Measuring the Atmospheric Neutrino Oscillation Parameters with IceCube DeepCore

by

Kayla Leonard DeHolton

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Physics)

at the

University of Wisconsin – Madison

2022

Date of final oral examination: 08/24/2022

The dissertation is approved by the following members of the Final Oral Committee:

Francis Halzen, Gregory Breit Professor and Hilldale Professor, Department of Physics

Albrecht Karle, Professor, Department of Physics

Vernon Barger, Vilas Professor and Van Vleck Professor, Department of Physics

Jessi Cisewski-Kehe, Assistant Professor, Department of Statistics

Abstract

For decades, the Standard Model of Particle Physics has stood the test of time, being one of the most comprehensive and reliable models ever proposed. One of the few exceptions to its robustness was the discovery of neutrino oscillations (and consequently the implication that neutrinos have mass). It is one of the only confirmed pieces of evidence of physics beyond the Standard Model. Since this discovery, there has been a worldwide effort on both the theoretical and experimental fronts to answer many questions that this discovery raised. Many neutrino experiments around the world seek to measure the model parameters that describe these oscillations within the current theoretical 3 neutrino framework.

The IceCube Neutrino Observatory is a kilometer-scale detector embedded in the Antarctic ice at the South Pole and detects Cherenkov light produced by neutrinos interacting in the ice. DeepCore is a densely instrumented sub-array within the full detector that observes interactions of atmospheric neutrinos down to 5 GeV. At these energies, Earth-crossing muon neutrinos have a high chance of oscillating to tau neutrinos. DeepCore is able to measure these oscillations with precision comparable to accelerator-based experiments, but it is also complementary to accelerator measurements because it probes longer distance scales and higher energies, peaking above the tau lepton production threshold.

This dissertation presents the effort involved in curating one of the largest neutrino oscillation datasets in the world, with over 200,000 events spanning almost 10 years (from April 2012 to January 2022) and a neutrino purity of over 99%. This sample is optimized for performing a measurement of the atmospheric neutrino oscillation parameters. The nearly unprecedented level of statistics also requires unprecedented precision in the treatment of systematic uncertainties. Compared to previous DeepCore analyses, this analysis benefits from improved background rejection, new simulation, a new reconstruction, a new machine learning particle classification algorithm, improved modeling of systematic uncertainties, and more years of data. Given the large number of changes compared to previous analyses, an extensive suite of quality and robustness checks were performed.

Even with the extensive testing program, our fit to data informed us that our modeling is not yet on par with the level of statistical precision achieved with the new sample; the goodness-of-fit between our data and simulation yielded a p-value of 0%. At the time of writing, there remains an ongoing effort to improve our models to describe the full dataset. This dissertation includes the unblinded results of an analysis performed with a sub-sample containing 20% of the full dataset.

Acknowledgements

I would like to express my sincerest thanks to all the people that have encouraged my lifelong interest in science and have supported my scientific endeavors every step of the way.

To my parents, Dan and Polly –
I am where I am today because of you. You fostered my inquisitiveness at an early age with a Fact-of-the-Day every night, more Eyewitness Books than I could count, trips to the science museums in Chicago, and so much more. Thank you for encouraging my interests and for your never-ending support.

To my sister, Tory –
You've been by my side through every phase of life. We've shared everything, from the literal sense of clothes and toys as kids, to experiences like Mr. Lee's science camp and Space Camp in Huntsville, and now careers in STEM. Thank you for being beside me through it all.

To my husband, Kyle –
Thank you for being my right hand man in everything I do. From answering questions about coding, to proof-reading papers and emails, listening to me rehearse slides, and staying up for all the late nights of work, you probably feel like you've been in grad school for the past 5 years too. I sincerely appreciate every time you asked with a smile "Is there anything I can do to help you?". You happily moved to Madison, and now to Pennsylvania, and who knows where it'll be after that, but I am deeply grateful to have found someone so supportive of my scientific pursuits. I'm so grateful for all the life adventures, past, present, and future, that we get to do together.

To Sarah, Will, Manuel, Austin, Marjon, Raamis, Alex, Grant, and all the WIPAC grad students –
You have made the past 5 years unforgettable. I will forever cherish the memories of having lunches on the Square, spending Friday nights at the Terrace, Will asking every day if we want to go get ice cream, and traveling the world with you all for collaboration meetings. Thank you for being the most amazing group of coworkers-turned-friends.

To Abby, Joelle, Susan, Adrian, Leah, Trevor, Rob, and all the other PGSC & GMaWiP officers –
These groups were a transformative part of my grad school experience and my personal and professional development. I appreciate the camaraderie we shared as part of that and for everything I learned from you all about leadership styles, open-mindedness, and being an advocate for good.

To those that first introduced me to research, especially my first two undergrad research mentors Sarah and André –
Thank you for introducing me to how interesting and rewarding research can be, and for being patient as I was first starting out. To André, working with you was the reason I switched my career trajectory from astronomy to neutrino physics, and I am so glad that you introduced me to this exciting field.

To Tom, Jason, Étienne, Tetiana, and the NBI IceCube group in Copenhagen –
The sample and analysis in this dissertation would quite literally be nothing without your teamwork, mentorship, and guidance. I am truly grateful to all of you for taking me under your wing.

To everyone who contributed to OscNext –
It has been so fulfilling to watch OscNext blossom from an idea on a whiteboard at a collaboration meeting all those years ago to something real. It's been a privilege being part of this team of amazing researchers, and learning from each person's expertise and contributions.

And to Francis –
Thank you for being the best graduate advisor I could ever have conceived of. Thank you for giving me scientific and academic freedom to explore the topics that interested me; for connecting me with the resources and mentors that I needed to learn and grow as a researcher; and for always saying yes to opportunities that I never expected to have as a graduate student. I will never forget the level of unconditional trust that you extended to me from the very beginning (before I had probably even earned it!) which constantly made me feel respected, motivated, and confident in my own abilities. I strive to internalize these qualities so I can be as good of a mentor to others one day, as you have been to me. My sincerest thanks for everything you have done for me over the last 5 years.

Table of Contents

List of Figures

List of Acronyms

BDT	boosted decision tree
BFP	best fit point
CC	charged current
CP	charge-parity
CSMS	Cooper-Sarker-Mertsch-Sarkar
DIS	Deep Inelastic Scattering
DOM	Digital Optical Module
KDE	Kernel Density Estimation
LLH	log-likelihood
MC	Monte Carlo
NC	neutral current
PID	particle identification
PMNS	Pontecorvo–Maki–Nakagawa–Sakata
PMT	photomultiplier tube

1

Introduction

1.1 Overview

The field of neutrino physics has flourished in the last two decades since the discovery of neutrino oscillations [1–3]. Although the existence of this phenomenon is considered well established, the precise values of the parameters that govern these oscillations are still active questions. Even the best estimates, obtained by combining data from many experiments around the world [4], have percent-level uncertainty on these neutrino parameters. This experimental precision pales in comparison to the precision with which other fundamental particles' properties have been measured (properties such as the mass or magnetic moment of particles like the electron and muon are often quoted in parts-per-million, billion, or even trillion [5, 6]). And for good reason: neutrinos are notoriously difficult to observe, so they (and their properties) have evaded precision measurement.

IceCube DeepCore is one of many experiments around the world that has measured these parameters [7, 8] and continues to seek improved precision on the measurement of these parameters. I, along with a group of collaborators, have spent the past few years developing a new DeepCore neutrino sample and new analysis techniques for measuring the properties of neutrino oscillations. The sample, which we refer to as "OscNext", is now being used for almost a dozen analyses. Much of the development has been a group effort and therefore this dissertation is not intended to be a comprehensive summary of *all* parts of the sample and analysis development. I have instead tried to focus this dissertation on my own personal contributions, while

still including enough relevant context about the sample and analysis at large. Therefore, in the overview that follows, I provide clarification as to what my personal contributions were to this large group effort.

The remaining portion of this chapter will cover the relevant background information on neutrino physics as well as the DeepCore detector. Chapter 2 will cover the event selection and sample; I contributed the work that is discussed in Section 2.3, and a large portion of the work discussed in Section 2.4. Chapter 3 covers a technique I developed to increase the efficiency of our background muon simulation; I then used this technique to generate the muon simulation used across our sample and analyses. Chapter 4 includes the analysis methods, techniques, and choices that are common to many of the analyses that use the sample; I contributed heavily to the work described in Section 4.1. Chapters 5 and 6 cover the analysis for which I was the lead analyzer, and worked in conjunction with another collaborator. Chapter 7 presents the results of a fit to a 20% sub-sample of the real data; I was the sole analyzer for this work. Chapter 8 provides some concluding remarks and outlook.

1.2 Introduction to Neutrino Oscillations

Neutrinos in the Standard Model

The Standard Model of Particle Physics is widely accepted as the most comprehensive and accurate description of almost all of the known fundamental particles and forces (except gravity). See Figure 1.1. The Standard Model predicts 3 massless neutrinos which only interact via the Weak Force. Each neutrino (ν) gets its name, or "flavor", from the charged lepton in its generation: electron (e), muon (μ), or tau (τ).

In the Standard Model description, neutrinos only interact via the Weak Force and have an extremely low probability of interacting (i.e. a small cross section). The interaction types get their names based on the charge of the boson mediating the interaction. Charged current (CC) interactions are mediated by the charged W boson. See Figure 1.2a for an example. Neutral current (NC) interactions are mediated by the neutral Z boson. See Figure 1.2b for an example. The examples shown in Figure 1.2 are examples of Deep Inelastic Scattering (DIS) in which the interaction occurs with an individual quark (u or d) within the nucleon (proton p or neutron n), and breaks apart the nucleon, producing a shower of additional particles. While there are other types of neutrino interactions that occur at lower energies, these are the main interaction types that will be relevant for this work.

One of the only experimentally confirmed pieces of evidence of physics beyond the Standard Model was the discovery of neutrino oscillations, which will be introduced in the next section [1–3]. This phenomenon requires that neutrinos have mass, whereas the Standard Model predicted that neutrinos have no mass. This discovery paved the way for what is now a very active field of research – neutrino physics – which aims to measure the parameters that govern these oscillations, measure the neutrino mass and determine the origin of neutrino mass, search for additional neutrinos outside the Standard Model, and more.

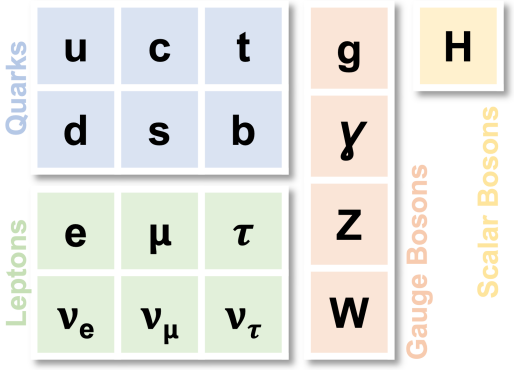
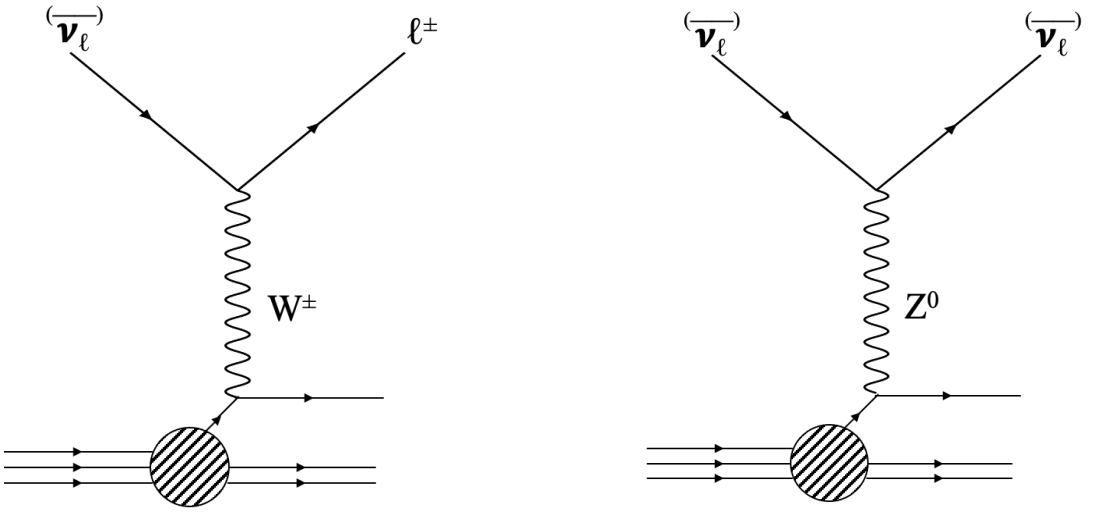


Figure 1.1: The Standard Model of Particle Physics consists of quarks, leptons, and bosons. The leptons consist of three neutrinos ν and three charged leptons e, μ, and τ.



(a) Charged current interaction (CC)

(b) Neutral current interaction (NC)

Figure 1.2: Diagram of Deep Inelastic Scattering (DIS) neutrino interactions. (a) Charged current interaction mediated by a W boson. The incoming particle is a neutrino; the outgoing particle is its associated charged lepton. (b) Neutral current interaction mediated by a Z boson. The incoming particle is a neutrino, and the outgoing particle is the same neutrino. The circle with the hatch pattern represents a nucleon (proton p or neutron n) that has been broken apart and then the interaction occurs with an individual quark. This process of breaking apart the nucleon is known as Deep Inelastic Scattering.

Neutrino Oscillations and the 3ν Paradigm

Most particles in the Standard Model have a definite mass; however, for the neutrinos the picture is more complicated. While there are the three flavors of neutrinos (ν_e, ν_μ, and ν_τ) and there are three neutrino mass states (ν_1, ν_2, and ν_3), there is not a 1:1 correspondence between these states. The mass states form one basis and the flavor states form a second basis. A neutrino in a definite state in one basis can be expressed as a quantum superposition of the states in the other basis. Figure 1.3 shows the contributions of each flavor state to each mass state. This mismatch of bases results in the phenomenon of "neutrino oscillations", whereby a neutrino can change flavor as it propagates over macroscopic distances.

Figure 1.3: Decomposition of each mass state (ν_1, ν_2, and ν_3) into the flavor states (ν_e, ν_μ, and ν_τ). Data from [4].

The matrix that relates these two bases is known as the PMNS matrix (named after B. Pontecorvo, Z. Maki, M. Nakagawa, and S. Sakata) and is shown in Equation 1.1. This matrix is often notated as U due to the fact that the matrix is assumed to be unitary in the standard 3ν description.

$$\begin{bmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{bmatrix} = \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} \tag{1.1}$$

This matrix can be decomposed into 3 rotation matrices, with each relating a pair of two states i and j. Each pair of states has a characteristic angle θ_{ij} that describing the mixing of those two states. There is one additional term introduced known as the "CP-violating phase". Charge-parity (CP) symmetry dictates that a particle should behave the same as an antiparticle with the opposite spin. It has not yet been experimentally confirmed whether there is CP-violation in the neutrino sector, so a CP-volating phase δ_{CP} is introduced to allow for this possibility. By convention [5], the matrix is expressed in the following way:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_{23}) & \sin(\theta_{23}) \\ 0 & -\sin(\theta_{23}) & \cos(\theta_{23}) \end{pmatrix} \begin{pmatrix} \cos(\theta_{13}) & 0 & \sin(\theta_{13})e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -\sin(\theta_{13})e^{i\delta_{CP}} & 0 & \cos(\theta_{13}) \end{pmatrix} \begin{pmatrix} \cos(\theta_{12}) & \sin(\theta_{12}) & 0 \\ -\sin(\theta_{12}) & \cos(\theta_{12}) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then, abbreviating $\cos(\theta_{ij})$ and $\sin(\theta_{ij})$ to c_{ij} and s_{ij} for simplicity, we find that

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta_{CP}} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta_{CP}} & c_{13}c_{23} \end{pmatrix}. \tag{1.2}$$

This matrix can be used to express the flavor state $\alpha \in \{e, \mu, \tau\}$ that a neutrino is created in as a linear combination of the mass states $i \in \{1, 2, 3\}$:

$$|\nu_\alpha\rangle = \sum_{i=1}^{3} U_{\alpha i}^* |\nu_i\rangle .$$

As the neutrino propagates, it evolves according to the standard time evolution operator, and thus the time-evolved state at some later time t can be expressed in terms of its energy E:

$$|\nu_\alpha(t)\rangle = \sum_{i=1}^{3} U_{\alpha i}^* e^{-iE_i t} |\nu_i(t=0)\rangle .$$

Note that this is expressed in natural units where $\hbar = 1$ and $c = 1$. We can then determine the probability P that a neutrino which started in the flavor α is now β after having some time to evolve:

$$P_{\alpha\beta} = \Big| \langle \nu_\beta | \nu_\alpha(t) \rangle \Big|^2 = \Big| \sum_{j=1}^{3} \sum_{i=1}^{3} U^*_{\alpha i} U_{\beta j} e^{-iE_i t} \langle \nu_j \rangle \, \nu_i(t) \Big|^2.$$

Then we apply the assumptions that neutrinos are ultrarelativistic: $E_i = \sqrt{p_i^2 + m_i^2} \approx p + \frac{m_i^2}{2E}$ and $t \approx \frac{L}{c} = L$ where L indicates the distance that the neutrino traveled during the time t. The probability simplifies to:

$$\begin{aligned} P_{\alpha\beta} = \delta_{\alpha\beta} \quad &- \quad 4 \sum_{i<j}^{3} \mathrm{Re} \left[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j} \right] \sin^2 \left(\frac{(m_i^2 - m_j^2) L}{4E} \right) \\ &+ \quad 2 \sum_{i<j}^{3} \mathrm{Im} \left[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j} \right] \sin \left(\frac{2(m_i^2 - m_j^2) L}{4E} \right) \end{aligned} \tag{1.3}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta (1 if $\alpha = \beta$, 0 if $\alpha \neq \beta$). To simplify notation, it is conventional to notate $m_i^2 - m_j^2$ as Δm_{ij}^2 and refer to these as the "mass splittings". We can see from this equation that the flavor changes are oscillatory in nature due to the sine term, and that the probabilities are given by the values of θ_{ij} in the PMNS matrix, Δm_{ij}^2, L, and E. Converting from natural units to units that are useful in an experimental setting, we find that the characteristic oscillatory pattern is driven by the sinusoidal term:

$$\sin^2 \left(1.267 \Delta m_{ij}^2 \frac{L}{E} \right) \tag{1.4}$$

when Δm_{ij}^2 is given in [eV^2] and L/E is in [km/GeV] (or equivalently, [m/MeV]).

Measuring Neutrino Oscillations

The physical parameters of interest to be measured are the mixing angles θ_{ij} from the PMNS matrix and the mass splittings Δm_{ij}^2. The observable parameters from the experiments are L and E. We see from Equation 1.4 that the value of Δm_{ij}^2 determines the relevant L/E scale. Two of the states are much closer together than the third, so for practical purposes, there are only two relevant mixing scales. The names of each mixing scale are derived from the historical anomaly that they solved. The first is the "solar" mixing (which solves the "solar neutrino problem", and is often abbreviated with the symbol $\odot$), with $\Delta m_{21}^2 = \mathcal{O}(10^{-5})$ eV^2 [5]. The second is the "atmospheric" mixing (which solves the "atmospheric neutrino anomaly"), with $\Delta m_{atm}^2 = \Delta m_{32}^2 \approx \Delta m_{31}^2 = \mathcal{O}(10^{-3})$ eV^2 [5]. Figure 1.4 shows the ability of many different experiments to measure various L and E scales. Some experiments have "fixed baselines" since the production sources and detector are both located at fixed points. Other experiments indicated in shaded

regions, such as DeepCore, can access a range of baselines; this will be discussed in more detail in the next section. The diagonal lines indicate the L/E scales that corresponded to the solar (dashed) and atmospheric (dot dashed) mixing. From this, we can see which type of mixing each experiment is sensitive to.

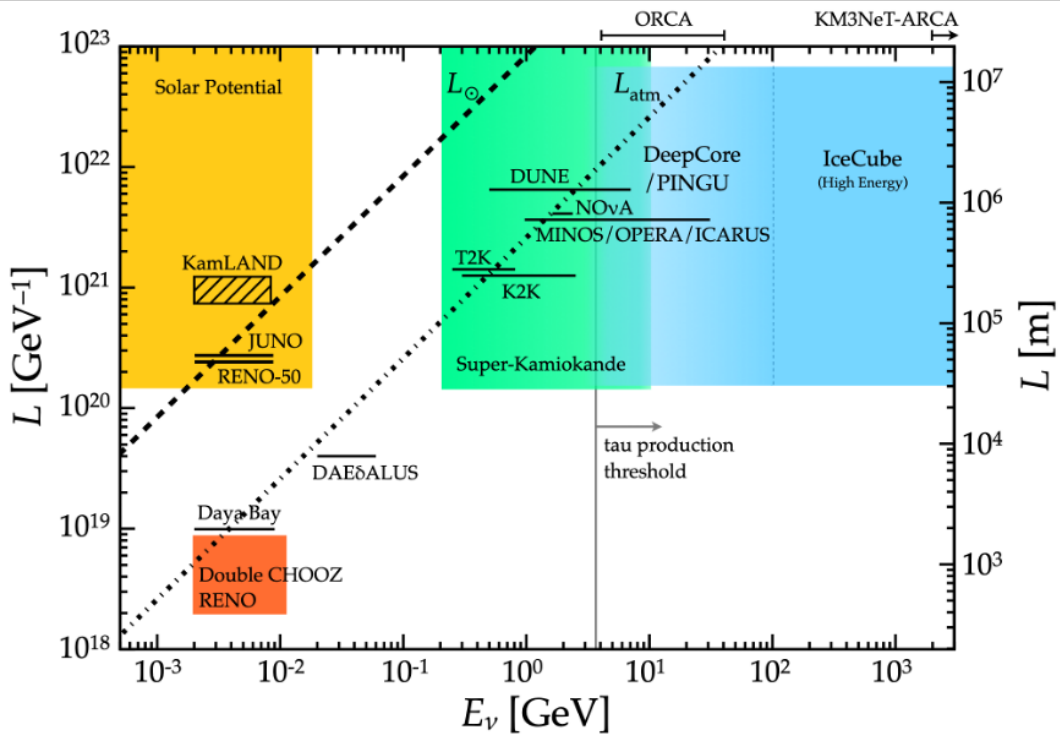


Figure 1.4: Experimental landscape for various L and E scales. Experiments marked with horizontal lines have "fixed baselines"; experiments with shaded/patterned regions, such as DeepCore, can access a range of baselines. The diagonal lines indicate the L/E scales that corresponded to the solar (dashed) and atmospheric (dot dashed) mixing. Figure Credit: [9].

Atmospheric Neutrino Oscillations

As mentioned in the previous section, the "atmospheric" parameters determine the mixing between the the third mass state and the first and second mass states. This was historically measured by experiments which detect neutrinos produced by the interaction of cosmic rays in Earth's atmosphere. These cosmic ray interactions produce mesons like pions and kaons, which subsequently can decay into neutrinos. The flux of atmospheric neutrinos is dominated by the muon flavor with a subleading contribution from the electron flavor. See Figure 1.5. Among each flavor, the neutrinos contribute a larger fraction than the antineutrinos.

These neutrinos are produced on a spherical shell across all of Earth with a median production height of $\sim$15 km above sea level [10]. This means that an experiment designed to observe atmospheric neutrinos can observe neutrinos produced $\sim$15 km away to neutrinos that were produced on the opposite side of the Earth and traveled an Earth diameter of $\sim 12,700$ km and any distance in between. Thus, as shown in Figure 1.6,

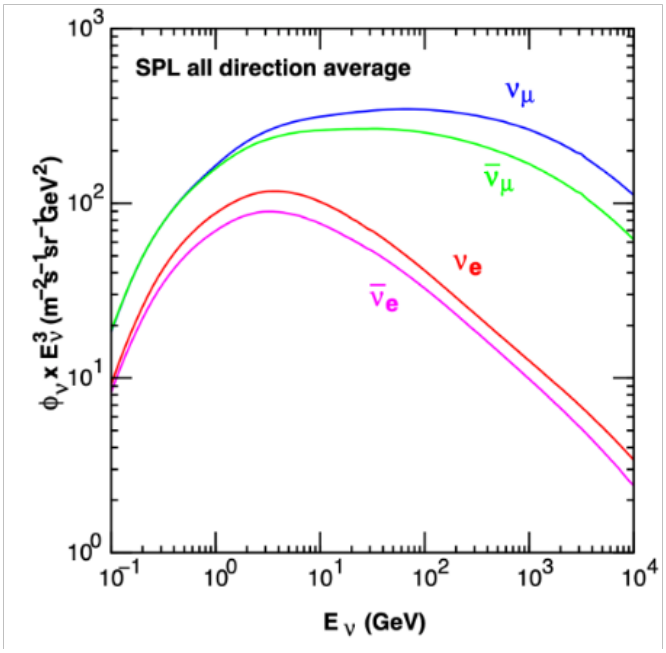


Figure 1.5: Predicted flux of atmospheric neutrinos at the South Pole, averaged over all directions. The flux is dominated by the muon flavor with a subleading contribution from the electron flavor. Figure Credit: [10].

the arrival direction provides the information about the baseline L for any atmospheric neutrino experiment. The baseline is related to the arrival direction by

$$L \approx -D_{Earth}\ \cos(\theta_{zenith})$$

where θ_{zenith} is the arrival direction defined as the angular deviation from downgoing with respect to a given detector, as shown in Figure 1.6.

The strongest oscillation signal relevant for neutrinos produced in the atmosphere is the disappearance of muon neutrinos. Equation 1.3 can be approximated as follows to describe the survival probability of atmospheric muon neutrinos:

$$P_{\nu_\mu \to \nu_\mu} \approx 1 - \sin^2(2\theta_{23})\sin^2\left(\frac{\Delta m^2_{32} L}{4E}\right) \tag{1.5}$$

With the baseline L converted to $\cos(\theta_{zenith})$, we can express the survival probabilities as a function of E and $\cos(\theta_{zenith})$. See Figure 1.7. We see that the first oscillation peak for neutrinos crossing the entire Earth diameter ($\cos(\theta_{zenith}) = -1.0$) occurs at about 25 GeV.

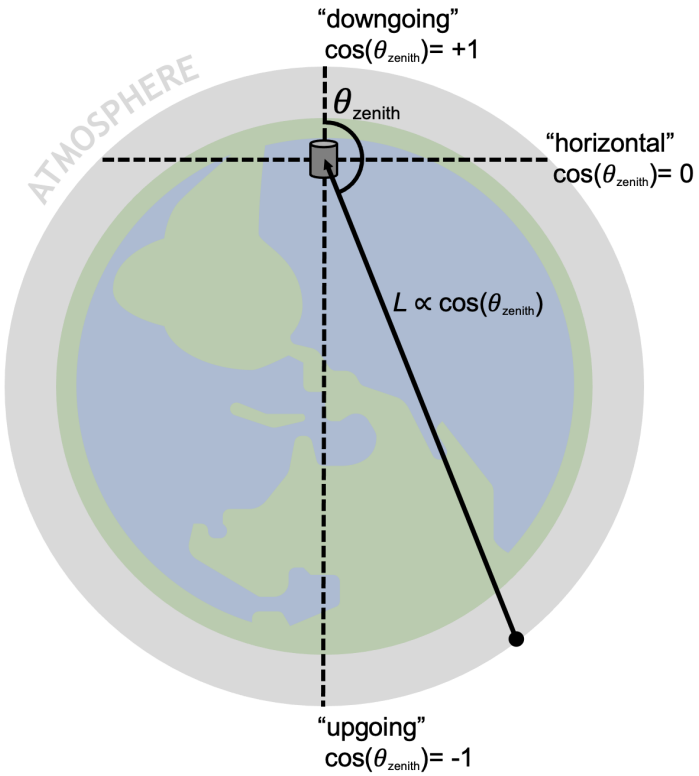


Figure 1.6: Diagram of the relationship between the neutrino arrival direction $\cos(\theta_{zenith})$ and the distance that the neutrino traveled. Note that the arrival direction is defined with respect to the detector's internal coordinate system.

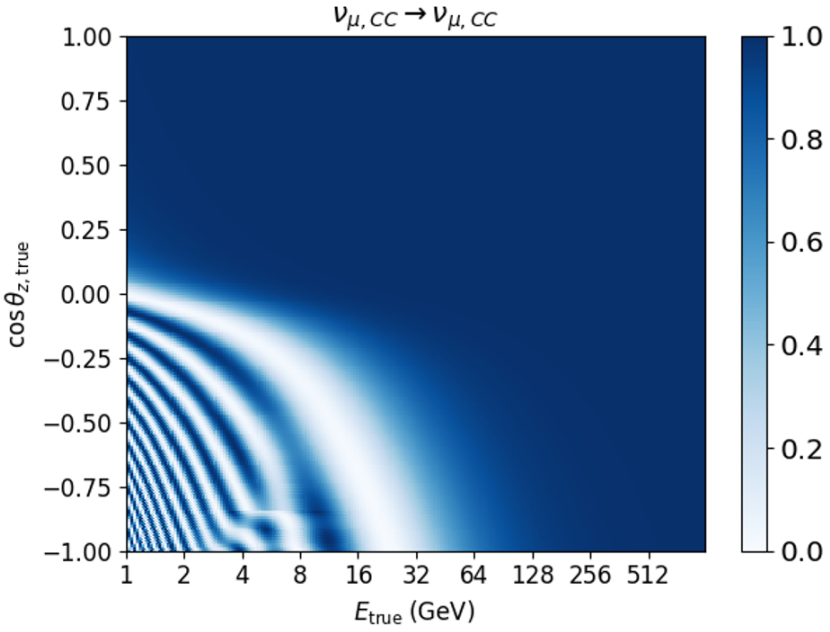


Figure 1.7: Muon neutrino survival probability as a function of true neutrino energy E and arrival direction $\cos(\theta_{zenith})$ (used as a proxy for the distance L that the neutrino traveled). For neutrinos crossing the entire Earth diameter ($\cos(\theta_{zenith}) = -1.0$), the first oscillation peak occurs at about 25 GeV.

1.3 Introduction to IceCube DeepCore

IceCube

The IceCube Neutrino Observatory is a cubic kilometer detector located at the South Pole, that consists of 5,160 modules spread across 86 strings arranged in a hexagonal grid with 125 meter spacing between adjacent strings [11]. See Figure 1.8. Each module, which is referred to as a Digital Optical Module (DOM), contains a 10 inch photomultiplier tube (PMT) that detects optical Cherenkov photons.

Cherenkov radiation refers to the process by which charged particles emit light along a coherent wavefront when traveling in a polarizable medium at a speed greater than the phase velocity of that medium, as illustrated in Figure 1.9. The light is emitted in a cone with a characteristic angle θ_c such that $\cos(\theta_c) = \frac{1}{\beta n}$, where n is the index of refraction of the material and $\beta = \frac{v}{c}$ is the particle's velocity relative to the speed of light in a vacuum. For ice, the index of refraction is $n = 1.31$ and therefore the Cherenkov angle is about $41°$ for highly relativistic particles.

As mentioned in previous sections, when neutrinos interact in the ice and undergo Deep Inelastic Scattering (DIS), a shower of particles is produced. Many of these particles are highly-energetic charged particles, which will emit Cherenkov light that can be detected by the DOMs.

DeepCore

DeepCore is a sub-array within IceCube located in the bottom center portion of the full detector [13]. See Figure 1.10. It consists of 8 dedicated DeepCore strings as well as 7 of the nearby IceCube strings. Many of the DOMs on the DeepCore strings have PMTs with higher quantum efficiency, which means they are able to detect more photons. The denser spacing allows us to observe events with lower energies, compared to a more sparse array like IceCube. While IceCube is designed to look for events in the TeV-PeV energy range, DeepCore can observe events with energies down to about 5 GeV.

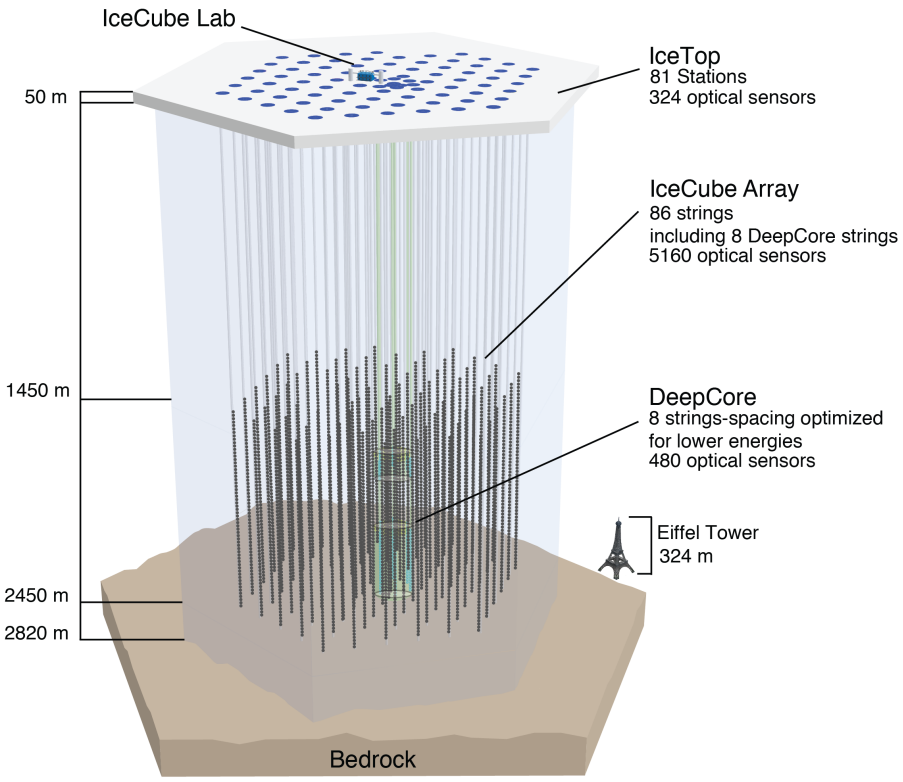


Figure 1.8: The IceCube Detector is a cubic kilometer array consisting of 5,160 modules buried in the Antarctic ice. Figure Credit: IceCube Internal Gallery.

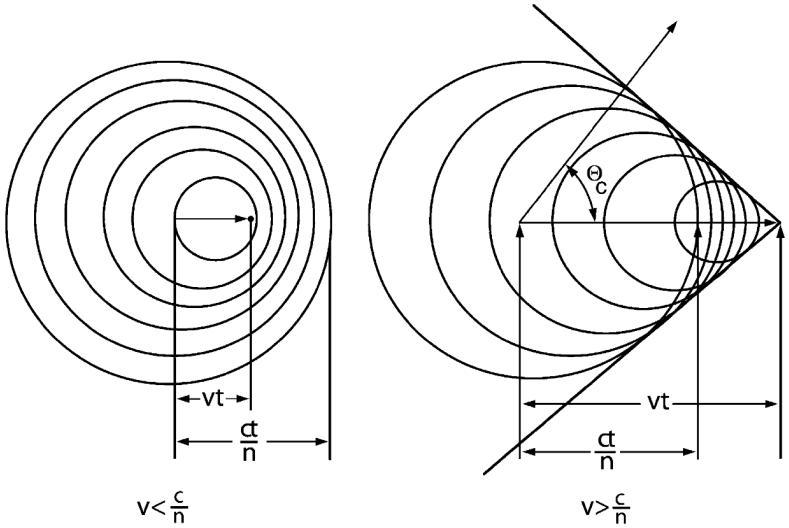


Figure 1.9: Diagram of Cherenkov light. A particle traveling with a velocity lower than (left) and higher than (right) the phase velocity of the medium. The Cherenkov effect only occurs for the particle on the right, because it is traveling with sufficient speed for a coherent wavefront of light to be formed. The light is emitted in a cone with a characteristic angle θ_c such that $\cos(\theta_c) = \frac{1}{\beta n}$, where n is the index of refraction of the medium and $\beta = \frac{v}{c}$ is the particle's velocity relative to the speed of light in a vacuum. Figure Credit: [12].

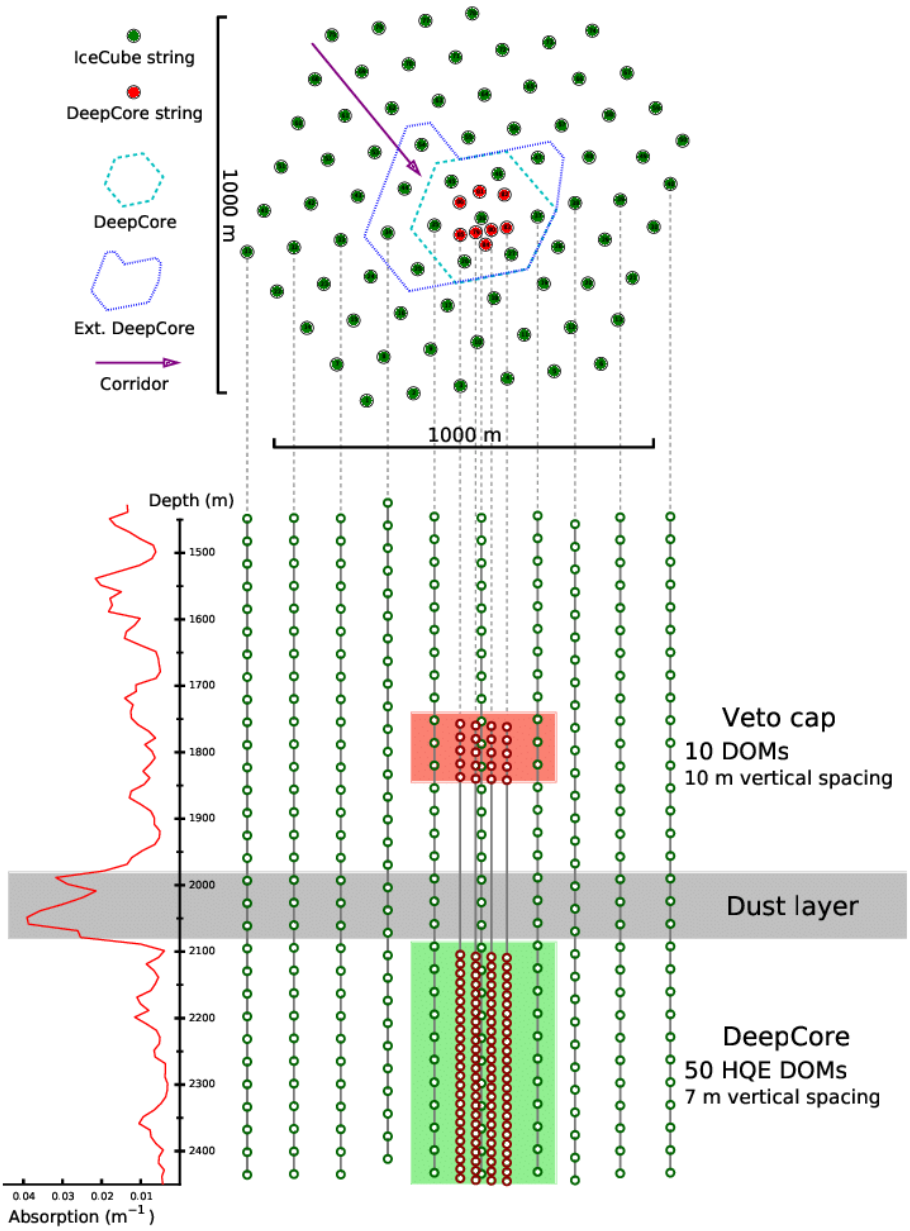


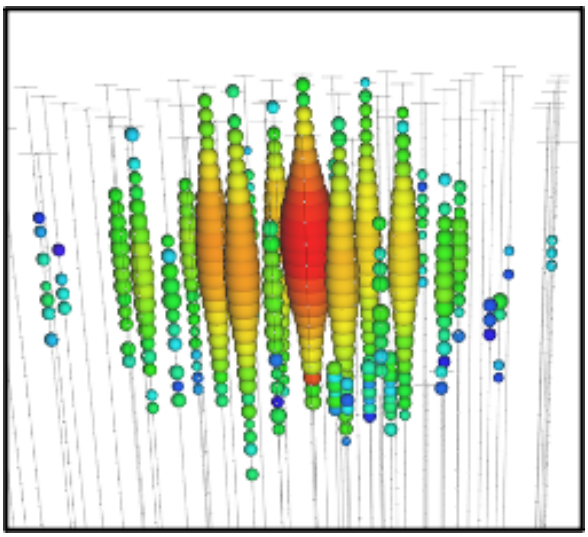
Figure 1.10: The IceCube DeepCore Detector is located at the bottom center of IceCube. It consists of 8 dedicated DeepCore strings along with 7 surrounding IceCube strings. Figure Credit: [8].

Event Types

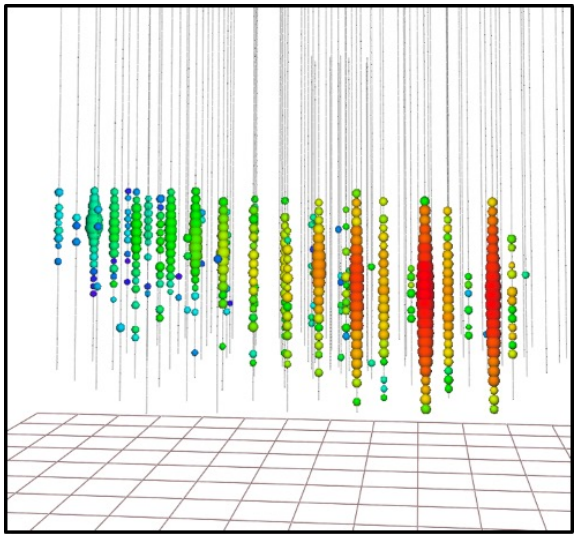
There are two main types of neutrino events in IceCube and DeepCore. The first event type is "cascades" which are roughly spherical in shape and are produced by showers of particles. All events have a hadronic shower component that occurs when the nucleus is broken apart by DIS. For NC events, this is all that is observed. The ν_e CC and ν_τ CC events have additional shower components due to the outgoing charged lepton (the interaction of e and the decay of τ), but these particles travel such short distances that the charged lepton component cannot be distinguished from the hadronic shower at the vertex. Thus, these types of events are all resolved as a single cascade. The second event type is "tracks" resulting from ν_μ CC interactions. The outgoing μ lepton can travel long distances which results in an elongated shape.

Figure 1.11 compares what these event types look like for high energy events in the full IceCube detector and low energy events in DeepCore. In these event displays, each colored dot indicates a single DOM that observed light. The color corresponds to the timing information, with red indicating the earlier hits and blue indicating the later hits. The size of the dot corresponds to how much light the DOM detected, with larger dots indicating more light. In the top two images showing the IceCube events, we see that the cascade is approximately spherical and the track is quite elongated. By contrast, the DeepCore events are much lower in energy and therefore have fewer modules hit and contain much less information. These events are also much harder to distinguish between the spherical and elongated shapes. We will return to this topic in Section 2.3 to discuss a new technique for differentiating between these event types in DeepCore.

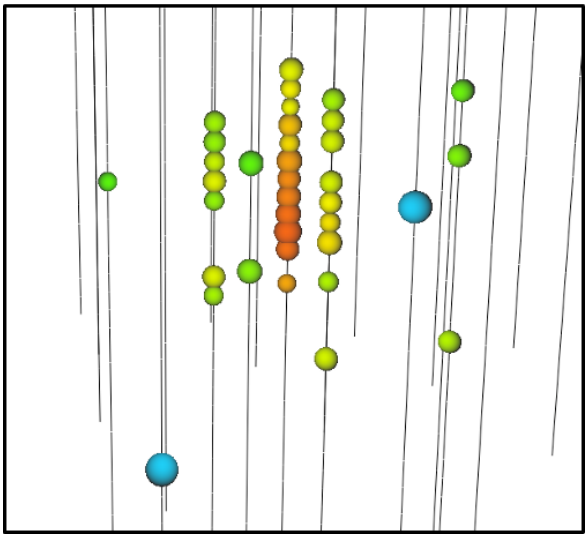
In addition to the neutrino events, we also have two main types of background events. The first is atmospheric muons. These muons leave light along their path creating an elongated signature (as is the case for secondary muons produced in ν_μ CC interactions). Therefore we use veto regions and restrict ourselves to events contained in DeepCore, since neutrinos leave no light in the detector prior to their interaction vertex. Figure 1.12 shows how a veto region is used to eliminate a muon entering the detector by identifying light in the veto region, and how a neutrino event will pass this veto region leaving no light until it interacts. The second main background is random noise, caused by dark noise from the PMTs or radioactive decays in the glass housings that encase the DOMs. These hits tend to be scattered more uniformly throughout the detector in space and time. We can use space-time causality to determine if the hits are likely a random coincidence or if they are consistent with a relativistic particle crossing the detector.



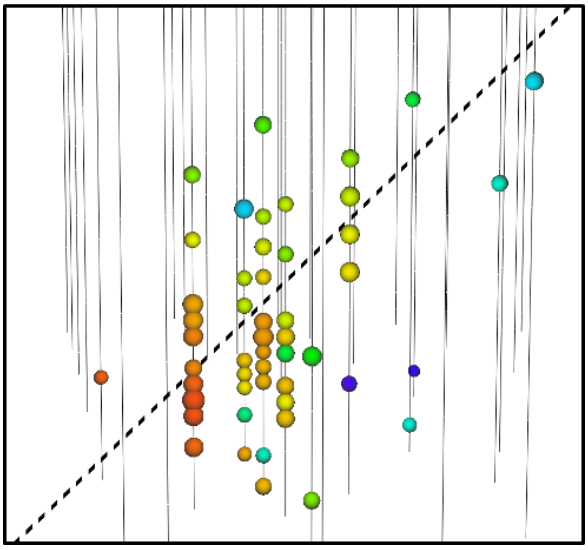
(a) IceCube cascade; 1 PeV



(b) IceCube track; 290 TeV

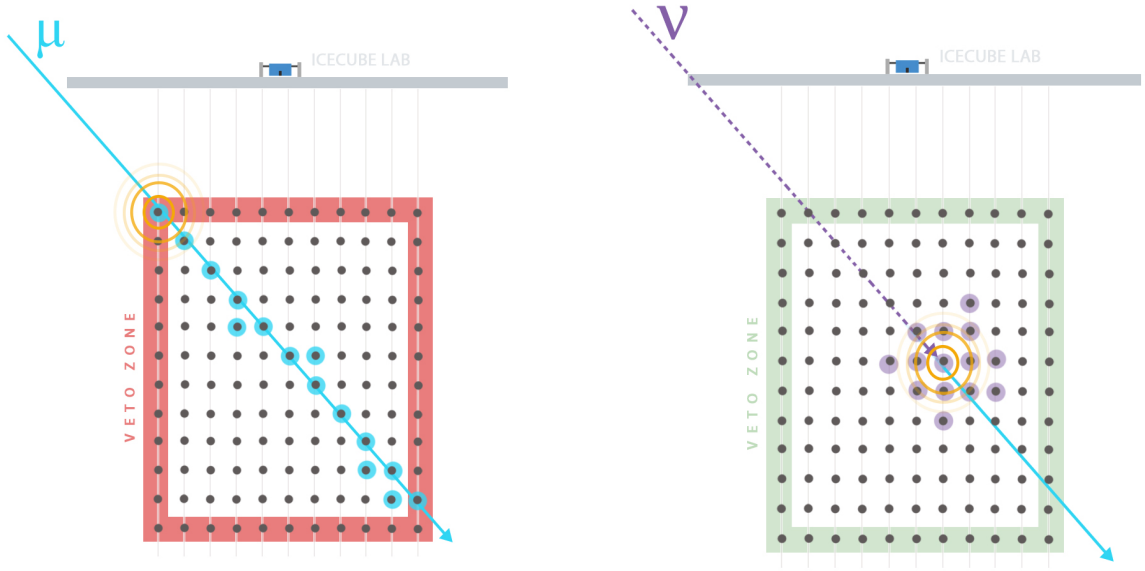


(c) DeepCore cascade; 25 GeV



(d) DeepCore track; 25 GeV

Figure 1.11: Event displays for cascades and tracks in IceCube and DeepCore. Each dot corresponds to an individual DOM that saw light. The color indicates the timing information with red indicating the first hits in time, and blue indicating the later hits in time. The size of the dot indicates the amount of light detected by the module (with a larger dot indicating more charge). For the IceCube events, the cascade is approximately spherical and the track is quite elongated. By contrast, the DeepCore events which are much lower energy have fewer modules hit and therefore have much less information. These events are also much harder to distinguish between the spherical and elongated shapes. The event shown in (a) and (b) correspond to real data events from [14] and [15], respectively.



(a) Atmospheric muon is vetoed.

(b) Neutrino passes the veto.

Figure 1.12: Diagram of the veto principle. (For illustration purposes only; for the work presented in later chapters, the allowed region is restricted only to DeepCore. The rest of the IceCube detector is used as the veto zone.) Figure Credit: IceCube Internal Gallery.

Detector and Ice Properties

Since the primary detector medium is natural ice, we have no control over the purity and optical properties of the ice. Therefore, a substantial internal calibration program is required to characterize and model the detector and the surrounding ice. Our detector observes the photons produced in a neutrino interaction, so we need to understand how these photons propagate through the ice, referred to henceforth as "bulk ice". As the light travels, it can be attenuated by the medium (described by an absorption coefficient) or can change direction (described by a scattering coefficient). The amount of scattering and absorption that occur change as a function of the depth of the detector, due to the stratified layers of ice that form a glaciological record. Figure 1.13 shows how the effective coefficients that describe the scattering and absorption change as a function of depth. The instrumented detector depths are shown as the red shaded region. In the middle of the detector is a dust layer with very poor optical properties (enhanced scattering and absorption).

When the detector was constructed, columns of ice were melted, the strings of DOMs were deployed down the water holes, and then the holes were allowed to refreeze. The ice in these holes, referred to as "hole ice", has therefore been altered and does not share its optical properties with the surrounding bulk ice. When the ice refroze, a column of air bubbles became trapped at the center of each hole which leads

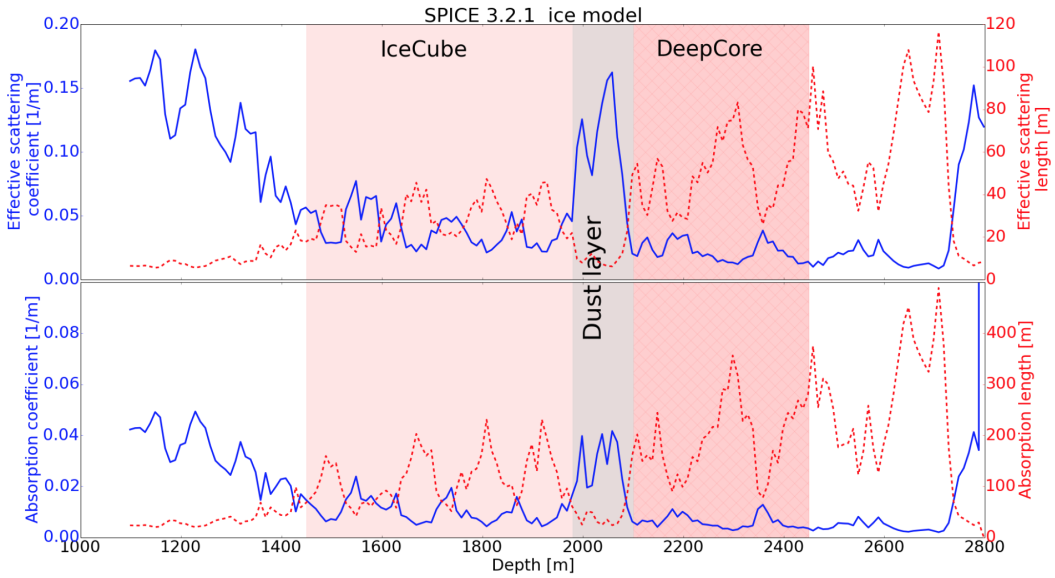


Figure 1.13: Effective scattering (top) and absorption (bottom) coefficients for photons in South Pole ice as a function of depth. The instrumented detector depths are indicated by the red shaded region. In the middle of the detector is a dust layer with very poor optical properties (enhanced scattering and absorption). Figure Credit: [16].

to increased scattering and uncertainty in this region of the ice. The effects of the hole ice are dependent on the incoming angle of the photon, since horizontal photons will travel through less hole ice than steeply inclined ones, as shown in Figure 1.14. In addition, we can see from the diagram that the PMT is oriented downwards so photons traveling upwards towards the face of the PMT are more likely to be detected than photons traveling downwards since there is no sensor on the top of each module. All of this information is combined to form a single model that describes how the response of the detector changes as a function of the incident angle η of the photon. The parameters used to describe this space are referred to as hole ice p_0 and p_1. Figure 1.15 shows how the angular response curve changes for a continuous sweep of values.

The final type of detector uncertainty relates to the efficiency of the PMT. The efficiency refers to the fraction of incident photons that are able to be detected. While this has been constrained by both lab and in situ calibration measurements, these measurements aren't perfect so we allow for additional uncertainty. We refer to this as "DOM Efficiency" and it is described using a scaling factor with a value of 1.0 corresponding to the calibrated model's central value. This effect is particularly important for low energy events, because it can change whether or not you see an event at all if it is near the detector's triggering threshold (in contrast to high energy events where you are simply changing the amount of light that is able to be detected for a given event).

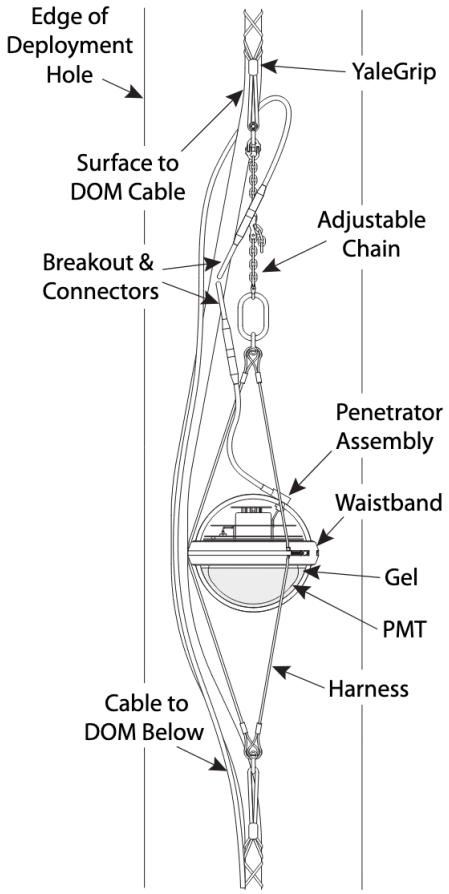


Figure 1.14: Diagram of DOM in deployed hole. Figure Credit: [11].

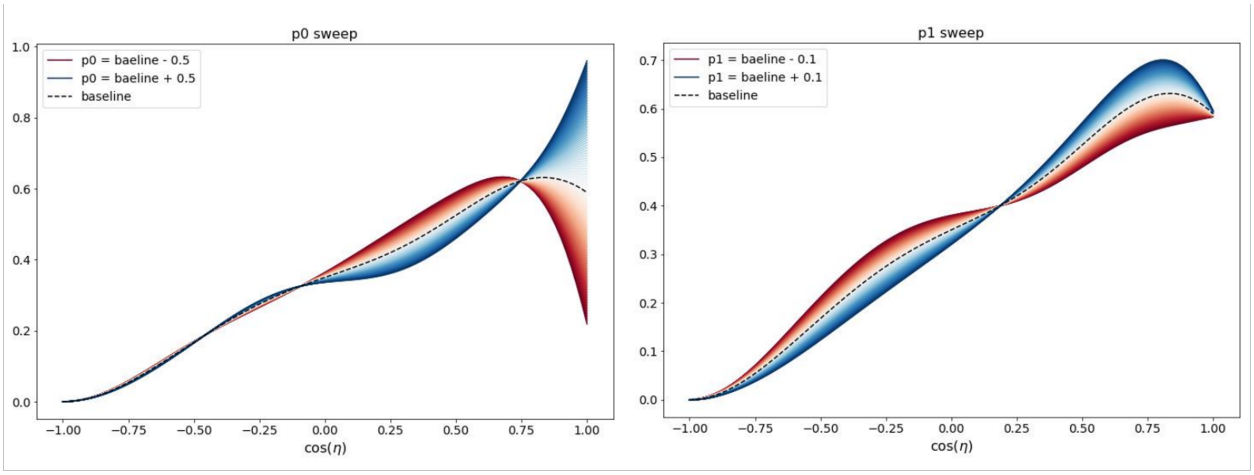


Figure 1.15: Relative acceptance of the photons as a function of the incident angle η of the photon. These plots show how the relative angular acceptance changes for various values of p_0 (left) and p_1 (right). cos(η)=1 indicates the bottom of the DOM and the face of the PMT. cos(η)=-1 indicates the top of the DOM where there is no PMT. Figure Credit: [17].

2

The OscNext Sample

Shiny new tools now available!

– Tom Stuttard

Performing a measurement of the atmospheric neutrino oscillation parameters requires curating a targeted sample of atmospheric neutrino data as well as generating accompanying simulation to model the data. Since the previous generation of DeepCore oscillation analyses, there have been many new updates made available. Over the past few years, about a dozen people from the IceCube Collaboration (including myself) have revisited almost every step involved in both developing an atmospheric neutrino sample and subsequently performing analyses with that sample. This effort, along with the resulting sample and suite of tools, are known internally as "OscNext". This chapter will cover the OscNext sample.

The data used in this sample spans just shy of a decade, from April 2012 to January 2022. We only use the portion of the data that has been cleared for use in physics analyses, which gives 9.28 years of detector livetime.

2.1 Simulation

Simulation is used for two primary purposes: first, most of the sample development and analysis development are performed with simulation so that the analysis remains blinded to the real data, and second, the analysis fit uses a binned maximum-likelihood method so we must have a model that gives us the expected

number of events for various parameter combinations in order to perform the binned maximum-likelihood fit. Our simulation is generated using Monte Carlo (MC) methods, and therefore the terms "simulation", "Monte Carlo", and "MC" may be used interchangeably throughout. We repeatedly simulate many events each seeded with a different random number. We generally try to simulate significantly more events than we have in data so that our statistical uncertainty is dominated by the real statistical uncertainty rather than the statistical uncertainty associated with our modeling.

We must simulate our neutrino signal and also our primary backgrounds, in order to develop techniques to distinguish between the two and suppress the backgrounds, as well as to account for any remaining background in the final sample when a fit is later performed. We use a package called GENIE for simulating neutrinos, which is a standard neutrino event generator used across the field of neutrino physics [18]. We use a modified version that has been adapted into the IceCube software framework. For our atmospheric muon background, we use an internal software known as MuonGun which generates muons at random points on a cylindrical shell around the detector. This is then weighted to a realistic energy spectrum. More details about MuonGun and improvements I made to its efficiency can be found in Chapter 3. The second background is noise which can be caused by dark noise from the PMT or radioactive decays in the glass housings that encase the DOMs. Noise is modeled using an internal software package known as Vuvuzela [19]. Vuvuzela is used both on its own for pure noise events and applied to other simulated event types to make the simulated neutrino and muon events more realistic.

Simulation Procedure

The follow steps are used to produce simulation:

- Step 1: Particle generation and propagation. Neutrinos and muons are produced and propagated through the ice. When these particles interact in the ice, any resulting energy deposits and/or secondary particles are recorded.

- Step 2: Photon propagation. Energy deposits are converted into photon particles and the photons are propagated through the ice. This step has been optimized to run on GPUs and is still one of the most computationally expensive parts of producing simulation.

- Step 3: Detector response. This step simulates the response of the detector's PMTs, noise, electronics, and triggering to generate output that reproduces the detector's output.

At this point, the aggregate simulation bears resemblance to real detector data, though admittedly does not have satisfactory agreement yet. Further levels of selection are necessary to ensure that the data and

simulation events that we use are high quality and well modeled.

Systematic Uncertainty of the Detector

For Steps 2 and 3, we have an appreciable amount of uncertainty in the parameters that describe the photon propagation in ice and the detector response. The effects of changing these parameters are not well described analytically a priori, so we must change the model parameters and directly resimulate under the new scenario. We resimulate our events for many discrete parameter values, and later on (see Section 4.2) we will introduce an interpolation method that allows us to describe the detector response analytically at the final level.

2.2 Event Selection

The purpose of an event selection is to get a "good" sample of signal events that can later be used for an analysis, but this must balance competing objectives: high statistics (keeping as many signal events as possible to reduce statistical uncertainty) and high purity (eliminating as many background events as possible). This is done through several successive levels of processing. Figure 2.1 shows the rates of noise, muons, and each neutrino flavor for all of the levels of processing between the DeepCore Filter and the final analysis level. The remainder of this section will go into more detail about each level of processing.

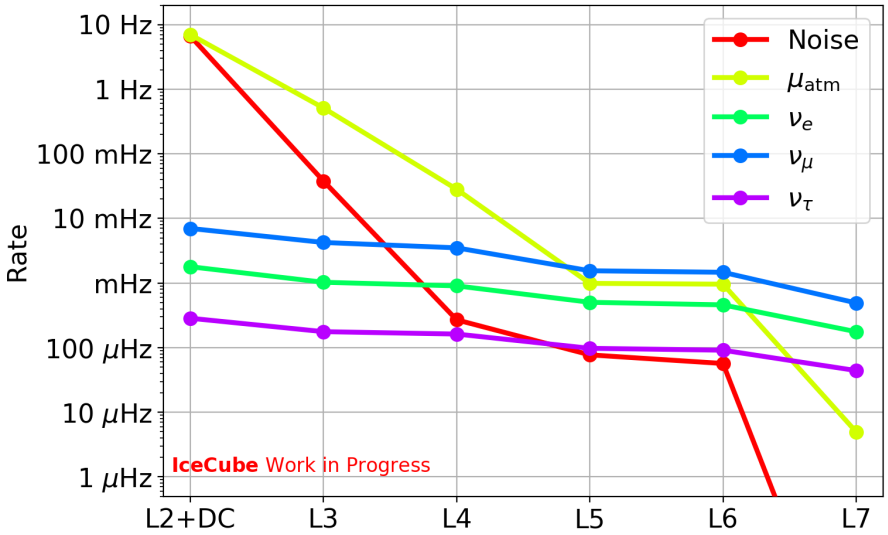


Figure 2.1: The OscNext Event Selection, beginning at the DeepCore Filter through the final analysis level. Overall, the atmospheric muon and noise backgrounds are suppressed by 6 and 8 orders of magnitude respectively.

Level 1

Level 1 is the detector trigger and is common across all of IceCube. It requires multiple DOMs near each other to see light and when this occurs, the data is read out for the full detector. More details on the trigger can be found in [11].

Level 2

At Level 2, there are many filters to identify and sort different types of events. The only filter relevant for this work is the DeepCore filter, which uses the veto technique introduced in Section 1.3. Since DeepCore is located in the bottom center portion of IceCube, it is able to use the upper and outer layers of IceCube beyond DeepCore as a veto region. A quick reconstruction estimates the vertex position to be the center-of-gravity of hits inside the DeepCore fiducial volume. Then, a relative velocity is calculated between hits in the veto region and the vertex estimate. If the velocity between the hits in the veto region and fiducial volume is consistent with a muon hypothesis, then the event will be removed by the filter. Although some of these events could be secondary muons from neutrino interactions outside the detector, this would be indistinguishable from our atmospheric muon background and thus we reject them anyway. More details on the DeepCore Filter can be found in [13].

Level 3

The objective of Level 3 is to use simple, fast cuts to remove other event types which we are not interested in and/or do not simulate, such as muon bundles in which two or more atmospheric muons enter the detector together, coincident events in which a neutrino and atmospheric muon enter the detector at the same time, and events which are pure noise. At this point, a cleaning procedure is performed to eliminate hits which are not causally connected: $(x_2 - x_1) > c(t_2 - t_1)$. That is, if the distance between hits is greater than the distance a particle moving the speed of light could travel in the same time window, then it is assumed to be noise. The cleaning procedure eliminates hits in an event which are likely noise and leaves behind hits which could be causally connected. Several cuts are made on these cleaned events to eliminate backgrounds; a representative sample of cuts are shown in Figure 2.2 and explained in this section, but note that this is not an exhaustive list of the cuts that are made at this level. Additional cuts are made in a similar spirit to target these same backgrounds.

If an event has fewer than 6 remaining DOMs that saw light after the cleaning procedure, the event is discarded as this is still consistent with a pure noise event. Figure 2.2a shows this cut along with the noise rates which peak below the cut value, and the signal rates. The neutrino events that are lost in this step

are likely too low energy to be well reconstructed anyway, so this is not of concern. Atmospheric muons are targeted using another veto cut. If there are 10 or more hits in the top portion of the detector, the event is discarded because this is consistent with the hypothesis of a down-going atmospheric muon. This cut is shown in Figure 2.2b; we see that this eliminates most of the simulated atmospheric muons, while most of the neutrino signal is retained. The third category of background events targeted at this level is unsimulated events such as coincidences, i.e. two events within the same trigger window. If there are two events occurring in a trigger, then the event will generally appear to be longer, therefore we can eliminate these types of events based on the time length (the difference in time between the first hit and the last hit). Figure 2.2c shows the cut on this variable. If an event lasts longer than 5,000 ns, it is discarded. We see that this eliminates many of the unsimulated events, but keeps our neutrino signal.

In total, this level eliminates 99.4% of simulated noise events and 93% of simulated atmospheric muons, and keeps about 61% of neutrino events. The overall event rate at this point is about 1 Hz.

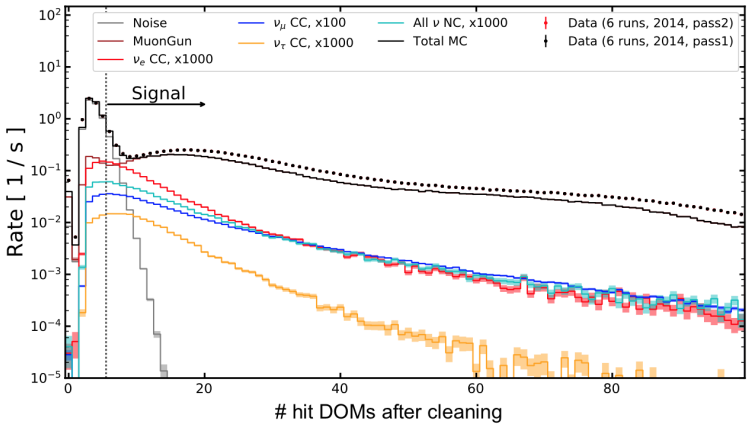
Level 4

Level 4 consists of two boosted decision trees (BDTs) designed to eliminate noise and muons. Each one is trained to differentiate its target event type from neutrinos, by returning a score from 0 to 1 indicating how neutrino-like (1.0) or background-like (0.0) the event is. The output score distributions of these two classifiers are shown in Figure 2.3. Cuts are placed at 0.7 for the noise classifier and 0.65 for the muon classifier. Only events which are more neutrino-like than the given threshold are kept.

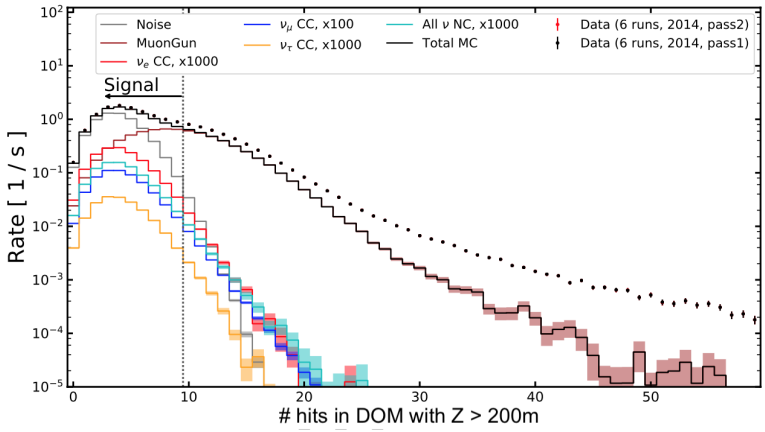
In total, this level eliminates more than 99% of the remaining simulated noise events and 94% of the remaining simulated atmospheric muons, and keeps over 80% of neutrino events.

Level 5

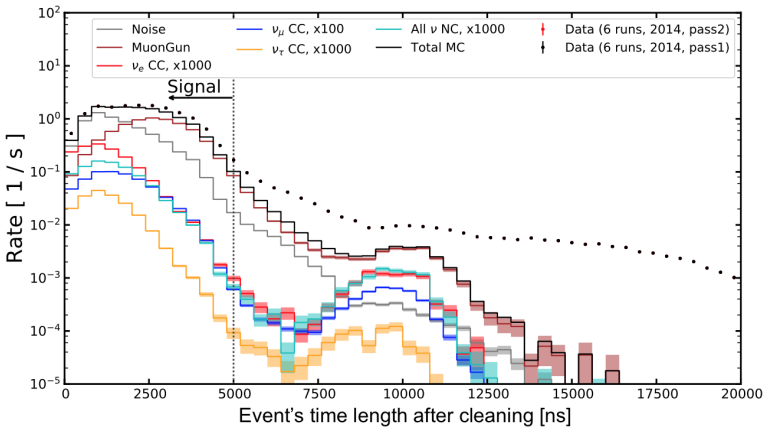
Level 5 further targets the muons in order to achieve a neutrino-dominated sample. The muons that remain do not look like the typical muon signature, which is why they were not eliminated by previous cut levels. Because IceCube is more sparse than DeepCore, it is possible for dim muons to travel between strings and have only very small amounts of light registered by the strings along the corridor, until reaching the more dense DeepCore region where light is detected by more strings. An example of one of these corridors is indicated by the purple arrow on Figure 1.10. We remove events which have more than 2 DOMs hit along an identified corridor. Additionally, we require events to have vertex estimates contained within DeepCore, to eliminate events that start just outside DeepCore. Finally, we tighten the cuts that were made on the Level 4 classifiers (to 0.85 for noise, and 0.9 for muons), so that we have a neutrino dominated sample.



(a) Number of DOMs remaining that saw light after the cleaning procedure. Events are discarded if they have fewer than 6 remaining DOMs.

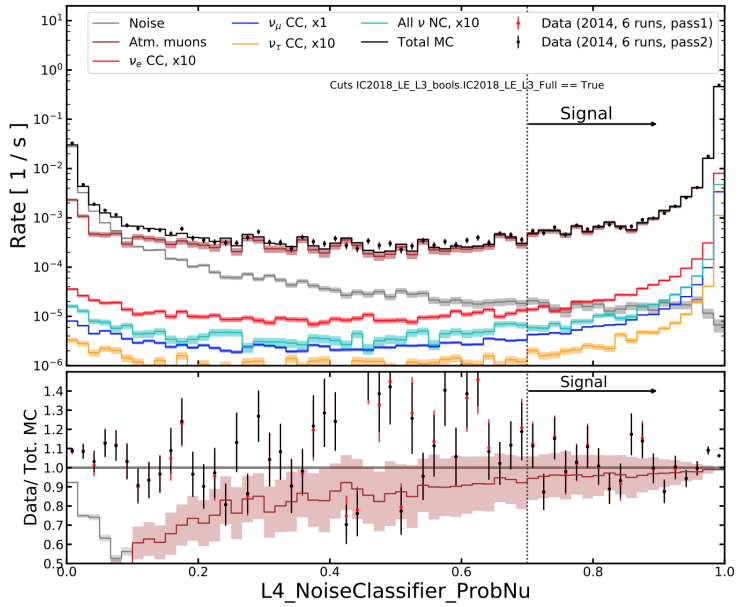


(b) Number of hits in the top portion of the detector. Events are discarded if they have 10 or more hits in this veto region.

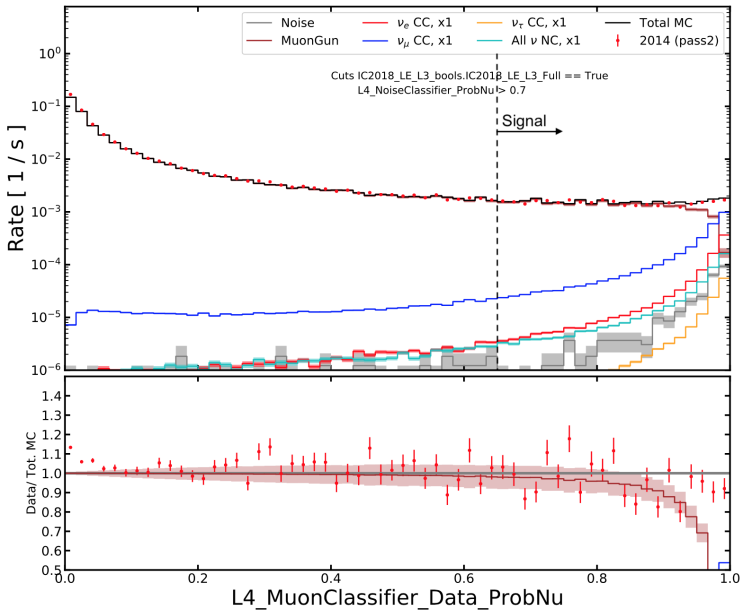


(c) Difference in time between the first hit and the last hit. Events longer than 5,000 ns are removed, which eliminates many unsimulated events that are likely coincident events.

Figure 2.2: Sample cuts made at Level 3 to eliminate noise, muons, and unsimulated coincident events. Figure Credit: [20].



(a) Output score distribution for the Level 4 noise BDT classifier. We see that the noise (grey) is peaked at 0, and the neutrino signal (red, blue, teal, gold) is peaked at 1.0. A cut is placed at 0.7 and only events which are more neutrino-like than this are kept.



(b) Output score distribution for the Level 4 muon BDT classifier. We see that the muons (maroon) are peaked at 0, and the neutrino signal (red, blue, teal, gold) is peaked at 1.0. A cut is placed at 0.65 and only events which are more neutrino-like than this are kept.

Figure 2.3: Cuts made at Level 4 to eliminate noise and muons. Figure Credit: [20].

Level 6 - Reconstruction

Once we have reach a point where our sample is dominated by neutrinos rather than backgrounds, i.e. after Level 5, it becomes computationally feasible to perform more sophisticated and computationally intensive reconstruction algorithms. We use a brand new reconstruction method described in [21] called RetroReco. It is a likelihood-based reconstruction with an 8-dimensional event parameterization to describe the neutrino interaction. The model, shown in Figure 2.4, takes into account the direction of incoming neutrino in zenith and azimuth $\Psi(\theta, \phi)$, the vertex of the neutrino interaction $V(x, y, z, t)$, the energy of the shower at the interaction vertex $E_{cascade}$, and the length of the track component L_{track}. The (x, y, z) coordinates are given in terms of the IceCube coordinates, where $(0, 0, 0)$ is defined to be the center of the IceCube detector. The reconstruction takes into account the detector response and determines what combination of parameters is most likely, given the pattern of hits observed in the detector.

There are a few additional pieces of information provided by the reconstruction that will be useful for the steps that follow. The total energy is the sum of energy in the track and cascade portions, where the track energy has been calculated from the track length using the assumption that the muon experiences a constant energy loss of 0.22 GeV/m. RetroReco also provides an estimate of how track-like an event is. This is determined by computing the difference in log-likelihood (LLH) of how well the model fits the data for the full 8-dimensional hypothesis, compared to a hypothesis where the track length $L = 0$. This quantifies how necessary the track component is to describing the event well. This quantity is referred to as "Zero dLLH" and is used in Section 2.3. In addition, RetroReco also returns uncertainties on each quantity, given as an upper and lower bound.

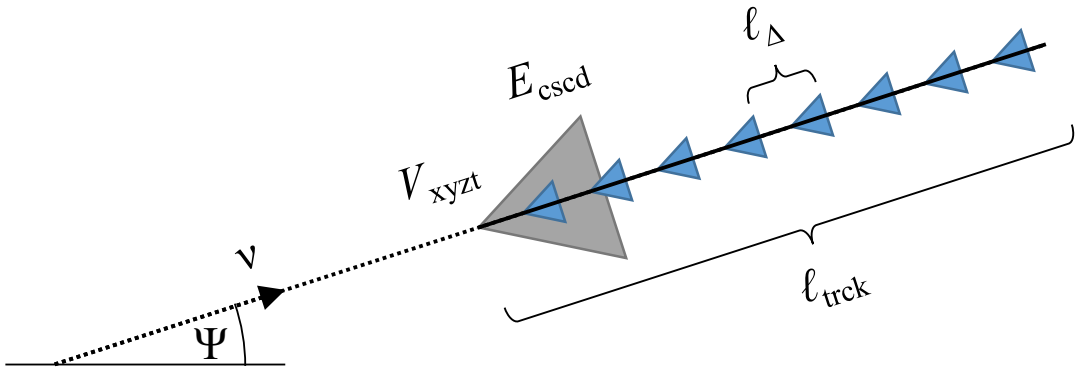


Figure 2.4: Diagram of the 8-dimensional event parameterization used for reconstruction: direction of incoming neutrino in zenith and azimuth $\Psi(\theta, \phi)$, vertex of the neutrino interaction $V(x, y, z, t)$, energy of the shower at the interaction vertex $E_{cascade}$, and length of the track component L_{track}. Figure Credit: [21].

Level 7

There is one final level of selection performed after the reconstruction. A series of cuts are used to target the coincident events that were first mentioned at Level 3. These cuts include the reconstructed vertex time, as well as the number of DOMs hit in the veto region (IceCube). We require that events have no light in the top 15 layers of IceCube, and fewer than 8 hits in the outer few layers of IceCube. Additionally, we require the reconstructed vertex time to be less than 14,500 ns.

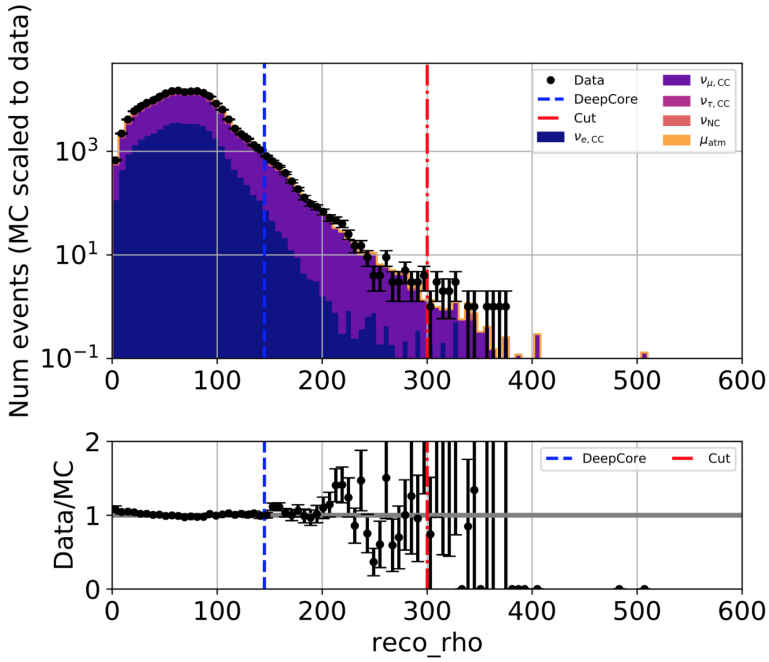
Next, we make cuts on the containment, similar to those performed at Level 5, but this time using the vertex position from the full reconstruction, rather than the quick estimates that were available previously. The vertex x and y positions are converted to a radial distance ρ. We define a cylindrical region around DeepCore and keep only events with a reconstructed vertex inside. Figure 2.5 shows the distribution of the reconstructed vertex coordinates along with the cut values for the ρ and z coordinates.

At Level 6, in addition to the full RetroReco reconstruction, we run an additional quick reconstruction that provides direction only. This quick reconstruction is called "SANTA" and more details of it can be found in [21]. It uses a hit cleaning algorithm to select only hits from unscattered photons, referred to as "direct hits". This is determined by applying causality cuts to determine which hits are consistent with a Cherenkov cone. We make a cut to require that events have at least 3 direct hits. In general, these hits are the most useful and are the highest quality since being unscattered means they are, by default, less influenced by ice properties. Additionally, events with few or no direct hits typically have larger errors on the reconstruction.

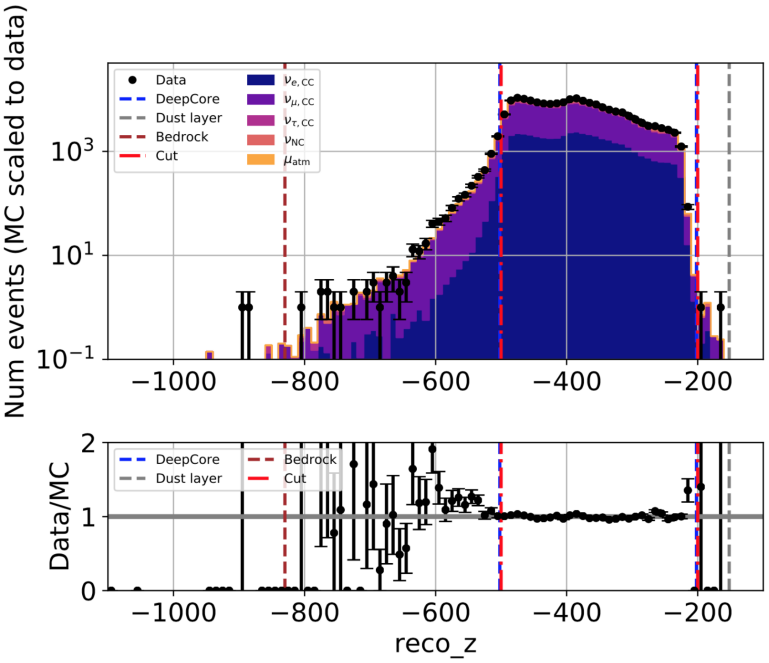
The main background rejection power at Level 7 comes from a BDT that is trained to distinguish the remaining muons from neutrinos. Figure 2.6 shows the output score distribution of this classifier. A cut is placed at 0.8 and we only keep events with scores above this value. The process for choosing this cut value is described in more detail in Section 4.1.

At Level 7, we also run a classifier to distinguish tracks versus cascades, but no selection cuts are made on this variable. More details of this classifier will be provided in the next section.

At this point, we have reached the final sample that will be used for analyses. We consider noise events to be completely removed at this point, and they are therefore not included in the analysis. No simulated noise events remain at final level, which given the statistics of the simulation implies that noise accounts for less than 0.03% of the sample. Table 2.1 lists the expected rates and number of events for each of the event types.



(a) Radial coordinate of reconstructed vertex in meters. Only events with $\rho < 300$ m are kept.



(b) Z-position reconstructed vertex in meters. Only events with $-500 < z < -200$ m are kept.

Figure 2.5: Distribution of reconstructed vertex coordinates, along with the containment cut values indicated by the red lines. Figure Credit: [22].

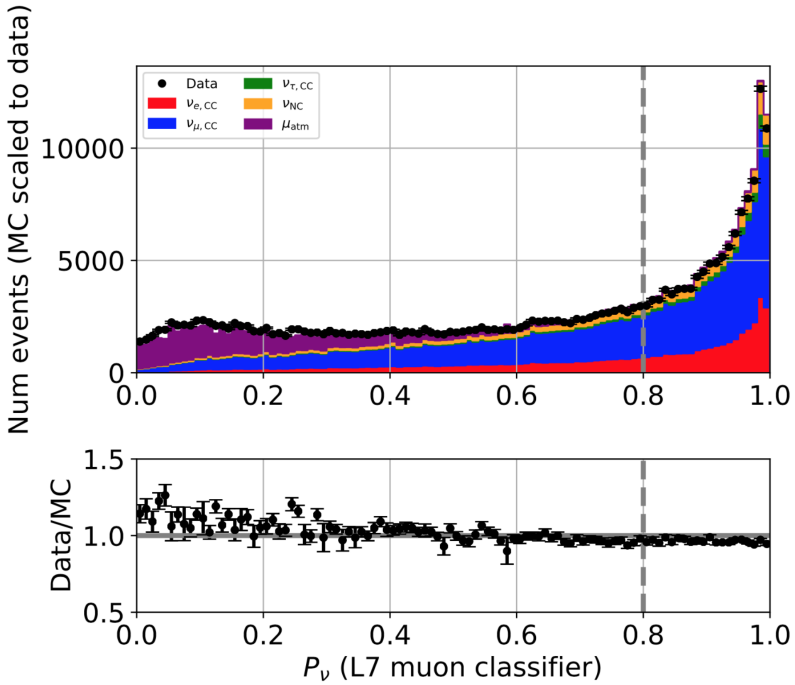


Figure 2.6: Output score distribution of Level 7 muon classifier. The atmospheric muon background (purple) peaks at low values near 0. The neutrinos (red, blue, green, yellow) peak near 1. A cut is placed at 0.8 and only events above this value are kept. Figure Credit: [22].

	Rate [mHz]	Num events [9.3 yr]	% of sample
$\nu_{e,\mathrm{CC}}$	0.165	48234 ± 75	23.0
$\nu_{\mu,\mathrm{CC}}$	0.436	127725 ± 129	61.0
$\nu_{\tau,\mathrm{CC}}$	0.033	9678 ± 21	4.6
ν_{NC}	0.076	22245 ± 51	10.6
μ_{atm}	0.005	1463 ± 87	0.7
Noise	-	-	< 0.03
Total	0.715	209346 ± 182	100

Table 2.1: Final level event rates after all cuts are applied. The quoted uncertainty comes from the finite MC statistics. The background simulation is more limited in statistics than the neutrino simulation. No simulated noise events remain at final level.

2.3 Particle Identification

Because the oscillations that we are studying are inherently flavor-dependent, we must have some way to identify events of different flavors. Unfortunately, at these low energies the only information we are currently able to get is from muons. Muon neutrino charged current interactions leave muons in the detector that may travel far enough to have visible tracks. Taus produced in tau neutrino charged current interactions decay to muons 17% of the time. All other interactions (NC events, NuE CC, and the other 83% of NuTau CC events) appear as single cascades because we cannot distinguish between hadronic and electromagnetic cascades. The process of distinguishing between tracks and cascades is referred to as "event classification" and as "particle identification (PID)". For this analysis, I developed a new event classification scheme. This is the first generation of DeepCore analyses to use machine learning techniques for this classification. I used a boosted decision tree (BDT) trained on simulation, in particular I used the XGBoost algorithm [23].

Input Variables

The following variables are used as inputs to the BDT:

- **Track Length** This variable is the most straightforward measure of how track-like an event is. It is simply the reconstructed track length. An event with a very long reconstructed track likely does have a true muon track present.

- **Cascade Energy** This variable is RetroReco's estimate of how much light is in the cascade portion of the event (as opposed to the track portion).

- **Zero *d*LLH** This variable is RetroReco's estimate of how necessary the track hypothesis is to explain the shape of the event. Upon completing the full 8-dimensional minimization of the reconstruction and determining the log-likelihood (LLH) of the event's hits with the best fit parameters, it forces the track length to 0 and determines the LLH of the hits having been produced by an event with the remaining 7 parameters. If the fit gets significantly worse by removing the track portion, then the event is likely a track because the track portion is required in order to get a good fit. If the LLH remains almost unchanged after eliminating the track component, then the event is likely a cascade because it does not require a track to get a good fit.

- **Zenith** We expect there to be some zenith dependency on the ability to identify track-like events due to the geometry of the detector. Because the spacing between adjacent DOMs in the vertical direction

is much smaller than in the x-y direction, we expect tracks that are nearly upgoing or downgoing to be easier to identify than horizontal ones.

- **Zenith Uncertainty** In addition to providing the most likely value for each reconstructed parameter, RetroReco also reports a lower and upper bound as a proxy for the uncertainty of that parameter. The uncertainty is computed by taking the difference between the upper and lower bounds for the zenith estimate of each event. Because tracks are typically associated with better pointing resolution due to their longer lever arm, we expect events with a smaller zenith uncertainty to be more track-like.

The distribution of these input variables, as well as the output score, for tracks and cascades can be seen in Figure 2.7, including the agreement between data and simulation for each variable.

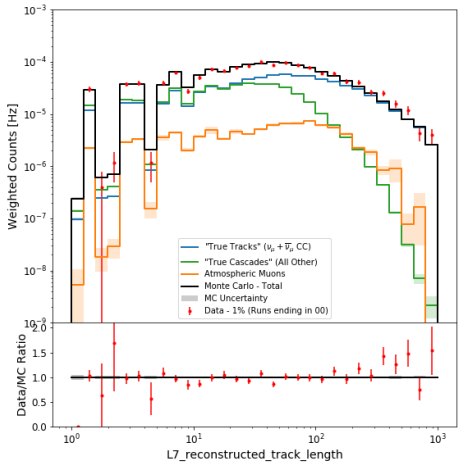
Classifier Training

The classifier was trained using events that passed the final level muon classifier. To get a good clean sample for training, the track-like events used for training were $\nu_{\mu,cc}$ and $\overline{\nu_{\mu}}_{,cc}$ with reconstructed total energy between 5 and 500 GeV, and the cascade-like events used for training were $\nu_{e,cc}$ and $\overline{\nu_{e}}_{,cc}$ again with reconstructed total energy between 5 and 500 GeV. The lower energy events are excluded from training because they do not have any distinguishable track features, so it would confuse the classifier to say that something that looks very cascade-like is actually a "track". ν_{τ} events are excluded from training because some produce tracks (i.e. the 17% of $\nu_{\tau,cc}$ interactions that decay into muons) while others produce cascades.

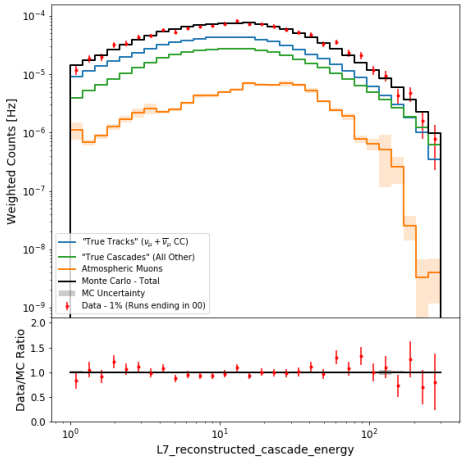
The weights used for training are an unoscillated atmospheric neutrino flux from [10]. We chose to use unoscillated weights because energy and $\cos(\theta_{zenith})$ are inputs into the classifier, and we wanted to prevent any influence of the oscillation pattern being used to assign whether events are more likely to be tracks or cascades. Sample balancing was also applied during training so that the sum of the track weights was equal to the sum of the cascade weights, to ensure equal priority is given to correctly identifying both track and cascade populations. Half of events are used for training and half are used for testing.

Performance

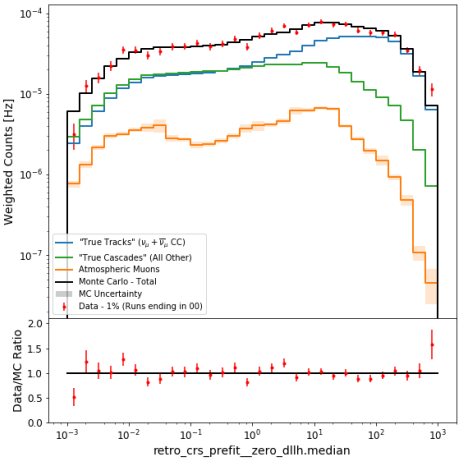
We can evaluate the performance of this method in different portions of phase space, namely as a function of energy and cos(zenith). Figure 2.8 shows the fraction of tracks and cascades correctly identified. A few general features can be seen. At low energies, fewer events are identified as tracks because there are no distinguishable muons. In tracks (left panel) this shows up as a deficit of events classified correctly, but in cascades (right panel) most low energy events are classified correctly. At high energies, both tracks and



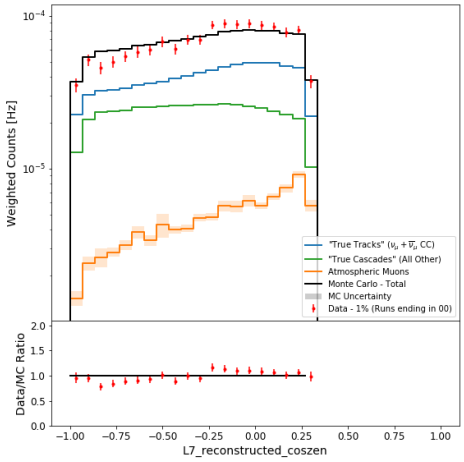
(a) Track Length



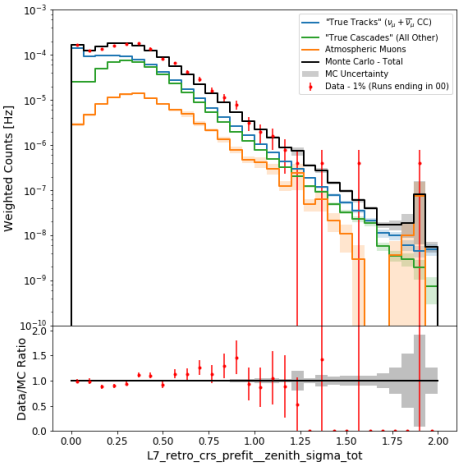
(b) Cascade Energy



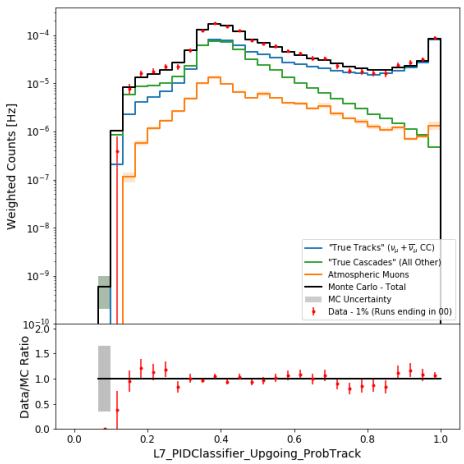
(c) Ratio of the goodness-of-fit between the cascade+track and cascade-only hypotheses



(d) $\cos(\theta_{zen})$

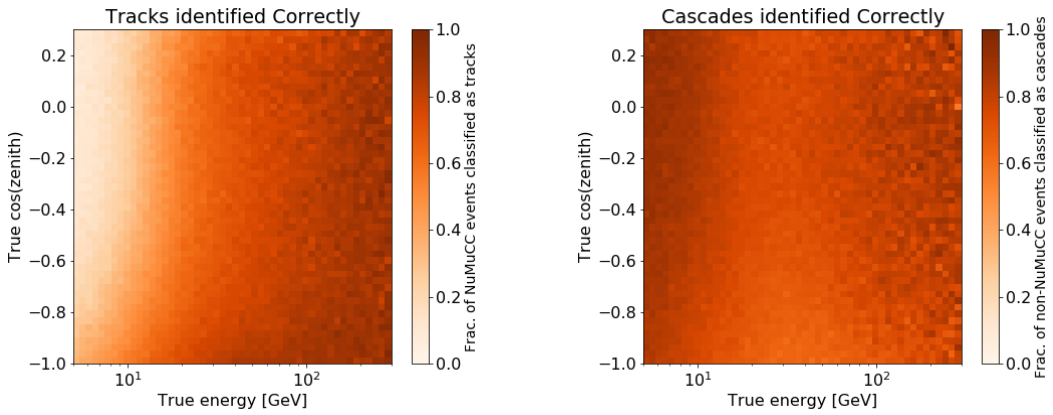


(e) Zenith Uncertainty



(f) Output score of PID classifier

Figure 2.7: Variable distributions and Data/MC comparisons for each classifier input variable and the classifier output variable.



(a) True Tracks ($\nu_{\mu,cc}$)

(b) True Cascades

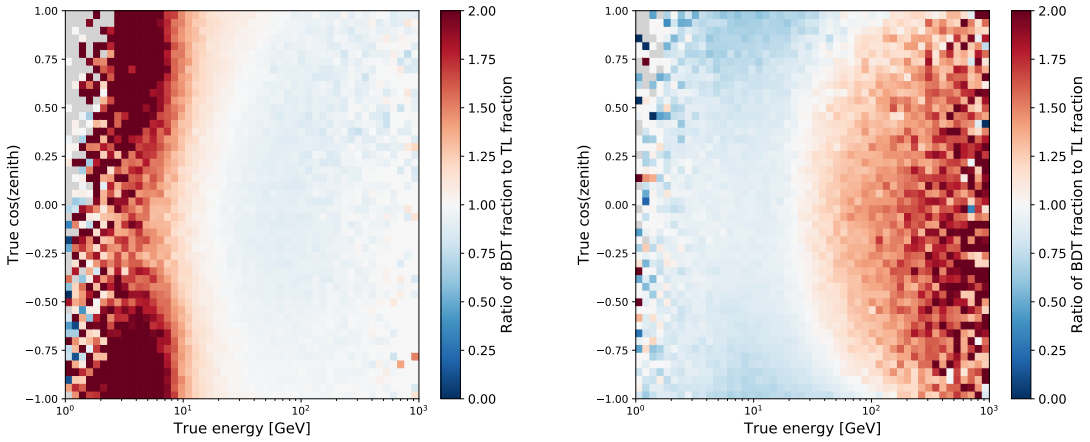
Figure 2.8: Fraction of events classified correctly as a function of energy and $\cos(\theta_{zenith})$. (a) shows $\nu_{\mu,cc}$ events classified as tracks, (b) shows all other particles classified as cascades.

cascades are identified correctly because they are easy to distinguish. At intermediate energies there is some confusion in both types. In the left panel, a zenith structure is also visible at intermediate energies, because upgoing and downgoing tracks look more track-like due to the closer DOM spacing and are more easily identified, whereas horizontal tracks are more difficult to identify. The structure visible in Figure 2.8a carries over into our final analysis templates, as will be seen in later chapters.

Once we have the output score distribution as shown in Figure 2.7f, we can determine where to place the cut boundaries to make the distinction in our analysis, but further discussion on this topic will be reserved for Section 4.1.

Comparison to previous methods

The previous generation of oscillation analyses [7, 8] before OscNext used univariate methods for classification; one analysis used reconstructed track length and a different analysis used the Zero dLLH quantity described earlier in this section. Figure 2.9 shows a ratio of the fraction of events classified correctly using this classifier method compared to the fraction of events classified correctly using the method that only relies on track length. The regions indicated in red indicate portions of phase space where the new classifier method performs better. In the left panel, you can see that the classifier is better at correctly identifying low energy tracks. This is because the track length only method requires events be high enough energy to produce a muon that can travel 35 meters, whereas the classifier method can pick out track-like shapes on a smaller scale. In the right panel, you can see that the classifier method performs better for high energy cascades because even if an event is reconstructed with a long track length, if the overall shape is still more spherical, then the classifier might determine it is a cascade.



(a) True Tracks ($\nu_{\mu,cc}$)

(b) True Cascades

Figure 2.9: Ratio of the fraction of events classified correctly using the classifier method compared to the track length only method. Red is where the classifier method performs better, blue is where the track length only method performs better. (a) shows the ratio of tracks classified correctly, (b) shows the ratio of cascades classified correctly.

2.4 Improvements to the Reconstructed Energy Proxies

RetroReco provides two proxies for energy. The first one is an estimated track energy, which at the time of reconstruction is assumed to scale linearly with track length using the assumption of a minimum ionizing muon with a constant energy loss of 0.22 GeV/m. The second is an estimated cascade energy, which is obtained by scaling the light produced by a 1 GeV electromagnetic (EM) cascade to the total light yield of the event. This estimate is known at the EM-equivalent cascade energy, E_{EM}. In reality, both of these are approximations.

Cascade events from neutrino interactions are never purely electromagnetic since inelastic scattering breaks apart the nuclei and generates a hadronic component to the cascade. The decay of secondary τ leptons can also produce hadronic showers. Since the hadrons produced are more massive and slower than the electrons and positrons created during a pure EM shower, less Cherenkov light is expected from hadronic cascades (HD for short). This means that E_{EM} is likely underestimating the true energy of the cascade. Additionally, there can be neutral shower products which do not produce Cherenkov emission. To account for this, a correction factor is applied based on GEANT4 simulation of EM and HD cascades in ice [24, 25]. In these simulations, proton and electron primaries of equal energies were simulated to interact in the ice, and the total light yield of the cascade was taken to be the sum of the track length of all charged decay products with a velocity larger than the Cherenkov threshold. From that, a ratio of light yield can be

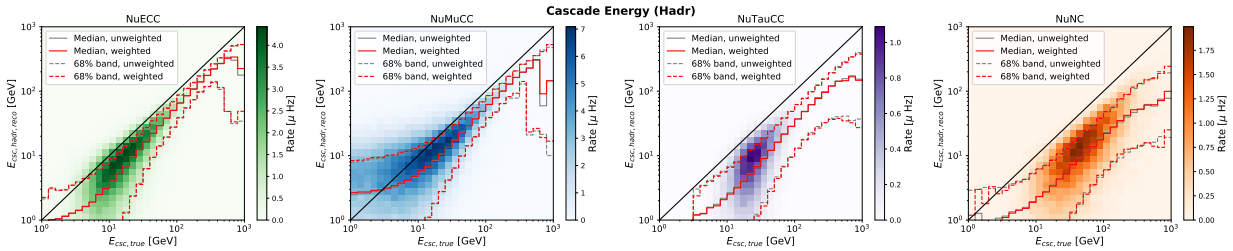
computed between EM and HD cascades at various energies:

$$F(E) = \frac{T_{HD}(E)}{T_{EM}(E)} \tag{2.1}$$

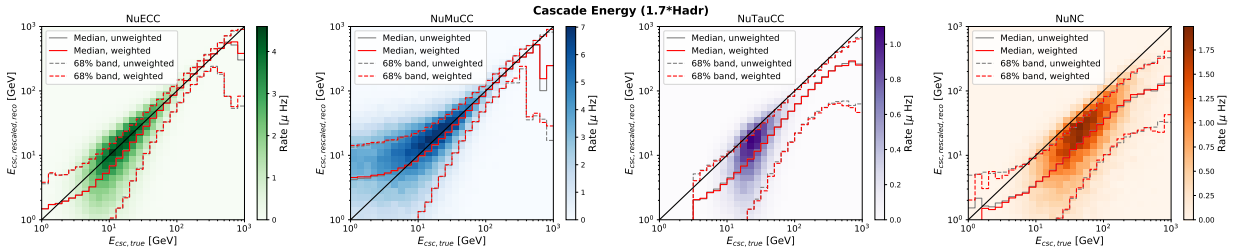
where $T_{HD}(E)$ and $T_{EM}(E)$ correspond to the respective Cherenkov track length (light yield) of the hadronic and electromagnetic cascade simulated with a primary of energy E. This relationship allows us to convert from the EM energy to HD energy.

For the tracks, we can get an improved estimate of the real track energy by using more realistic energy losses rather than a linear approximation. The track energy can be recalculated from the track length using an interpolation of the muon energy loss data tabulated in [26].

Even after these changes, the resulting reconstructed cascade energies and track lengths do not perfectly match their truth parameters in simulation. A linear scaling factor is introduced to make the reconstructed parameters better match the truth. Figure 2.10 shows the hadronic cascade energy before and after a scaling factor of 1.7 is applied, and Figure 2.11 shows the track length before and after a scaling factor of 1.45 is applied.



(a) Hadronic cascade energy before rescaling



(b) Hadronic cascade energy after rescaling

Figure 2.10: Hadronic cascade energy before and after a linear scaling factor of 1.7 is applied. This bias correction is applied to make the reconstructed energy be a better estimator of true cascade energy. The red lines show the median and 68% containment regions. The black line is the 1:1 line where the reconstructed value exactly matches the truth. The correction factor pulls the median to more closely follow the 1:1 line. The different panels correspond to the different interaction types: ν_e CC, ν_μ CC, ν_τ CC, and all NC.

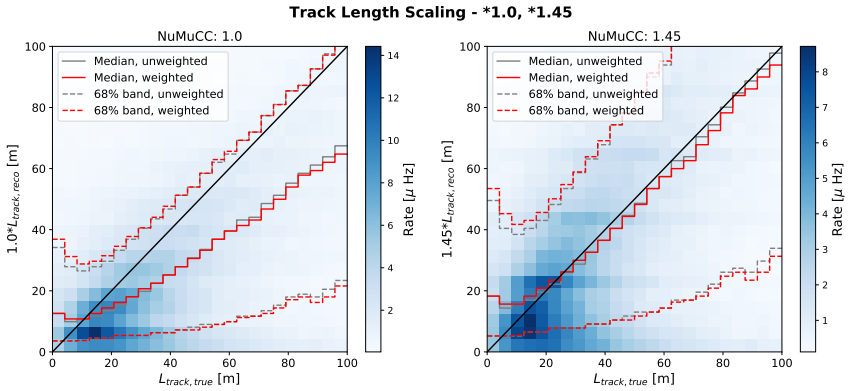
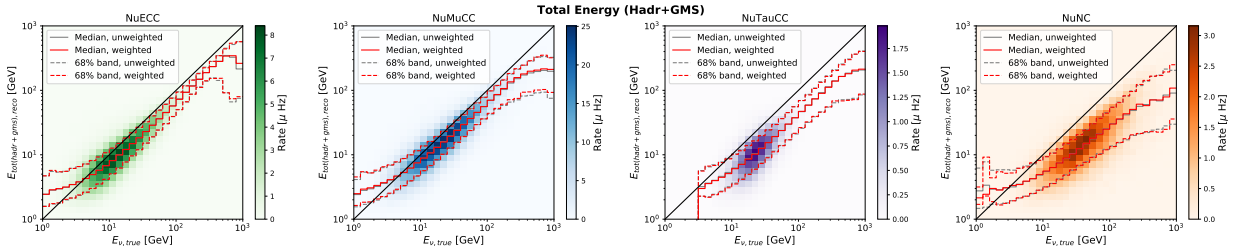
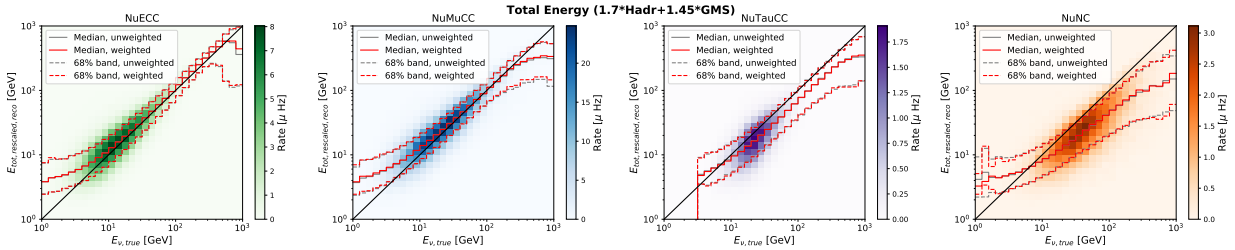


Figure 2.11: Reconstructed track length before and after a linear scaling factor of 1.45 is applied. This bias correction is applied to make the reconstructed track length be a better estimator of true track length. The red lines show the median and 68% containment regions. The black line is the 1:1 line where the reconstructed value exactly matches the truth. The correction factor pulls the median to more closely follow the 1:1 line. Only ν_μ CC is shown because other flavors do not have true tracks.

The total energy is again computed by summing the contributions from the cascade energy and track energy, this time with the adjusted values. The effects on the total energy with and without these correction factors applied can be seen in Figure 2.12. For ν_e CC events and ν_μ CC events, the total reconstructed energy matches the neutrino energy quite well for most of our events. For the ν_τ CC and Neutral Current events, the reconstructed energy is lower than the initial neutrino energy, because there is energy carried away by an outgoing neutrino, not all of the incoming neutrino energy is visible. Therefore we wouldn't expect these to have perfect energy resolution. We can remove the "missing" energy due to the outgoing neutrino to get a quantity we refer to as the "deposited energy", and then we can perform the same comparison. This is shown in Figure 2.13. We find that this procedure provides a reliable method for determining the true deposited energy in the energy range of 5 to 300 GeV. This energy estimator is used for the analysis in the chapters that follow.



(a) Total reconstructed energy before rescaling



(b) Total reconstructed energy after rescaling

Figure 2.12: Total reconstructed energy before and after linear scaling factors are applied. This bias correction is applied to make the reconstructed energy be a better estimator of true energy. The red lines show the median and 68% containment regions. The black line is the 1:1 line where the reconstructed value exactly matches the truth. The correction factor pulls the median to more closely follow the 1:1 line. The different panels correspond to the different interaction types: ν_e CC, ν_μ CC, ν_τ CC, and all NC.

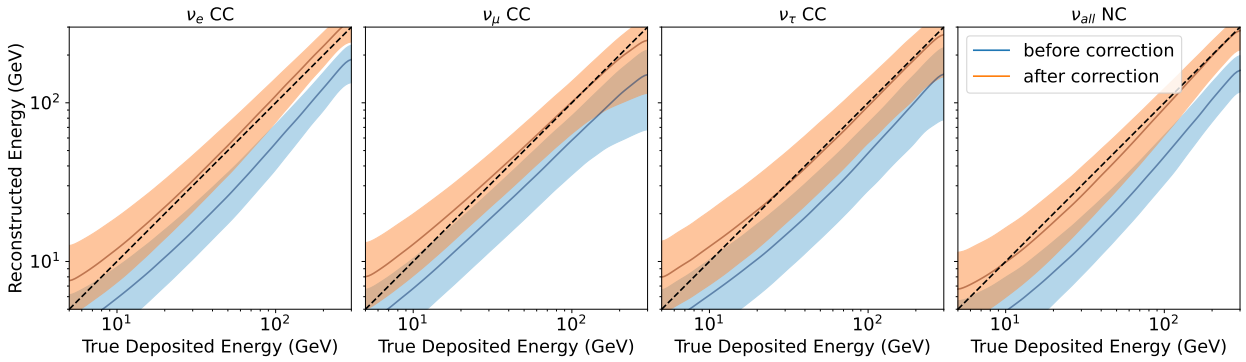


Figure 2.13: Reconstructed deposited energy before and after the scaling factors are applied. This procedure makes the reconstructed energy a better estimator of true deposited energy. The shaded regions show the 68% containment regions, and the solid lines show the median. The dashed, black line is the 1:1 line where reconstructed energy exactly matches the true deposited energy. The correction pulls the median to more closely follow the 1:1 line. Figure Credit: [21].

3

Optimization of Muon Simulation

Twice as fast, and this isn't even my final form.

– Juan Pablo Yáñez, describing the optimization discussed in this chapter

Atmospheric muons, which are generated in the same type of air showers that produce atmospheric neutrinos, are one of our largest backgrounds. As shown in Figure 2.1, atmospheric muons initially outnumber neutrinos by several orders of magnitude. When trying to perform an analysis, it is crucial to have trustworthy modeling of our backgrounds to understand any potential contamination that could be present in our final sample. This requires large amounts of Monte Carlo (MC) simulation with enough simulated background events in the final level to understand what types of events are making it into the final sample. Generating enough MC at the final sample poses significant computational challenges because the event selection, by design, eliminates almost all of the muons. For every 1,000,000 muons that pass the DeepCore Filter, only 1 will survive to the final level. To make it computationally feasible to generate enough muon MC necessary for an analysis, I developed a new scaling technique using Kernel Density Estimation (KDE) and importance sampling to improve the efficiency of our simulation, which will be covered in this chapter.

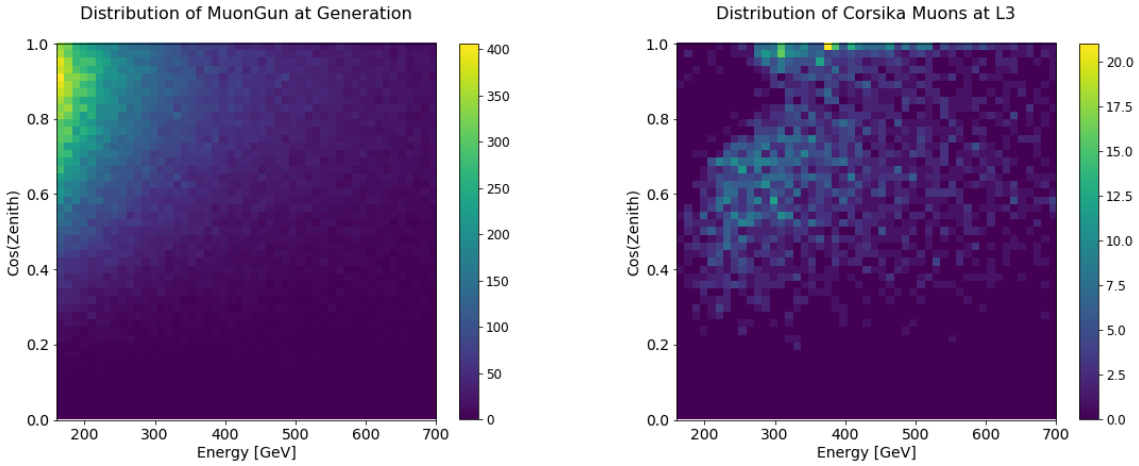
3.1 Generating a Target Kernel

Atmospheric muons can be simulated using an air-shower simulation package known as Corsika [27], however this is very computationally expensive and, for our purposes, quite inefficient. Instead, we use

an internal tool known as MuonGun which generates muons on the surface of a cylinder surrounding the detector and weights this to match air shower simulation like Corsika, rather than re-simulating the full air showers. The muon energies in this chapter refer to the energy at the surface of the generation cylinder.

The events initially generated by MuonGun look very different than the resulting distributions after a few levels of the event selection. Notice that at the MuonGun generation level (Figure 3.1a), most of the muons begin in the upper left corner around $\cos(\theta_{zenith})$ of 0.8-1.0 and less than ~250 GeV, whereas for Corsika Level 3 (Figure 3.1b), essentially no muons remain in that same region of phase space. This means that we are simulating many, many muons that are almost immediately discarded. This is very computationally inefficient. In order to improve the efficiency of this simulation, this new technique modifies the distribution of muons when they are generated, to more closely match the distribution after a few levels, so that more events survive to the later levels.

For the purposes of developing this technique, the "target" distribution that we use is Level 3 Corsika simulation. Even though the low statistics leads to a patchy distribution, we observe from Figure 3.1b that two distinct populations are starting to form, a cluster near $\cos(\theta_{zenith}) = 1.0$ and another cluster spanning $\cos(\theta_{zenith})$ from about 0.4 to 0.8. This is the result of a geometrical detector effect due to the detector being approximately cylindrical. We can separate these populations based on whether they enter through the top of the cylinder or through the side of the cylinder; therefore we will refer to these populations as "top" and "side". (It is extremely unlikely for a muon to enter through the bottom of the cylinder because Earth acts as a shield.) Figure 3.2 shows a diagram of this geometry (not to scale). Figure 3.3 shows the events divided into these two populations.



(a) Generation Level

(b) Level 3

Figure 3.1: Simulated distributions of muons when they are first generated and after a few levels of processing. For the region of phase space where most muons are generated, almost no muons survive to later levels.

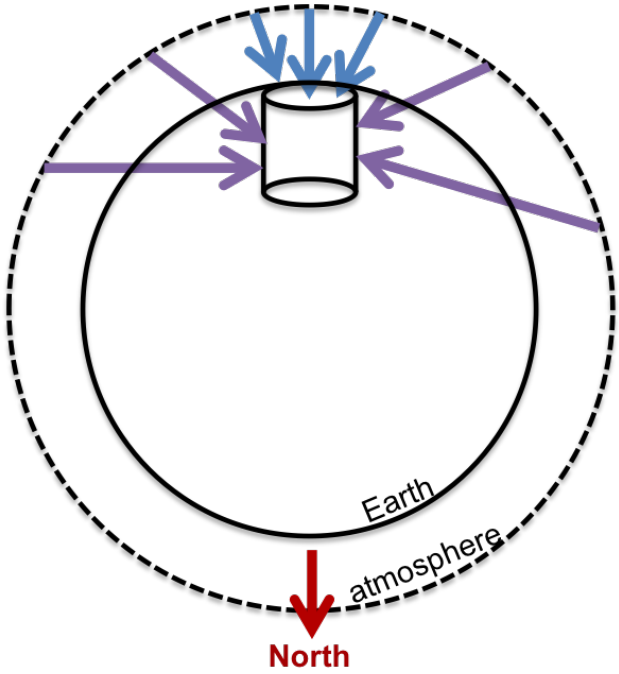
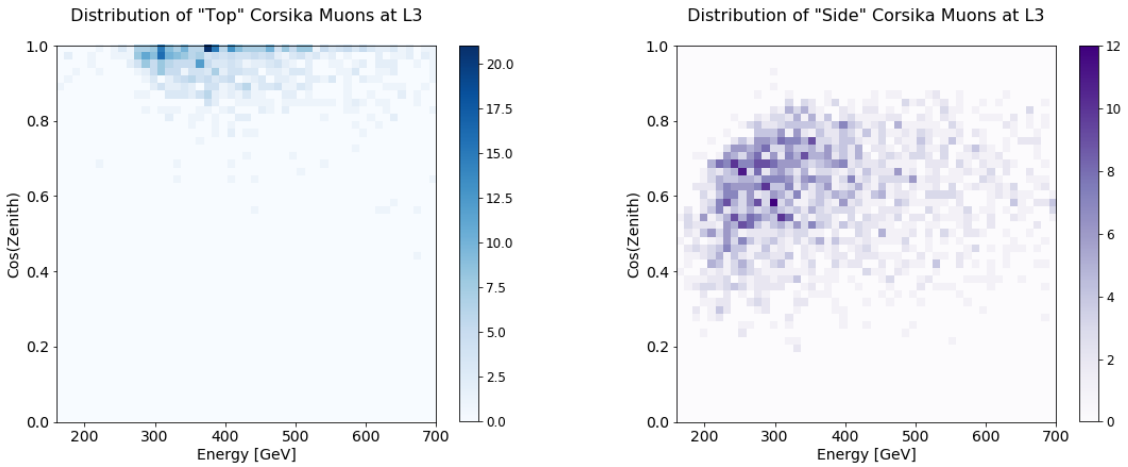


Figure 3.2: Diagram (not to scale) showing the cylindrical shell around detector. We can separate these populations based on whether they enter through the top of the cylinder (blue) or through the side of the cylinder (purple). The distributions of each event type is shown in Figure 3.3.

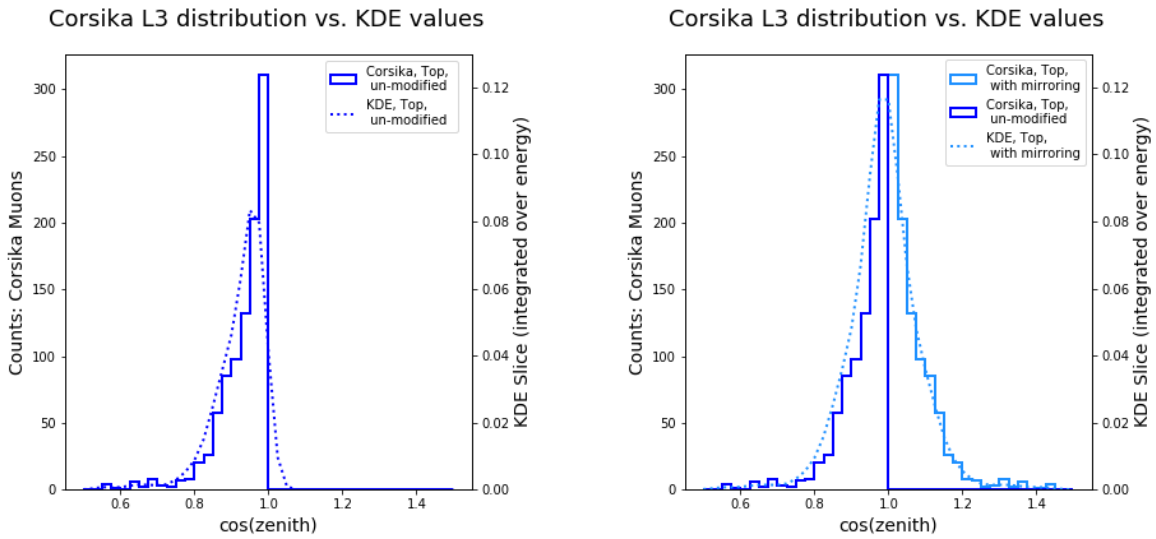


(a) Top.

(b) Side

Figure 3.3: Simulated Level 3 distributions of muons which entered through the top (left) or side (right) of the detector.

Before generating the kernels, there is one more modification necessary. Because the kernel smooths out the distribution, it does not handle sharp boundaries well. It smears the top bin above $\cos(\theta_{zenith})$=1 and does not give the correct behavior in the last bin (See Figure 3.4a). This can be circumvented by creating the kernel from a distribution without sharp boundaries, by making a mirror copy of the data and flipping it over the $\cos(\theta_{zenith})$=1 boundary. Of course a value such as $\cos(\theta_{zenith})$=1.1 is unphysical, but it creates the correct distribution in our region of interest (0 to 1). This is only done to the muons in the "top" distribution because the "side" distribution does not have a sharp drop at a boundary; it is a continuous distribution centered far from $\cos(\theta_{zenith})$=1. This duplicate distribution is created using the mapping $x \to 1 + (1 - x)$ where x is each event's $\cos(\theta_{zenith})$. Figure See Figure 3.4b shows a 1d projection of the original distribution, the mirrored distribution, and the resulting kernel. We see that it now exhibits the correct behavior in the region of interest, i.e. up to $\cos(\theta_{zenith})$=1.



(a) Before applying mirror

(b) After applying mirror

Figure 3.4: Distribution of simulated events and the resulting KDE values, with and without mirroring applied. When mirroring is applied, we correct the behavior in the last bin (just below 1.0).

After getting our desired L3 simulation and making the necessary modifications, we create the 2-dimensional kernels using `scipy.stats.gaussian_kde`. There is one kernel for the top events and one kernel for the side events, each kernel is generated in 2-d using energy and $\cos(\theta_{zenith})$, as was shown in Figure 3.3. Figure 3.5 shows 1-dimensional slices compared to the histogram of the simulated data used to create it. The KDEs trace a smoothed version of the histograms which is exactly what we want and expect.

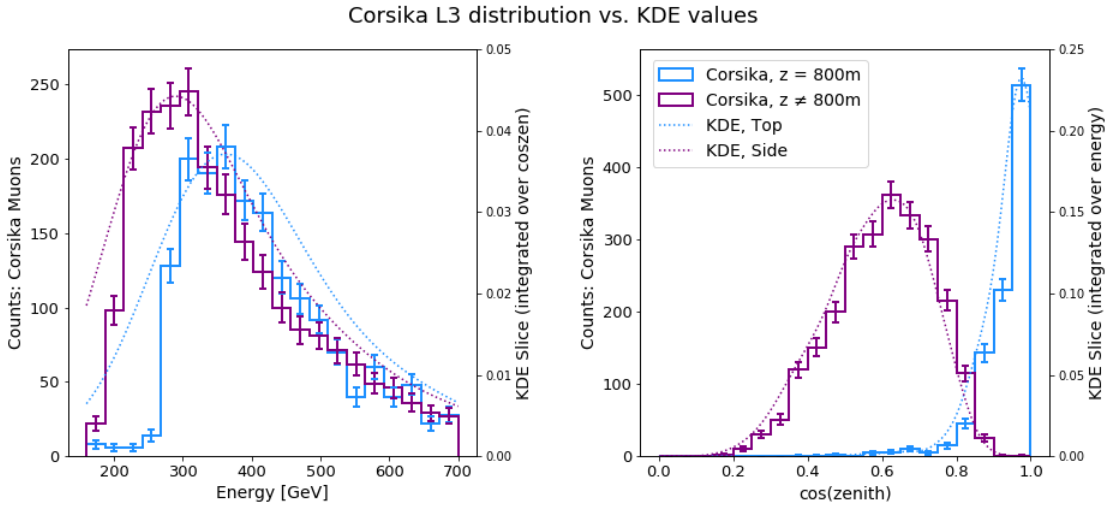


Figure 3.5: 1-dimensional slices of the KDE compared to the histogram of the simulated data used to create it. As expected, the KDEs trace a smoothed version of the histogram.

3.2 Rejection Sampling

Once we have our desired target distribution, we use it at generation level to determine which events to keep and eliminate using a technique called rejection sampling. For each generated muon, the value of the KDE at that point is determined based on its energy, $\cos(\theta_{zenith})$, and whether it entered through the top or side. The maximum KDE value is used to normalize this and turn it into a probability of keeping the event. For example, an event at the point where the KDE is at its maximum would have a 100% chance of being kept, while a smaller KDE value would correspond to a lower chance of being kept. Then, a random number is generated and we use the accept-reject sampling method to determine if the event should be kept. If the random number is less than the probability of being kept, we keep the event. If the random number is greater, then we discard the event. A 1-dimensional slice (for illustration purpose only) is shown in Figure 3.6. One caveat is that we didn't want to completely eliminate any portions of phase space and potentially introduce unexpected effects down the road, so we implemented a floor at 1% so that if the probability of keeping an event is less than 1%, we use 1% as its probability instead. This ensures that we keep about 1 in 100 events even in the most unlikely survival regions, rather than completely eliminating these regions from our simulation.

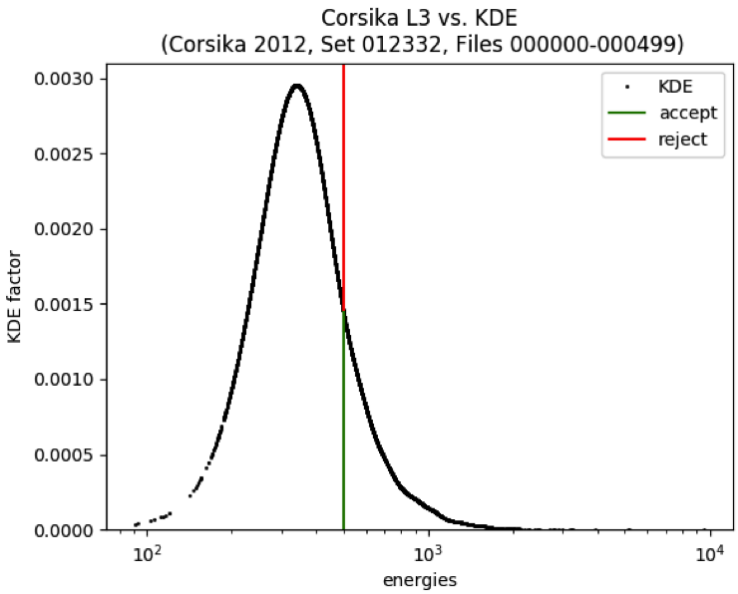
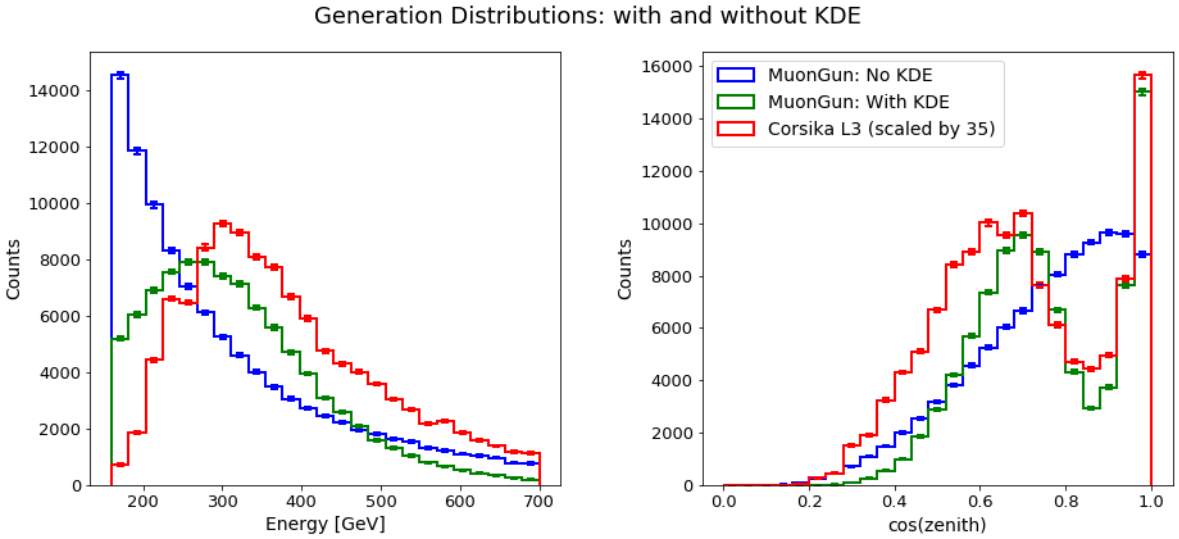


Figure 3.6: Example 1-dimensional slice of the KDE. For a given event, the corresponding KDE value is determined and then normalized by the maximum KDE value to determine the probability of being kept. A random number is drawn between 0 and 1. If the random number is less than the probability of being kept, we keep the event (shown in green). If the random number is greater, then we discard the event (shown in red).

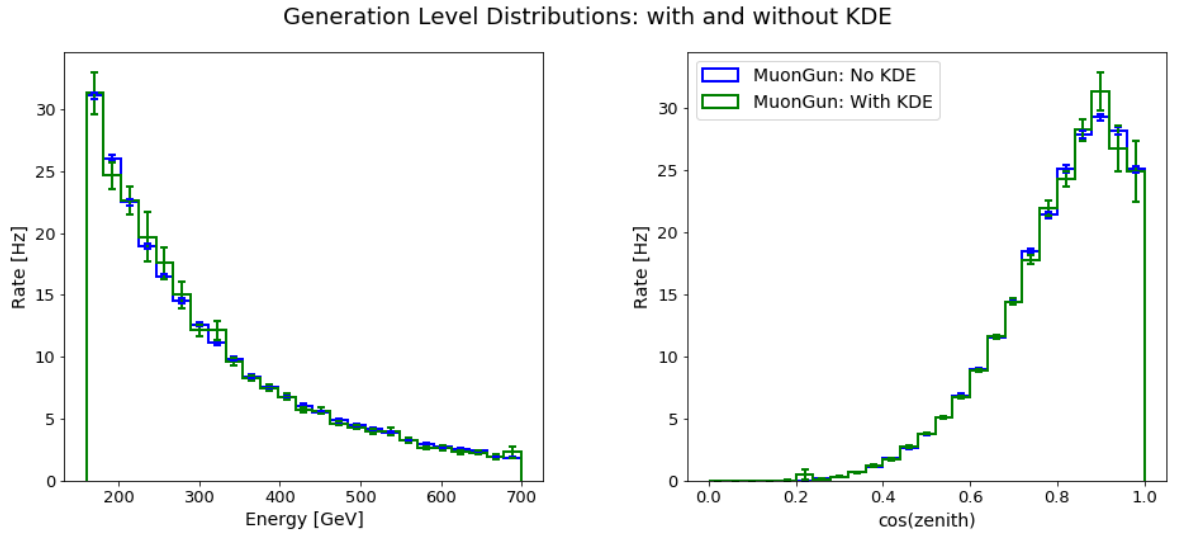
3.3 Re-weighting

Finally, because this is a technique to change the efficiency of simulation, there should be no net change on the physical weighted outcome. Therefore we must re-weight our events to take into account the way we changed the underlying simulated distribution. To do this, we divide each event's weight by the probability of having kept it during rejection sampling. Therefore, events that were unlikely to be kept are up-weighted since they effectively need to represent the events similar to it that were discarded. Figure 3.7 compares the generated distributions with and without the KDE, for both the unweighted and weighted scenario. We see that the unweighted distribution is now a closer match to later levels of processing, but the weighted physical distributions are still unchanged.

I performed a study to determine the improvement from this new method and found that it improves the speed of simulation by about 40% while simultaneously allowing for about 40% more events to be kept through later levels of processing.



(a) Unweighted distributions



(b) Weighted distributions

Figure 3.7: Generated distributions with (green) and without (blue) the KDE, for both the unweighted (top) and weighted (bottom) scenario. We see that the unweighted distribution is now a closer match to later levels of processing (red), but weighted physical distributions are still unchanged.

4

Analysis Methods for Measuring the Atmospheric Neutrino Oscillation Parameters

4.1 Binned Analysis

As mentioned in earlier chapters, the primary goal of this work is to measure the atmospheric neutrino oscillation parameters, namely θ_{23} and Δm^2_{32}. As seen in Figure 1.7, this presents as the disappearance of muon flavored neutrinos in certain portions of the energy-cos(θ_{zenith}) phase space. Because of this, we perform a binned analysis whereby the events are binned in the following parameters: reconstructed energy, reconstructed cos(θ_{zenith}), and the flavor/particle identification (PID) described in Section 2.3.

The boundaries and binning for these parameters were chosen to balance competing interests: eliminating regions of phase space where we do not properly simulate events versus keeping as much of the phase space as possible that contributes to improving our sensitivity to the oscillation parameters of interest. These boundaries and binnings were also chosen in conjunction with the boundary for the final level muon cut using the procedures outlined in the following sub-sections. Ultimately, the following binning scheme is used:

- **Energy**: 12 $\log_{10}$-spaced bins between 5 and 300 GeV
- **cos**(θ_{zenith}): 10 linearly-spaced bins between -1.0 and 0.3

- **PID**: 3 bins with edges at [0,0.5,0.85,1.0]

Final Level Muon Cut, $\cos(\theta_{zenith})$ cut, and the atmospheric muon background

The first decision to make is with regard to the amount of atmospheric muons that we allow into our final sample. Due to the difficulties with simulating an adequate number of atmospheric muons, it can be a problematic background to model properly. Because of this, we ultimately choose a conservative approach that uses stringent cuts to reduce the number of atmospheric muons in our final sample. Three scenarios of binning cuts in zenith were tested:

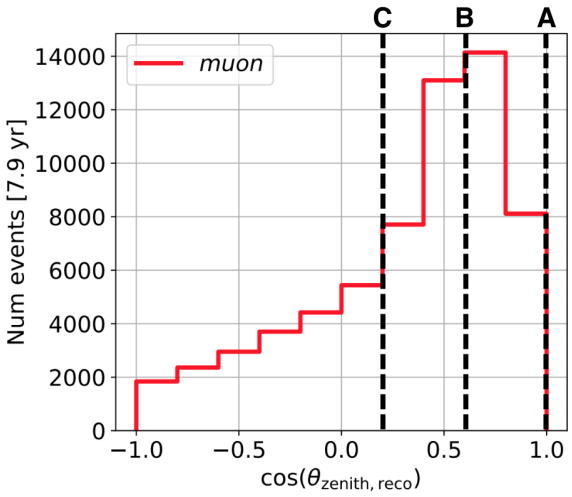
- **Scenario A:** Full sky (no cuts)
- **Scenario B:** Mild downgoing event cut (keep $\cos\theta_{zenith} < 0.6$)
- **Scenario C:** Harsh downgoing event cut (keep $\cos\theta_{zenith} < 0.2$)

In addition to zenith angle cuts, three choices of cuts on the final level muon classifier were tested:

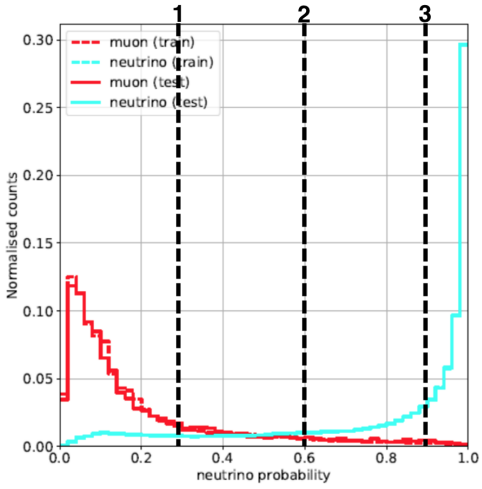
- **Scenario 1:** Higher statistical power (keep classifier output score > 0.3)
- **Scenario 2:** Medium statistical power (keep classifier output score > 0.6)
- **Scenario 3:** Lower statistical power (high neutrino purity) (keep classifier output score > 0.9)

Distributions of both of these variables along with the cut values indicated are shown in Figure 4.1.

The different scenarios of muon classifier cuts were tested to evaluate the cost/benefits of sample statistics vs. signal purity in the context of the resulting analysis sensitivity they provide. All combinations of zenith and classifier cut scenarios were tested, and evaluated using the outcome of a sensitivity test (see Section 5.2 for a description of such test). The outcome of this optimization study is illustrated in Figure 4.2. We ultimately chose the restrictive muon cuts in order to simplify the treatment of muons in later stages of the analysis; the cut values chosen were $\cos\theta_{zenith} < 0.3$ and muon classifier score > 0.8 which is closest to the "C3" scenario tested here. Although this does correspond to the worst sensitivity of the scenarios tested, the impact was small enough that it was still worthwhile in order to allow us to use the simplified muon treatments as will be described Section 4.2.



(a) Zenith distribution of muon background events, showing the three cut values tested: 1.0, 0.6, 0.2.



(b) L7 muon classifier distribution for muon and neutrino events, showing the three cut values tested: 0.3, 0.6, 0.9.

Figure 4.1: Final level cut values tested during optimization of $\cos(\theta_{zenith})$ (left) and Level 7 muon classifier (right).

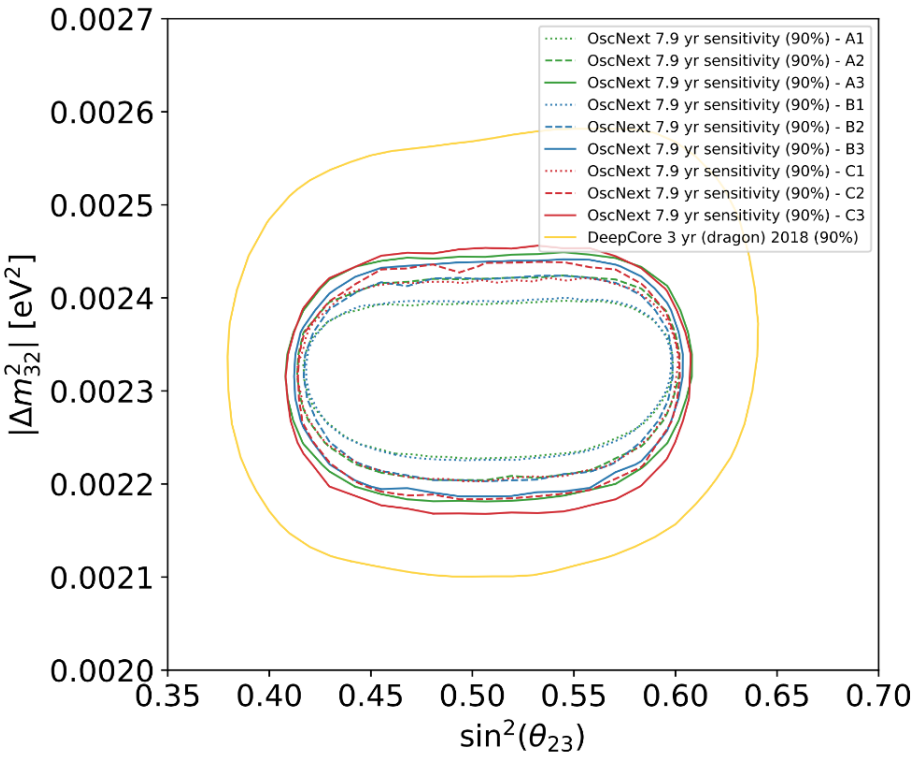


Figure 4.2: Sensitivities to the atmospheric neutrino oscillation parameters for various combinations of the Level 7 muon classifier and $\cos(\theta_{zenith})$ cut configurations. The colors indicate the zenith cut (A=green, B=blue, C=red). The line styles indicate the muon classifier cut (1=dotted, 2=dashed, 3=solid). The chosen cut was closest to C3 (solid red). The yellow line shows the result from a previous generation of analysis [7] for comparison.

PID Binning

A similar test was performed to determine the effect of the number and location of PID bins on the resulting sensitivity. Four different PID binning configurations were explored:

- **Scenario 2:** Two PID bins only (boundaries: [0, 0.5, 1.0])
- **Scenario 3a:** Three PID bins option A (boundaries: [0., 0.35, 0.5, 1.0])
- **Scenario 3b:** Three PID bins option B (boundaries: [0., 0.5, 0.85, 1.0])
- **Scenario 4:** Four PID bins (boundaries: [0., 0.35, 0.5, 0.85, 1.0])

Distributions of the PID variable along with these cut values are shown in Figure 4.3.

Performing tests of the effects of these cut configurations on the sensitivity (see Figure 4.4), we found that Scenarios 2 and 3a yielded similar sensitivities, and Scenarios 3b and 4 yielded similar sensitivities. We decided to go to with Scenario 3b to maximize sensitivity without adding unnecessary bins that don't contribute towards improving the sensitivity. The chosen binning scheme results in the $\nu_{\mu,cc}$ purities for each bin listed in Table 4.1.

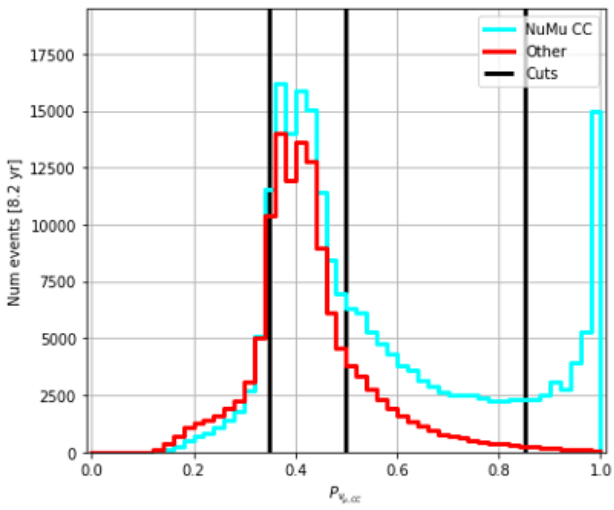


Figure 4.3: PID distribution and bin configurations tested. The 2-bin scenario uses only the middle cut at 0.5. The 3a-bin scenario uses two lower cuts (0.35 and 0.85). The 3b-bin scenario uses the two upper cuts (0.5,0.85). The 4-bin scenario uses all three cuts.

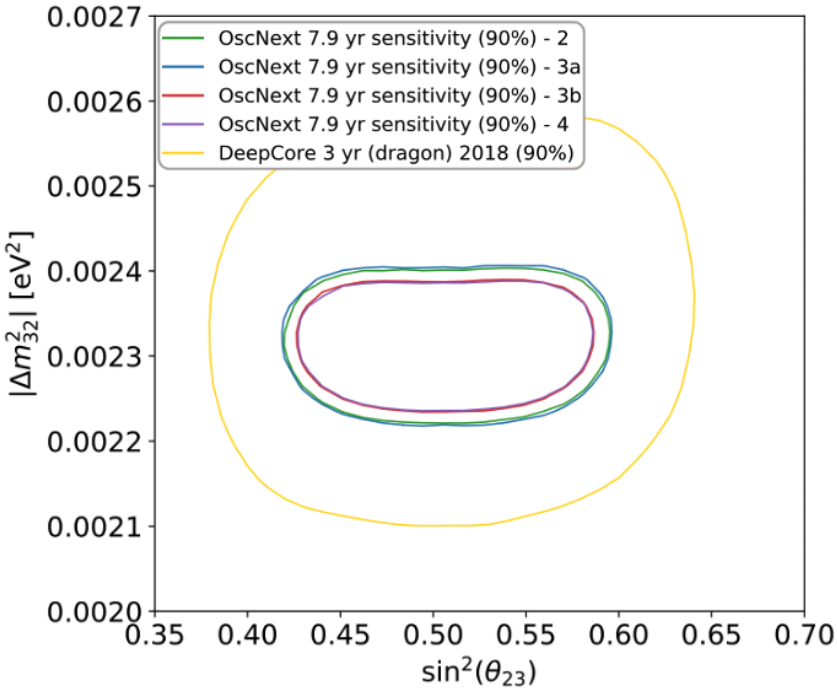


Figure 4.4: Sensitivities to atmospheric neutrino oscillation parameters for various PID binning configurations. Scenarios 2 (green) and 3a (blue) yielded similar sensitivities, and Scenarios 3b (red) and 4 (purple) yielded similar sensitivities. The yellow line shows the result from a previous generation of analysis [7] for comparison.

Bin label	PID Score Range	$\nu_{\mu,cc}$ Purity
Cascades	0.0-0.5	48%
Low purity tracks	0.5-0.85	72%
High purity tracks	0.85-1.0	97%

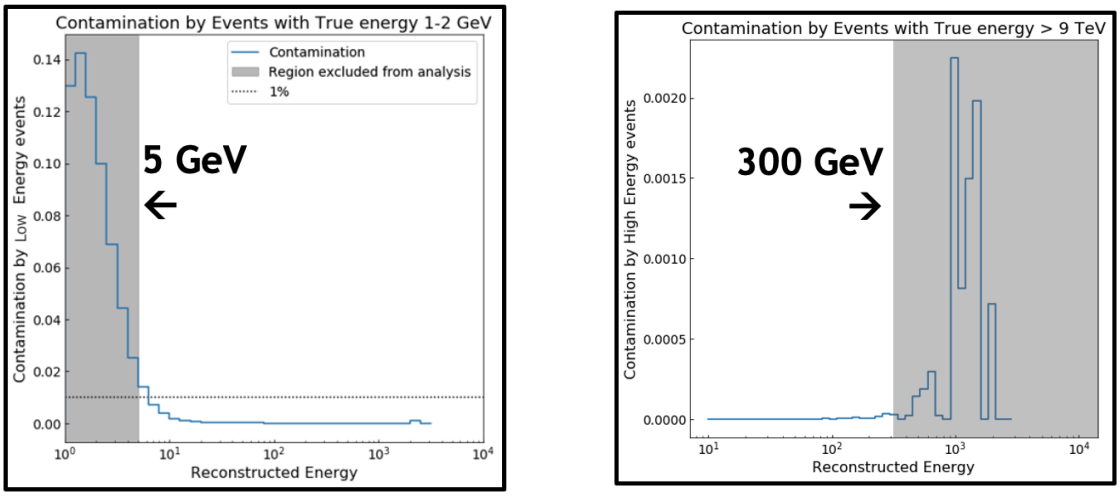
Table 4.1: PID bin boundaries and the resulting $\nu_{\mu,cc}$ purity

Energy Boundaries

To complete the analysis binning optimization, the reconstructed energy range was chosen to include events solely within the range $[5, 300]$ GeV. These boundaries were chosen to limit the contributions of unsimulated misreconstructed events into our sample, which could lead to data/MC disagreement. We simulate events in the true energy range of 1 GeV to 10 TeV, but at both extremes there is some spread in the reconstructed energies that this corresponds to. To estimate the potential contamination of *unsimulated* events reconstructing into our analysis region, we look at the next best thing – the very extreme values of events we do simulate. At the lower range, we look at events with true energies of 1-2 GeV and at the upper end we look at events with true energies of 9-10 TeV. Comparing the reconstructed energy distribution of all

events versus the events in these extreme regions, we can estimate the potential contamination. Figure 4.5 shows the fraction of contamination due to low/high energy events into the reconstructed energy distribution, along with the proposed reconstructed energy boundaries of 5 GeV and 300 GeV.

In the lowest energy bin around reconstructed energy of 5 GeV, about 1-1.5% of events in that bin are events with true energies of 1-2 GeV (see Figure 4.5a). Therefore, we estimate that the contribution of events into that lowest bin due to unsimulated events (events with true energy below 1 GeV) to be at the level of less than about 1%. At the highest energies, the contamination is much smaller, on the order of 0.01% (see Figure 4.5b). But there are also much fewer statistics at high energy, so we place the cut at 300 GeV to avoid regions where statistical fluctuations are large. Our resulting sample therefore is restricted to events with reconstructed energies of 5-300 GeV.



(a) Lower bound

(b) Upper bound

Figure 4.5: Contamination of events near the edges of the simulated energy regions into the reconstructed analysis space is used as an upper bound estimate on the possible contamination due to events outside the simulated energy bands.

Number of bins in energy and $\cos(\theta_{zenith})$

Once the external boundaries/edges are determined, we also must choose the internal spacing of the bins. We want enough bins that we can resolve the oscillation peak, but a binning too fine would reduce the number of events in each bin and increase statistical error in each bin. In the end, we have 12 $\log_{10}$-spaced bins in energy and 10 linear bins in $\cos(\theta_{zenith})$.

Bin Masking

There is one final choice made with respect to the bins used in the analysis. Bins with very few expected counts can introduce unwanted behavior in the oscillation fits if a data event is found in a bin with $<<1$ expected events. Because these bins don't contribute much to the analysis and to avoid any issues during the fit, we have chosen to mask off bins with very few events. Or in other words, these bins are excluded from the analysis. They are mostly low energy bins in more track-like PID regions (since it is difficult/impossible to confidently identify low energy $\nu_{\mu,CC}$ events as tracks since the μ produced is low energy and does not travel far enough to be clearly identified). The bins we have chosen to mask are shown in Figure 4.6 as the grey shaded boxes.

Expected Event Maps

Putting together all of the binning components described in this section, we end up with the expected distribution of events (sometimes referred to as "templates" or "maps") shown in Figure 4.6 according to our nominal simulation. We can also quantify the effect that our physics parameters of interest will have on these templates. Figure 4.7 shows the change in the expected number of events in each bin that occurs if the physics parameters are shifted from $\theta_{23} = 45°$ to $\theta_{23} = 50°$ (top), and from $\Delta m^2_{31} = 2.48 \times 10^{-3}$ eV^2 to $\Delta m^2_{31} = 2.38 \times 10^{-3}$ eV^2 (bottom). The effects are strongest in the track bins and occur in the upgoing region around 20-50 GeV, similar to the expected effect from Figure 1.7.

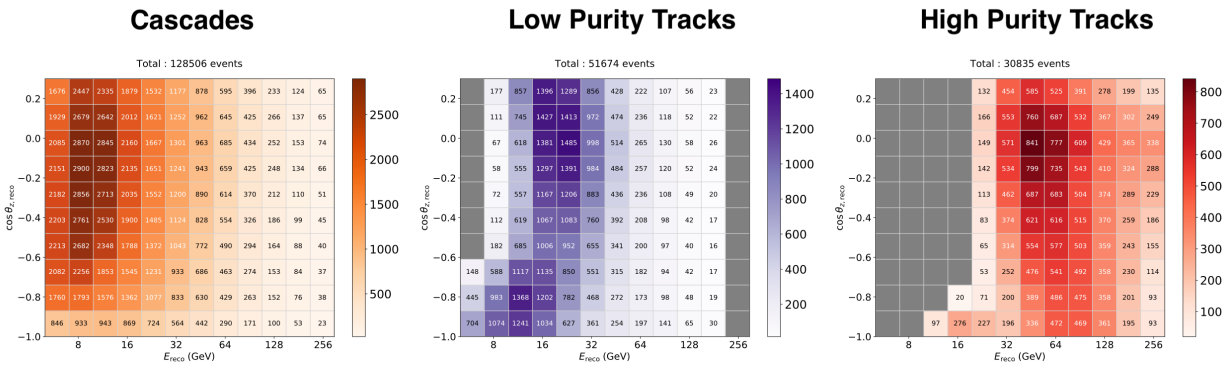
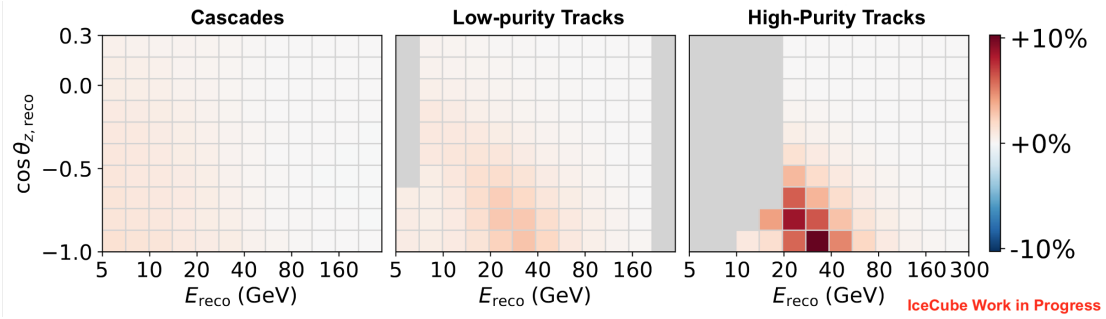
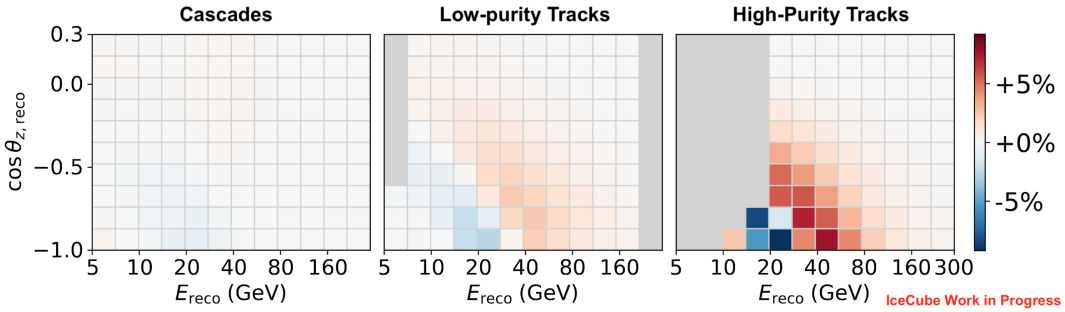


Figure 4.6: Expected distribution of events in the final analysis binning. The grey boxes indicate the bins that have been masked.



(a) Effect of shifting θ_{23} from 45° to 50°.



(b) Effect of shifting Δm^2_{31} from 2.48×10^{-3} eV^2 to 2.38×10^{-3} eV^2.

Figure 4.7: Change in the expected number of events in each bin, as the physics parameters are shifted.

4.2 Treatment of Systematic Uncertainties

Flux Uncertainties

The uncertainties associated with the production of neutrinos and muons in the atmosphere include the overall normalizations of the rates of neutrinos and muons, the spectral index of the neutrino flux, as well as a suite of parameters introduced by Barr et al. to describe hadron production uncertainties [28]. Experimental data is used to estimate the uncertainty in the neutrino production for various portions of E and x_{lab} phase space, for production from both pions and kaons, where x_{lab} is the amount of energy that the secondary receives from its parent particle. Figure 4.8 shows the phase space divided into regions we refer to as "blocks". Figure 4.9 shows the uncertainty associated with each block.

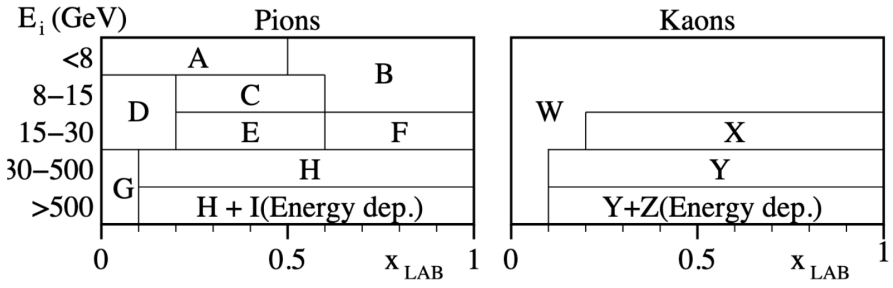


Figure 4.8: Barr blocks used for parameterizing flux uncertainties. Each block corresponds to a certain portion of phase space in E and x_{lab}. Figure Credit: [28].

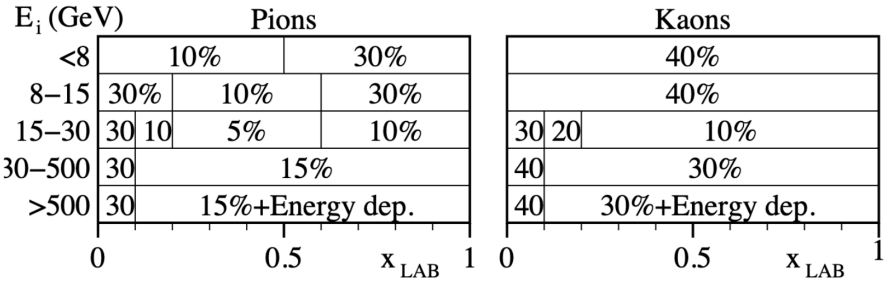


Figure 4.9: Uncertainty on atmospheric neutrino flux production corresponding to each Barr block. Figure Credit: [28].

Cross Section Uncertainties

The dominant cross section for our energy range is Deep Inelastic Scattering (DIS), but we also have subleading contributions from charged-current quasi-elastic (CC QE) events and charged-current resonance events (CC RES). For each interaction type, we include the axial mass M_A as a nuisance parameter. Changing the axial mass changes the associated form factor, which in turn changes the number of neutrinos that undergo that type of interaction.

To treat uncertainties in the DIS cross-section, we use a bespoke treatment that allows for interpolation between discrete models. The Genie package [18] that we use for generating simulated neutrinos is generally considered quite reliable at lower energies, but less so at high energies. At high energies, the CSMS model [29], named after A. Cooper-Sarker, P. Mertsch, and S. Sarkar, is the model that would typically be used, but they only report down to 100 GeV in their paper. This analysis sits right in the intermediate region, so we introduce an interpolation scheme that allows the data to choose if it prefers a more Genie-like or CSMS-like model. This is implemented as a scaling factor where 0 corresponds to Genie, and 1 corresponds to CSMS. We refer to this parameter as "DIS CSMS".

Treatment of Detector Uncertainties

The various types of detector uncertainties were introduced in Section 1.3 such as bulk ice, hole ice, and DOM efficiency, and in Section 2.1 we discussed the technique for simulating the detector response for several discrete parameter values. At this point, we can use this information to quantify the detector response at the analysis level. We begin by binning the events in energy, $\cos(\theta_{zen})$, and PID using the scheme described in Section 4.1. Then, we treat every bin independently. For each bin, we collect the parameter values used to make the simulation $\vec{x} = (x_1, x_2, x_3, x_4, x_5)$, where x_1 is the value of the DOM efficiency parameter, x_2 is the value of the bulk ice scattering parameter, etc. for each of the 5 detector uncertainty parameters, and $y = N_{counts}$ (the number of counts in that bin for that set). We then fit a hyperplane to these $(\vec{x}_i, y_i)$, where i is each simulated systematic set. The output of this fit is a slope for each systematic parameter and an intercept. Figure 4.10 shows the input and output of this procedure for the case of two systematic parameters (for visualization purposes only). The actual hyperplane fit uses all 5 detector systematic parameters.

Figure 4.11 shows an example of a hyperplane fit projected into only 1-dimension, hole ice p_1. A horizontal line would indicate that that parameter does not strongly influence the number of events in that bin; a steep line would indicate that the number of events in that bin has a strong dependence on the parameter value. Therefore, we see that for the example shown in Figure 4.11, the tracks are slightly more sensitive to the hole ice p_1 parameter than the cascades are, for that particular bin. We can then use these hyperplanes to

estimate the number of events expected in a bin, even for values of the detector systematic parameters that were not directly simulated.

Originally we had planned to use this procedure for both neutrinos and muons, but we encountered issues with the muon fits due to the low number of muons in our final sample. Ultimately, we decided to make a harsher cut on the muons so that the impact of muons in the final sample was sufficiently negligible that we no longer needed to account for this type of uncertainty on the muons. Therefore the hyperplane procedure described here is only used for neutrinos.

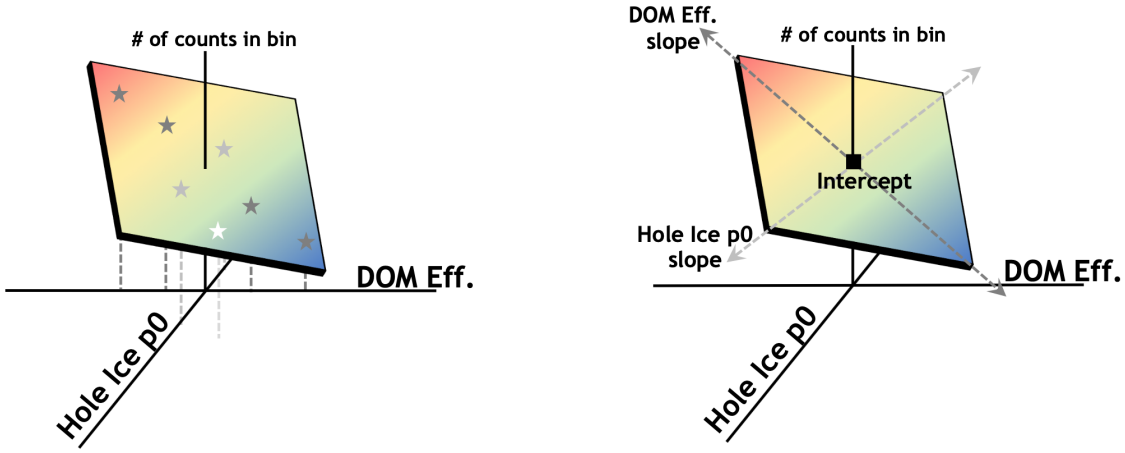


Figure 4.10: Diagram of the hyperplane fit used for modeling the response of the detector to systematic uncertainties. The left diagram shows the input, i.e. the number of events in that bin for a given combination of detector systematic parameter values. The right diagram shows the output, i.e. the slopes associated with each parameter and the intercept. This diagram shows only 2 detector parameters for visualization; the real hyperplane fit uses all 5 detector parameters.

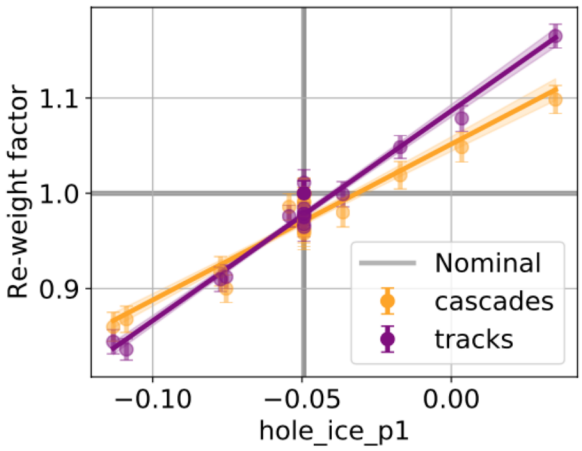


Figure 4.11: Example hyperplane fit projected into 1-dimension, hole ice p_1. The scatter points indicate the discrete simulated sets, and the line indicates the outcome of the hyperplane fit. For this particular bin, the tracks are slightly more sensitive to the hole ice p_1 parameter than the cascades.

Final List of Nuisance Parameters

There were additional types of systematic uncertainty studied that were determined to have a negligible impact on the outcome of a fit. Table 4.2 shows the final list of nuisance parameters included in the analysis, along with their nominal values, the type of prior imposed, and the allowed range.

Category	**Param**	**Nominal**	**Prior**	**Range**
Flux	$\Delta\gamma_\nu$	0	gaussian, $\sigma = \pm 0.1$	[-0.5, 0.5]
Flux	Barr, g_π^+	0	gaussian, $\sigma = \pm 0.3$	[-1.5, 1.5]
Flux	Barr, h_π^+	0	gaussian, $\sigma = \pm 0.15$	[-0.75, 0.75]
Flux	Barr, i_π^+	0	gaussian, $\sigma = \pm 0.122$	[-0.61, 0.61]
Flux	Barr, w_K^+	0	gaussian, $\sigma = \pm 0.4$	[-2, 2]
Flux	Barr, y_K^+	0	gaussian, $\sigma = \pm 0.3$	[-1.5, 1.5]
Cross-sections	$M_{A,QE}$	0	gaussian, $\sigma = \pm 1$	[-2, 2]
Cross-sections	$M_{A,res}$	0	gaussian, $\sigma = \pm 1$	[-2, 2]
Cross-sections	DIS	0	gaussian, $\sigma = \pm 1$	[-3, 3]
Normalization	N_ν	1	uniform	[0.1, 2]
Normalization	N_μ	1	uniform	[0.1, 3]
Detector/Ice	DOM Efficiency	1	gaussian, $\sigma = \pm 0.1$	[0.8, 1.2]
Detector/Ice	Hole ice, p_0	0.102	uniform	[-0.6, 0.5]
Detector/Ice	Hole ice, p_1	-0.0493	uniform	[-0.15, 0.05]
Detector/Ice	Ice absorption	1	gaussian, $\sigma = \pm 0.05$	[0.9, 1.1]
Detector/Ice	Ice scattering	1.05	gaussian, $\sigma = \pm 0.1$	[0.85, 1.25]

Table 4.2: Table of nuisance parameters used in fit.

4.3 Test Statistic

Our analysis is performed using a binned Poisson likelihood as the optimization metric. For each bin i, we can compute the probability of observing $k_{data,i}$ data events given the expected number of events λ_i for that bin:

$$L = \prod_{i\epsilon\{\text{bins}\}} \text{Pois}(k_{\text{data},\,i}; \lambda_i) = \prod_{i\epsilon\{\text{bins}\}} \frac{\lambda^{k_{\text{data},\,i}} e^{-\lambda}}{(k_{\text{data},\,i})!} \tag{4.1}$$

or taking the log:

$$LLH = \sum_{i\epsilon\{\text{bins}\}} \log\left(\frac{\lambda^{k_{\text{data},\,i}} e^{-\lambda}}{k_{\text{data},\,i}!}\right) \tag{4.2}$$

In the context of Monte Carlo simulation, the expected number of events λ_i for a given bin is given by the sum of all the weighted MC events falling into that bin. We also include penalty terms for the nuisance parameters which have Gaussian prior constraints imposed. The penalty term is given by

$$\sum_{j\epsilon\{\text{sys}\}} \frac{(s_j - \hat{s}_j)^2}{\sigma_{s_j}^2} \tag{4.3}$$

where s_j and σ_{s_j} are the j^{th} nuisance parameter's nominal value and standard deviation, and $\hat{s}_j$ is the estimated value of the parameter during the fit. Putting these together, our final test statistic $-LLH$ is given by:

$$-LLH = -\sum_{i\epsilon\{\text{bins}\}} \log\left(\frac{\lambda^{k_{\text{data},\,i}} e^{-\lambda}}{k_{\text{data},\,i}!}\right) + \sum_{j\epsilon\{\text{sys}\}} \frac{(s_j - \hat{s}_j)^2}{\sigma_{s_j}^2} \tag{4.4}$$

4.4 Analysis Framework

The analysis is performed using an open-source software framework called PISA; more details on this framework can be found in [30]. This software provides a staged approach for folding together information from many components to produce expected event distributions, as well as performing the binned likelihood minimization as described above. The number of expected events depends on the flux Φ, the oscillation probabilities P_{osc}, the cross sections σ, and the detector response $\mathcal{E}$:

$$N = \Phi \cdot P_{osc} \cdot \sigma \cdot \mathcal{E} \tag{4.5}$$

Each of these is computed in a dedicated stage and can be folded together. There are additional stages available for other specialized tasks, such as template smoothing, that can be used if needed.

5

Pre-fit Tests for Evaluating Analysis Performance

The name of the Asimov data set is inspired by the short story "Franchise", by Isaac Asimov. In it, elections are held by selecting the single most representative voter to replace the entire electorate.

– G. Cowan, K. Cranmer, E. Gross, O. Vitells [31]

After all analysis choices are finalized, a number of tests are performed prior to fitting to real data. They are meant to evaluate the potential of the analysis to perform as expected, the robustness of the fitting strategy, and its capacity to run smoothly in the idealized case where the data is derived directly from MC simulation, referred to as "pseudodata". A number of tests refer to "Asimov" pseudodata sets [31], which refer to an artificial data set generated using the expected values of the parameters (as opposed to, for example, one where statistical fluctuations have been applied). This chapter contains information for each test about its purpose, how the test is performed, and the conclusions that can be drawn from the outcome.

5.1 Asimov Inject-Recover Test

Purpose: Ensure that the minimizer is capable of converging to the correct true value of injected physics parameters.

Test procedure:

1. The MC template is weighted to specific values of the physics parameters, and provided as Asimov pseudodata for the fit.
2. Perform the fit to pseudodata.
3. Check whether the recovered (i.e. fitted) values match the injected truth.
4. Repeat for several different values of the physics parameters.

Outcome: Figure 5.1 shows the result of an inject-recover test for various injected values of the physics parameters. For every injected combination of physics parameters, the minimizer recovers the truth. This can be seen in the figure as the fitted values (purple x's) falling exactly on top of the true injected parameters (yellow circles).

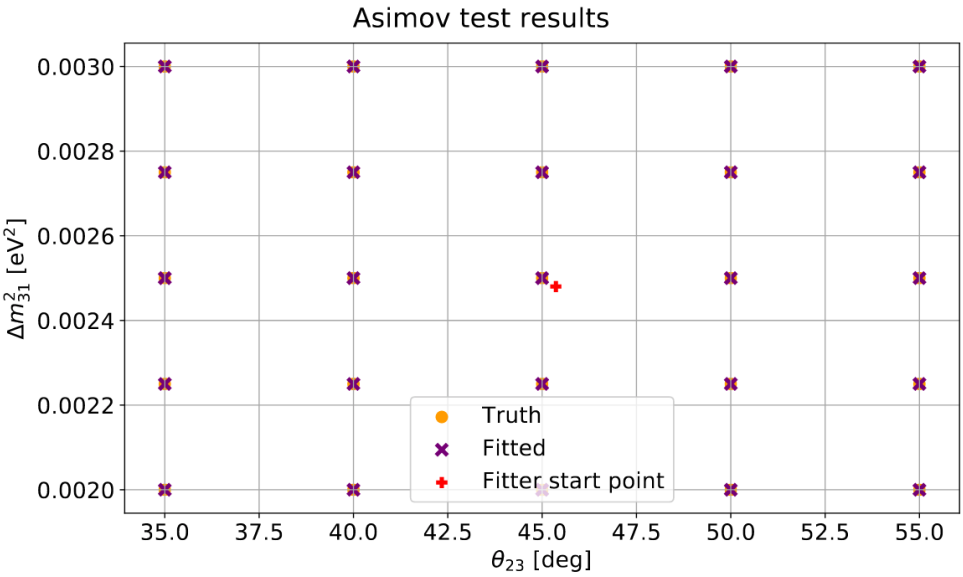


Figure 5.1: Results of the inject-recover test with Asimov pseudodata. For each of the scan points shown in yellow, a fit was performed where the pseudodata is generated with that combination of true physics parameter values. In every case, the analysis is able to recover the correct values of the true physics parameters.

5.2 Sensitivity Test

Purpose: Quantify the precision with which the analysis is able to measure the physics parameters of interest (for a given combination of true physics parameters). Essentially this is trying to answer: for a given combination of true physics parameters, what other combinations of physics parameters can we reject at the 90% confidence level?

Test procedure:

1. The Monte Carlo template is weighted to nominal values of the physics parameters, and provided as Asimov pseudodata for the fit. (No statistical fluctuations applied.)
2. Perform a free fit to pseudodata, to determine the test statistic LLH value at the best fit point.
3. Perform the fit to pseudodata again with the physics parameters fixed to a certain off-nominal value during the fit.
4. Quantify the difference in LLH between the fit with the off-nominal physics parameters and the free fit to the best fit point.
5. Repeat steps 3 & 4 with fixing the physics parameters to many different combinations, to map out the likelihood space.
6. Determine the test statistic critical values that corresponds to 1σ and/or 90% confidence level using Wilks' theorem [32].
7. Draw a contour where the difference in LLH equals the test statistic critical values found above.

Outcome: Following the fits described above, the test statistic value is calculated for every tested combination of physics parameters, and from that, a 90% contour is drawn. The sensitivity can also be compared to the currently reported measurement contours of the oscillation parameters from other experiments. This is done in Fig. 5.2. We see that the analysis is expected to yield a result that constrains the parameters of interest with similar precision to other existing neutrino experiments.

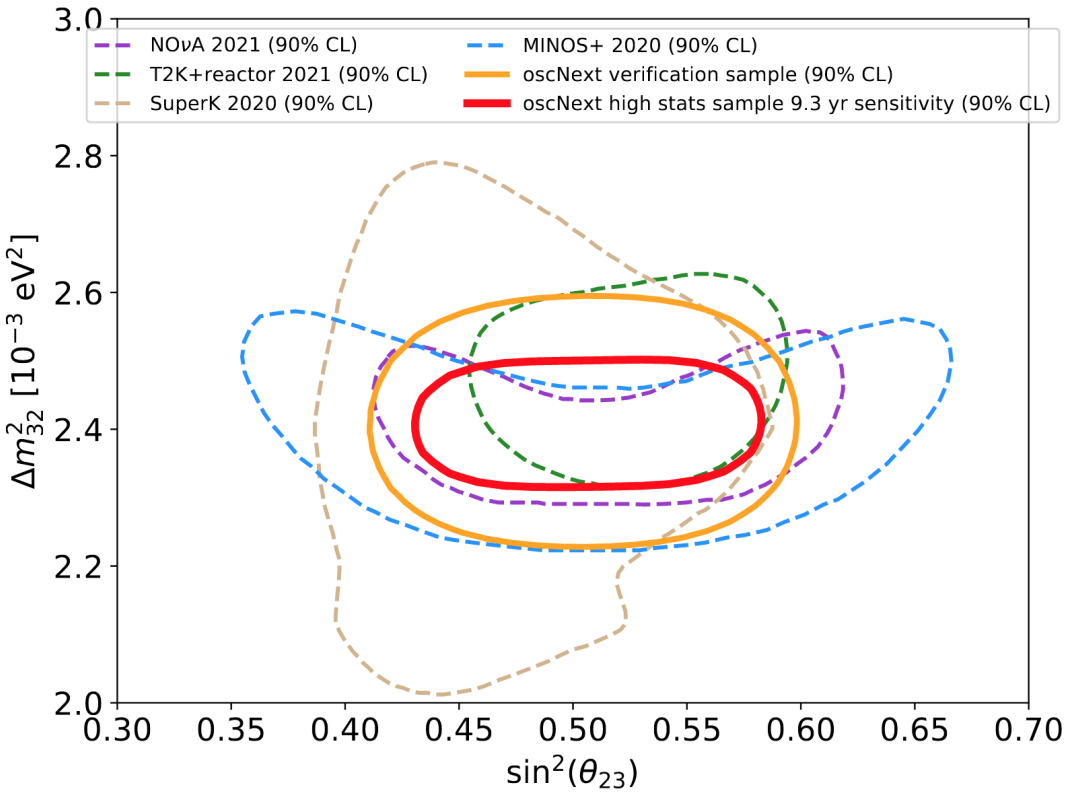


Figure 5.2: Outcome of sensitivity test: 90% *expected* contour (in red) in comparison to the 90% confidence level measurements of another DeepCore measurement (gold) and a few other experiments' published results (dashed lines) [33–36].

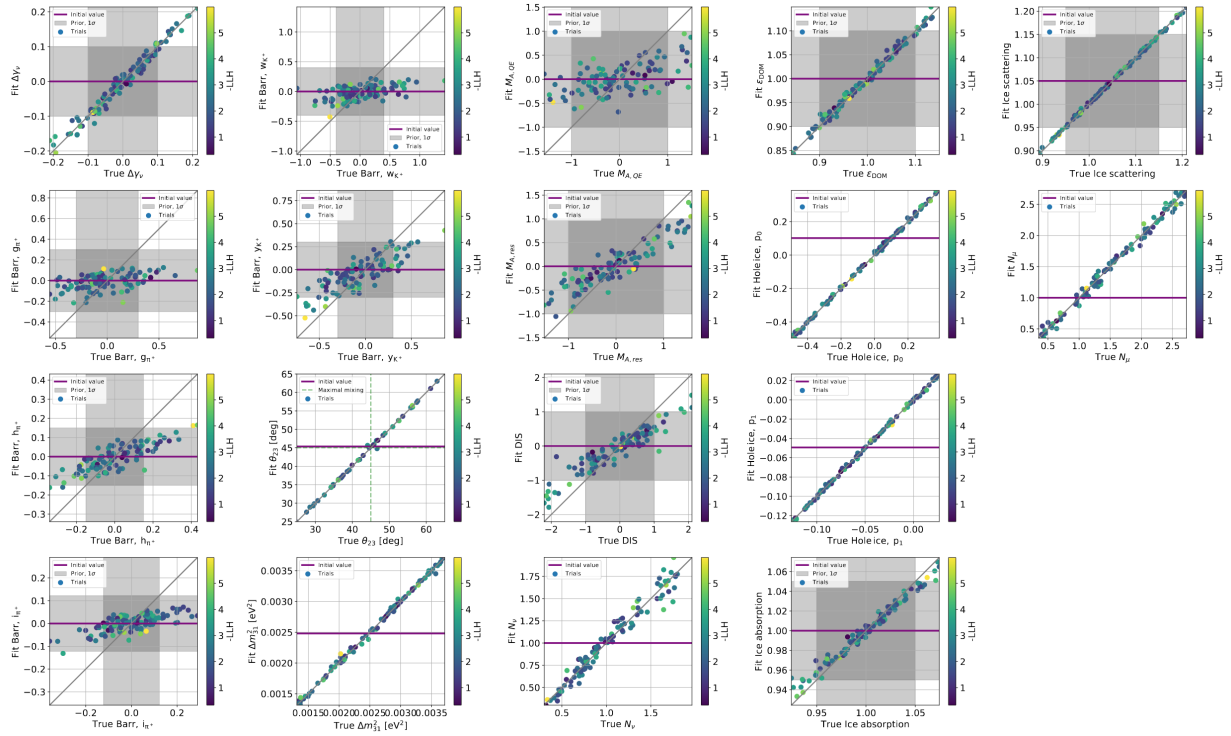
5.3 Parameter Ensemble Test

Purpose: Ensure that the analysis fit behaves well for off-nominal values of the nuisance parameters, and that those parameters are able to be accurately recovered in the fit. This is another form of inject-recover test. (Note: the previously mentioned tests investigate how the analysis responds to changes in *physics* parameters, and so this test is designed to investigate behavior for *all* parameters, i.e. both the *systematic/nuisance* parameters and the physics parameters.)

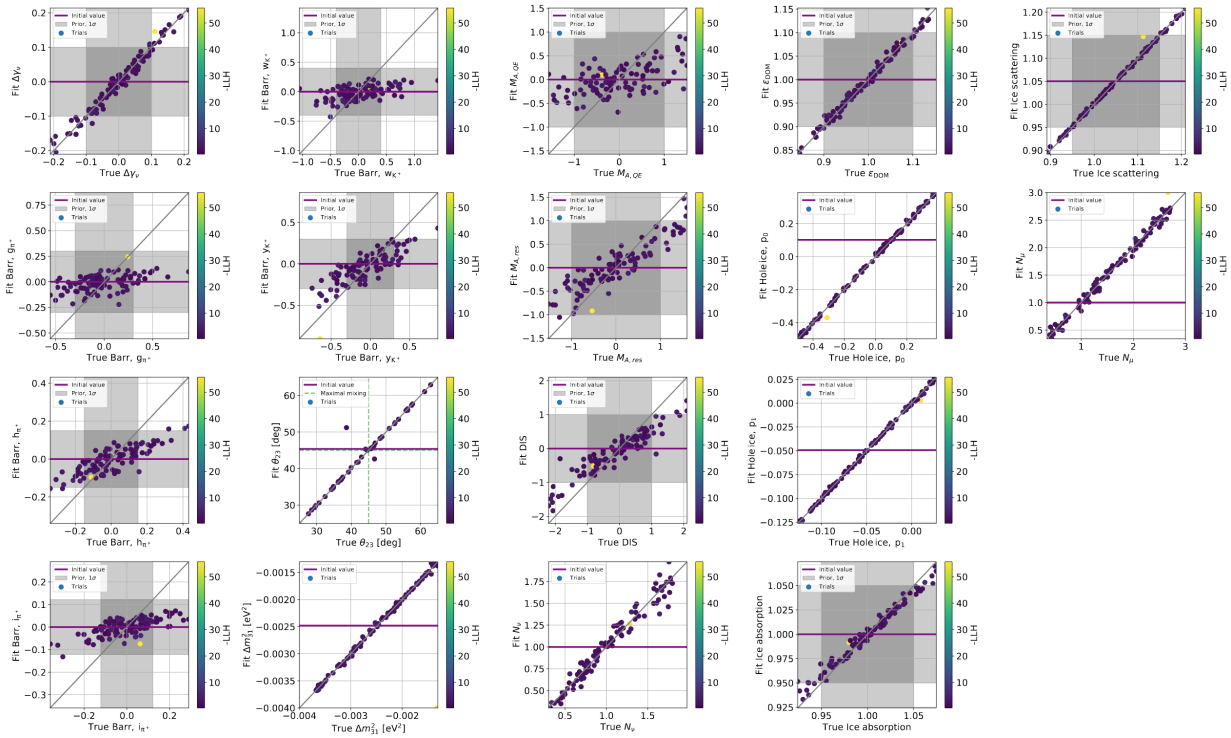
Test procedure:

1. Draw a single possible realization of the various combinations of physics and nuisance parameters, according to their given priors and ranges (i.e. Gaussian or uniform, as indicated in Table 4.2).
2. Re-weight the MC template to the chosen values of the physics and nuisance parameters, and provide it as pseudodata for the fit. (No statistical fluctuations are applied.)
3. Perform the fit to the pseudodata.
4. Repeat for many different realizations of the nuisance and physics parameters.
5. Check whether the recovered (i.e. fitted) values match the injected truth.

Outcome: Figure 5.3 shows the results of the parameter ensemble tests. The minimizer is able to recover the true parameters. For some of the parameters to which we only have a weak dependence (such as the Barr parameters and the axial mass terms), the injected value is only partially recovered due to the effects of the prior pulling the parameter towards its nominal value. In other words, for these parameters, the power of the data to constrain the parameter is comparable to the prior imposed.



(a) Normal Ordering



(b) Inverted Ordering

Figure 5.3: Outcome of parameter ensemble test for each ordering. Each subplot shows the injected parameters (on the x-axis) and the recovered parameters (on the y-axis), for 100 trials. In general, most parameters are able to be recovered. For some parameters (such as the Barr parameters and the axial mass terms), the fit value is influenced by the Gaussian prior.

5.4 Ensemble Test

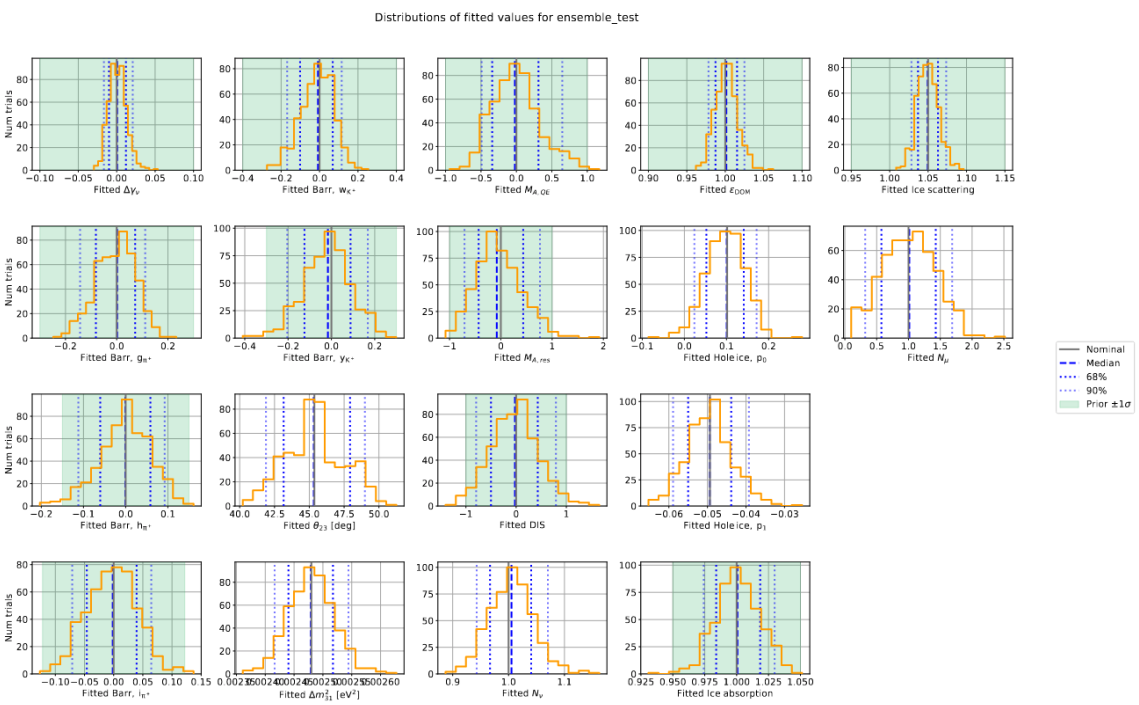
Purpose: Ensure that the analysis behaves correctly when statistical fluctuations are applied to the MC template.

Test procedure:

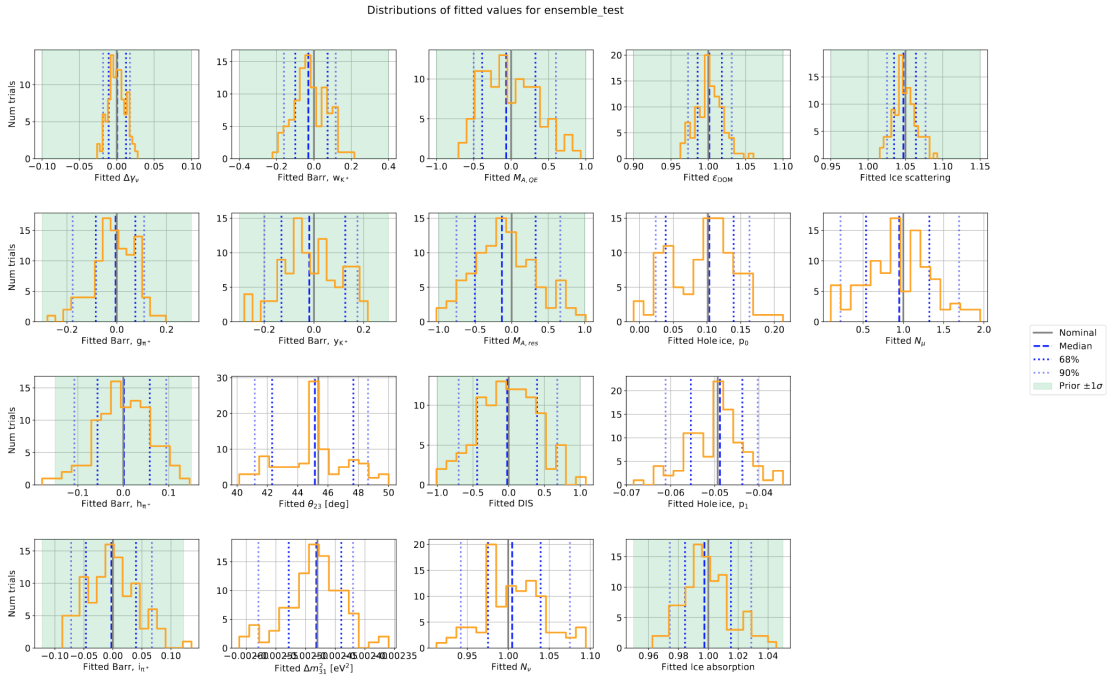
1. A number of pseudo-experiment templates are created by statistically fluctuating the bin counts of the nominal simulation template.

2. For each trial, fit the pseudo-data using the standard analysis procedure.

Outcome: Fig. 5.4 shows the distribution of fitted parameters resulting from the trials in this test. For all parameters, we can see that the distributions of fitted values (gold) are peaked around the true injected values (grey vertical line), as expected.



(a) Normal Ordering



(b) Inverted Ordering

Figure 5.4: Distribution of fitted parameters resulting from 500 pseudodata templates for normal ordering (100 trials for inverted ordering), generated by applying statistical fluctuations to the nominal template. The fitted values (gold) are peaked around the true injected values (grey vertical line), as expected. If a parameter has a Gaussian prior, then its 1σ band is shown by the green shaded region.

6

Blind Fits to Data

What you really learn in a life of science is the vastness of what we don't know.

– David Eagleman

As is customary in many scientific experiments, our analysis is performed in a blind way. This means that all analysis choices should be settled prior to revealing any information that could indicate (either directly or indirectly) what the results of the analysis are. In this case, the results refer to the preferred values of Δm^2_{32} and $\sin^2(\theta_{23})$. The previous chapters introduced how the analysis choices were made and how the analysis set-up is tested for robustness before a fit to data is ever performed. This chapter will discuss the next phase in which we actually perform the fit to data, but again refrain from revealing any information about the best fit values of Δm^2_{32} and $\sin^2(\theta_{23})$. We refer to this as a "Blind Fit" because, as the name suggests, we have performed a fit but still remain blinded to the physics parameters of interest.

As the epigraph of this chapter alluded to, the blind fits revealed that we do not yet know all there is to know about our data sample. This analysis has a factor of $\sim 3\times$ more events than any previous DeepCore oscillation analysis, and as we continue to increase the number of events and decrease the amount of statistical uncertainty present in our analyses, any potential systematic effects which had not been detectable in previous analyses can become important. This chapter will discuss the outcome of our blind fits and will touch on some of the investigation that followed.

6.1 Original Blind Fit

The blind fits are performed with a phased approach and the procedures must be approved in advance by members of the collaboration. We want to ensure that we receive enough information from the fit to determine if the results are trustworthy, but without looking at any information that could compromise the blindness of the study. The fits are run as follows:

1. **Run fits** - Fit both normal ordering and inverted ordering, but for blind fits do not report which one is preferred.

2. **Pseudotrials at best fit** - Generate many realizations of psuedodata assuming the Best Fit Point (BFP). To do this, we inject the preferred mass ordering and best fit parameter values and then apply statistical fluctuations to the simulated templates. These pseduodata templates are then fit using the standard procedure. (These scripts are automated so we don't view the best fit parameters.)

Nuisance Pulls

We begin by looking at how far the *nuisance* parameters are pulling from their priors. We set a soft boundary at 2σ and say that if we get any parameter beyond two sigma, we will pause and undergo further investigation. This might seem too strict at first glance given that we have 12 parameters with Gaussian priors, but we find this to be a reasonable choice for a few reasons. These parameters have constraints placed on them by reliable external data, and therefore values of the nuisance parameters that contradict the external data would warrant further investigation anyway. Additionally, we could naively calculate how often we expect to get "unlucky" and randomly get a parameter beyond 2σ given that we have 12 parameters, but in practice, we expect a parameter to "randomly" fall beyond 2σ far less than this calculation yields. From Figures 5.3 and 5.4, we see that in our MC simulation we have only weak sensitivity to some parameters, and that the prior is more constraining than the data for those parameters. We see that the simulated trials fall within the shaded 1σ bands far more than 68% of the time. In fact, it can be up to 90-100% of the time for some parameters. This tells us that regardless of the true value or statistical fluctuations, we expect a parameter to fall outside 2σ quite rarely.

When we observe the nuisance pulls for our blind fit (see Figure 6.1), we find that two parameters are pulling beyond 2σ. Additionally we find that two more parameters, Barr g and h, are pulling beyond 1.5σ even though our earlier tests showed that we generally always expect those parameters to fit back within 1σ (as seen in Figure 5.3). This tells us that there is something unexpected occurring in the data that we are not able to reproduce with statistical fluctuations or by varying the nuisance parameters in our simulation.

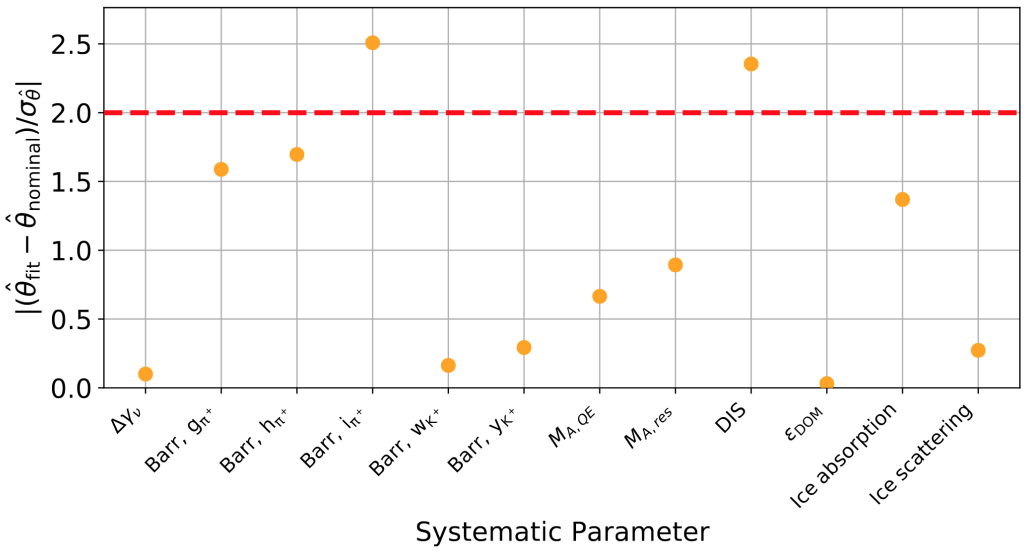


Figure 6.1: Parameter pulls for systematic nuisance parameters for the blind fit. Two parameters fall beyond the $> 2\sigma$ threshold.

Goodness-of-fit

While the nuisance pulls in the previous section suggest that there may be something occurring in data that we didn't anticipate from the MC simulation, it doesn't necessarily provide the full picture. The next metric we evaluate uses the pseudotrials at the best fit point (BFP) to examine the goodness-of-fit. We assume that the BFP is truly the correct underlying combination of physics and nuisance parameters, and evaluate what range of LLHs would be reasonable given our statistical uncertainty. As defined in Section 4.3, values closer to 0 indicate that the (pseudo)data is well described by the template, and values farther from 0 indicate that the (psuedo)data is not described well by the template.

When we plot this distribution comparing the observed and expected LLH values, we find that observed LLH value is far worse than the expected LLH values (see Figure 6.2). This indicates that our simulation, even with the extreme pulls in nuisance parameters, is not describing the data template well.

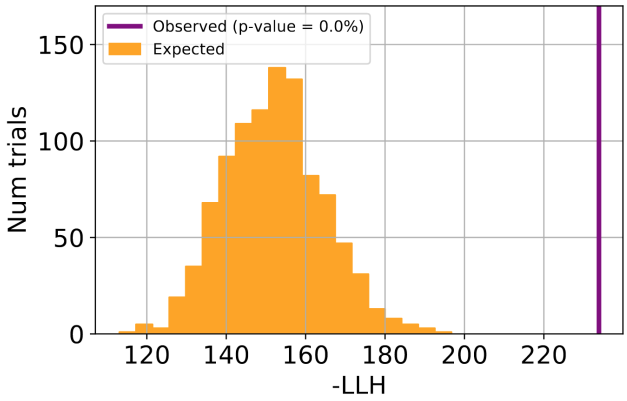


Figure 6.2: Outcome of goodness-of-fit test. The yellow distribution represents the expected likelihood distribution for psuedotrials generated at the BFP with statistical fluctuations applied. The purple line indicates the observed likelihood value. We see the observed value is not consistent with statistcal fluctautions.

Minimization checks

An obvious question to ask next is whether or not the minimizer has gotten stuck. Although we perform tests of the minimizer ahead of time across a range of physics and nuisance parameters to make sure the minimizer is capable of handling many values, it could be possible that in the real data fit, it either fails right away or fails to converge after a long time. We use our pseudotrials mentioned in previous sections of this chapter to determine what a "reasonable" number of fit iterations is. As shown in Figure 6.3, almost all of the trials fall between 1,000–3,000 iterations, so we check whether the real data fit also finished within the range of 1,000–3,000, and we find that it did. This indicates that it is likely not a catastrophic failure of the minimizer.

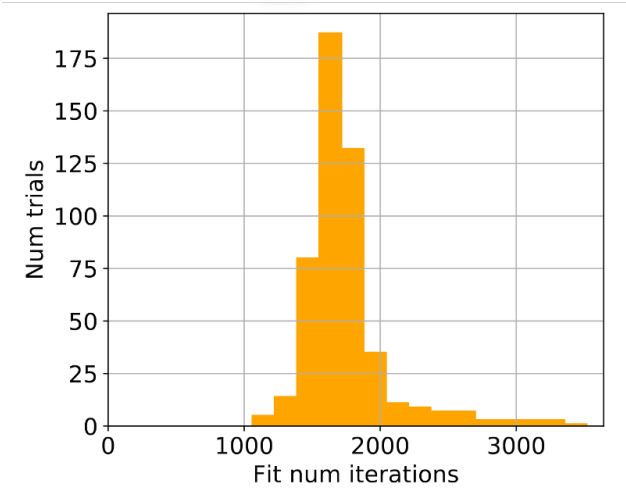


Figure 6.3: Expected number of fit iterations based on many pseudotrials. We confirmed that the fit to real data also converged within 1,000–3,000 iterations, which is in line with this expectation.

6.2 Further Investigations

Based on our knowledge of the parameters with extreme pulls and knowledge of issues that had been encountered in prior DeepCore studies, there were several possible leads that warranted further investigation.

Removing high energy bins

The parameters that pulled to extreme values in the initial fit, DIS CSMS and the Barr parameter corresponding to the I block, are both relevant for describing the high energy events in our sample ($> \sim$100 GeV). Furthermore, the high energy region has historically been excluded from analyses; previous analyses like [7, 8] only used reconstructed energies up to 56 GeV. This motivated a study to determine if it was a discrepancy at high energies causing the effect. To test this, we removed the high energy bins and performed the blind fit again (see Figure 6.4.) Unfortunately, this did not resolve our tension as we still see pulls $> 2\sigma$.

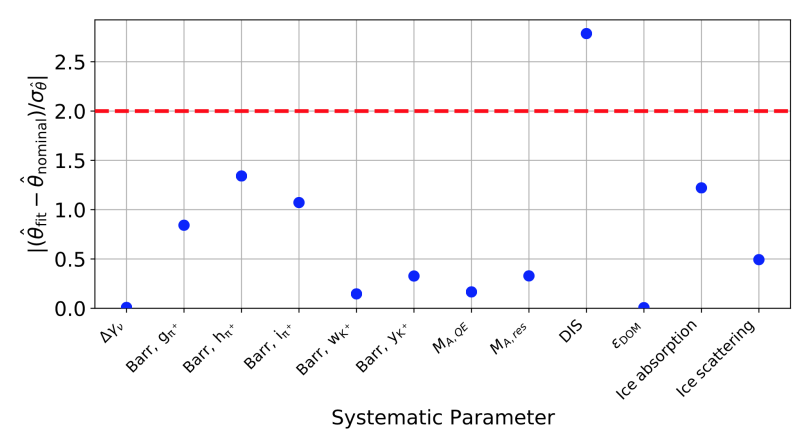


Figure 6.4: Nuisance parameter pulls for blind fit with high energy bins removed.

Removing downgoing bins

Similarly, there have historically been concerns in IceCube analyses about the atmospheric muon background. As mentioned in previous chapters, the muons are largely downgoing and horizontal. The reconstructed zenith distribution of muons was shown in Figure 4.1a. We tested a more restrictive cut on $\cos(\theta_{zenith})$ to test this muon hypothesis, but again found that this did not resolve the tensions in the fit. See Figure 6.5.

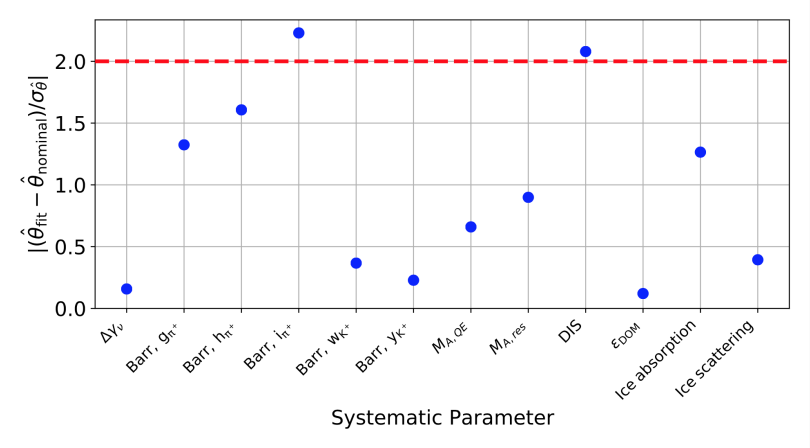


Figure 6.5: Nuisance parameter pulls for blind fit with horizontal/downgoing bins removed.

Independent PID fits

To narrow down where the discrepancies are occuring, we performed the fit treating each individual PID bin separately. So rather than having one set of nuisance parameters that is the best fit across the entire sample, each PID bin is allowed to dictate its own preferred value. The results of this test are shown in Figure 6.6. Since the nuisance parameter pulls pass the 2σ threshold at this point, we also run the ensemble tests to determine the goodness-of-fit for each PID bin's fit. See Figure 6.7. We see that by loosening this constraint, the disagreement between data and MC is improved and the observed test statistic is now consistent with the expected distribution. We see that the low-purity PID bin still has a nuisance parameter pulling quite close to 2σ, but for all the PID bins, the goodness-of-fit is consistent with the expected distributions. This indicates the tension in the full fit could be arising from trying to describe multiple particle types at once, or from a yet-to-be-identified systematic effect that affects different event types in different ways.

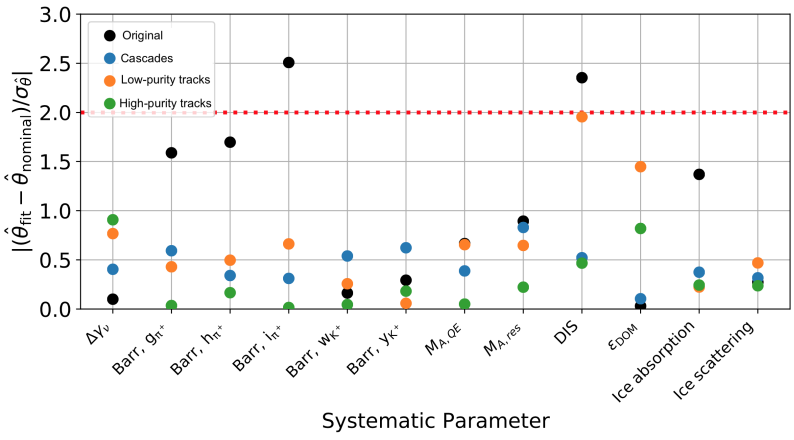
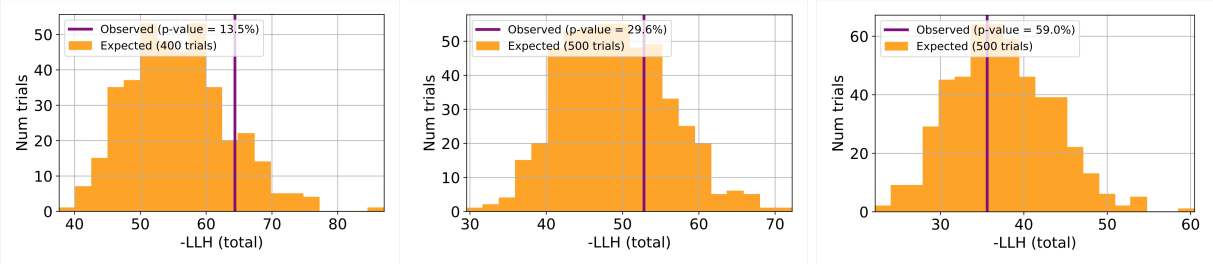


Figure 6.6: Nuisance parameter pulls for blind fits where each PID bin fit independently.



(a) Cascades (b) Low-purity tracks (c) High-purity tracks

Figure 6.7: Goodness-of-fit tests for the blind fits where each PID bin is fit independently.

Treatment of Systematic Uncertainties

One of our largest sources of systematic uncertainty, as mentioned previously, is the ice properties. Cascade events are generally more susceptible to ice systematics than the other event types. The calibration program in IceCube is very active, and as such, is always working to further our understanding and improve our ice models. There have been new updates to the ice models in recent years, such as a model that takes into account the birefringence of light due to the ice's crystalline structure [37]. This model corrects a known discrepancy in the azimuthal data/MC agreement, but since our analysis does not use azimuth, this should not have a large effect to first-order. However, this model could have second-order effects that do impact our observable parameters. Our analysis was underway before this model became available, so it was not originally included, but there is currently an effort to include this ice model in our analysis. This would allow us to determine if the birefringence modeling is the missing piece necessary for our model to describe our data well.

Also under investigation are the flux and cross section systematics. The parameters that generated the large pulls in the original fit are new parameters that were not included in previous generation analyses. Therefore, it could be possible that the new parameterization or implementation are flawed. There are still many open avenues to explore, but is very likely that whatever the resolution may be, we will learn something new about our detector and/or our models.

7

20% Sub-sample Fit

Sometimes the half is better than the whole.

– Hesiod, ancient Greek poet

7.1 Defining a Sub-sample

Due to the unresolved issues with the full sample's fit, it became necessary to move forward with only a sub-sample for the remainder of this work. IceCube collects data in increments of ∼8 hours, which are known as "runs". Each run is given an associated incremental number known as a "Run ID". The sub-sample used in this chapter uses only runs with Run IDs ending in 0 and 5. This ensures that the sub-sample is an accurate representation of the full sample spanning all data years. All other parts of the analysis remain the same as introduced in earlier chapters. Figure 7.1 shows the expected event distribution for the sub-sample.

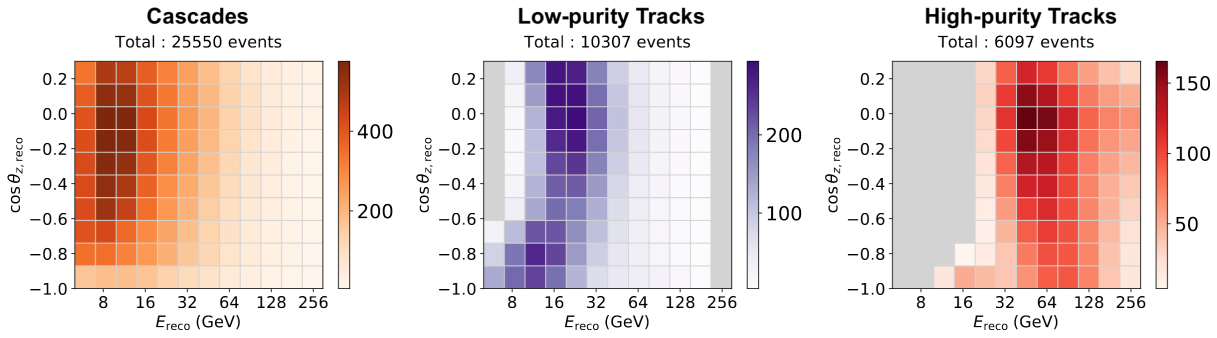


Figure 7.1: Expected nominal event distribution for the 20% sub-sample.

7.2 Pre-fit Tests

Before performing the fit, all pre-fit tests described for the full sample were repeated with the sub-sample to evaluate the performance of the analysis with the reduced statistics. This section shows the outcome of these tests; for the description and procedures associated with each test, refer to Sections 5.1–5.4.

Asimov Test (inject-recover test with no fluctuations)

Figure 7.2 shows the result of an inject-recover test for various injected values of the physics parameters. For every injected combination of physics parameters, the minimizer recovers the truth. This can be seen in the figure as the fitted values (purple x's) falling exactly on top of the true inject parameters (yellow circles).

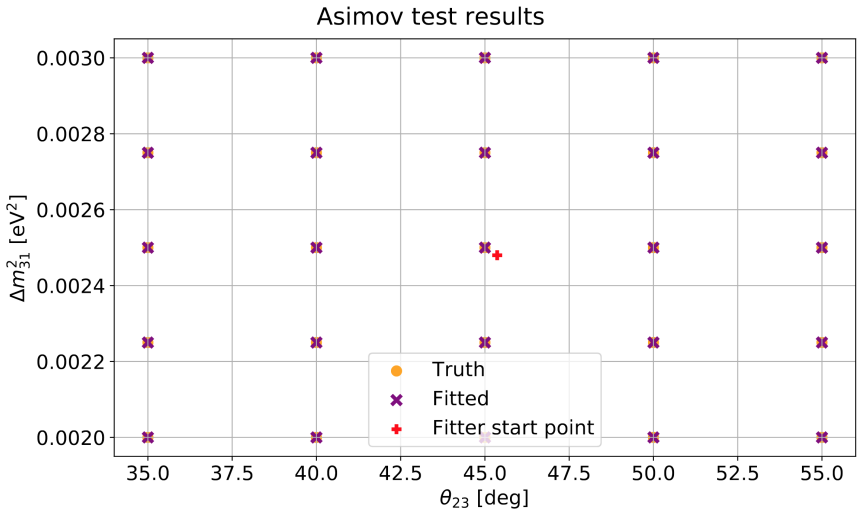


Figure 7.2: Results of the inject-recover test on Asimov pseudodata for the sub-sample. For each of the scan points shown in yellow, a fit was performed where the pseudodata is generated with that combination of true physics parameter values. In every case, the analysis is able to recover the correct values of the true physics parameters.

Parameter Ensemble Test (inject-recover test with nuisance param. fluctuations)

Figure 7.3 shows the result of a "parameter ensemble test" in which the physics and nuisance parameters are varied and injected as pseudodata. In general, the minimizer recovers the truth values. This can be seen in the figure by the trials falling approximately on the 1:1 line. This test tells us how well the analysis recovers off-nominal values of the nuisance parameters. For some parameters that we only have weak sensitivity too, the fit value is influenced by the Gaussian prior, indicated by the grey shaded region.

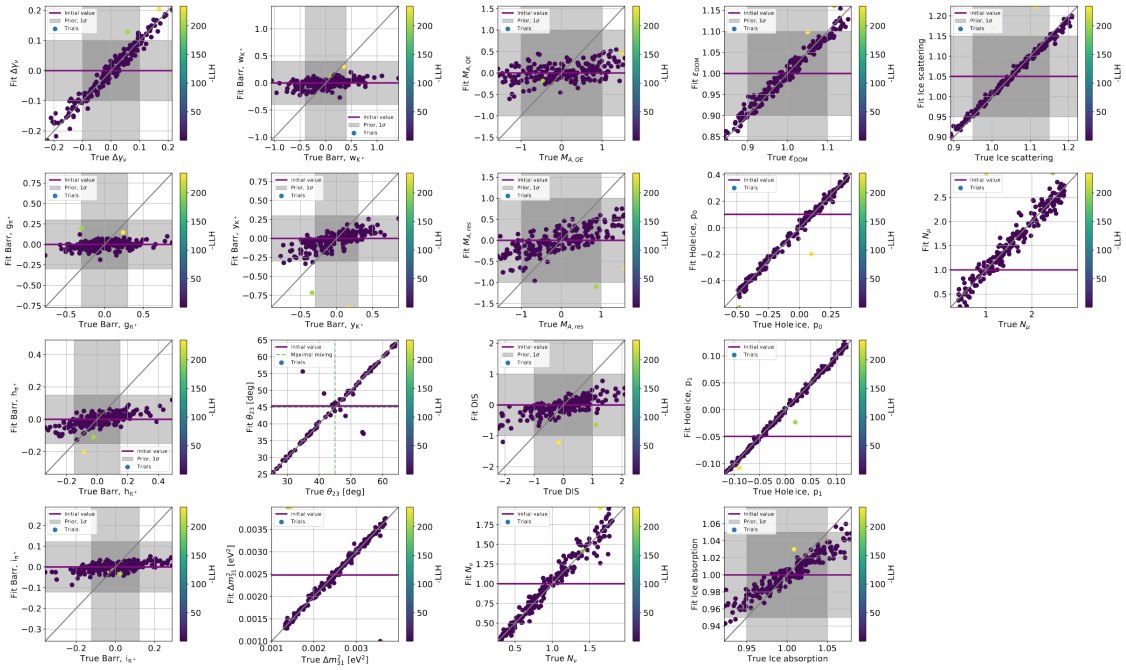


Figure 7.3: Outcome of parameter ensemble test for the 20% sub-sample. Each subplot shows the injected parameters (on the x-axis) and the recovered parameters (on the y-axis), for many trials. In general, most parameters are able to be recovered. For some parameters, the fit value is influenced by the Gaussian prior.

Ensemble Test (inject-recover test with statistical fluctuations)

Figure 7.4 shows the result of an "ensemble test" in which statistical fluctuations are applied and injected as pseudodata. Each sub-figure shows the distribution of fitted values for a given physics or nuisance parameter. For most parameters, we see that the distributions of fitted values (gold) are peaked around the true injected values (grey vertical line), as expected.

One notable exception is the muon normalization N_μ which represents a scaling factor on the amount of muon background events present in the sample, with 1 corresponding to the nominal Monte Carlo prediction. We see that although the true value for this parameter is 1 in all the trials, the fit prefers values close to 0 when statistical fluctuations are applied. This differs from the outcome we saw with the full sample where the true value of 1 was recovered more accurately (see Figure 5.4). In a standalone analysis, this may have warranted further investigation, however in order to keep this analysis as similar to the full sample analysis as possible, no changes were made. This behavior is not particularly worrisome due to the very low muon content in the sample ($< 1\%$). Figure 7.5 shows the nominal expected distribution of background muons in the 20% sub-sample, and we see that most of the analysis bins have on the order of single events with at most about 6 expected events. Compare to Figure 7.1 which shows the expected templates for all events in the sample, which have $\mathcal{O}(100\text{s})$ of events per bin.

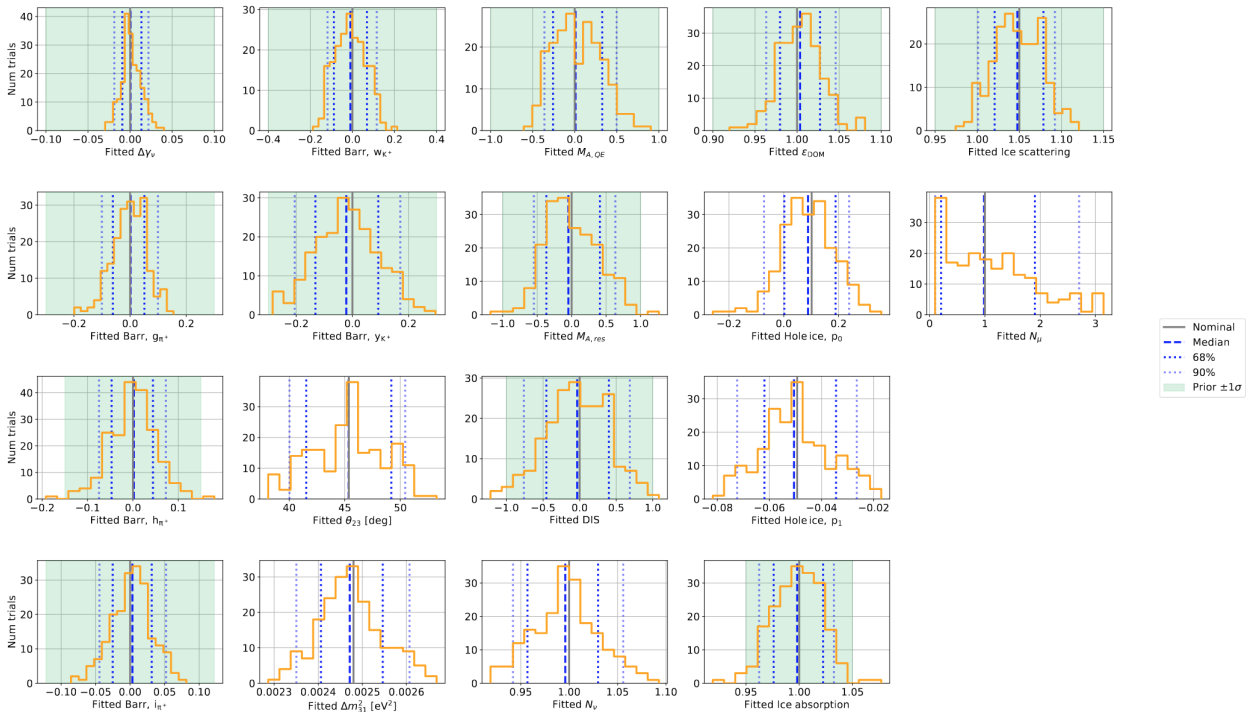


Figure 7.4: Distribution of fitted parameters resulting from many pseudodata templates, generated by applying statistical fluctuations to the nominal template. If a parameter has a Gaussian prior, then its 1σ band is shown by the green shaded region.

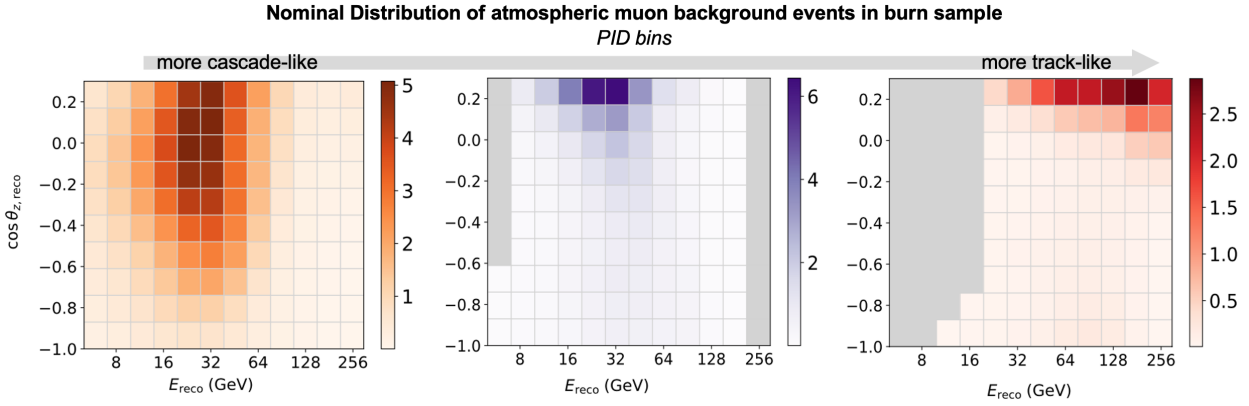


Figure 7.5: Expected nominal distribution of muons for the 20% sub-sample. All bins have very few expected events.

7.3 Blind Fits

The blind fit procedure introduced in Section 6.1 was repeated using this 20% sub-sample. First, the fits are run for both mass orderings. Then, that BFP is used to generate many pseudotrials.

Nuisance Parameter Pulls & Goodness-of-fit

We begin by looking at the same information as before: the nuisance parameter pulls and the goodness-of-fit. With this sub-sample, we find that no nuisance parameters are pulling beyond the pre-defined 2σ threshold, as shown in Figure 7.6. Additionally, our observed test statistic is in good agreement with the the expected test statistic distribution, as shown in Figure 7.7. This indicates that our model is describing our data to a level consistent with statistical fluctuations.

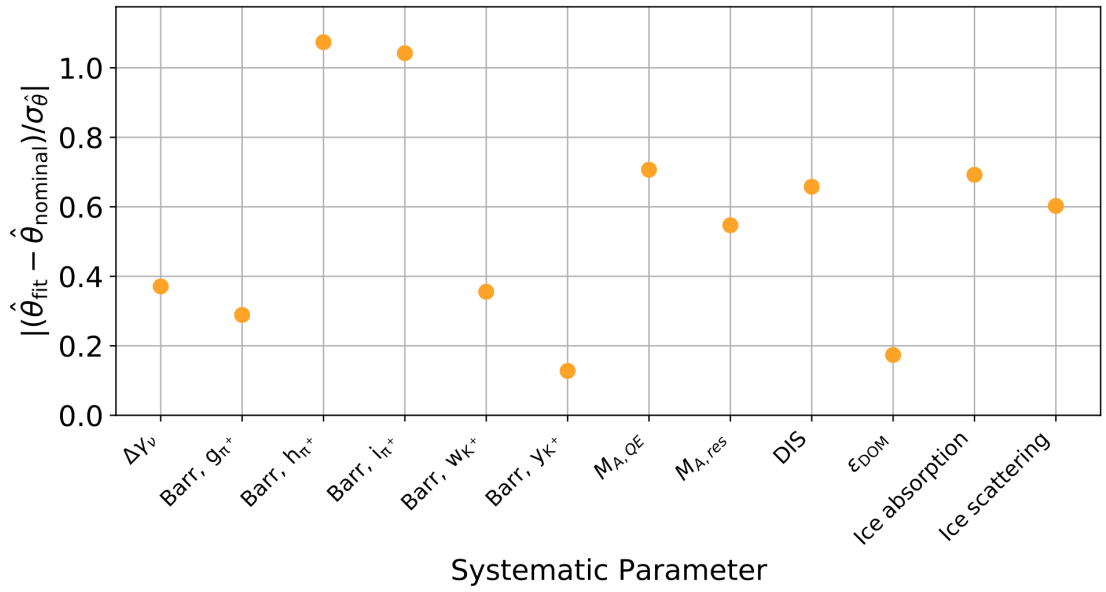


Figure 7.6: Parameter pulls for systematic nuisance parameters for the sub-sample blind fit.

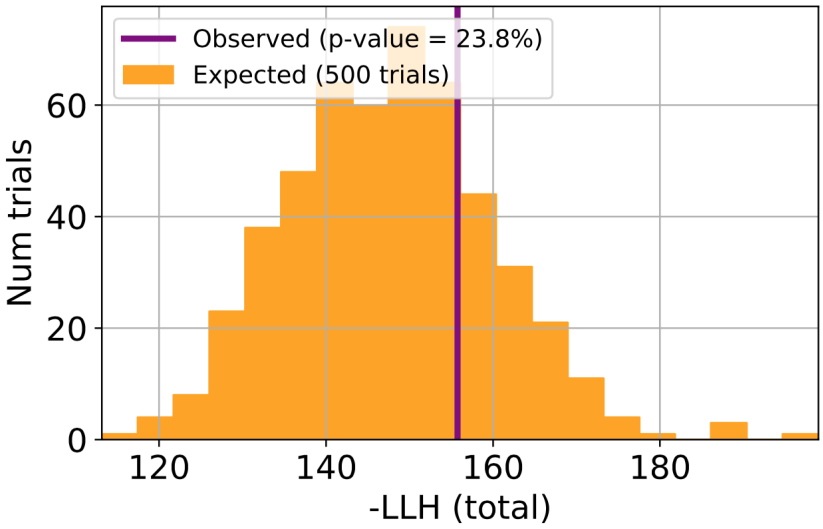
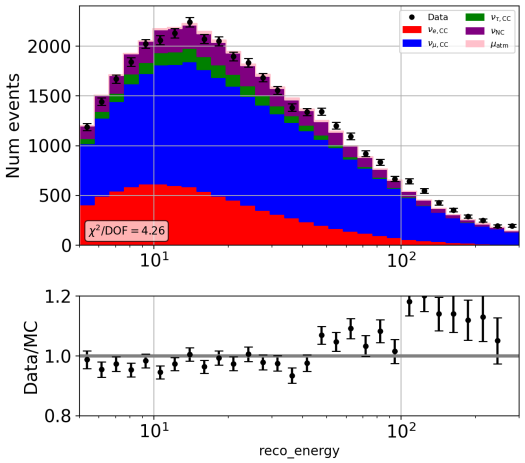


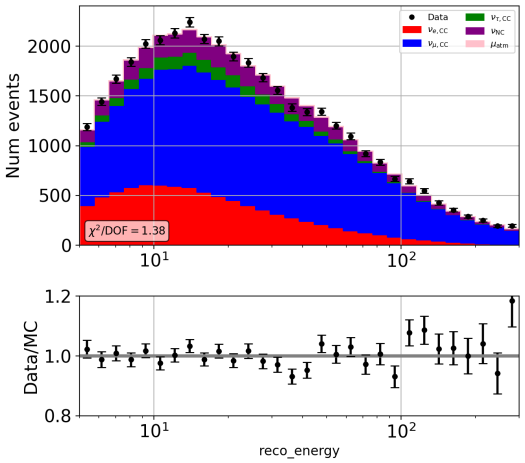
Figure 7.7: Outcome of goodness-of-fit test. The yellow distribution represents the expected likelihood distribution for 500 psuedotrials at the BFP with statistical fluctuations applied. The purple line indicates the observed likelihood value.

Data/MC Agreement in 1-dimensional projections

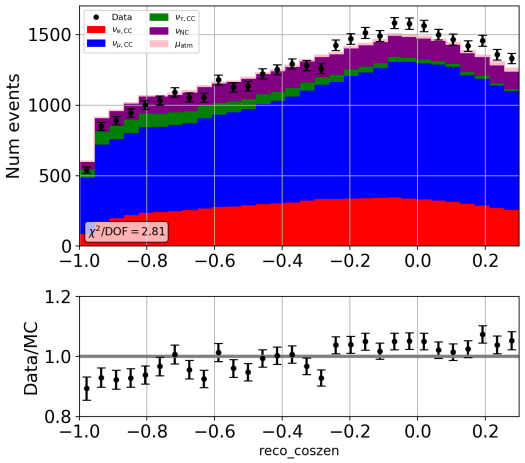
After these positive indications that the fit appears to be successfully describing our data, we move on to some additional checks to again evaluate how well the fit has performed, without compromising any information that could indicate where the best fit physics parameters lie. The oscillation pattern that defines our signal is visible in the energy-cos(θ_{zenith}) plane for the track-like PID bins (as shown in Figure 1.7). Because we need to slice our events in 3 dimensions to observe the oscillation pattern, it is generally accepted that conclusions about the oscillation values cannot be drawn from 1-dimensional projections. Therefore we are able to look at the data/MC agreement pre- and post-fit in the blind stage. Figure 7.8 shows the pre- and post-fit data/MC agreement for reconstructed energy, cos(θ_{zenith}), and the output BDT score from the PID classifier. We see that the fit corrects the slight tilts in both energy and cos(θ_{zenith}).



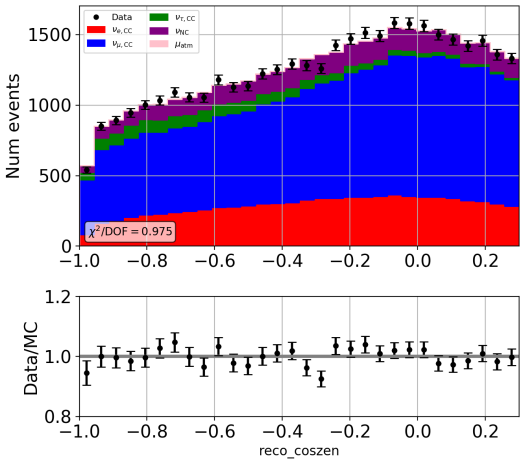
(a) Pre-fit reconstructed energy



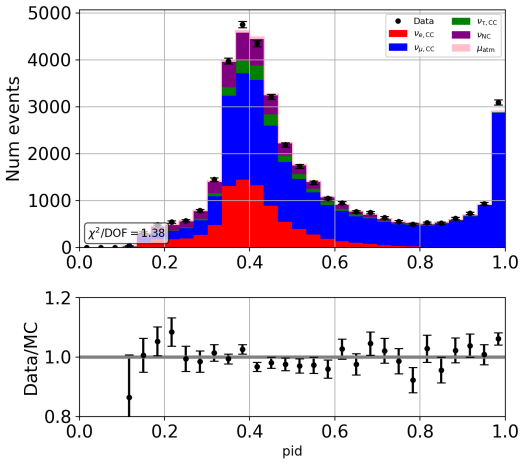
(b) Post-fit reconstructed energy



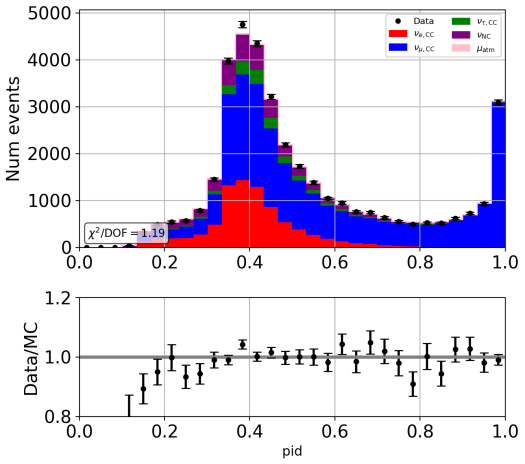
(c) Pre-fit reconstructed $\cos(\theta_{zenith})$



(d) Post-fit reconstructed $\cos(\theta_{zenith})$



(e) Pre-fit PID score



(f) Post-fit PID score

Figure 7.8: Data/MC agreement for each of our observable variables before (left) and after (right) the fit.

Evaluation of binned likelihood

Because the fit is performed using a binned likelihood, we can also look at the contributions from the individual bins to determine if they are distributed in agreement with statistical fluctuations. To do this, we compare the best fit MC template with the data template. For each bin, we compare the difference between these two templates to the total statistical uncertainty (the data statistical uncertainty is defined as $\sqrt{N}$ and the MC statistical uncertainty is calculated from the MC weights). If the bins are perfectly statistically distributed, we expect the distribution to follow a Normal distribution. Figure 7.9 shows the distribution of the bins, along with a Gaussian fit. We find that the Gaussian fit yields parameters quite close to $\mu = 0$ and $\sigma = 1$.

Next, we look at a map of the LLH values for each bin to look for any potential clustering. See Figure 7.10. The entry for each bin corresponds to the log of the Poisson likelihood introduced in Equation 4.2. This value is quantifying the probability that the number of observed data events in a given bin is a plausible realization, if the best fit point is assumed to be true. We can look for regions that have many unlikely realizations, which could indicate that there is a region of phase space that is not being modeled correctly or that there is potentially an unsimulated background in a certain region. Looking at Figure 7.10, we see that there are not regions which have significant clustering. In conjunction with Figure 7.9, we conclude that any variations between the data template and the MC template seem consistent with statistical fluctuations.

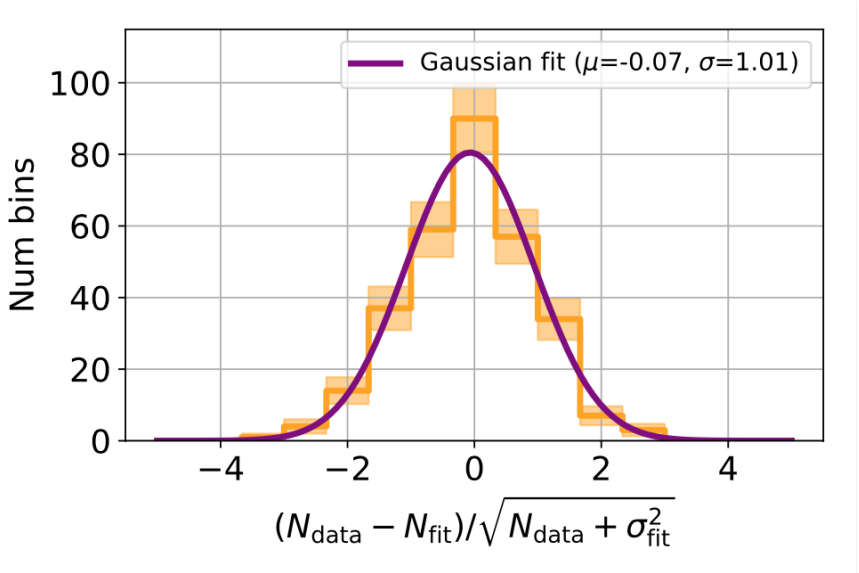


Figure 7.9: Distribution comparing the data/MC agreement (gold). Each entry in the histogram corresponds to an individual bin in the 3d templates. The purple line shows a Gaussian fit, which we see has parameters quite close to a normal distribution.

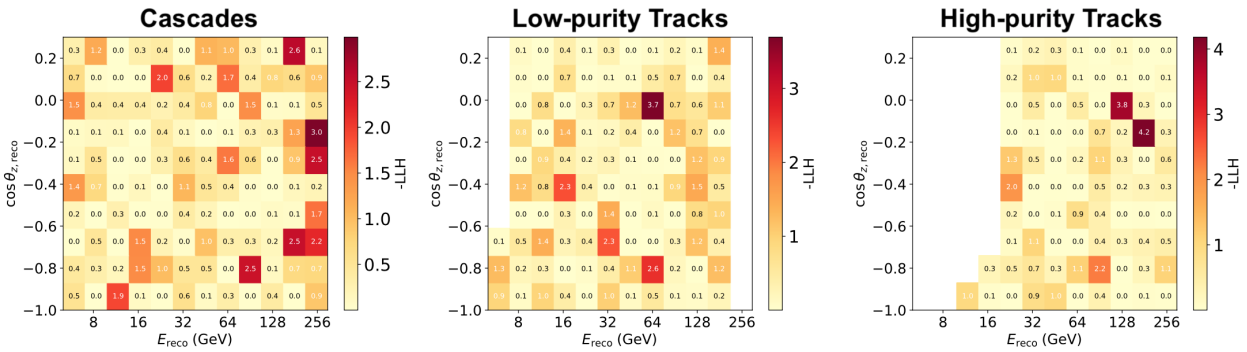


Figure 7.10: Contribution of each bin to the overall test statistic. A value near 0 indicates that the number of data events is a likely Poisson realization given the expected number of events from the best fit MC; A value far from 0 indicates that the number of data events is an unlikely Poisson realization given the expected number of MC events. We see that there is no strong evidence of clustering.

7.4 Results

Best fit parameter values

After determining that everything seen in the blind fit phase is satisfactory, we proceed with unblinding all parameters. The best fit values for all the fitted parameters are shown in Table 7.1. The best fit values of the atmospheric oscillation parameters of interest are $\theta_{23} = 49.26°$ and $\Delta m^2_{31} = 2.483110^{-3}$ eV^2, which corresponds to $\sin^2(\theta_{23}) = 0.5742$ and $\Delta m^2_{32} = 2.4092 \times 10^{-3}$ eV^2, as shown in Table 7.2.

Category	Param	Nominal	Prior	Range	Best Fit
Flux	$\Delta\gamma_\nu$	0	gaussian, $\sigma = \pm 0.1$	[-0.5, 0.5]	0.0371
Flux	Barr, g_π^+	0	gaussian, $\sigma = \pm 0.3$	[-1.5, 1.5]	-0.0866
Flux	Barr, h_π^+	0	gaussian, $\sigma = \pm 0.15$	[-0.75, 0.75]	-0.1611
Flux	Barr, i_π^+	0	gaussian, $\sigma = \pm 0.122$	[-0.61, 0.61]	0.1271
Flux	Barr, w_K^+	0	gaussian, $\sigma = \pm 0.4$	[-2, 2]	0.1422
Flux	Barr, y_K^+	0	gaussian, $\sigma = \pm 0.3$	[-1.5, 1.5]	-0.0383
Cross-sections	$M_{A,QE}$	0	gaussian, $\sigma = \pm 1$	[-2, 2]	0.7068
Cross-sections	$M_{A,res}$	0	gaussian, $\sigma = \pm 1$	[-2, 2]	0.5469
Cross-sections	DIS	0	gaussian, $\sigma = \pm 1$	[-3, 3]	0.6578
Normalization	N_ν	1	uniform	[0.1, 2]	0.9294
Normalization	N_μ	1	uniform	[0.1, 3]	0.1
Detector/Ice	DOM Efficiency	1	gaussian, $\sigma = \pm 0.1$	[0.8, 1.2]	1.017
Detector/Ice	Hole ice, p_0	0.102	uniform	[-0.6, 0.5]	0.2042
Detector/Ice	Hole ice, p_1	-0.0493	uniform	[-0.15, 0.05]	-0.1025
Detector/Ice	Ice absorption	1	gaussian, $\sigma = \pm 0.05$	[0.9, 1.1]	0.9654
Detector/Ice	Ice scattering	1.05	gaussian, $\sigma = \pm 0.1$	[0.85, 1.25]	1.1103
Oscillations	θ_{23} [deg]	45.4	uniform	[0, 90]	49.26
Oscillations	Δm^2_{31} [10^{-3}eV^2]	2.48	uniform	[1.0, 5.0]	2.4831

Table 7.1: Table of best fit parameter values for nuisance and physics parameters.

Parameter	Best Fit
$\sin^2(\theta_{23})$	0.5742
Δm^2_{32}	2.4092 10^{-3}eV^2

Table 7.2: Results for the atmospheric neutrino oscillation parameters, converted into $\sin^2(\theta_{23})$ and Δm^2_{32}.

90% confidence allowed region & comparison to other experimental results

Finally, we generate our 90% confidence level contour using a procedure similar to the sensitivity test described in Section 5.2. Scanning over a grid of $\sin^2(\theta_{23})$ and Δm^2_{32} values, the template is fixed to those values, the data is fit, and then we determine the difference in test statistic between the best fit point and the scan point. Assuming Wilks' theorem (see Section 7.5 for testing the validity of this assumption), we can determine what test statistic value corresponds to the 90% confidence level, and then draw a contour at that level. The contour is shown in Figure 7.11, along with the 90% result contours for several other global neutrino experiments [33–36].

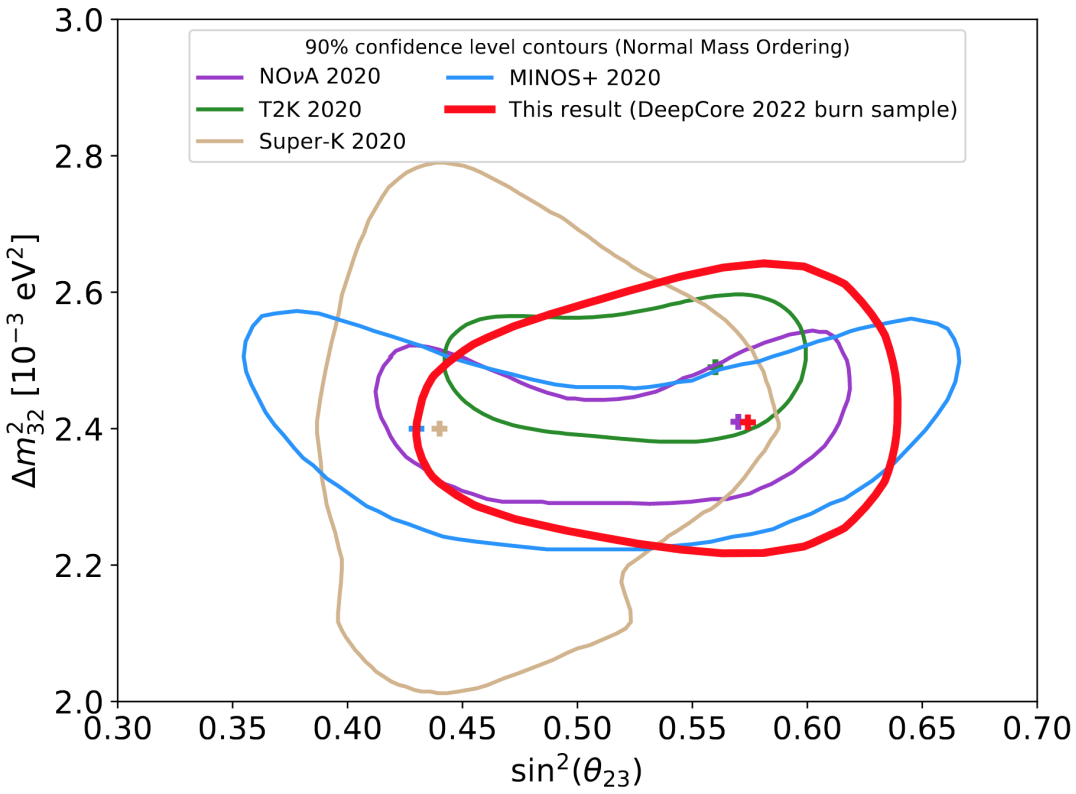


Figure 7.11: Contour showing the 90% confidence level allowed region for the physics parameters, along with results from several other global neutrino experiments. The crosses show the best fit values from each experiment. Other experimental results from [33–36].

7.5 Post-unblinding Checks

Examining best fit detector systematic parameters

As mentioned in Section 4.2, the detector systematic parameters are simulated in discrete sets and then interpolated between the discrete sets. To ensure that the best fit point (BFP) is in the interpolation region (rather than being an extrapolation), we investigate how the BFP compares to the simulated sets. Figure 7.12 shows the values for hole ice p_0 and p_1 for the nominal set, best fit point, and simulated systematic sets. We see that the best fit point is not more extreme than the simulated sets in either parameter. Figure 7.13 shows the angular acceptance as a function of the incident photon angle η on the DOM for the nominal set, the best fit point, and the simulated systematic sets. Similarly for the bulk ice, Figure 7.14 shows the values for bulk ice scattering and absorption for the nominal set, best fit point, and simulated systematic sets. We again see that the best fit point is not more extreme than the simulated sets in either parameter.

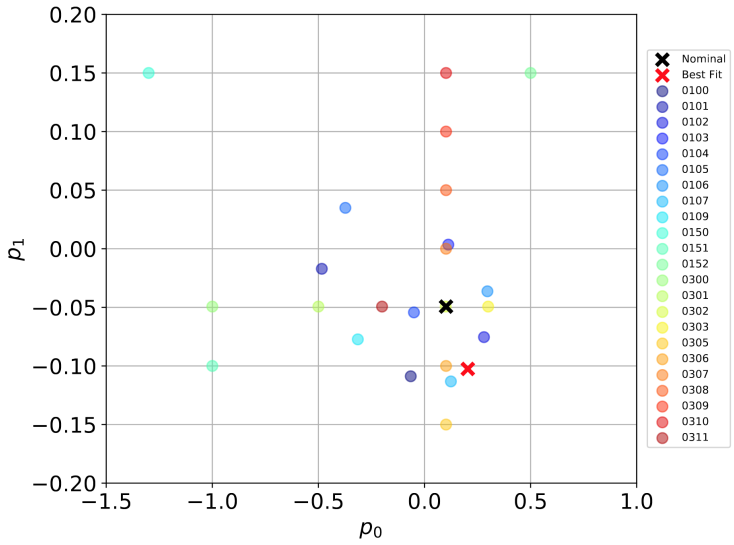


Figure 7.12: Values of hole ice systematic parameters p_0 and p_1 for the nominal set (black x), best fit point (red x), and various simulated systematic sets (colored circles).

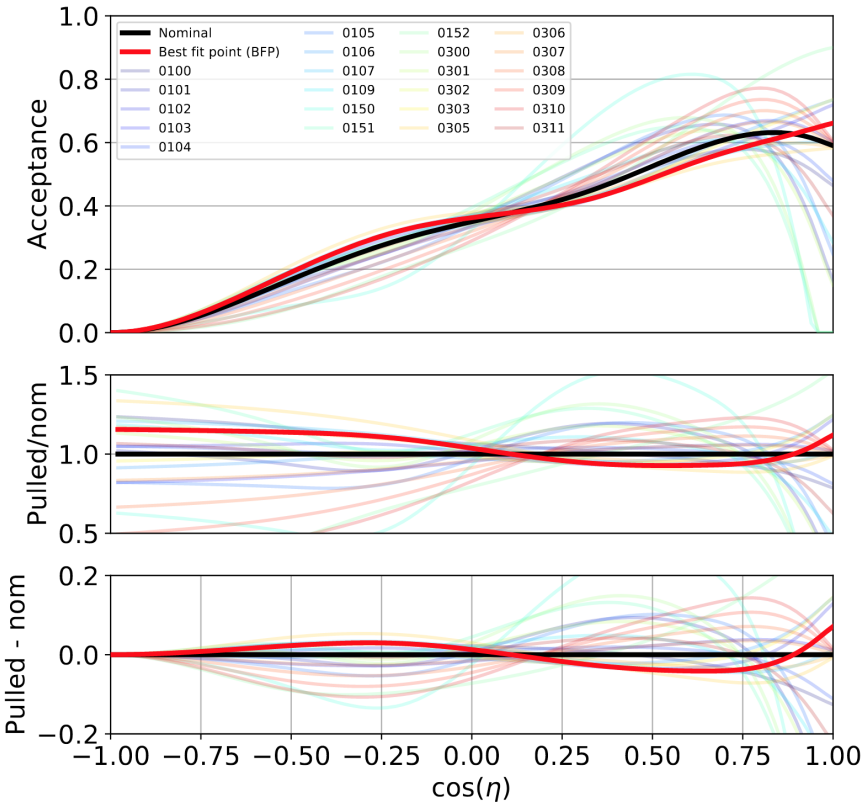


Figure 7.13: The angular acceptance as a function of the incident photon angle on the DOM for the nominal set (black), the best fit point (red), and the simulated systematic sets (various colors). cos(η)=1.0 corresponds to a photon hitting the face of the PMT, cos(η)=-1.0 corresponds to a photon hitting the top of a DOM opposite from the PMT face.

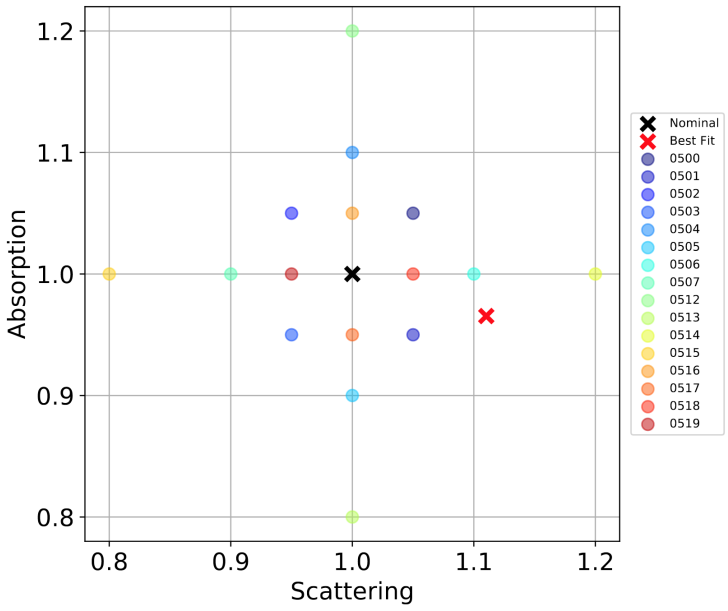


Figure 7.14: Values of bulk ice systematic parameters scattering and absorption for the nominal set (black x), best fit point (red x), and various simulated systematic sets (colored circles).

Data/MC Agreement for L/E

As originally shown back in Equation 1.4, the primary variable driving the oscillations is L/E. We examined the data/MC agreement of L and E (using $\cos(\theta_{zenith})$ as a proxy for length) after the fit in Figures 7.8b and 7.8d. We can use that same information to determine the distribution of the derived quantity L/E after the fit. Figure 7.15 shows the post-fit data/MC agreement, along with the no oscillation comparison. We find that this variable has good agreement between data and simulations after the fit. Additionally, we see that the oscillations have a large effect on the distribution, particularly in the range 100–1000 km/GeV.

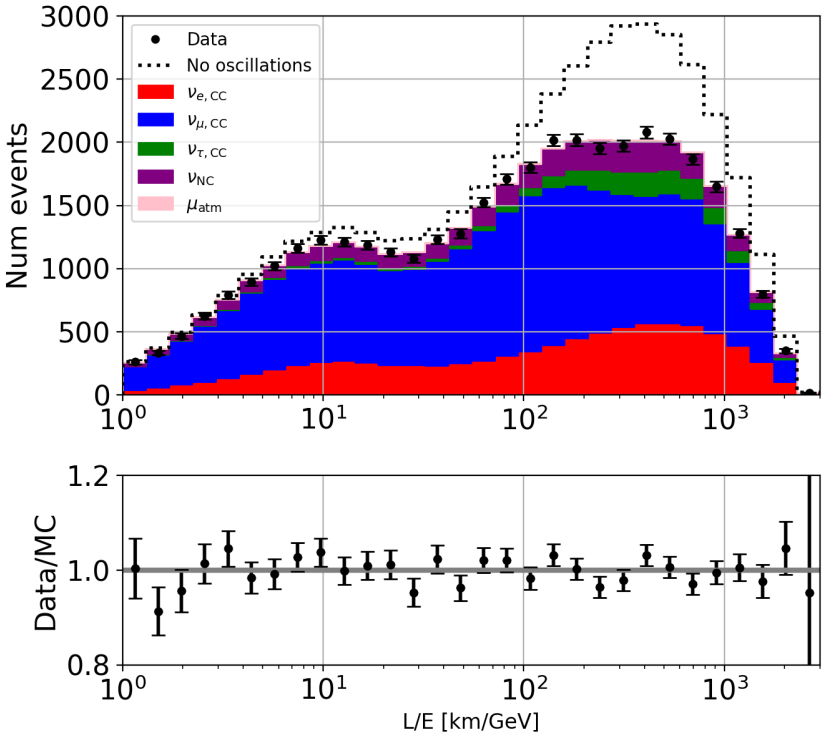


Figure 7.15: Post-fit data/MC agreement, along with the no oscillation comparison (dotted), for L/E which is the characteristic variable that the oscillations are sensitive to.

Assumption of Wilks' Theorem

As mentioned in Section 7.4, we assumed Wilks' theorem for determining the 90% confidence level contour. Wilks' theorem states that for a sufficiently large sample size, if the null hypothesis is true, then the distribution of the test statistic $-2\Delta LLH$ will approach the chi-squared χ^2 distribution for the number of degrees of freedom equal to the difference in the number of degrees of freedom between the null hypothesis and the hypothesis being tested [32]. To test the validity of this assumption, we use the pseudotrials generated as part of the blind fit stage (used in Section 7.3) to compare the observed $-2\Delta LLH$ distribution to the χ^2 distribution as suggested by Wilks' theorem. See Figure 7.16. Because the gold line is slightly above the 1:1 line, this indicates that we have slight overcoverage. This means that the contour reported in Section 7.4 is conservative. In other words, we expect 90% of pseudotrials to fall below the critical value, but there are actually more than 90% of trials below the threshold.

Given that this sub-sample fit is intended to be testing the validity of our analysis procedure, this conservative coverage is determined to be adequate. Once the fit with the full sample is successfully unblinded, if the trials indicate that there is still overcoverage, it would be appropriate to perform Feldman-Cousins corrections [38] on the result contour to ensure the accuracy of the stated 90% interval.

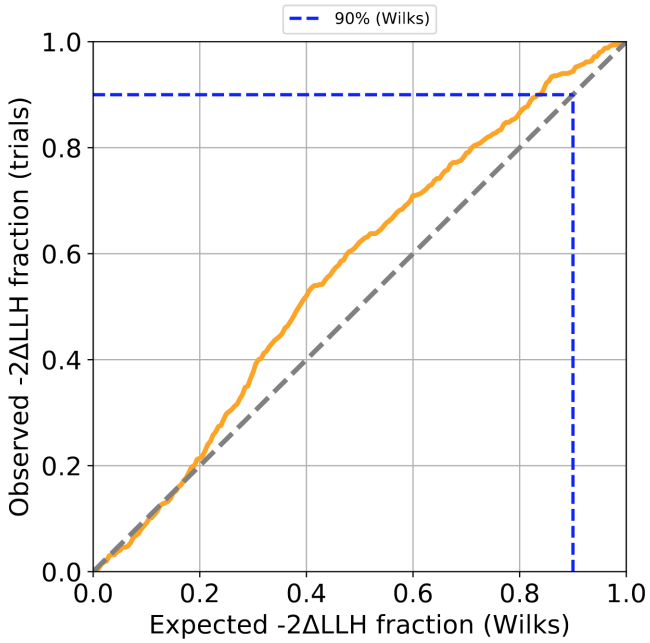


Figure 7.16: Comparison of the observed number of trials with a given $-2\Delta LLH$ to the expected number of trials predicted by Wilks' theorem. Because the gold line is slightly above the 1:1 line, this indicates that we have slight overcoverage. This means that the contour reported in Section 7.4 is conservative.

8

Conclusions & Outlook

In this work, we covered a new sample that was developed to measure neutrino oscillations in IceCube DeepCore using data from 2012–2022. Almost all aspects of the analysis are new compared to previous generations of analyses, such as the event selection, the reconstruction, the use of machine learning classification, the simulation, the majority of the data, and the treatment of many of the systematic uncertainties. I discussed my own work on particle identification, energy estimators, muon simulation, binning optimization, and more. I presented the status of my analysis to measure the atmospheric neutrino oscillation parameters.

Although the full data fit remains an ongoing work in progress as of the time of writing this, the statistical power that this sample will bring upon completion will allow it to measure the parameters θ_{23} and Δm^2_{32} with precision comparable to the current world-leading experiments. It also provides complementarity to these other experiments since we are sensitive to very different sources of systematic uncertainties; our detector observes naturally-occurring atmospheric neutrinos and uses a natural detection medium, whereas the other leading experiments are accelerator-based so they have very different methods of production and have fully human-made detectors.

There are many routes being investigated as possible sources of the observed discrepancy in the full fit, but along the way we continue to learn new things about the detector and our sources of systematic uncertainties. Everything learned here will be useful knowledge to carry forward to future generations of analyses, and eventually future detectors like the IceCube Upgrade [39]. Finally, using the 20% sub-sample, we confirmed the validity of our methods, simulation, and data to perform as expected and we measured the atmospheric neutrino oscillation parameters which appear consistent with other experimental measurements.

Bibliography

[1] The Super-Kamiokande Collaboration, "Evidence for Oscillation of Atmospheric Neutrinos", Phys. Rev. Lett. **81**, 1562–1567 (1998), DOI: 10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.

[2] The SNO Collaboration, "Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory", Phys. Rev. Lett. **87**, 071301 (2001), DOI: 10.1103/PhysRevLett.87.071301, arXiv:nucl-ex/0106015.

[3] The SNO Collaboration, "Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory", Phys. Rev. Lett. **89**, 011301 (2002), DOI: 10.1103/PhysRevLett.89.011301, arXiv:nucl-ex/0204008.

[4] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou, "NuFit 5.1", Journal of High Energy Physics 09 (2020) 178, DOI: 10.1007/JHEP09(2020)178, arXiv:2007.14792, www.nu-fit.org.

[5] Particle Data Group, "Chapter 14: Neutrino Masses, Mixing, and Oscillations in Review of Particle Physics", Progress of Theoretical and Experimental Physics **2020**, 083C01, DOI: 10.1093/ptep/ptaa104, https://pdg.lbl.gov/.

[6] Muon g-2, "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm", Phys. Rev. Lett. **126**, 141801 (2021), DOI: 10.1103/PhysRevLett.126.141801, arXiv:2104.03281.

[7] The IceCube Collaboration, "Measurement of Atmospheric Neutrino Oscillations at 6-56 GeV with IceCube DeepCore", Phys. Rev. Lett. **120**, 071801 (2018), DOI: 10.1103/PhysRevLett.120.071801, arXiv:1707.07081.

[8] The IceCube Collaboration, "Measurement of Atmospheric Tau Neutrino Appearance with IceCube DeepCore", Phys. Rev. D **99**, 032007 (2019), DOI: 10.1103/PhysRevD.99.032007, arXiv:1901.05366.

[9] The IceCube Collaboration, "PINGU: a vision for neutrino and particle physics at the South Pole", Journal of Physics G: Nuclear and Particle Physics **44**, 054006 (2017), DOI: 10.1088/1361-6471/44/5/054006, arXiv:1607.02671.

[10] M. Honda, M. Sajjad Athar, T. Kajita, K. Kasahara, and S. Midorikawa, "Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model", Phys. Rev. D **92**, 023004 (2015), DOI: 10.1103/PhysRevD.92.023004, arXiv:1502.03916 [astro-ph.HE].

[11] The IceCube Collaboration, "The IceCube Neutrino Observatory: Instrumentation and Online Systems", Journal of Instrumentation **12**, P03012–P03012 (2017), DOI: 10.1088/1748-0221/12/03/p03012, arXiv:1612.05093.

[12] S. Tavernier, *Experimental Techniques in Nuclear and Particle Physics.* (Springer, 2010).

[13] The IceCube Collaboration, "The Design and Performance of IceCube DeepCore", Astroparticle Physics **35**, 615–624 (2012), DOI: 10.1016/j.astropartphys.2012.01.004, arXiv:1109.6096.

[14] The IceCube Collaboration, "First Observation of PeV-Energy Neutrinos with IceCube", Phys. Rev. Lett. **111**, 021103 (2013), DOI: 10.1103/PhysRevLett.111.021103, arXiv:1304.5356.

[15] The IceCube Collaboration, Fermi-LAT, MAGIC, et al., "Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A", Science **361**, eaat1378 (2018), DOI: 10.1126/science.aat1378, arXiv:1807.08816.

[16] M. Rongen, "Calibration of the icecube neutrino observatory", PhD thesis (RWTH Aachen University, 2019), arXiv:1911.02016.

[17] P. Eller, *Unified angular acceptance model*, IceCube Internal: Reconstruction & Systematics Phone Call (18 June 2019).

[18] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher, G. Perdue, J. Yarba, *The GENIE Neutrino Monte Carlo Generator: Physics and User Manual*, arXiv:1510.05494, http://genie-mc.org.

[19] M. Larson, "A Search for Tau Neutrino Appearance with IceCube-DeepCore", PhD thesis (Niels Bohr Institute, University of Copenhagen, 2018).

[20] *oscNext*, IceCube Internal Technical Note., 2021.

[21] The IceCube Collaboration, "Low Energy Event Reconstruction in IceCube DeepCore", Submitted to EPJ-C, (2022), arXiv:2203.02303.

[22] *oscNext High Stats Sample*, IceCube Internal Technical Note., 2022.

[23] T. Chen and C. Guestrin, "XGBoost", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794 (2016), DOI: 10.1145/2939672.2939785, arXiv:1603.02754, https://xgboost.readthedocs.io/.

[24] M. Kowalski, "Search for neutrino-induced cascades with the amanda-ii detector", Thesis available here, PhD thesis (Humboldt University of Berlin, 2003).

[25] L. Raedel, "Simulation Studies of the Cherenkov Light Yield from Relativistic Particles in High-Energy Neutrino Telescopes with Geant4", Master's thesis (RWTH Aachen University, 2012).

[26] M. V. Garzelli, S. Moch, and G. Sigl, "Lepton fluxes from atmospheric charm revisited", Journal of High Energy Physics **2015**, 115 (2015), DOI: 10.1007/JHEP10(2015)115, arXiv:1507.01570.

[27] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, *CORSIKA: A Monte Carlo code to simulate extensive air showers*, Tech. Rep. FZKA 6019 (1998), DOI: 10.5445/IR/270043064, https://publikationen.bibliothek.kit.edu/270043064/3813660.

[28] G. D. Barr, S. Robbins, T. K. Gaisser, and T. Stanev, "Uncertainties in atmospheric neutrino fluxes", Phys. Rev. D **74**, 094009 (2006), DOI: 10.1103/PhysRevD.74.094009, arXiv:astro-ph/0611266.

[29] A. Cooper-Sarkar, P. Mertsch, and S. Sarkar, "The high energy neutrino cross-section in the Standard Model and its uncertainty", Journal of High Energy Physics **2011**, (2011), DOI: 10.1007/JHEP08(2011)042, arXiv:1106.3723.

[30] The IceCube Collaboration, "Computational techniques for the analysis of small signals in high-statistics neutrino oscillation experiments", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **977**, 164332 (2020), DOI: 10.1016/j.nima.2020.164332, arXiv:1803.05390, https://github.com/icecube/pisa.

[31] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", The European Physical Journal C **71**, 1554 (2011), DOI: 10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727.

[32] 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), DOI: 10.1214/aoms/1177732360.

[33] The NOvA Collaboration, "Improved measurement of neutrino oscillation parameters by the NOvA experiment", Phys. Rev. D **106**, 032004 (2022), DOI: 10.1103/PhysRevD.106.032004, arXiv:2108.08219.

[34] P. Dunne, "Latest Neutrino Oscillation Results from T2K", XXIX International Conference on Neutrino Physics and Astrophysics (Neutrino 2020), (2020), DOI: 10.5281/zenodo.3959558, https://t2k-experiment.org/results/oscillation-results-from-neutrino-2020/.

[35] V. Takhistov, "Atmospheric Neutrino Oscillations with Super-Kamiokande", 40th International Conference on High Energy Physics, (ICHEP 2020), https://indico.cern.ch/event/868940/contributions/3817174/attachments/2082629/3498300/ICHEP2020_SK_Takhistov.pdf.

[36] MINOS+ Collaboration, "Precision Constraints for Three-Flavor Neutrino Oscillations from the Full MINOS+ and MINOS Dataset", Phys. Rev. Lett. **125**, 131802 (2020), DOI: 10.1103/PhysRevLett.125.131802, arXiv:2006.15208.

[37] M. Rongen and D. Chirkin for the IceCube Collaboration, *A Novel Microstructure-Based Model to Explain the IceCube Ice Anisotropy*, Proceedings for ICRC 2021, arXiv:2107.08692.

[38] G. J. Feldman and R. D. Cousins, "Unified approach to the classical statistical analysis of small signals", Phys. Rev. D **57**, 3873 (1998), DOI: 10.1103/PhysRevD.57.3873, arXiv:physics/9711021.

[39] A. Ishihara for the IceCube Collaboration, *The IceCube Upgrade – Design and Science Goals*, Proceedings for ICRC 2019, arXiv:1908.09441.