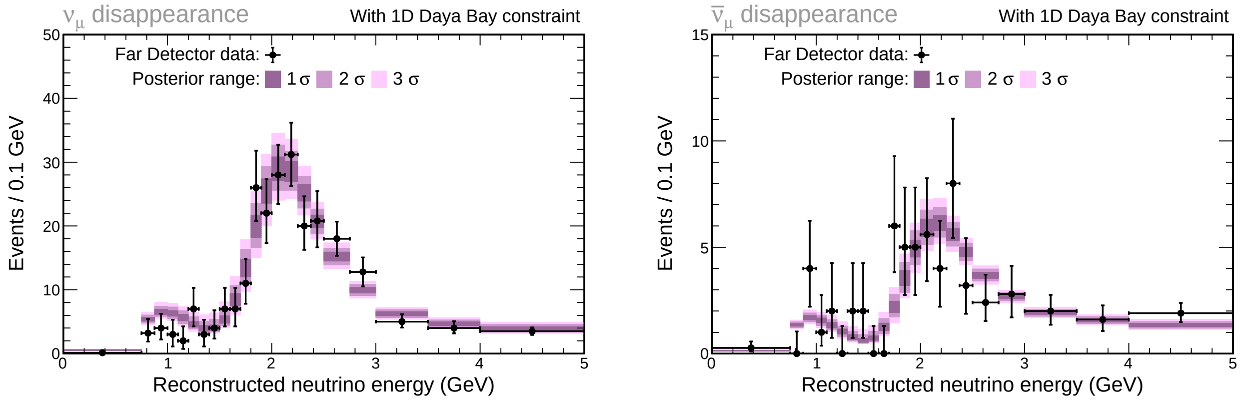


Figure 6.25: Posterior predictive ranges for ν_μ samples, with data points overlaid with Poisson errors

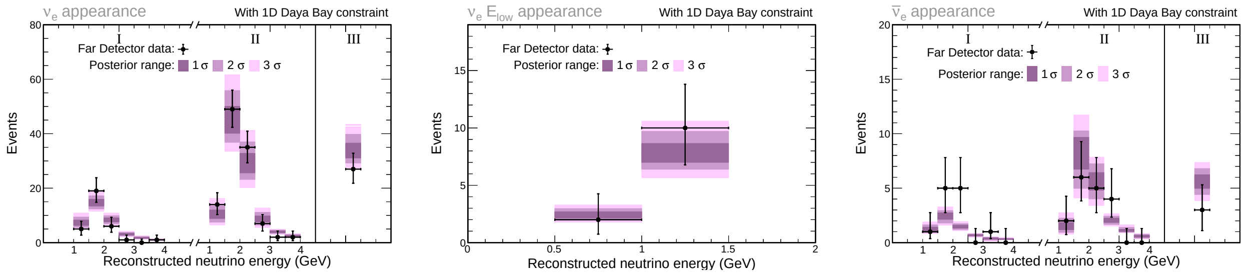


Figure 6.26: Posterior predictive ranges for ν_e samples, with data points overlaid with Poisson errors

The posterior range for the "bi-event" plot, comparing the ν_e counts versus $\bar{\nu}_e$ counts, is shown in figure 6.27. These have overlaid ellipses showing the average breakdown for various values of δ_{CP} in both the NO and IO. This is a powerful illustration of the δ_{CP} preferred values for NO and IO shown in figure 6.13, as well as the ordering preference.

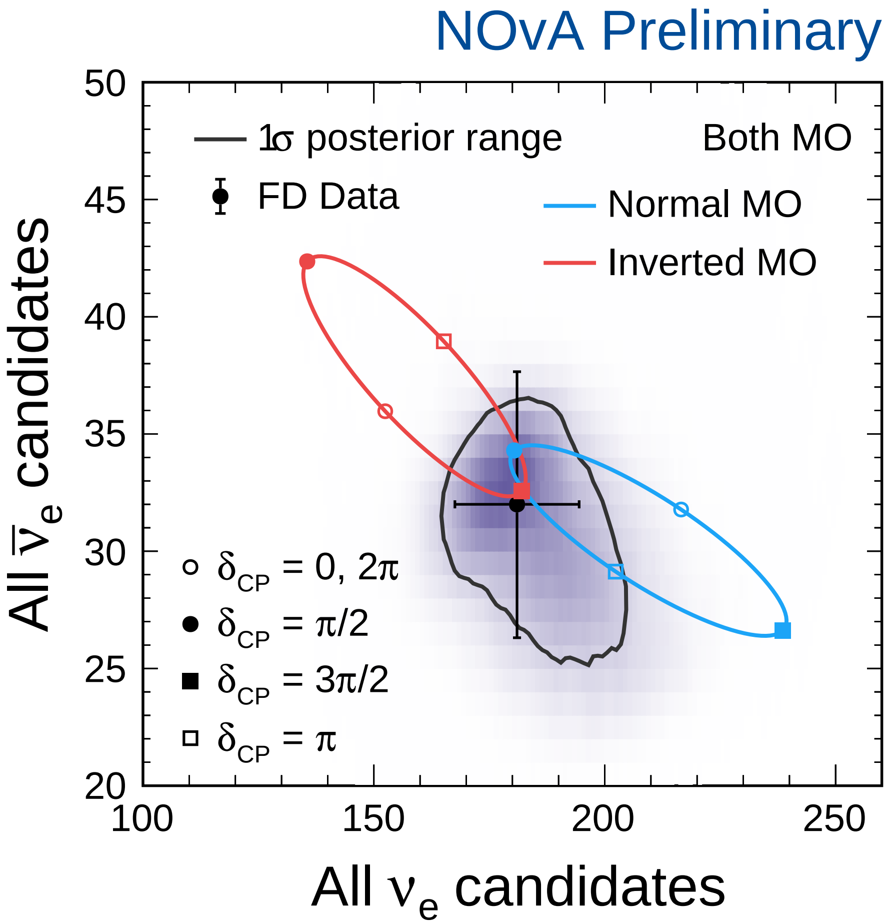


Figure 6.27: Posterior predictive distribution of ν_e counts versus $\bar{\nu}_e$ counts with the 1D reactor constraint applied

6.3.7 Comparison To Other Experiments

Finally, we can compare the measurements of oscillation parameters from this result to other results [12, 113, 51, 5, 6, 25, 18, 114, 11]. First, Figure 6.28 shows NOvA's measurement of Δm^2_{32} compared to the measurement by other experiments, for NO and IO separately. NOvA's measurement is the most precise single experiment measurement. The value is roughly consistent with other measurements. For the absolute value, the central value is higher than IceCube's, and lower than T2K's.

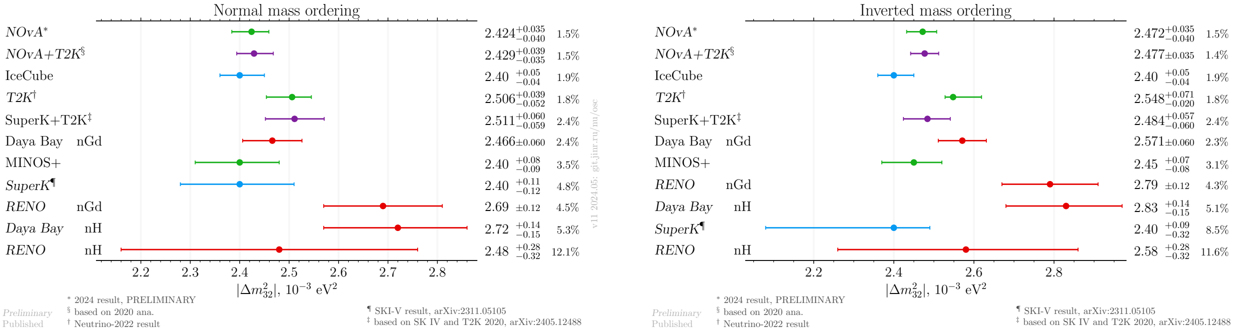


Figure 6.28: Δm^2_{32} 1 σ intervals for this result compared to other experiments, NO (left) and IO (right)

Figure 6.29 shows the $sin^2\theta_{23}$ measurement for NOvA compared to others. For both NO and IO, NOvA's measurement is in line with T2K's, SuperK+T2K's, and IceCube's.

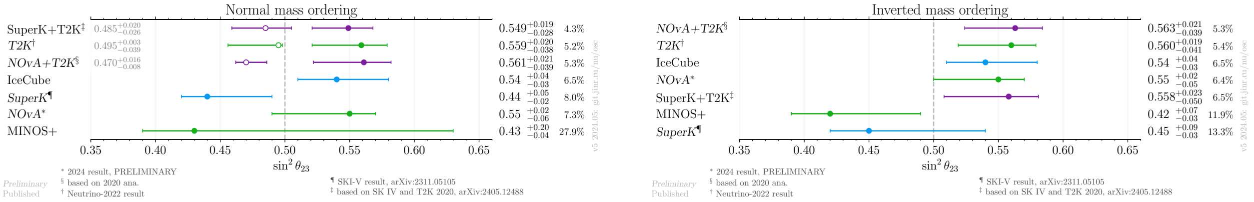


Figure 6.29: $sin^2\theta_{23}$ 1 σ intervals for this result compared to other experiments, NO (left) and IO (right)

In figure 6.30, NOvA's δ_{CP} measurement is compared with other's. In the NO, NOvA's credible interval has 2 segments, unlike other experiments. For the most part, NOvA's preference is lower than other measurements. In the IO, NOvA is very in line with everyone else.

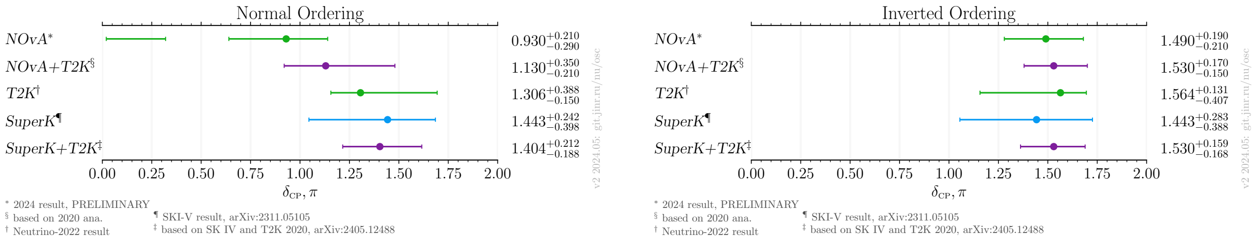


Figure 6.30: δ_{CP} 1 σ intervals for this result compared to other experiments, NO (left) and IO (right)

Figure 6.31 shows the δ_{CP} vs $sin^2\theta_{23}$ posterior compared to other recent results. The trends are mostly similar to the 2020 NOvA analysis. The NO posterior again prefers a separate region of δ_{CP} than T2K's, and T2K+SuperK. There is some overlap with T2K's 2022 measurement near π. For IO, they all agree well.

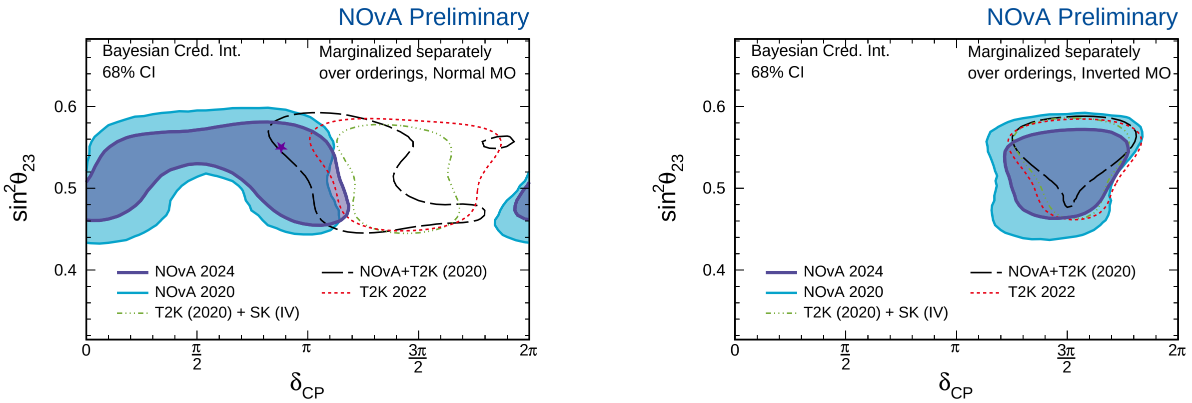


Figure 6.31: δ_{CP} vs $sin^2\theta_{23}$ 1 σ intervals for this result compared to other experiments, NO (left) and IO (right)

Lastly, figure 6.32 shows NOvA's $sin^2\theta_{23}$ vs Δm^2_{32} credible interval compared to other experiments. These contours show very nice agreement for both NO and IO. Compared to NOvA's in NO, T2K is slightly above, though still significantly overlaps. It has even greater overlap in IO.

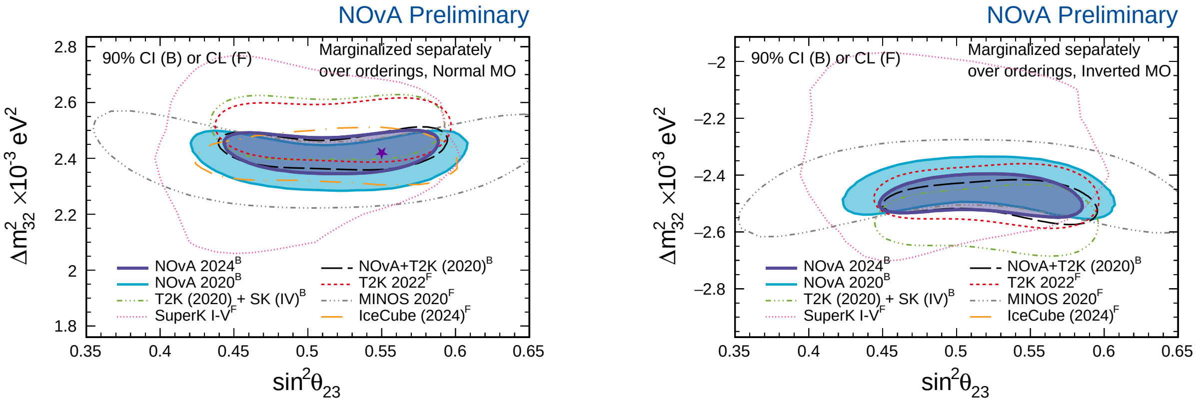


Figure 6.32: $sin^2\theta_{23}$ vs Δm^2_{32} 90% intervals for this result compared to other experiments, NO (left) and IO (right)

6.3.8 Validation

Figures 6.33 - 6.41 show consistency between NOvA's two MCMC samplers: Stan and ARIA. Overall, the two have nearly flawless consistency, which provides strong validation for the results from the Bayesian fitting of NOvA's data for the 2024 analysis.

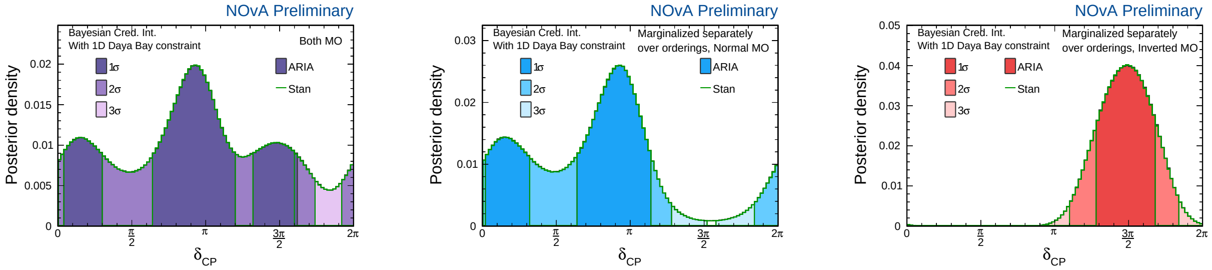


Figure 6.33: δ_{CP} with the 1D reactor constraint applied, comparing Stan and ARIA

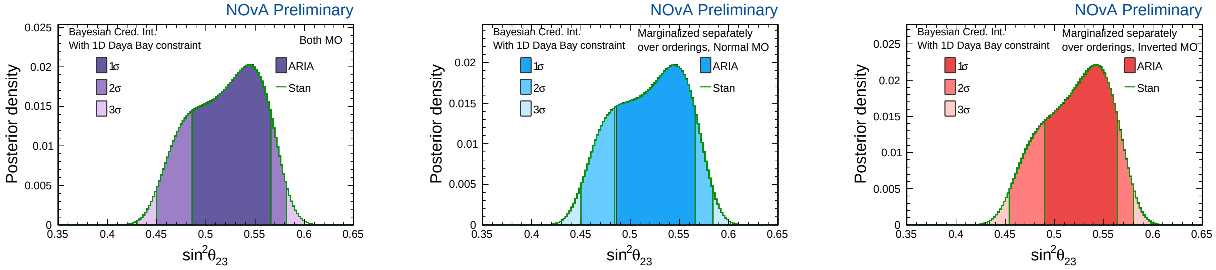


Figure 6.34: $sin^2\theta_{23}$ with the 1D reactor constraint applied, comparing Stan and ARIA

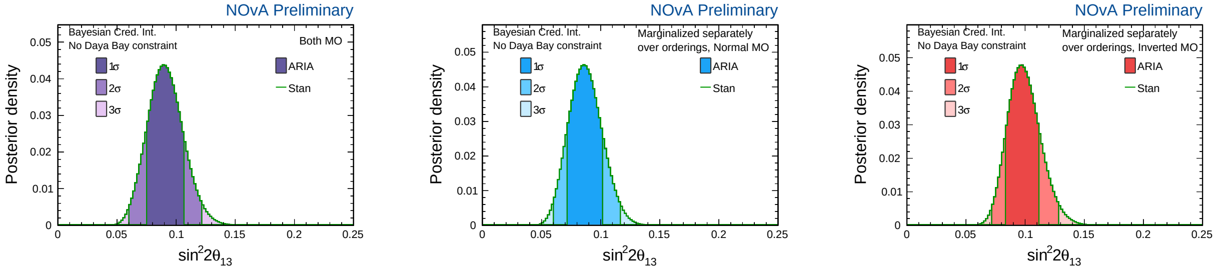


Figure 6.35: $sin^2 2\theta_{13}$ with no reactor constraint applied, comparing Stan and ARIA

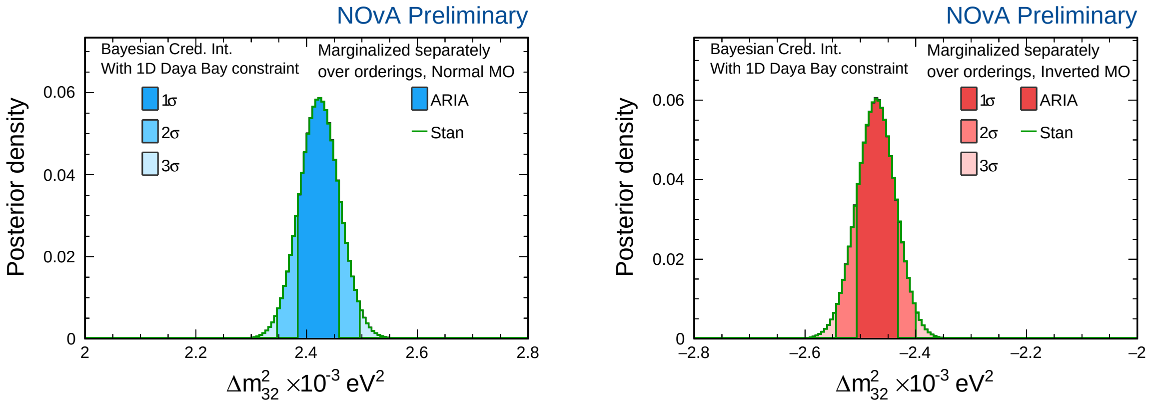


Figure 6.36: Δm^2_{32} with the 1D reactor constraint applied, comparing Stan and ARIA

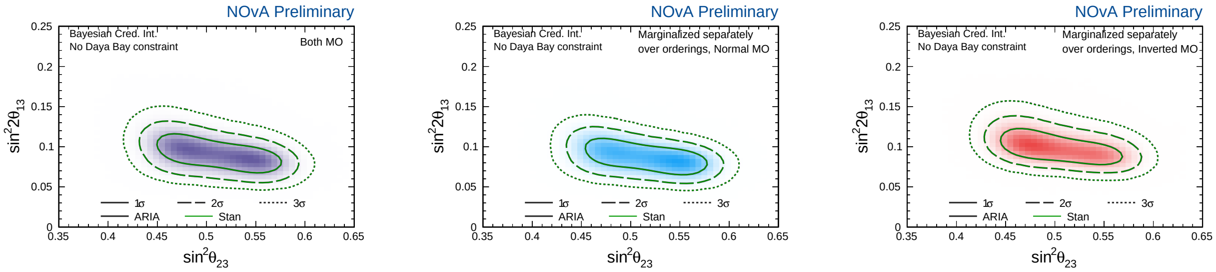


Figure 6.37: $sin^2\theta_{23}$ vs $sin^2 2\theta_{13}$ with no reactor constraint applied, comparing Stan and ARIA

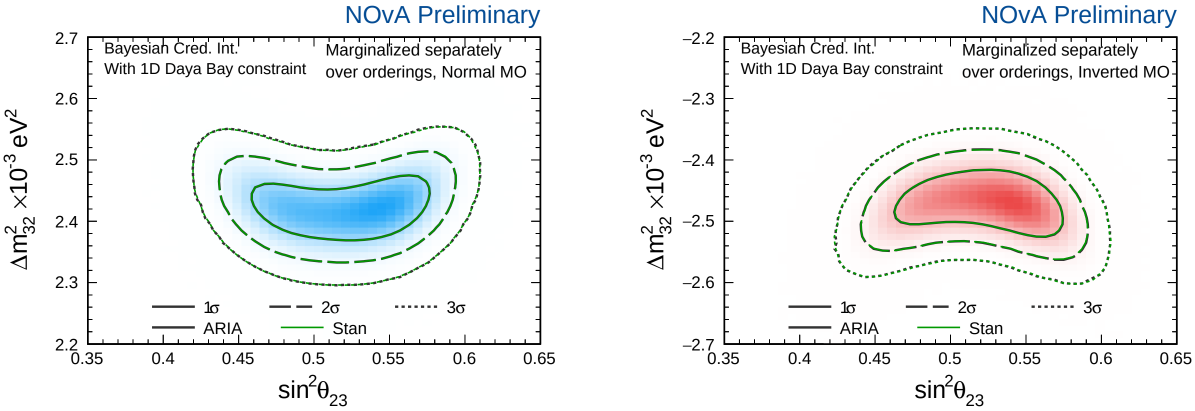


Figure 6.38: $sin^2\theta_{23}$ vs Δm^2_{32} with the 1D reactor constraint applied, comparing Stan and ARIA

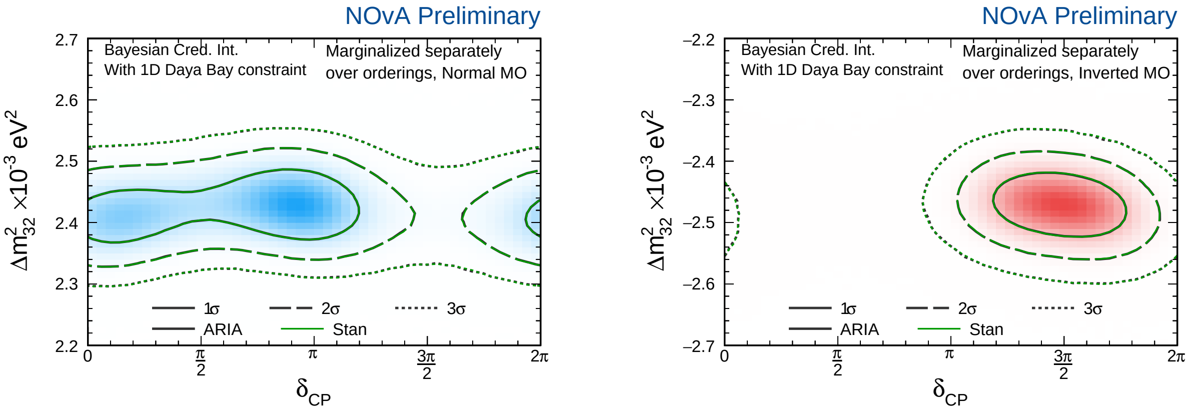


Figure 6.39: δ_{CP} vs Δm^2_{32} with the 1D reactor constraint applied, comparing Stan and ARIA

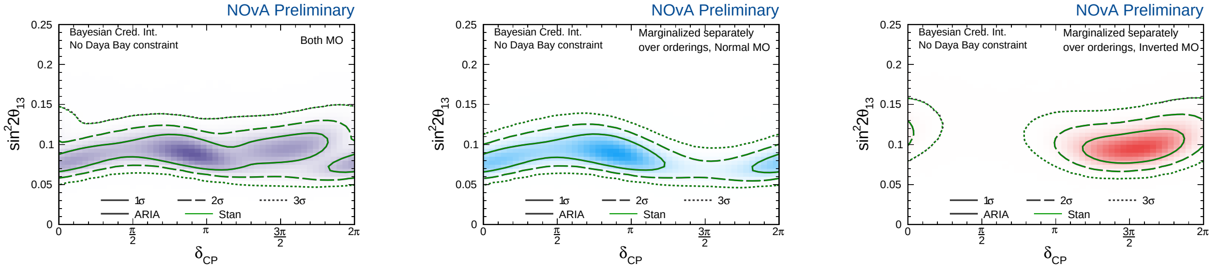


Figure 6.40: δ_{CP} vs $sin^2 2\theta_{13}$ with no reactor constraint applied, comparing Stan and ARIA

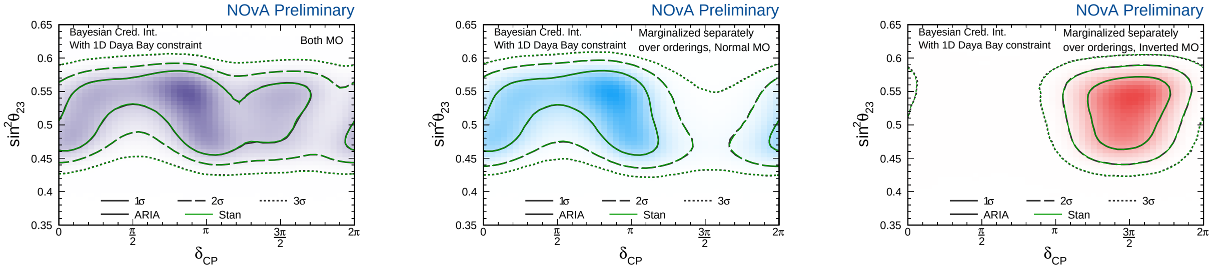


Figure 6.41: δ_{CP} vs $sin^2\theta_{23}$ with the 1D reactor constraint applied, comparing Stan and ARIA

Chapter 7

Conclusions

This dissertation describes the measurement of neutrino oscillation parameters using muon neutrino disappearance and electron neutrino appearance from 26.6 $\times 10^{20}$ POT of neutrino beam exposure and 12.5 $\times 10^{20}$ POT of antineutrino beam. This POT was the result of 10 years of NOvA data taking, and 95.6% more neutrino beam exposure since the last NOvA analysis.

This analysis yielded the following results for the oscillation parameters for the 1D reactor constraint:

Assuming Normal Ordering:

$$\Delta m^2_{32} = 2.424^{+0.035}_{-0.040} \times 10^{-3}\mathrm{eV}^2$$

$$sin^2\theta_{23} = 0.55^{+0.02}_{-0.06}$$

$$\delta_{CP} = 0.930^{+0.210}_{-0.290}\,\pi \,\cup\, 0.150^{+0.150}_{-0.110}\,\pi$$

Assuming Inverted Ordering:

$$\Delta m^2_{32} = 2.472^{+0.035}_{-0.040} \times 10^{-3}\text{eV}^2$$

$$sin^2\theta_{23} = 0.55^{+0.02}_{-0.05}$$

$$\delta_{CP} = 1.490^{+0.190}_{-0.210}\,\pi$$

with 76% preference for normal ordering.

This is currently the most precise measurement of Δm^2_{32} made by any single experiment. Consistency with reactor experiment's $sin^2(2\theta_{13})$ measurement is also used in different ways, and by using a 2D constraint from Daya Bay from their measurement of $sin^2(2\theta_{13})$ vs Δm^2_{32}, the preference for the normal ordering increases to 87%.

This author worked on many aspects of this analysis, including the energy estimation for both the ν_μ and ν_e events, systematic uncertainties, including the flux and beam uncertainty modeling, cross section uncertainty modeling, and the implementation of a new uncertainty into the analysis for Geant4 modeling uncertainty, and the Bayesian fit. This author also worked on several methods to improve future NOvA analysis, including many CNN based reconstruction algorithms, and helped build the framework to implement systematic uncertainties for these and other deep learning based algorithms.

Bibliography

[1] M. Abadi et al. TensorFlow: A system for large-scale machine learning, 2016.

[2] J. N. Abdurashitov et al. Solar neutrino flux measurements by the Soviet-American Gallium Experiment (SAGE) for half the 22 year solar cycle. *J. Exp. Theor. Phys.*, 95:181–193, 2002.

[3] K. Abe et al. Indication of Electron Neutrino Appearance from an Accelerator-Produced Off-Axis Muon Neutrino Beam. *Physical Review Letters*, 107(4), July 2011.

[4] K. Abe et al. Hyper-Kamiokande Design Report, 2018.

[5] K. Abe et al. Measurements of neutrino oscillation parameters from the T2K experiment using

$$3.6 \times 10^{21}$$

protons on target. *The European Physical Journal C*, 83(9), Sept. 2023.

[6] S. Abe et al. First joint oscillation analysis of Super-Kamiokande atmospheric and T2K accelerator neutrino data, 2024.

[7] Y. Abe et al. Indication of Reactor $\bar{\nu}_e$ Disappearance in the Double Chooz Experiment. *Physical Review Letters*, 108(13), Mar. 2012.

[8] B. Abi et al. Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I: Introduction to DUNE, 2020.

[9] P. Abratenko et al. Search for an Excess of Electron Neutrino Interactions in MicroBooNE Using Multiple Final State Topologies, 2022.

[10] A. Abusleme et al. JUNO physics and detector. *Prog. Part. Nucl. Phys.*, 123:103927, 2022.

[11] M. A. Acero, A. A. Aguilar-Arevalo, and D. J. Polo-Toledo. Neutrino Oscillation Analysis of 217 Live Days of Daya Bay and 500 Live Days of RENO. *Advances in High Energy Physics*, 2020:1–13, Jan. 2020.

[12] M. A. Acero et al. Improved measurement of neutrino oscillation parameters by the NOvA experiment. *Phys. Rev. D*, 106(3):032004, 2022.

[13] M. A. Acero et al. The Profiled Feldman-Cousins technique for confidence interval construction in the presence of nuisance parameters. 2022.

[14] M. A. Acero et al. Expanding neutrino oscillation parameter measurements in NOvA using a Bayesian approach. 11 2023.

[15] C. Achar et al. Detection of muons produced by cosmic ray neutrinos deep underground. *Physics Letters*, 18(2):196–199, 1965.

[16] P. Adamson et al. Combined Analysis of ν_μ Disappearance and $\nu_\mu \to \nu_e$ Appearance in MINOS Using Accelerator and Atmospheric Neutrinos. *Phys. Rev. Lett.*, 112:191801, May 2014.

[17] P. Adamson et al. The NuMI neutrino beam. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 806:279–306, Jan. 2016.

[18] P. Adamson et al. Precision Constraints for Three-Flavor Neutrino Oscillations from the Full MINOS+ and MINOS Dataset. *Physical Review Letters*, 125(13), Sept. 2020.

[19] N. Agafonova et al. Final Results of the OPERA Experiment on ν_τ Appearance in the CNGS Neutrino Beam. *Physical Review Letters*, 120(21), May 2018.

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

[21] A. Aguilar et al. Evidence for Neutrino Oscillations from the Observation of Electron Anti-neutrinos in a Muon Anti-Neutrino Beam. *Physical Review D*, 64(11), Nov. 2001.

[22] A. A. Aguilar-Arevalo et al. Updated MiniBooNE neutrino oscillation results with increased data and new background studies. *Physical Review D*, 103(5), Mar. 2021.

[23] Q. R. Ahmad et al. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, Jun 2002.

[24] M. H. Ahn et al. Measurement of neutrino oscillation by the K2K experiment. *Physical Review D*, 74(7), Oct. 2006.

[25] F. P. An et al. Precision Measurement of Reactor Antineutrino Oscillation at Kilometer-Scale Baselines by Daya Bay. *Phys. Rev. Lett.*, 130:161802, Apr 2023.

[26] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator: Physics and User Manual, 2015.

[27] B. Armbruster et al. Upper limits for neutrino oscillations muon-antineutrino to electron-antineutrino from muon decay at rest. *Physical Review D*, 65(11), June 2002.

[28] Y. Ashie et al. Evidence for an Oscillatory Signature in Atmospheric Neutrino Oscillations. *Phys. Rev. Lett.*, 93:101801, Sep 2004.

[29] A. Aurisano et al. The NOvA simulation chain. *J. Phys. Conf. Ser.*, 664(7):072002, 2015.

[30] A. Aurisano et al. A convolutional neural network neutrino event classifier. *Journal of Instrumentation*, 11(09):P09001–P09001, Sept. 2016.

[31] D. S. Ayres et al. The NOvA Technical Design Report. 10 2007.

[32] M. Baird et al. Event Reconstruction Techniques in NOvA. *Journal of Physics: Conference Series*, 664(7):072035, dec 2015.

[33] G. Bak et al. Measurement of Reactor Antineutrino Oscillation Amplitude and Frequency at RENO. *Physical Review Letters*, 121(20), Nov. 2018.

[34] V. Balas, R. Kumar, and R. Srivastavs. *Recent Trends and Advances in Artificial Intelligence and Internet of Things*. 01 2020.

[35] P. Baldi et al. Improved energy reconstruction in NOvA with regression convolutional neural networks. *Physical Review D*, 99(1), Jan. 2019.

[36] K. Bays et al. Ana2024 Cross-Section Modeling Technical Note. NOvA Internal Document 61559, 2024.

[37] J. Bian et al. RegCVN - Nue Energy Estimator Technote. NOvA Internal Document 42363, 2019.

[38] S. Brooks et al. *Handbook of Markov Chain Monte Carlo*. CRC Press, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742, 2011.

[39] R. N. Cahn et al. White Paper: Measuring the Neutrino Mass Hierarchy, 2013.

[40] J. Calcutt et al. Geant4Reweight: a framework for evaluating and propagating hadronic interaction uncertainties in Geant4. *Journal of Instrumentation*, 16(08):P08042, Aug. 2021.

[41] B. Carpenter et al. Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1):1–32, 2017.

[42] E. Catano-Mur et al. 3-Flavor changes between Prod 5 and Prod 5.1. NOvA Internal Document 61532, 2024.

[43] E. Catano-Mur et al. 3-Flavour event selection for Ana2024 analysis. NOvA Internal Document 61640, 2024.

[44] E. Catano-Mur et al. Data Quality Technical Note Prod5.1/2023. NOvA Internal Document 59876, 2024.

[45] E. Catano-Mur et al. Prod 5.1 Detector Systematics Executive Summary. NOvA Internal Document 53225, 2024.

[46] E. Catano-Mur, R. Nichol, and Z. Vallari. 3-Flavor Asimov Points. NOvA Internal Document 59328, 2023.

[47] E. Catano-Mur, R. Nichol, and Z. Vallari. Three-Flavour Executive Summary 2024. NOvA Internal Document 61655, 2024.

[48] E. Catano-Mur, R. Nichol, and Z. Vallari. Three flavour Ana2024: Box opening request. NOvA Internal Document 62364, 2024.

[49] K. Chadwick et al. FermiGrid—experience and future plans. *Journal of Physics: Conference Series*, 119(5):052010, jul 2008.

[50] B. T. Cleveland et al. Measurement of the Solar Electron Neutrino Flux with the Homestake Chlorine Detector. *The Astrophysical Journal*, 496(1):505, mar 1998.

[51] I. Collaboration. Measurement of atmospheric neutrino oscillation parameters using convolutional neural networks with 9.3 years of data in IceCube DeepCore, 2024.

[52] G. Davies and J. Vasel. A Tech Note on Cosmic Overlays for NOvA Analyses. NOvA Internal Document 23378, 2017.

[53] R. Davis, D. S. Harmer, and K. C. Hoffman. Search for Neutrinos from the Sun. *Phys. Rev. Lett.*, 20:1205–1209, May 1968.

[54] P. B. Denton and S. J. Parke. Simple and precise factorization of the Jarlskog invariant for neutrino oscillations in matter. *Physical Review D*, 100(5), Sept. 2019.

[55] P. B. Denton and R. Pestes. The impact of different parameterizations on the interpretation of CP violation in neutrino oscillations. *Journal of High Energy Physics*, 2021(5), May 2021.

[56] S. Duane et al. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, Sept. 1987.

[57] J. C. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3):32–57, 1973.

[58] A. Dye, A. Lister, and J. Wolcott. Ana2024: FD NuMI POT accounting. NOvA Internal Document 62111, 2024.

[59] P. Désesquelles et al. Cross talk and diaphony in neutron detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 307(2):366–373, 1991.

[60] K. Eguchi et al. First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance. *Physical Review Letters*, 90(2), Jan. 2003.

[61] G. J. Feldman and R. D. Cousins. Unified approach to the classical statistical analysis of small signals. *Physical Review D*, 57(7):3873–3889, Apr. 1998.

[62] Y. Fukuda et al. Solar Neutrino Data Covering Solar Cycle 22. *Phys. Rev. Lett.*, 77:1683–1686, Aug 1996.

[63] Y. Fukuda et al. Evidence for Oscillation of Atmospheric Neutrinos. *Physical Review Letters*, 81(8):1562–1567, Aug. 1998.

[64] Y. Fukuda et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998.

[65] Y. Fukuda et al. Measurements of the Solar Neutrino Flux from Super-Kamiokande's First 300 Days. *Phys. Rev. Lett.*, 81:1158–1162, Aug 1998.

[66] M. Fukugita and T. Yanagida. Baryogenesis Without Grand Unification. *Phys. Lett. B*, 174:45–47, 1986.

[67] Artificial Neural Networks and its Applications. `https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/`. Accessed: 2024-07-12.

[68] A. Gelman et al. *Bayesian Data Analysis.* CRC Press, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742, 2014.

[69] I. J. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* MIT Press, Cambridge, MA, USA, 2016. `http://www.deeplearningbook.org`.

[70] C. Hagmann, D. Lange, and D. Wright. Cosmic-ray shower generator (CRY) for Monte Carlo transport codes. In *2007 IEEE Nuclear Science Symposium Conference Record*, 10 2007.

[71] W. Hampel et al. GALLEX solar neutrino observations: results for GALLEX IV. *Physics Letters B*, 447(1):127–133, 1999.

[72] K. He et al. Deep Residual Learning for Image Recognition, 2015.

[73] A. Himmel, A. Moren, and E. Niner. Early Permanent Cosmic Rejection with a CNN. NOvA Internal Document 42766, 2020.

[74] K. Hirata et al. Observation of a small atmospheric ν_μ/ν_e ratio in Kamiokande. *Physics Letters B*, 280(1):146–152, 1992.

[75] S. Hochreiter and J. Schmidhuber. Long Short-term Memory. *Neural computation*, 9:1735–80, 12 1997.

[76] M. D. Hoffman and A. Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. page 1111.4246, 11 2011.

[77] P. V. C. Hough. *Method and means for recognizing complex patterns.* U.S. Patent 3,069,654, Dec 1962.

[78] A. G. Howard et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017.

[79] B. Jargowsky. NuMu Hadronic Energy CNN Technote for Prod5.1. NOvA Internal Document 48360, 2020.

[80] B. Jargowsky. Numu Energy Regression CNN Update. NOvA Internal Document 60180, 2024.

[81] B. Jargowsky. PCA TNs: Flux, Genie, & Geant4. NOvA Internal Document 62099, 2024.

[82] B. Jargowsky and R. Nichol. Prod5.1 Numu Energy Estimators Technote. NOvA Internal Document 61064, 2024.

[83] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[84] C. W. Kim and A. Pevsner. *Neutrinos in physics and astrophysics*, volume 8. 1993.

[85] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization, 2017.

[86] G. Knoll. *Radiation Detection and Measurement (4th ed.)*. John Wiley, Hoboken, NJ, 2010.

[87] L. Kolupaeva, O. Samoylov, and I. Shandrov. Matter effect in neutrino oscillations for long-baseline experiments. *Phys. Part. Nucl. Lett.*, 14(7):975–980, 2017.

[88] R. K. Kutschke. art: A framework for new, small experiments at Fermilab. *J. Phys. Conf. Ser.*, 331:032019, 2011.

[89] J. Lee, W. Sung, and J.-H. Choi. Metamodel for Efficient Estimation of Capacity-Fade Uncertainty in Li-Ion Batteries for Electric Vehicles. *Energies*, 8(6):5538–5554, 2015.

[90] M. Mandal. Introduction to Convolutional Neural Networks (CNN). `https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/`. Accessed: 2024-07-12.

[91] Metropolis, Nicholas and others. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6):1087–1092, June 1953.

[92] A. Mislivec. Tech Note: Near-to-Far Extrapolation in Transverse Momentum for the NOvA 2020 3-Flavor Analysis. NOvA Internal Document 44401, 2020.

[93] N. Nayak. Flux Systematics for the 2018 NOvA Oscillation Analyses. NOvA docid=27884, 2018.

[94] T. Nosek. Genie Cross Section Systematics for the 2020 3Flavor Analysis Using PCA. NOvA docid=44274, 2020.

[95] M. Ohlsson. Extensions and explorations of the elastic arms algorithm. *Computer Physics Communications*, 77(1):19–32, 1993.

[96] A. Paszke et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library, 2019.

[97] D. Pershey, J. Huang, and M. Judah. TDSlicer Technote. NOvA Internal Document 27689, 2018.

[98] M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory.* Addison-Wesley, Reading, USA, 1995.

[99] J. Porter. ReMId retraining 2019 technote. NOvA Internal Document 42277, 2019.

[100] F. Psihas. *Measurement of Long Baseline Neutrino Oscillations And Improvements From Deep Learning.* PhD thesis, Indiana University, 2018.

[101] F. Psihas et al. Context-Enriched Identification of Particles with a Convolutional Network for Neutrino Events. *Phys. Rev. D*, 100(7):073005, 2019.

[102] M. Rabelhofe. MENATE neutron file systematic. NOvA Internal Document 59999, 2024.

[103] A. Radovic. LSTMs For Energy Estimation a Proof of Concept. NOvA Internal Document 18801, 2017.

[104] C. Rubbia et al. Underground operation of the ICARUS T600 LAr-TPC: first results. *Journal of Instrumentation*, 6(07):P07011–P07011, July 2011.

[105] W. Shorrock. Numu Binning. NOvA Internal Document 60579, 2024.

[106] G. Song. Model Tests of Inclusive CC Proton Production Over Full 5D Phase Space. 2018.

[107] A. Sousa et al. Implementation of Feldman-Cousins Corrections and Oscillation Calculations in the HPC Environment for the NOvA Experiment. *EPJ Web Conf.*, 214:05012, 2019.

[108] M. Srednicki. *Quantum field theory.* Cambridge University Press, 1 2007.

[109] C. Sweeney. Technical not for Geant4 reweighting systematic. NOvA Internal Document 56031, 2022.

[110] C. Szegedy et al. Going Deeper with Convolutions, 2014.

[111] D. Torbunov. NuMu Energy Estimator Technote for Prod3 MC. NOvA Internal Document 23342, 2017.

[112] D. Torbunov. LSTM NuMu Energy Estimator Technote for Prod5. NOvA Internal Document 42304, 2020.

[113] Z. Vallari. NOvA-T2K Joint Analysis Results. Joint Experimental-Theoretical Physics Seminar, Fermilab, 2024.

[114] T. Wester et al. Atmospheric neutrino oscillation analysis with neutron tagging and an expanded fiducial volume in Super-Kamiokande I–V. *Physical Review D*, 109(7), Apr. 2024.

[115] Wikipedia contributors. Principal component analysis — Wikipedia, the free encyclopedia, 2016. [Online; accessed 17-July-2024].

[116] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1):60 – 62, 1938.

[117] J. Wolcott. New neutrino oscillation results from NOvA with 10 years of data. XXXI International Conference on Neutrino Physics and Astrophysics, 2024.

[118] S. Wu. LSTM Energy Estimation in NOvA Experiment. NOvA Internal Document 55321, 2022.

[119] Y. Xiao. Measurement of Neutrino-electron Elastic Scattering in the NOvA Near Detector. WIN 2023, 2023.

[120] T. Zimmerman. The MASDA-X chip - A New multichannel ASIC for readout of pixelated amorphous silicon arrays. 12 1998.

[121] P. A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

Appendix A

Deep Learning Methods

A.1 Neural Networks

Neural Networks (NNs), also called artificial neural networks, are computational networks that map an input to an output using a number of "nodes" with tuned weights [69]. See figure A.1 for an illustration. The "input layer" can be thought of as a list of numbers input, one for each node shown. Each line in the figure represents a connection between nodes, where the numbers from the column of nodes on the left is transformed into the adjacent column. Each line can be represented by a transformation on the last node, $wx + b$, where w is a weight, b is a bias, and x is the value of the last node. For a node i in one layer, the value, x'_i, is the sum of all the connections coming from the last layer, x_j, passed through an "activation function", f:

$$x'_i = f(\sum_j w_{i,j} x_j + b_i) \tag{A.1}$$

. The activation function can be anything, but common choices are linear or “ReLU”:

$$ReLU(x) = \begin{cases} x & \text{if} \quad x > 0 \\ 0 & \text{otherwise} \end{cases}$$

.

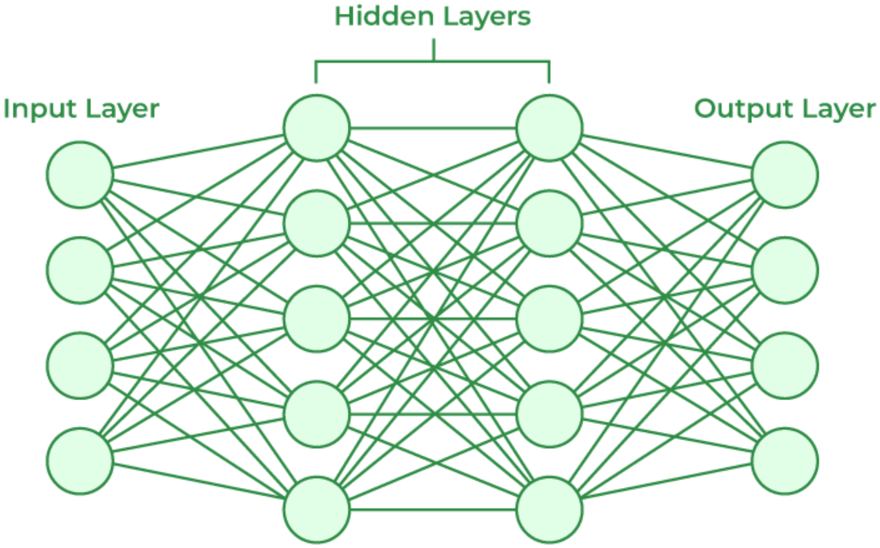


Figure A.1: Illustration of fully connected neural network. Image taken from Reference [67].

The output layer is a specially treated layer, as here you need to match the format of the output data you are trying to fit to. In the case of regression, where we want the output to follow some possibly continuous spectrum of values, we have one node with a linear activation function in the case of one regression output, or any number of nodes to match the number of regression outputs needed. For the case of classification, we want to output one or more numbers between 0 and 1 to represent how confident the network is the input is or is not in a certain class. For binary classification, we want one number between 0 and 1, so we have one output node with a sigmoid function, σ, activation:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

which will map an input x to a number between 0 and 1, in a smooth fashion, which has a non-zero derivative everywhere, which will be important for learning, discussed in section A.1.1. For multiple classes, the "softmax" function will be used with as many nodes as classes, which will jointly normalize these numbers between 0 and 1, such that they add up to 1.

The figure here shows a fully connected network, but more specialized networks with different connection schemes also exist and are more common to use than a simple network of this form, but the ideas shown here will be the same.

A.1.1 Gradient Descent Learning

We've seen that neural networks have tuned parameters to achieve a mapping between inputs and outputs, but this is only a useful concept if we have a way to tune them to achieve such a mapping. We call the process of updating these weights "training". For a supervised network, such as the ones we'll discuss here, we have a training dataset that has the inputs as well as the desired outputs. We then have a "loss function" that tells us how close the inputs and outputs are, to quantify if we are getting better or worse.

We update these weights by going through the training dataset, and using gradient descent to moving the weights in the direction which minimizes the loss function. The gradient here is taken in parameter space, which then gives the negative of the direction to update the weights. Many software frameworks exist to implement neural networks and optimized training procedures, such as TensorFlow [1] and PyTorch [96].

A.2 Convolutional Neural Networks (CNN)

CNNs are a specialized type of NN that use a number of “filters” on an input image to capture characteristics in a way that is translationally invariant across the image [34], see figure A.2. NOvA uses CNNs in several applications, as the NOvA detector can see events as two “images”, one for each view, both for regression and classification tasks [30, 101, 35]. There are many CNN architectures, which use different constructions of CNN layers combined with other layer types. NOvA uses MobileNet [78], GoogLeNet [110], and ResNet [72].

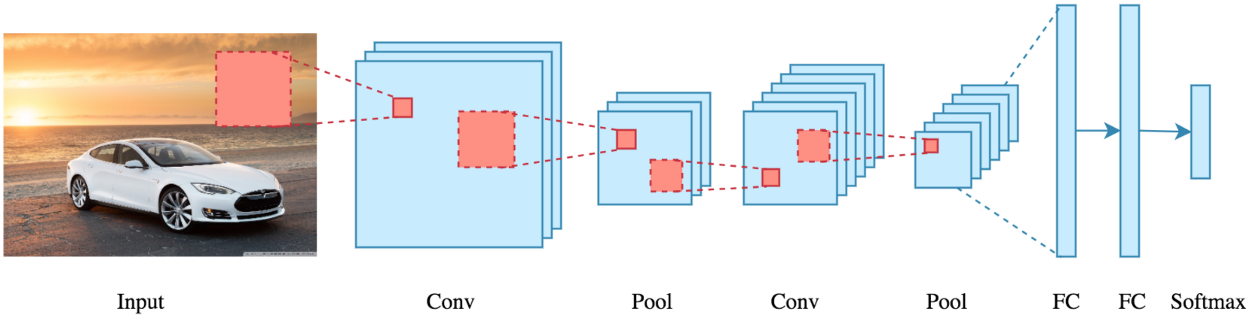


Figure A.2: Illustration of a CNN. Image taken from Reference [90].