

MEASUREMENT OF ATMOSPHERIC MUON NEUTRINO DISAPPEARANCE WITH  
ICECUBE USING CONVOLUTIONAL NEURAL NETWORK RECONSTRUCTIONS

By

Jessie Micallef

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Physics – Doctor of Philosophy  
Computational Mathematics, Science, and Engineering – Dual Degree

2022

## **ABSTRACT**

### **MEASUREMENT OF ATMOSPHERIC MUON NEUTRINO DISAPPEARANCE WITH ICECUBE USING CONVOLUTIONAL NEURAL NETWORK RECONSTRUCTIONS**

By

Jessie Micallef

Neutrinos are neutral, fundamental particles that oscillate, or change flavor, while they travel. This phenomenon means that neutrinos deviate from the Standard Model prediction that they are massless, creating an exciting window into physics exploration beyond the Standard Model. Understanding neutrino oscillation and constraining their behavior is crucial to furthering the understanding of these particles and how they fit, or do not fit, into the Standard Model. The IceCube Neutrino Observatory has been detecting neutrinos for more than 10 years, leading to a large sample of atmospheric neutrino data available for studying neutrino oscillations. Reconstructing the neutrino interactions, such as the neutrino's energy and direction, are key to constraining neutrino oscillation. A fast and robust reconstruction method was developed using convolutional neural networks (CNNs) and optimized to reconstruct parameters necessary to both reconstruct and isolate a pure atmospheric neutrino sample using the IceCube detector. This work compares the performance of this reconstruction to the current likelihood-based reconstruction currently used in IceCube. An analysis of the muon neutrino disappearance is then pursued using 9.28 years of neutrino data. The analysis shows competitive projected sensitivity, the ability to account for numerous systematics, and robust recovery of the physics parameters under statistical and systematic variations. While the 9.28 year sample is still under collaboration review, one year of the total data is used to perform a confirmatory study on the CNN reconstruction and sample. The oscillation parameter constraints from this one year analysis are in alignment with past IceCube analyses and other neutrino experiments within one sigma. This opens the pathway to use the CNN reconstruction for future analyses studying low energy neutrinos on IceCube.

To all the those who inspired me,  
And to all those whom I hope to inspire

## ACKNOWLEDGMENTS

For Professor Tyce DeYoung, who encouraged and supported my pursuit of a unique path in physics and computation.

For Professors Claudio Kopper, Kendall Mahn, and Kirsten Tollefson who always had their doors open and supported me and my work.

For Dr. Brian Clark, who taught me what it means to be a postdoc, a great scientist, and mentor. And who reminded me to breathe.

For Dr. Shiqi Yu, who joined me on this adventure that is oscillation analysis and answered my many neutrino questions.

For both ICER and the CMSE department, who maintained the high performance computing cluster and encouraged student access to it. This was one of the greatest, unexpected perks of my graduate career that has shaped my research trajectory.

For Dr. Devyn Cantu, who was the luckiest part of my graduate school experience to find such a kindred soul in the same research group. Thanks for showing me around my first collaboration meeting, graduating first, and generally showing me how grad school is done.

For Dr. Alison Peisker, who was my partner in grad school from the qualifier to TAing to research to pandemic walks.

For Dr. Joao Pedro Athayde Marcondes de Andre and Dr. Joshua Hignight, who answered my many questions as I entered grad school research and were wonderfully supportive postdocs.

For my IceCube collaborators, who provided an evolving community for me to grow into as a researcher.

For Professor Myron Campbell and the KOTO research group, who shaped me into a curious, competent researcher ready to tackle the challenges of grad school research. And for inspiring me to climb mountains.

For Laura Wood, my friend in so many adventures from comic cons to walking the Mackinac Bridge, who reminded me that life is what you make of it.



For Claire Kopenhafer, who took on a whole conference with me and reminded me of the joys of  
random, epic hobbies.

For my graduate peers and cohort, who answered so many of my questions outside of the  
classroom and created a supportive community of physicists.

For the Grand Canonical Ensemble, i.e. physics choir, and all those in it, who brightened my days  
with music and friendship.

For Strange Matter, Biggby, Koala Bakery and Cafe, and the Lansing Lugnuts for providing the  
best thesis-writing environments.

For Allison Drechny, who listened to me rant about these strange little neutrino particles and the  
roller coaster ride that is research. And who never failed to make me laugh.

For my parents, who support me in every way possible. I don't even know how to acknowledge all  
you do for me.

For my sister and brother-in-law, who celebrated with me through my struggles and triumphs.

Thanks for throwing me the biggest mid thesis writing party (i.e. your wedding).

For my cat, who can solve any problem by crawling into my arms.

For my family, who have always supported me, no matter what, no matter when. I love you all.

# TABLE OF CONTENTS

<b>LIST OF TABLES</b> . . . . .	<b>x</b>
<b>LIST OF FIGURES</b> . . . . .	<b>xii</b>
<b>CHAPTER 1 NEUTRINO PROPERTIES AND OSCILLATION</b> . . . . .	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Neutrinos in the Standard Model . . . . .	2
1.2.1 Neutrino Mass Hierarchy . . . . .	2
1.2.2 Neutrino Oscillation in Vacuum . . . . .	5
1.2.3 Neutrino Oscillation in Matter . . . . .	7
1.3 Neutrino Interactions . . . . .	8
1.3.1 Elastic and Quasi-Elastic Scattering . . . . .	10
1.3.2 Resonant Scattering . . . . .	10
1.3.3 Deep Inelastic Scattering . . . . .	11
1.4 Atmospheric Neutrino Sources . . . . .	11
1.4.1 Neutrino Flux . . . . .	12
1.5 Oscillation Physics with Atmospheric Neutrinos . . . . .	14
1.5.1 Expected Oscillation Signature . . . . .	14
1.5.2 Status of Global Oscillation Measurements . . . . .	15
1.5.3 Status of IceCube Oscillation Measurements . . . . .	16
<b>CHAPTER 2 NEUTRINO DETECTION TECHNIQUES</b> . . . . .	<b>17</b>
2.1 IceCube Neutrino Observatory . . . . .	17
2.1.1 Geometry of the Main Detector . . . . .	17
2.1.2 Digital Optical Module . . . . .	18
2.1.3 DOM Efficiency . . . . .	18
2.2 Neutrino Detection in Ice . . . . .	18
2.2.1 Cherenkov Radiation . . . . .	20
2.2.2 Detection Signatures in IceCube . . . . .	21
2.2.2.1 Topology of Cascade-Like Events . . . . .	21
2.2.2.2 Topology of Track-Like Events . . . . .	23
2.3 Optical Properties of the Ice . . . . .	23
2.3.1 Bulk Ice Properties . . . . .	24
2.3.1.1 Depth Dependence . . . . .	24
2.3.1.2 Azimuthal Dependence . . . . .	25
2.3.2 Hole Ice Properties . . . . .	25
<b>CHAPTER 3 LOW ENERGY SIMULATION AND DATA SAMPLE PROCESSING</b> . . . . .	<b>27</b>
3.1 Monte Carlo Samples . . . . .	27
3.1.1 Neural Network Training Samples . . . . .	27
3.1.2 Analysis Sample . . . . .	28

3.2	Monte Carlo Generation . . . . .	29
3.2.1	Neutrino Events . . . . .	29
3.2.2	Muon Events . . . . .	30
3.2.3	Noise Hits . . . . .	31
3.2.4	Noise Events . . . . .	32
3.3	Propagating MC Photons and Detector Response . . . . .	32
3.4	Data Processing . . . . .	32
3.4.1	Level 2 . . . . .	33
3.4.2	Level 3 . . . . .	34
3.4.3	Level 4 . . . . .	35
3.4.4	Level 5 . . . . .	36
3.4.5	Level 6 . . . . .	36
<b>CHAPTER 4 FAST LOW ENERGY RECONSTRUCTION USING CONVOLUTIONAL NEURAL NETWORKS . . . . .</b>		<b>37</b>
4.1	Architecture . . . . .	38
4.2	Preparing the Training Samples . . . . .	42
4.2.1	Data Agreement . . . . .	42
4.2.2	Preparing Input Features . . . . .	43
4.2.3	Sample Cleaning Differences between Training Sets . . . . .	44
4.2.3.1	Neutrino vs. Noise BDT . . . . .	45
4.2.4	Number of Pulses and DOMs Hit . . . . .	45
4.2.4.1	Containment . . . . .	46
4.2.5	Training Each CNN . . . . .	46
4.2.5.1	Energy . . . . .	48
4.2.5.2	Zenith Angle . . . . .	49
4.2.5.3	Vertex . . . . .	51
4.2.5.4	Track Classifier . . . . .	52
4.2.5.5	Muon Background Classifier . . . . .	53
4.3	Performance . . . . .	54
4.3.1	Reconstruction Runtime . . . . .	56
4.3.2	Energy . . . . .	57
4.3.3	Zenith . . . . .	58
4.3.4	Vertex . . . . .	60
4.3.5	Track Classifier . . . . .	62
4.3.6	Atmospheric Muon Classifier . . . . .	63
4.4	Stability Against Detector Systematics . . . . .	64
4.4.1	Reconstructing Different DOM Efficiencies . . . . .	65
4.4.2	Reconstructing Different Hole Ice Parameterizations . . . . .	66
4.4.3	Reconstructing Different Bulk Ice Properties . . . . .	68
4.4.4	Conclusion of Robustness to Systematics . . . . .	69
<b>CHAPTER 5 FLERCNN ANALYSIS SAMPLE SELECTION . . . . .</b>		<b>71</b>
5.1	Resolution Motivated Cuts . . . . .	71
5.1.1	Starting Vertex . . . . .	71

5.1.2	Number of hit DOMs . . . . .	71
5.1.3	Energy Range . . . . .	72
5.2	Data/MC Agreement Motivated Cuts . . . . .	72
5.2.1	Identifying Atmospheric Muons . . . . .	72
5.2.1.1	Straight Veto Cut for Muons . . . . .	73
5.2.1.2	Muon BDT . . . . .	73
5.2.2	Direct Pulses . . . . .	75
5.2.3	Cosine Zenith Range . . . . .	76
5.3	Application of Cuts . . . . .	77
5.4	Analysis Distribution Template . . . . .	78
5.4.1	PID . . . . .	79
5.4.2	Binning . . . . .	81
5.4.3	Muon Sample KDE Template . . . . .	82
5.4.4	Masking . . . . .	84
<b>CHAPTER 6 ANALYSIS OF MUON NEUTRINO DISAPPEARANCE WITH 9.28</b>		
	<b>YEARS OF DATA . . . . .</b>	<b>85</b>
6.1	Overview of Analysis Procedure . . . . .	85
6.2	Test Statistic . . . . .	86
6.3	Verifying Data Consistency . . . . .	87
6.3.1	Data Quality Check Between Years . . . . .	87
6.3.2	Data/MC Agreement . . . . .	90
6.4	Systematic Uncertainties . . . . .	93
6.4.1	Atmospheric Flux . . . . .	93
6.4.1.1	Cosmic Ray Uncertainty . . . . .	93
6.4.1.2	Uncertainty on Meson Production . . . . .	94
6.4.1.3	Normalization Terms . . . . .	95
6.4.2	Oscillation Modeling and Parameters . . . . .	95
6.4.3	Cross Sections . . . . .	96
6.4.4	Detector Systematics . . . . .	96
6.4.4.1	Hypersurfaces . . . . .	97
6.4.5	Assessment of Systematic Parameters' Impact . . . . .	99
6.4.6	Summary of All Systematic Parameters . . . . .	101
6.5	Verifying the Analysis Pipeline . . . . .	103
6.5.1	Asimov Inject-Recover Test . . . . .	103
6.5.2	Projected Sensitivity To Oscillation Parameters . . . . .	104
6.5.3	Parameter Ensemble Test . . . . .	105
6.5.4	Ensemble Test . . . . .	107
6.6	Blind Fit . . . . .	108
6.6.1	Blind Fit Procedure . . . . .	109
6.6.2	Blind Fit Result . . . . .	110
6.6.3	Potential Contributions to the Blind Fit Result . . . . .	112
<b>CHAPTER 7 ANALYSIS OF MUON NEUTRINO DISAPPEARANCE WITH 0.92</b>		
	<b>YEARS OF DATA . . . . .</b>	<b>116</b>

7.1	One Year Sample . . . . .	116
7.2	Changes in Analysis with One Year Sample . . . . .	116
7.2.1	Adjusting the Test Statistic . . . . .	118
7.2.2	Summary of Changes . . . . .	118
7.3	Adjusted Expectations with One Year of Livetime . . . . .	119
7.3.1	Sensitivity Projection for One Year . . . . .	119
7.3.2	Impact of Systematics on One Year Analysis . . . . .	119
7.3.3	Pre-Fit Data/MC Agreement for One Year . . . . .	120
7.4	Blind Fit . . . . .	123
7.4.1	Data/MC Agreement using Best Fit Values . . . . .	123
7.5	Best Fit Result . . . . .	126
7.5.1	Best Fit Values and Contours . . . . .	127
<b>CHAPTER 8</b>	<b>CONCLUSIONS . . . . .</b>	<b>133</b>
8.1	Potential for Future Work on FLERCNN . . . . .	134
8.1.1	Optimize Current Architecture . . . . .	134
8.1.2	Expand FLERCNN Reconstruction Output . . . . .	135
8.1.3	Broaden FLERCNN Applications . . . . .	136
8.2	Potential for Future Work on IceCube Neutrino Oscillations . . . . .	136
<b>BIBLIOGRAPHY</b>	<b>. . . . .</b>	<b>138</b>

## LIST OF TABLES

Table 3.1:	Simulation settings used to generate neutrinos for the analysis sample. . . . .	30
Table 4.1:	Describes the loss function and activation used for each network. . . . .	42
Table 4.2:	Describes the transformations that the input features undergo during preprocessing.	44
Table 4.3:	Describes the additional training sample cuts. For ease of viewing, cells highlighted in green indicate that the cut in that column was applied to the training sample for the network in that row. . . . .	45
Table 4.4:	Describes the training samples for each network, which more details expanded in the following sections. . . . .	47
Table 4.5:	Describes the number of files used for training, the model chosen in terms of the raw training pass and epoch. . . . .	47
Table 4.6:	Describes the training samples learning rate and changes. . . . .	48
Table 4.7:	FLERCNN and Retro Reconstruction sample cut comparison after Level 5. All analysis cuts that apply to the FLERCNN sample are described in more detail in Sections 5.1 and 5.2. . . . .	55
Table 4.8:	Runtime and implications of applying the different reconstructions to the MC sample. . . . .	56
Table 5.1:	Describes cuts on the neutrino MC analysis sample and their individual impact when added to the final sample. . . . .	78
Table 5.2:	Describes the final rates, including re-weighting with interpolated hypersurfaces and muon KDE smoothing described in Sections 6.4.4.1 and 5.4.3, predicted by the MC simulation. . . . .	79
Table 5.3:	Describes the analysis binning using the CNN reconstructed and classified variables. . . . .	82
Table 6.1:	Settings for the systematic sets used to generate the hypersurfaces, with each row describing a new set. Cells highlighted in green indicate settings that vary from their baseline values (row 1). Rows highlighted in gray indicate systematic sets that are currently under study with potential mismatched simulation (Section 7.2). . . . .	98

Table 6.2: Systematic parameters, including those fixed during fit, along with the priors and bounds on the parameters that are free during fit. Parameters that are fixed have their row highlighted in gray. . . . .	102
Table 6.3: Stop conditions for the blind fit plan. . . . .	110
Table 7.1: Total counts with 0.92 years of livetime predicted by the updated MC simulation.	117
Table 7.2: Changes made for the one year analysis to handle updates and reduced in statistics.	119
Table 7.3: Agreement between the data and post-fit MC using the best fit values for the physics and systematic parameters, quantified by the $p$ -value pre- vs. post-fit. . .	126
Table 7.4: Best fit values for the physics and systematic parameters using one year of livetime with FLERCNN sample. The pull describes the difference in $\sigma$ of the parameters with Gaussian priors. . . . .	127

## LIST OF FIGURES

Figure 1.1: Standard Model of particle physics, including the particles and their mediators. The lepton sector (green) and the bosons (red) are the critical regions for this work. . . . .	2
Figure 1.2: The distances and energies necessary to probe the two mass squared differences ( $\Delta m_{21}^2$ in the dashed line and $\Delta m_{32}^2$ in the dotted-dashed line). Plot from [1]. . . . .	4
Figure 1.3: Feynman diagrams for the charged current (left) and neutral current (right) interactions. The $\ell$ represents the leptons ( $e$ , $\mu$ , or $\tau$ ). This particular example is of deep inelastic scattering mechanism (Section 1.3.3). . . . .	9
Figure 1.4: Contributions of each of the interaction cross sections as a function of energy (plots from [2]). Neutrinos are in the left plot and anti-neutrinos in the right plot. The general trend is the same for both neutrino and anti-neutrino, with QE dominating at the lowest energies ( $< 2$ GeV), resonance at a few GeV, and DIS at the 10 GeV-scales. . . . .	9
Figure 1.5: Description of atmospheric neutrino sources from cosmic ray interactions . . .	12
Figure 1.6: The neutrino flux ( $\Phi_\nu$ ) with all directions averaged as a function of energy (left plot), ratio of the flux for the difference seasons as a function of energy (middle plot) and the neutrino flux as a function of $\cos \theta_z$ (plots from [3]). . . .	13
Figure 1.7: Simulated probability of $\nu_e$ , $\nu_\mu$ , $\nu_\tau$ (left to right) $\rightarrow \nu_\mu$ at various distances (in cosine zenith) and energies assuming normal ordering. Notice that the mass effects are visible around $\cos \theta_z = -0.8$ in all three plots, in the region of 2-16 GeV for the first plot and 4-8 GeV in the middle and left plots. . . . .	15
Figure 1.8: Result contours for global experimental results, including IceCube 7.5 year result and likelihood-based 9.28 year analysis' sensitivity projection [4, 5, 6, 7, 8, 9]. . . . .	16
Figure 2.1: Top and side view of the IceCube detector (modified from [10]). The left side plot shows the variation in the scattering and absorption as a function of depth (Section 2.3.1). Two footprints are highlighted in the top view: the DeepCore array circled in a dotted blue line and the strings used to train the CNN reconstruction (Chapter 4) circled in purple. . . . .	19
Figure 2.2: Schematic of the digital optical module (DOM). . . . .	20



Figure 2.3: Different types of interactions, their secondaries, and the expected signatures for incident neutrino energies < 100 GeV (modified from [11]). . . . .	22
Figure 2.4: Scattering and absorption coefficients (left scale in blue line) and lengths (right scale in red) as a function of depth. Smaller values of the coefficients indicate cleaner ice [12]. . . . .	25
Figure 2.5: The charge difference between MC and data measured as a function of the azimuth angle, showing the impact of the tilt and flow anisotropy [13]. . . . .	26
Figure 3.1: Flow chart describing the processing for simulation (left) and data (right). . . . .	28
Figure 4.1: Top view of the IceCube detector, with DeepCore strings in red, IceCube strings in gray, and a blue highlight on all strings used for the low-energy CNN. . . . .	39
Figure 4.2: Detailed network architecture for FLERCNN with the legend for colored layers on the right. The notation at the top indicates that for each event, there are 8 DeepCore strings with 60 DOMs per string and 5 summary variables per DOM fed into one branch of the network. The other branch of the network sees the 19 IceCube strings with their 60 DOMs per string and 5 summary variables per DOM. . . . .	41
Figure 4.3: Shows the rate of the default discriminator in blue (from the SPE templates used to generate FLERCNN data and MC) compared to data as a function of charge. Agreement is seen above 0.25 PE [14]. . . . .	43
Figure 4.4: Energy training sample and loss progression. . . . .	49
Figure 4.5: Zenith training sample and loss progression. . . . .	50
Figure 4.6: X position training sample and Y position training sample. . . . .	51
Figure 4.7: Z training sample and vertex loss progression. . . . .	52
Figure 4.8: PID training sample and loss progression. . . . .	53
Figure 4.9: True cosine zenith and energy distribution used to train the muon classifier. The true energy for the muon simulation is the energy of the muon at generation, which is why it is different than the energy of the neutrino, which is shown at the interaction point. The $\nu_\mu$ and $\nu_e$ histograms are stacked, in comparison with the muon histogram (not stacked). . . . .	54
Figure 4.10: Loss vs. epoch progression as the muon classification network trained, using the binary crossentropy loss function. . . . .	54

Figure 4.11: Testing distributions for the $\nu_\mu$ CC and $\nu_e$ CC events. The energy and zenith cuts are not applied in this case, since they will be applied based on the individual reconstructions' outputs during testing. All other cuts applied are the CNN Cuts described in Table 5.1. . . . .	56
Figure 4.12: Resolution shown for true energy for the $\nu_\mu$ CC and $\nu_e$ CC events. With the invisible energy carried off by the outgoing neutrinos from NC interactions and some $\nu_\tau$ CC interactions, the resolution is shown vs. deposited energy for these selections. . . . .	59
Figure 4.13: Resolution shown for reconstructed cosine zenith for the $\nu_\mu$ CC tracks and remaining cascade events. . . . .	59
Figure 4.14: Resolution shown for reconstructed radius from the center string (string 36) of the starting vertex for the $\nu_\mu$ CC tracks and remaining cascade events. . . . .	61
Figure 4.15: Resolution shown for reconstructed z position of the starting vertex for the $\nu_\mu$ CC tracks and remaining cascade events. . . . .	61
Figure 4.16: Confusion matrix for vertex reconstruction, exploring efficiency of cutting at 200m, proposed for oscillation analysis. . . . .	61
Figure 4.17: Performance of Track classifier with binary classification histogram on the left and the ROC on the right. . . . .	62
Figure 4.18: Performance of muon classifier with binary classification histogram on the left and the ROC on the right. . . . .	64
Figure 4.19: Bias and spread for the FLERCNN performance on various detector systematic sets for DOM efficiency. The FLERCNN reconstruction is shown in the solid lines and the likelihood-based reconstruction (retro) is shown in the dotted lines. . . . .	65
Figure 4.20: Bias and spread for the FLERCNN performance on various detector systematic sets for hole ice. The FLERCNN reconstruction is shown in the solid lines and the likelihood-based reconstruction (retro) is shown in the dotted lines. . . . .	67
Figure 4.21: Bias and spread for the FLERCNN performance on various detector systematic sets for bulk ice. The FLERCNN reconstruction is shown in the solid lines and the likelihood-based reconstruction (retro) is shown in the dotted lines. . . . .	68
Figure 5.1: The data/MC disagreement for the number of hits in the top 15 layers of IceCube DOMs. The left plot shows before the cut (keeping events < 0.5 hits) and the right plot shows the adjusted agreement after these events are removed. . . . .	73

Figure 5.2: The data/MC disagreement for the number of hits in the outermost IceCube strings. The left plot shows before the cut (keeping events with $< 7.5$ hits) and the right plot shows the adjusted agreement after these events are removed. . . . .	74
Figure 5.3: Performance of Muon BDT, with the BDT predicted neutrino probability on the left and the rate with various proposed threshold cuts on the right (plots created by Dr. Shiqi Yu). . . . .	76
Figure 5.4: Number of DOMs hit with direct pulses, identified using the SANTA algorithm. The left plot shows before the cut (keeping events with $< 3$ direct hits) and the right plot shows the adjusted agreement after these events are removed. The MC scaling to the data means that removing the mismatched bin improves the overall normalization for the data/MC agreement. . . . .	77
Figure 5.5: Rates of the different selections comparison from the earlier levels to the final level FLERCNN sample after analysis cuts, including re-weighting with interpolated hypersurfaces and muon KDE smoothing described in Sections 6.4.4.1 and 5.4.3. . . . .	79
Figure 5.6: Result of the CNN track-like classifier on the nominal MC, binned more finely in the top plot and binned at the proposed 3-bin boundaries for analysis in the bottom plot. Note that the neutrino (and muon) flavors are all stacked. . . . .	80
Figure 5.7: Percentages for each flavor's fraction of events located in that PID bin as a function of energy. For example, the $\nu_\mu$ CC sample shows that 80% of the $\nu_\mu$ CC events $> 60$ GeV are classified as "tracks" with the chosen PID binning. . . . .	81
Figure 5.8: Composition of each PID bin as a function of true energy, with the relative percent of the each flavor's contribution to the events in each PID bin. . . . .	81
Figure 5.9: Number of expected events in each bin over the given 9.28 years of livetime, with neutrino weights adjusted by interpolated hypersurfaces to account for detector systematics and the expected muon distribution smoothed by KDE. . . . .	82
Figure 5.10: Muon distribution in the analysis binning before the KDE smoothing (left column), with sharp features due to low statistics, and after KDE smoothing applied (middle column) along with the pull in each bin (right column). The gray bins are where $\sigma_{\text{hist}} = 0$ . . . . .	83
Figure 5.11: Percentage of muons expected in the sample per analysis bin after muon KDE smoothing. . . . .	84
Figure 5.12: Analysis binning after masking, with the masked bins shown in white. . . . .	84
Figure 6.1: Energy distribution compared between years of data. . . . .	88

Figure 6.2: Zenith angle distribution compared between years of data. . . . .	88
Figure 6.3: PID distribution compared between years of data. . . . .	88
Figure 6.4: Muon BDT distribution compared between years of data. . . . .	89
Figure 6.5: Rates compared between years of data, split up by month. . . . .	90
Figure 6.6: Data/MC agreement for reconstructed energy and zenith angle. . . . .	91
Figure 6.7: Data/MC agreement for PID and muon BDT classifiers. Note that the muon BDT classifier has a cut applied at 0.8, which accounts for the behavior from 0.4-0.8. . . . .	92
Figure 6.8: Data/MC agreement for starting vertex $z$ and $\rho_{36}$ reconstructed variables (radial position relative to the center IceCube string #36). . . . .	92
Figure 6.9: Barr parameters labels, chart from [15]. . . . .	94
Figure 6.10: Example of hypersurface fit in one analysis bin projected on the DOM effi- ciency axes, using the discrete sets to create a continuous fit (plot from [8]). The lighter data points indicate systematic sets that have additional detector parameters pulled in addition to the DOM efficiency. The darker data points are systematic sets with only DOM efficiency changed. . . . .	99
Figure 6.11: Relative effect of each systematic parameter on the analysis, determined by pulling the parameter from nominal to generate pseudo-data but fixing it to nominal when fitting. Bars in orange are systematic parameters that are free for the proposed analysis fit, while the bars in purple are proposed to be fixed. . . . .	101
Figure 6.12: Scan over physics parameters showing the inject values for the pseudo-data in the orange dot, the fitted value in the purple cross, and the fitter starter point in red plus (the latter is the same for all scans). All fitted physics parameters recover their injected truth parameters. . . . .	104
Figure 6.13: Sensitivity of the FLERCNN and $LLH$ -based analysis at the verification sample best fit point, compared to global results. . . . .	106
Figure 6.14: Quantifying the ability of the fitter to recover systematic parameters injected at various off-nominal values, with a perfect recovery on the diagonal 1:1 gray line and the results for various fits shown in the dots. The color of the dots corresponds to the fit's total $-LLH$ result, where a smaller value (purple) of $-LLH$ indicates a better fit than a larger value (yellow). This test is run assuming normal ordering of the neutrino masses. . . . .	107

Figure 6.15: Fit results for all systematic and physics parameters after 500 ensemble test trials, assuming normal ordering. The distribution of fitted values for each parameter, including the distribution's median value which should ideally be aligned with the nominal value and the distribution's $1\sigma$ which should be within the range of the prior. . . . .	108
Figure 6.16: Fit results for all systematic and physics parameters after 500 ensemble test trials, assuming inverted ordering. The distribution of fitted values for each parameter, including the distribution's median value which should ideally be aligned with the nominal value and the distribution's $1\sigma$ which should be within the range of the prior. . . . .	109
Figure 6.17: Systematic pulls for the nine year FLERCNN sample blind fit shown in terms of $\sigma$ for parameters with a Gaussian prior. None of the parameters hit their stop condition, with all Gaussian pulls under $1\sigma$ . . . . .	111
Figure 6.18: Pulls in $\sigma$ for each of the analysis bins between the data and the MC at the fitted physics and systematic parameters for the nine year FLERCNN sample. None of the bins hit the stop conditions for the pull, with all bins pulling less than $3\sigma$ . . . . .	111
Figure 6.19: Test statistic result for nine years of data fit shown in purple line with a distribution of ensemble test TS results in yellow showing the expected TS spread for 499 trials at the data best fit point. The data TS falls within the ensemble distribution of TS results with a p-value of 0.6% which hits the stop condition requiring it to be $> 5\sigma$ . . . . .	112
Figure 7.1: Analysis binning after masking, with the masked bins shown in white, for the one year of livetime. . . . .	117
Figure 7.2: Projected sensitivity for the one year analysis (red) vs. the nine year analysis (black), where the expected contour is wider with less statistics. . . . .	120
Figure 7.3: New systematic mismodeling plot is on the left for the one year sample, with comparison to the right plot (from Figure 6.11) for the nine year sample. . . . .	121
Figure 7.4: Data/MC pre-fit agreement for one year of livetime with the reconstructed energy (left) and reconstructed zenith angle (right). . . . .	121
Figure 7.5: Data/MC pre-fit agreement for one year of livetime with the CNN classified PID (left) and BDT Muon classification (right). . . . .	122
Figure 7.6: Data/MC pre-fit agreement for one year of livetime with the starting vertex reconstructed $\rho_{36}$ (left) and reconstructed z positions (right). . . . .	122

Figure 7.7: Systematic pulls for the one year blind fit shown in terms of $\sigma$ for parameters with a Gaussian prior. . . . .	123
Figure 7.8: Pulls in $\sigma$ for each of the analysis bins between the data and the MC at the fitted physics and systematic parameters. . . . .	124
Figure 7.9: Test statistic result for the data fit shown in purple line with a distribution of ensemble test TS results in yellow showing the expected TS spread for 499 trials at the data best fit point. The data TS falls within the ensemble distribution of TS results with a p-value of 31.7%. . . . .	124
Figure 7.10: Post-fit data/MC agreement for the energy (left) and zenith angle (right). . . . .	125
Figure 7.11: Post-fit data/MC agreement for the PID score (left) and background muon BDT score (right). . . . .	125
Figure 7.12: Post-fit data/MC agreement for the $\rho_{36}$ (left) and $z$ (right). . . . .	126
Figure 7.13: Result FLERCNN one year fit shown on the explored range (range of number line plot), Gaussian prior if relevant (blue shaded region), nominal value (green dashed line), stop condition boundary (orange lines), and fitted value (red cross). For historical reasons, the analysis software fits for $\Delta m_{31}^2$ , solving for $\Delta m_{32}^2 = \Delta m_{31}^2 - \Delta m_{21}^2$ . . . . .	129
Figure 7.14: Result of the one year analysis, with 90% contour in red, $1\sigma$ contour in blue, and the best fit point in the black X. The 1D projections for the physics parameters are show on top of and to the right of the main plot, with the $\chi_{\text{mod}}^2$ value shown for each physics parameter. . . . .	131
Figure 7.15: Result FLERCNN one year contours compared to other IceCube $\nu_\mu$ disappearance anaylses (left) [16, 17, 8] and global experimental results (right) [4, 5, 6, 7]. . . . .	131
Figure 7.16: Projected sensitivity of the one year analysis using the one year result BFP. . . . .	132

# CHAPTER 1

## NEUTRINO PROPERTIES AND OSCILLATION

### 1.1 Introduction

Particle physics uses the smallest, most fundamental building blocks to answer the universe's biggest questions. While the Standard Model of particle physics lays the foundation for fundamental interactions, several discrepancies in the neutrino sector, such as the existence of neutrino masses and flavor mixing, leave open critical questions. Thus, neutrinos provide a rich opportunity to explore the Standard Model and to probe many of the open physics questions. Constraining the neutrino oscillation parameters gives us a better understanding of the remaining neutrino mysteries. The IceCube Neutrino Observatory offers a large atmospheric neutrino sample with which to perform an analysis measuring the neutrino oscillation phenomena, which will be explored in this work.

As data sets in neutrino physics get larger, there is a need for computational methods that can handle increased sample complexity and quickly process the results. Machine Learning has been at the forefront of computational advances in the 21st century and has shown promising results when applied to other neutrino and particle physics experimental applications [18, 19, 20]. IceCube's data and MC samples already exceed 100 million events, such that utilizing fast and accurate computational solutions are vital. This work will describe a series of convolutional neural networks developed expressly for reconstruction of 10 GeV-scale atmospheric neutrinos in the IceCube detector. This new reconstruction is then applied to a muon neutrino disappearance analysis to constrain the oscillation parameters and place these results in the broader context of other neutrino oscillation experiments.

## 1.2 Neutrinos in the Standard Model

Neutrinos are fundamental, spin 1/2 particles that are part of the lepton sector in the Standard Model (Figure 1.1). The neutrinos are neutral, but correspond to a charged lepton pair, so that there are three known neutrino flavors: electron neutrino ( $\nu_e$ ), muon neutrino ( $\nu_\mu$ ), and tau neutrino ( $\nu_\tau$ ). The Standard Model predicts that neutrinos are massless, but due to the fact that neutrinos have been observed to oscillate [21, 22], at least two of the neutrino states must have non-zero mass.

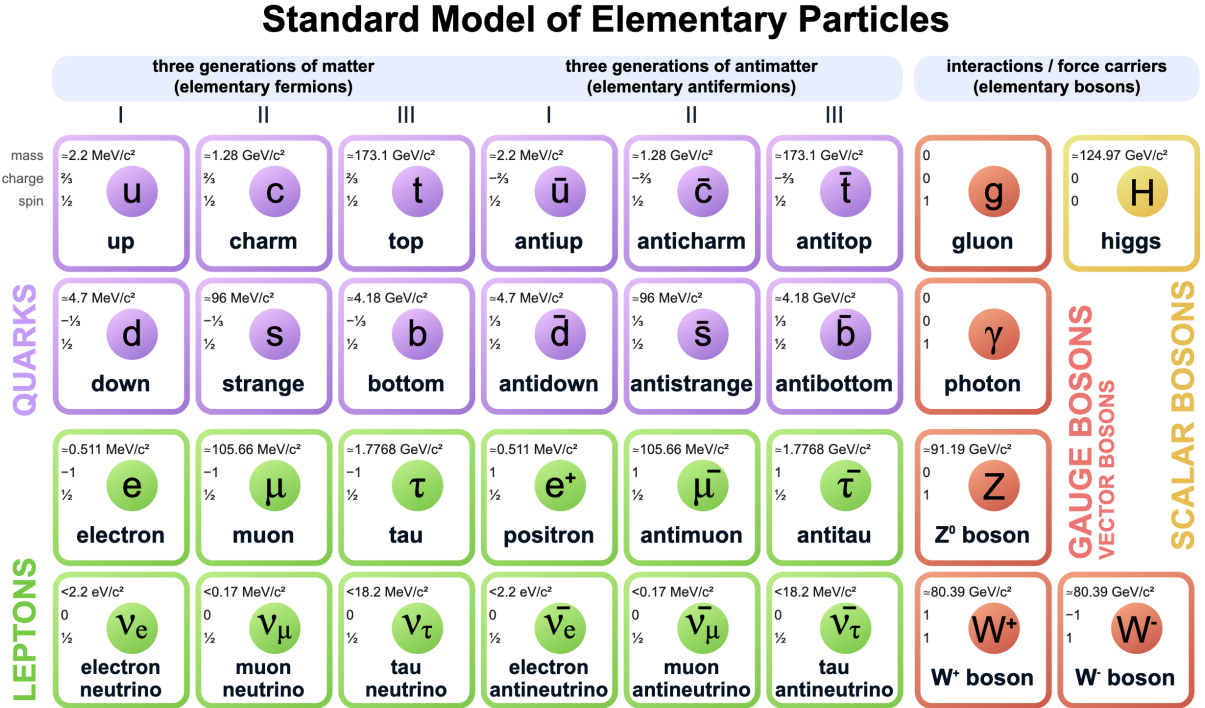


Figure 1.1: Standard Model of particle physics, including the particles and their mediators. The lepton sector (green) and the bosons (red) are the critical regions for this work.

### 1.2.1 Neutrino Mass Hierarchy

Each of the neutrino *flavor states* ( $\nu_{\alpha=e,\mu,\tau}$ ) are described using linearly independent superpositions of their *mass states* ( $\nu_{i=1,2,3}$ ). The absolute masses are still a source of experimental measurements, with the sum of the three masses estimated to be  $< 1 \text{ eV}$  from measurements of beta decay, supernova, and atmospheric sources [23, 24, 25]. While the sign of the difference between the  $\nu_1$  and  $\nu_2$  mass states is known, where ( $m_{\nu_2} > m_{\nu_1}$ ), the sign of  $m_{\nu_3}$  is not yet known [23]. The case where



$(m_{\nu_3} > m_{\nu_2} > m_{\nu_1})$  is called the *normal ordering*. The case where  $(m_{\nu_2} > m_{\nu_1} > m_{\nu_3})$  is called the *inverted ordering*.

Neutrinos are generated and interact in their flavor states, but propagate in their mass states, which can lead to *neutrino oscillation* or flavor changing between the neutrino's generation and interaction stages. Sections 1.2.2 and 1.2.3 discuss the factors that influence the probability of oscillations. Equation 1.1 shows the mathematical representation of the flavor states written in terms of the mass eigenstates.

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i} |\nu_i\rangle \quad (1.1)$$

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U_{\text{PMNS}} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (1.2)$$

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (1.3)$$

The unitary matrix  $U$  is called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) mixing matrix and describes the mixing between the neutrino mass and flavor states [26, 27].

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{23} & \sin \theta_{23} \\ 0 & -\sin \theta_{23} & \cos \theta_{23} \end{pmatrix} \begin{pmatrix} \cos \theta_{13} & 0 & e^{-i\delta_{\text{CP}}} \sin \theta_{13} \\ 0 & 1 & 0 \\ e^{i\delta_{\text{CP}}} \sin \theta_{13} & 0 & \cos \theta_{13} \end{pmatrix} \begin{pmatrix} \cos \theta_{12} & \sin \theta_{12} & 0 \\ -\sin \theta_{12} & \cos \theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.4)$$

where  $\theta$  is the mixing angle and  $\delta_{\text{CP}}$  is the Charge-Parity (CP) violating phase. Each matrix in Equation 1.4 is typically probed experimentally by using neutrinos from different sources: atmospheric for  $\theta_{23}$ , reactor for  $\theta_{13}$ , and solar for  $\theta_{12}$  respectively. This is due to the values of the PMNS matrix and mass differences between the neutrinos, which nature has fortuitously given us such that the oscillations between the  $\nu_2$  and  $\nu_3$  states essentially mix fully first (on shorter distance

scales) before engaging in mixing with the  $\nu_1$  state. This is not necessarily motivated by any theory, but measurements have indicated that this is the case. Using the analysis in this work, energies from 5-100 GeV are considered and the distances atmospheric neutrinos can travel range from 20-12,500 km depending on if it is generated above the detector in atmospheric at the South Pole or generated at the North Pole and travels through the Earth's core to the detector. Figure 1.2 shows the regions of energy ( $E$ ) and distance ( $L$ ) that different neutrino experiments probe, and therefore which mass squared differences (such as  $\Delta m_{32}^2 = m_3^2 - m_2^2$ ) they are sensitive to. IceCube's blue shaded region intersects with the dotted-dashed  $m_{32}^2$  line in Figure 1.2, indicating that it is sensitive to measuring the atmospheric  $m_{32}^2$  term. How the neutrino's energy, distance travelled, and mass squared difference factor into the probability of oscillation are explored in more detail in Section 1.2.2.

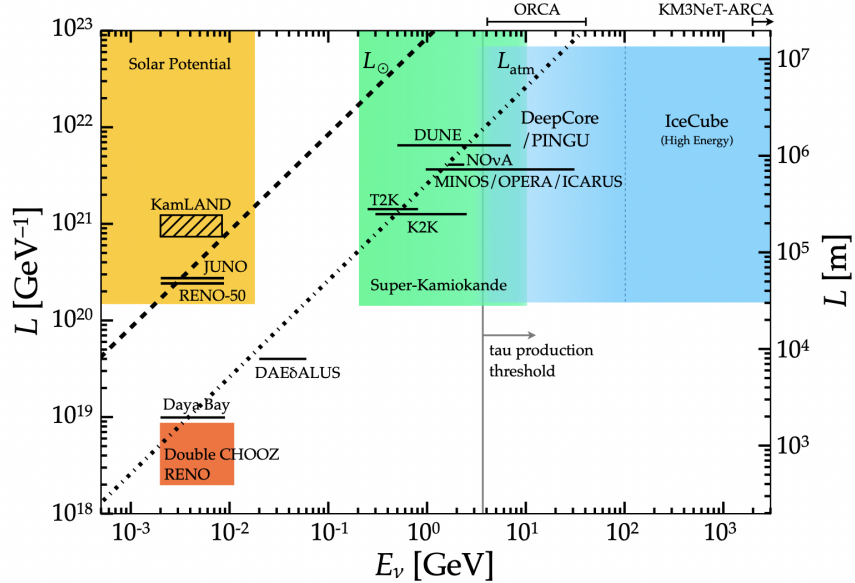


Figure 1.2: The distances and energies necessary to probe the two mass squared differences ( $\Delta m_{21}^2$  in the dashed line and  $\Delta m_{32}^2$  in the dotted-dashed line). Plot from [1].

This work studies atmospheric neutrino oscillation using the IceCube detector, and thus will focus on the  $\theta_{23}$  mixing angle. Another experiment that probes the  $\theta_{23}$  mixing angle using atmospheric neutrinos is Super-Kaminokande (or Super-K) shown in the bottom of the green region in Figure 1.2 [6]. Accelerator experiments at certain energies and distances are also sensitive to

$\theta_{23}$ , so that experiments like T2K, NOvA, and MINOS+ will be used for comparison (Section 1.5.2), also denoted in the green region in Figure 1.2 at smaller energies and distances than IceCube probes [5, 4, 7].

### 1.2.2 Neutrino Oscillation in Vacuum

As the neutrino propagates, the evolution is described as using the mass eigenstates:

$$|\nu(t)\rangle = \sum_i U_{\alpha i} e^{-iE_i t} |\nu_i\rangle \quad (1.5)$$

where the energy is  $E_i = \sqrt{\vec{p}^2 + m_i^2}$ , where  $\vec{p}$  is the momentum and  $m$  is the mass.

We then look at the probability of observing the neutrino in each of the flavor states:

$$P_{\nu_\alpha \rightarrow \nu_\beta}(t) = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2 \quad (1.6)$$

$$= \left| \sum_i U_{\beta i} U_{\alpha i}^* e^{-iE_i t} \right|^2 \quad (1.7)$$

$$= \sum_{i,j} U_{\beta i}^* U_{\alpha i} U_{\beta j} U_{\alpha j}^* e^{-i(E_i - E_j)t} \quad (1.8)$$

Some approximations can be made due to the fact we know the neutrino masses are small and we are in the relativistic regime, so that  $E \gg m_i$  and  $t \approx L$ , where  $L$  is the distance the neutrino has travelled [28].

$$E_i \approx E + \frac{m_i^2}{2E} \quad (1.9)$$

Therefore, to find  $E_i - E_j$ , we can substitute in this approximation so that

$$E_i - E_j \approx E + \frac{m_i^2}{2E} - E - \frac{m_j^2}{2E} \quad (1.10)$$

$$\approx \frac{m_i^2 - m_j^2}{2E} \quad (1.11)$$

$$\approx \frac{\Delta m_{ij}^2}{2E} \quad (1.12)$$

where  $\Delta m_{ij}^2$  is referred to as the *mass squared splitting*. Putting this back into the probability:

$$P_{\nu_\alpha \rightarrow \nu_\beta}(E, L) = \sum_{i,j} U_{\beta i}^* U_{\alpha i} U_{\beta j} U_{\alpha j}^* e^{-i(\frac{\Delta m_{ij}^2}{2E})L} \quad (1.13)$$

Euler's formula can be used to expand the exponential in Equation 1.13 along with using the unitary nature of U, so that:

$$\begin{aligned} P_{\nu_\alpha \rightarrow \nu_\beta}(E, L) = & \delta_{\alpha\beta} - 4 \sum_{i < j} \text{Re} \left( \sum_i U_{\beta i}^* U_{\alpha i} U_{\beta j} U_{\alpha j}^* \right) \sin^2 \left( \frac{\Delta m_{ij}^2 L}{4E} \right) \\ & + 2 \sum_{i < j} \text{Im} \left( \sum_i U_{\beta i}^* U_{\alpha i} U_{\beta j} U_{\alpha j}^* \right) \sin \left( \frac{\Delta m_{ij}^2 L}{2E} \right) \end{aligned} \quad (1.14)$$

Note that this is specific to neutrinos, with the imaginary term subtracted for the antineutrino case. The oscillation probability relies on the fundamental properties of the mixing angles from the PMNS matrix elements, which determine the amplitude of the oscillations, and the mass squared difference of the mass states, which determines the frequency of the oscillations. The probability is dependent on the energy of the neutrino ( $E$ ) and its distance traveled ( $L$ ), which can be measured experimentally. In total, there are seven free terms that govern neutrino oscillations:  $\theta_{23}, \theta_{13}, \theta_{12}, \Delta m_{31}^2, \Delta m_{32}^2, \Delta m_{21}^2, \delta_{CP}$ .

A simplification of this is to consider the two neutrino approximation, where there are only two flavors ( $\alpha = e, \mu$ ) and mass states ( $i = 1, 2$ ). In this case,  $\Delta m^2$  accounts for the only two mass states ( $m_1^2 - m_2^2$ ) and U becomes the 2x2 matrix described as

$$U = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (1.15)$$

Thus, the two flavor approximation for oscillation probability simplifies to

$$P_{\nu_\alpha \rightarrow \nu_\beta}(E, L) = \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \quad (1.16)$$

From this simplified view in Equation 1.16 it is clearer to see that to measure the mixing angle ( $\theta$ ) and the mass squared difference ( $\Delta m^2$ ), one needs to experimentally determine the neutrino's flavor (both  $\alpha$  and  $\beta$ ), energy ( $E$ ), and distance travelled from generation to interaction ( $L$ ). For atmospheric neutrino oscillation measurements, this means understanding the expected neutrino flux per neutrino flavor to predict  $\alpha$  (Section 1.4). For  $\beta$ ,  $E$ , and  $L$ , these will be determined in the detector itself using a given reconstruction method (Chapter 4).

In addition, using Equation 1.16, one can note that the mass squared difference determines the frequency of the oscillation probability and the mixing angle controls the amplitude of the oscillation probability. Due to the relative amplitude of the  $\theta_{23}$  term and frequency of the  $\Delta m_{32}^2$  term at the distances ( $L$ ) and energies ( $E$ ) that IceCube probes with atmospheric neutrinos, the two flavor simplification is a good, first-order approximation of the expected oscillation features (Section 1.4). The  $\nu_\mu \rightarrow \nu_e$  oscillation features are sub-dominant and not resolvable for the IceCube atmospheric  $\nu_\mu$  disappearance analyses (left plot of Figure 1.7), thus the oscillation between the  $\nu_\mu \rightarrow \nu_\tau$  is the main contribution to the oscillation pattern observed (right plot of Figure 1.7). Even so, the full three flavor oscillation picture is considered computationally in the analysis in this work along with the matter effects described in the following section (Section 1.2.3).

### 1.2.3 Neutrino Oscillation in Matter

As neutrinos pass through matter, they have the potential to interact with leptons and nucleons that is dependent on the density of the matter (Section 1.3). These interactions add an additional potential to the Hamiltonian for the neutrino oscillation:

$$H = H_0 + V \quad (1.17)$$

where  $H_0$  is the Hamiltonian for the kinetic energy, as defined in a vacuum (Section 1.2.2) and  $V$  is the matter interaction potential.

For the neutrinos traveling through the earth, the only surviving term that impacts the potential from lepton and nucleon interactions is from charged current interactions with electrons (Section 1.3), so that  $V = \text{diag}(V_e, 0, 0)$  [29].

$$V_e = \pm \sqrt{2} G_F n_e(x) \quad (1.18)$$

where  $G_F$  is the Fermi coupling constant and  $n_e(x)$  is the electron number density in the medium [29, 30]. The electron number density is not constant in the earth and must be properly accounted for in the analysis (Section 6.4.2). This adds a density-dependent term to the oscillation calculation that is specific to the path of the neutrino. This is accounted for in the computational modeling with estimates of the electron density of the earth, so it remains that an oscillation experiment must measure the same parameters (final flavor, energy, and path of the neutrino). Accounting for the mass effects means that the modeling gets more complicated, where one must account for the electron density estimation in the earth when considering systematic parameters that could affect the experiment. Note that for normal ordering, the matter effects are seen in the neutrino channel (visible in Figure 1.7). For inverted order, the matter effects are seen in the antineutrino channel.

### 1.3 Neutrino Interactions

Neutrinos interact only via the weak force, which has two types of mediators: the neutral  $Z^0$ -boson and the charged  $W^\pm$ -boson. The Feynman diagrams for the interactions with these mediators are described in Figure 1.3. charged current (CC) interactions are mediated by the  $W^\pm$  boson, and with an incident neutrino resulting in a charged lepton that is associated the incident neutrino's flavor (i.e.  $\nu_e \rightarrow e$ ). Neutral current (NC) interactions are mediated by the  $Z^0$  boson, and with an incident neutrino resulting in an output neutrino of the same flavor ( $\nu_e \rightarrow \nu_e$ ).



Figure 1.3: Feynman diagrams for the charged current (left) and neutral current (right) interactions. The  $\ell$  represents the leptons ( $e$ ,  $\mu$ , or  $\tau$ ). This particular example is of deep inelastic scattering mechanism (Section 1.3.3).

At energies from 1-200 GeV, which encompasses the energy range relevant for the  $\nu_\mu$  disappearance analysis in Chapter 6, there are three mechanisms for the neutrino scattering. Figure 1.4 shows the cross section as a function of energy, where the expected scattering is either quasi-elastic (QE), resonant (RES), or deep inelastic scattering (DIS) as energy increases. Note that the antineutrino cross section has a smaller total magnitude (right plot in Figure 1.4), with roughly the same overall shape as the neutrino cross section so that the ratio between them is about 2 across all energies. The following sections explain the process for each of the three aforementioned scattering mechanisms.

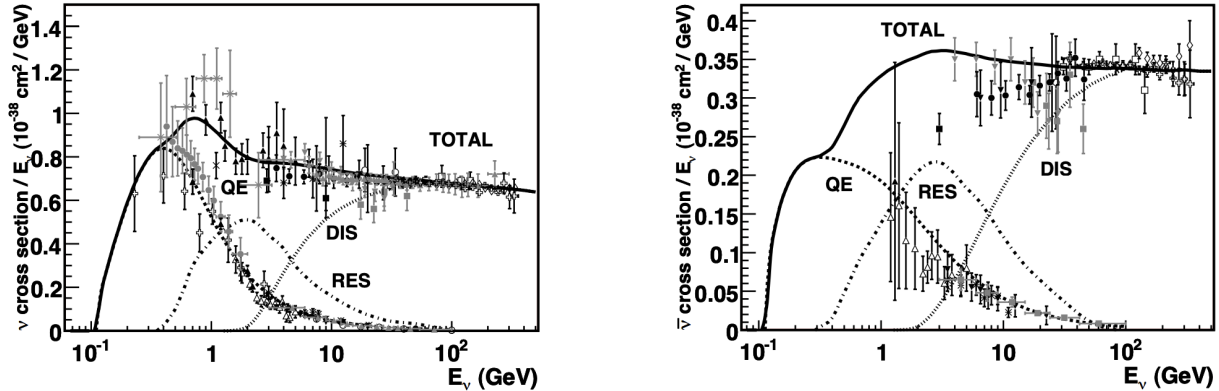


Figure 1.4: Contributions of each of the interaction cross sections as a function of energy (plots from [2]). Neutrinos are in the left plot and anti-neutrinos in the right plot. The general trend is the same for both neutrino and anti-neutrino, with QE dominating at the lowest energies ( $< 2$  GeV), resonance at a few GeV, and DIS at the 10 GeV-scales.

### 1.3.1 Elastic and Quasi-Elastic Scattering

This interaction occurs when a neutrino scatters off of an entire nucleon and frees a nucleon or multiple nucleons. Elastic scattering can occur for neutral current interactions and quasi-elastic (QE) scattering for charged current interactions. The cross section for this mechanism dominates at energies  $< 2$  GeV [2]. The analysis in this work will focus on reconstructed energies  $> 5$  GeV, so this process should only be relevant for a small fraction of the total sample. For QE neutrino CC interactions, the neutron is converted to a proton. For QE antineutrino CC interactions, the proton is converted to a neutron. For QE NC interactions (regardless of neutrino or antineutrino), there is no conversion, such that the proton will remain a proton after scattering. Calculating the cross section depends on some constants that have been constrained experimentally to high precision, the neutrino energy and four-momentum transfer, along with an axial mass term ( $M_A$ ) [2]. This term has had a number of experimental measurements [2]. For the purposes of this analysis, the GENIE simulation framework is used since it is optimized to model low energy neutrino interactions, and it defines  $M_A = 0.99$  GeV [31].  $M_{A,QE}$  is also included as a systematic parameter (Sections 3.2 and 6.4.3).

### 1.3.2 Resonant Scattering

Resonance describes when the neutrino strikes the nucleon into an excited state, creating a nucleon and typically one or more pions, with photons and kaons also possible but less likely [2]. The cross section calculation typically used the Rein–Sehgal model, which is also employed by the GENIE simulation [2, 32]. Note that there is also an axial mass  $M_A$  term in the resonant scattering cross section, but it is set to  $M_A = 1.12$  GeV in GENIE [31]. There is a separate  $M_{A,res}$  systematic parameter included in the analysis to account for this (Section 6.4.3). The energy range used for this analysis, starting at reconstructed energies of 5 GeV, means that resonance will also be a subdominant mechanism that affects a small fraction of the events. The majority of the analysis sample will be composed of events that have undergone deep inelastic scattering.



### 1.3.3 Deep Inelastic Scattering

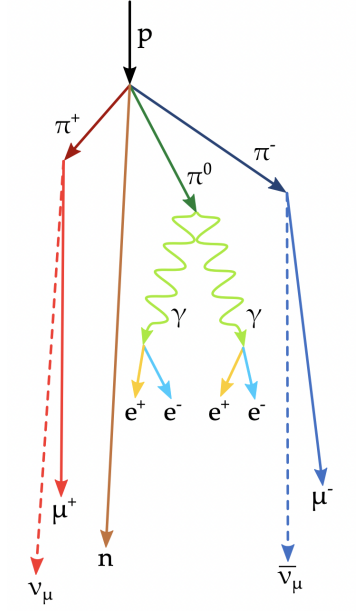
Deep Inelastic Scattering (DIS) occurs when the neutrino has enough energy to scatter off of a quark in the nucleon. This mechanism leads to a hadronic shower along with the final state lepton, depicted by the Feynman diagram examples for CC and NC in Figure 1.3. DIS is expected to dominate at energies  $> 10$  GeV as seen in Figure 1.4, and its cross section increases approximately linearly with energy in this region [2]. Thus, this will be the main mechanism for events in the analysis sample outlined in Chapters 5, due to the events selected so that their reconstructed energy is between 5-100 GeV.

The DIS cross section has multiple theoretical models that have been calculated with a focus on different energy ranges. The main neutrino interaction simulation framework used in this work, GENIE, is optimized for  $< 100$  GeV energies [31, 33]. While this is mostly aligned with the proposed energy range of the analysis in this work, there is a noted mismatch between some more current models of the DIS cross section, such as the CSMS [34]. CSMS calculation is known to not produce reliable results  $< 100$  GeV, but does model the cross section well at higher energies. The mismatch in the cross section between these two models will be accounted for in the systematics of the analysis (Section 6.4).

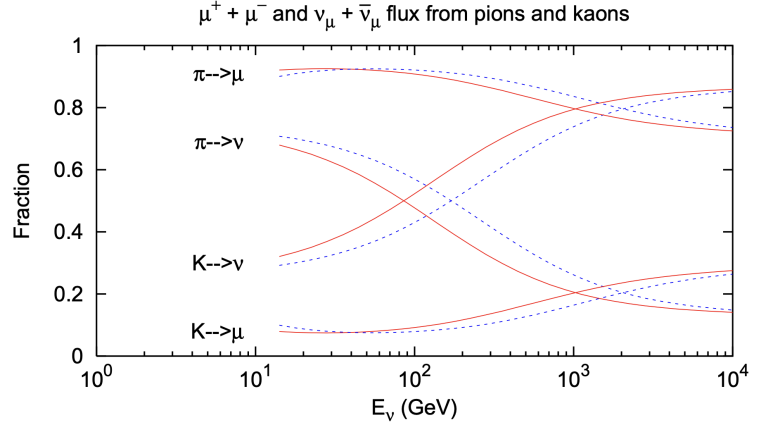
## 1.4 Atmospheric Neutrino Sources

One prevalent, natural source of neutrinos is from the cosmic ray interactions in the atmosphere. High energy nuclei and protons from outer space, or cosmic rays, interact with air molecules in the upper atmosphere, creating a hadronic shower (*hadrons* are particles made of two or more quarks). The cosmic ray shower is depicted in Figure 1.5a, showing the pions and kaons produced in the shower. The decay of these hadrons is the primary source of the atmospheric neutrinos [35]. Figure 1.5b shows the fraction of neutrinos expected from the pions vs. kaons, which is both energy dependent (x-axis) and incident zenith angle dependent (solid line indicates  $\theta_z = 0^\circ$  and dotted line  $\theta_z = 60^\circ$ ) [35].

This means that the neutrino flux is dependent on the meson production modeling, which in-



(a) Diagram of the particle shower from cosmic ray interaction (from [36]).



(b) Fraction of muons and muon neutrinos expected from the pion and kaon decay. The solid line indicates  $\theta_z = 0^\circ$  (vertical) and the dotted line  $\theta_z = 60^\circ$  (plot from [35]).

Figure 1.5: Description of atmospheric neutrino sources from cosmic ray interactions

cludes factors such as the cosmic ray flux, hadronic interaction modeling of the shower, atmospheric density and temperature, and the Earth's magnetic field [3]. Some of these will be considered as systematic parameters during the analysis, with more details in Section 6.4.

### 1.4.1 Neutrino Flux

Some of the factors that influence the meson production modeling, and in turn the neutrino flux, have been mentioned. Numerous models have attempted to quantify the neutrino flux, with the Honda model [37] used as a reference for this work. To look closer at the behavior of the neutrino flux, Figure 1.6 depicts the flux at the South Pole Lab (SPL) across various dependencies.

The left plot shows the direction averaged flux as a function of neutrino energy. The gray dashed lines indicate the maximum region of interest for the  $\nu_\mu$  disappearance analysis (Section 6), denoting 1-200 GeV. The colored lines show that the largest expected flux at all energies is  $\nu_\mu$ . The  $\nu_\mu$  flux is more consistent than the  $\nu_e$  flux in this range, but broadly it can be said that the  $\nu_\mu$  flux

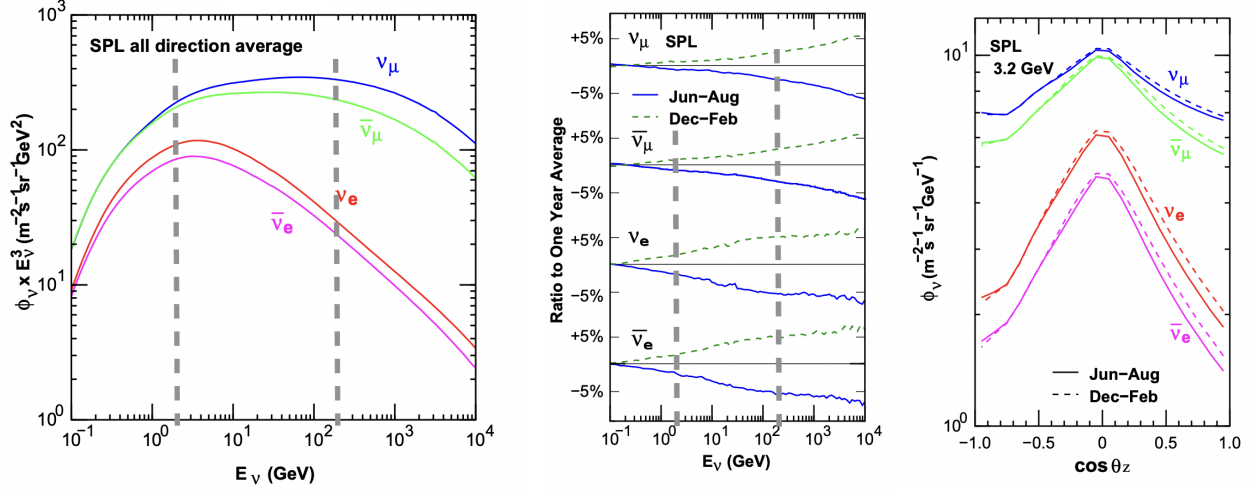


Figure 1.6: The neutrino flux ( $\Phi_\nu$ ) with all directions averaged as a function of energy (left plot), ratio of the flux for the difference seasons as a function of energy (middle plot) and the neutrino flux as a function of  $\cos \theta_z$  (plots from [3]).

is at least 2 but up to 100 times larger than the  $\nu_e$  flux. Both of the electron and muon neutrinos are typically twice as large compared to the expected flux of their respective antineutrino.

The middle plot shows the seasonal dependence of the flux, where the y-axis has the ratio of the flux to the one year average. Note that there are 4 plots stacked on top of each other, showing a range of  $\pm 5\%$  deviation from the average. There exists a clear seasonal difference in the flux, with Dec-Feb having an above average flux and Jun-Aug below average. This is due to seasonal change in the altitude of cosmic ray interactions [3]. Note that this difference is also energy dependent, with less seasonal variation at low energy compared to even the 200 GeV point marked with the dashed gray line. This contributes to the seasonal variations discussed in 4.2.1.

The right plot in Figure 1.6 shows the flux dependence on the zenith angle at a single energy (3.2 GeV). The seasonal variation (shown between the dashed and solid lines) is more severe for neutrinos that are vertically downgoing at the South Pole ( $\cos \theta_z = 1$ , explained further in Section 1.5). It is proposed that this is due to the muon energy loss in the atmosphere with the seasonal variations [3]. Overall, the shape of the flux is influenced by the geomagnetic field influencing the path of the charged leptons [3].

What we learn from these plots is that the neutrino flux has many factors that can cause

significant changes to the expected neutrino yield. Among these factors are influences that have energy, directional, seasonal, and flavor dependence. Section 6.4.1 discusses how these factors are accounted for in the oscillation analysis along with parameters that are used as systematics during the fit to account for potential deviations from the estimated models.

## 1.5 Oscillation Physics with Atmospheric Neutrinos

Atmospheric neutrinos can be detected at various baseline distances ( $L$ ) and energies ( $E$ ) to provide a measure of the oscillation over the 2D parameter space. The distance is measured by using the angle that the neutrino travels through the detector, relative to the surface about the detector. Thus, a neutrino that is generated in the atmosphere directly above the surface of the IceCube detector at the South Pole and travels down through the detector is called "downgoing" and is denoted with  $\theta_z = 0$  or  $\cos \theta_z = 1$ . An "upgoing" neutrino is created on the other side of the earth, at the North Pole, and travels through the length of the earth to reach IceCube ( $\theta_z = 180$  or  $\cos \theta_z = -1$ ). The distance  $L$  can then be calculated using the angle to find the distance through the earth that the neutrino has to travel to reach the detector. Since the zenith angle is the parameter measured using the detector, this will be reported such that  $\cos \theta_z \sim L$ .

### 1.5.1 Expected Oscillation Signature

Figure 1.7 shows the probability of oscillation to  $\nu_\mu$  for each of the neutrino flavors. The  $\nu_\mu$  disappearance ( $\nu_\mu \rightarrow \nu_\mu$ ) is expected at  $\cos(\theta_z) < 0.1$  and  $E < 100$  GeV, where the probability of a muon neutrino remaining a muon neutrino is near zero. This is the region that experiments must be able to resolve to measure the disappearance signature. The value of the mass squared difference determines where the oscillation bands are located relative to  $L$  and  $E$ —as the frequency of the oscillations. The mixing angle controls the depth of the oscillation signal shown in the color bar of Figure 1.7—as the amplitude of the probability. Note that the dominant consideration is for  $\nu_\mu \rightarrow \nu_\tau$ , which has a much stronger amplitude for the oscillation signal than  $\nu_\mu \rightarrow \nu_e$ . This is why  $\nu_\mu$  disappearance can be approximated with the two flavor neutrino picture, since the oscillation

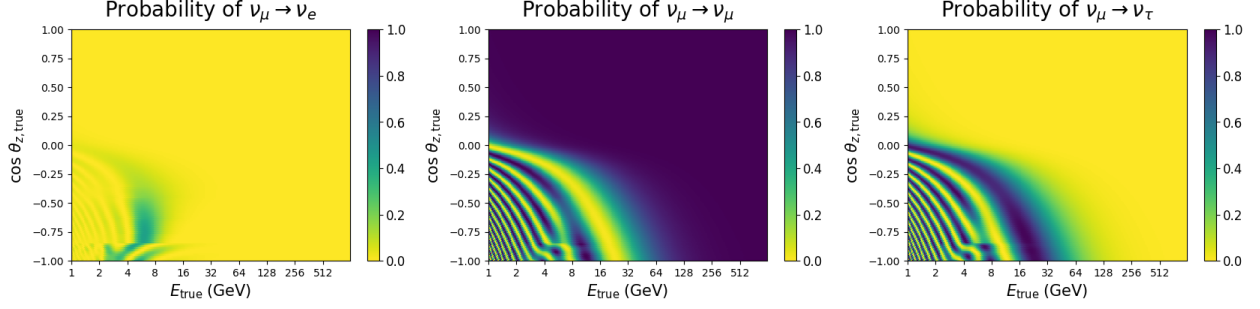


Figure 1.7: Simulated probability of  $\nu_e$ ,  $\nu_\mu$ ,  $\nu_\tau$  (left to right)  $\rightarrow \nu_\mu$  at various distances (in cosine zenith) and energies assuming normal ordering. Notice that the mass effects are visible around  $\cos \theta_z = -0.8$  in all three plots, in the region of 2-16 GeV for the first plot and 4-8 GeV in the middle and left plots.

into  $\nu_e$  is a sub-dominant effect. The analysis in this work is still performed computationally including the full three flavor oscillation paradigm, along with accounting for the matter effects.

### 1.5.2 Status of Global Oscillation Measurements

For  $\Delta m_{23}^2$  and  $\sin^2(\theta_{32})$ , numerous experiments have recently probed these parameters using atmospheric neutrinos like in IceCube and Super-K ([17, 16, 38]) and accelerator-based neutrinos like in NOvA, MINOS+, and T2K [4, 7, 5]. Figure 1.8 has the current results for all of the aforementioned experiments compared to IceCube’s 7.5 year golden neutrino sample (Section 1.5.3).

The advantage of IceCube’s ability to measure neutrino oscillation is the high statistics of atmospheric neutrino data available due to the detector’s large volume and the rate of atmospheric flux. IceCube has yet to publish a result taking advantage of the 10 years of livetime available with an analysis method capable of utilizing the full available sample. The goal of this work is to develop a reconstruction method to quickly process the high statistics sample and to analyze the sample using the IceCube  $\nu_\mu$  disappearance analysis procedure.

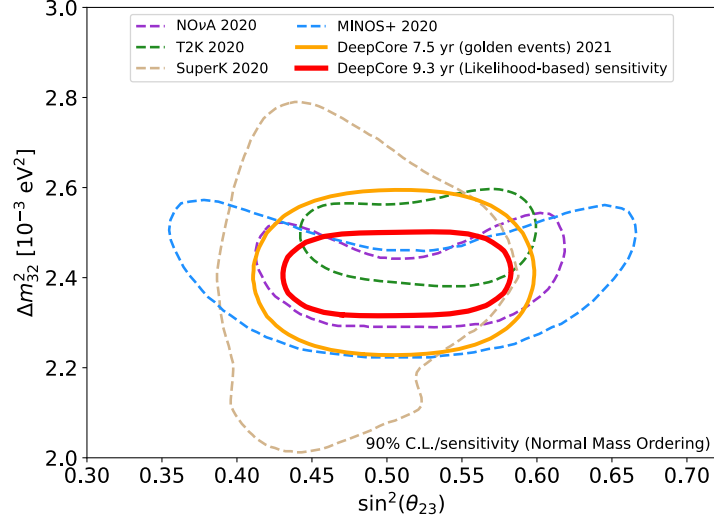


Figure 1.8: Result contours for global experimental results, including IceCube 7.5 year result and likelihood-based 9.28 year analysis' sensitivity projection [4, 5, 6, 7, 8, 9].

### 1.5.3 Status of IceCube Oscillation Measurements

The most recent IceCube  $\nu_\mu$  disappearance result uses 7.5 years of data. Due to the method used for this analysis, it only uses "golden" neutrino events, or a harsh selection criteria [10], such that the events must have at least 5 hits that come from direct light (i.e. unscattered light, see Section 2.3.1). Even using the tight requirement, the  $\nu_\mu$  disappearance result with the golden sample provides improved contours compared to the previous three year livetime IceCube results. The contours and best fit result for  $\sin^2(\theta_{23})$  and  $\Delta m_{32}^2$  will be used as a comparison for this analysis.

In addition to the 9.28 year livetime sample discussed in this work, there is another 9.28 year livetime sample that is reconstructed and optimized using a likelihood-based reconstruction [10]. Since the likelihood-based reconstructed sample analysis is not yet completed, it is only used in this work for comparisons on reconstruction resolution and projected sensitivity to the oscillation parameters.

## CHAPTER 2

### NEUTRINO DETECTION TECHNIQUES

#### 2.1 IceCube Neutrino Observatory

The IceCube Neutrino Observatory instruments a cubic kilometer of natural glacial ice at the South Pole. Deployed in columns of melted and then refrozen ice, IceCube is comprised of 5,160 Digital Optical Modules (DOMs) attached to 86 strings of cabling. The cables provide both power and communication capabilities to send readouts from the DOMs to the surface of the ice, which are then processed and recorded by computers in the IceCube Laboratory. Since the full deployment in the ice in 2011, the survival percentage of DOMs has exceeded 99% [39]. This stability has helped IceCube to record more than 10 years of livetime of data.

##### 2.1.1 Geometry of the Main Detector

Of the 86 strings, 78 are called "IceCube" strings. These strings are placed in a hexagonal array, with 125 m between them in  $xy$  position. Each IceCube string has 60 DOMs distributed between 1450 m to 2450 m in depth with a 17 m spacing between the modules [39]. The IceCube strings are shown in green in Figure 2.1.

The DeepCore strings are the center 8 strings shown in red in Figure 2.1. These are less regularly spaced in the  $xy$  plane, ranging from 41 m to 105 m apart, all of which are closer together than the IceCube string spacing. The DeepCore strings have 60 total DOMs, but the top 10 are used for vetos and span depths between 1750 m to 1850 m (red highlight in side view of Figure 2.1), while the bottom 50 DOMs are between 2100 m and 2450 m in depth (light green highlight in side view of Figure 2.1) [39]. The veto DOMs are spaced every 10 m vertically, whereas the core 50 DOMs on the DeepCore string are 7 m apart in  $z$  position. The majority of the DOMs on the DeepCore strings also have photomultiplier tubes (PMTs) with a higher quantum efficiency ( $\sim 35\%$ ) [39]. This aids in making the DeepCore DOMs more sensitive to low light, low energy

events.

### 2.1.2 Digital Optical Module

The fundamental detection unit on IceCube, the DOM, contains a 10" PMT pointing downward into the ice (Figure 2.2). Along with the PMT, each DOM houses an embedded high-voltage generator, a LED flasher board for calibration, and a main board that processes the analogue and digital signals from the PMT. The components of the DOM are protected by a 13" glass pressure sphere that withstands the high pressure deployment in the ice.

The PMT in the DOM captures photons from the Cherenkov light produced by charged particles traveling faster than the speed of light in ice (Section 2.2.1). The PMT amplifies the signal of electrons ejected from the photocathode due to the photon capture, creating a voltage drop at the anode that creates a pulse. The pulse's raw waveform is digitized and read out if the signal surpasses the threshold of 0.25 photoelectrons (p.e.) [39].

### 2.1.3 DOM Efficiency

The **DOM efficiency** describes the scaling of the expected PMT efficiency in the DOMs. The PMT's photon detection efficiency was measured before deployment into the ice for 13 DOMs, which informs the calibrated "expected" efficiency used for all DOMs. The DOM efficiency is a systematic parameter that allows the scaling on the calibrated efficiency to be adjusted. This is one of the systematic parameters for the analysis outlined Section 6.4.

## 2.2 Neutrino Detection in Ice

Neutrinos are detected through indirect methods, due to the fact that they only interact weakly, so one must measure the light produced by secondary particles and their successive interactions. This information is then used to reconstruct, or estimate, the energy, direction, and interaction vertex of the incident neutrino. Due to the different secondaries expected from the CC and NC neutrino interactions for the different flavors, there are two main classes of signatures expected.



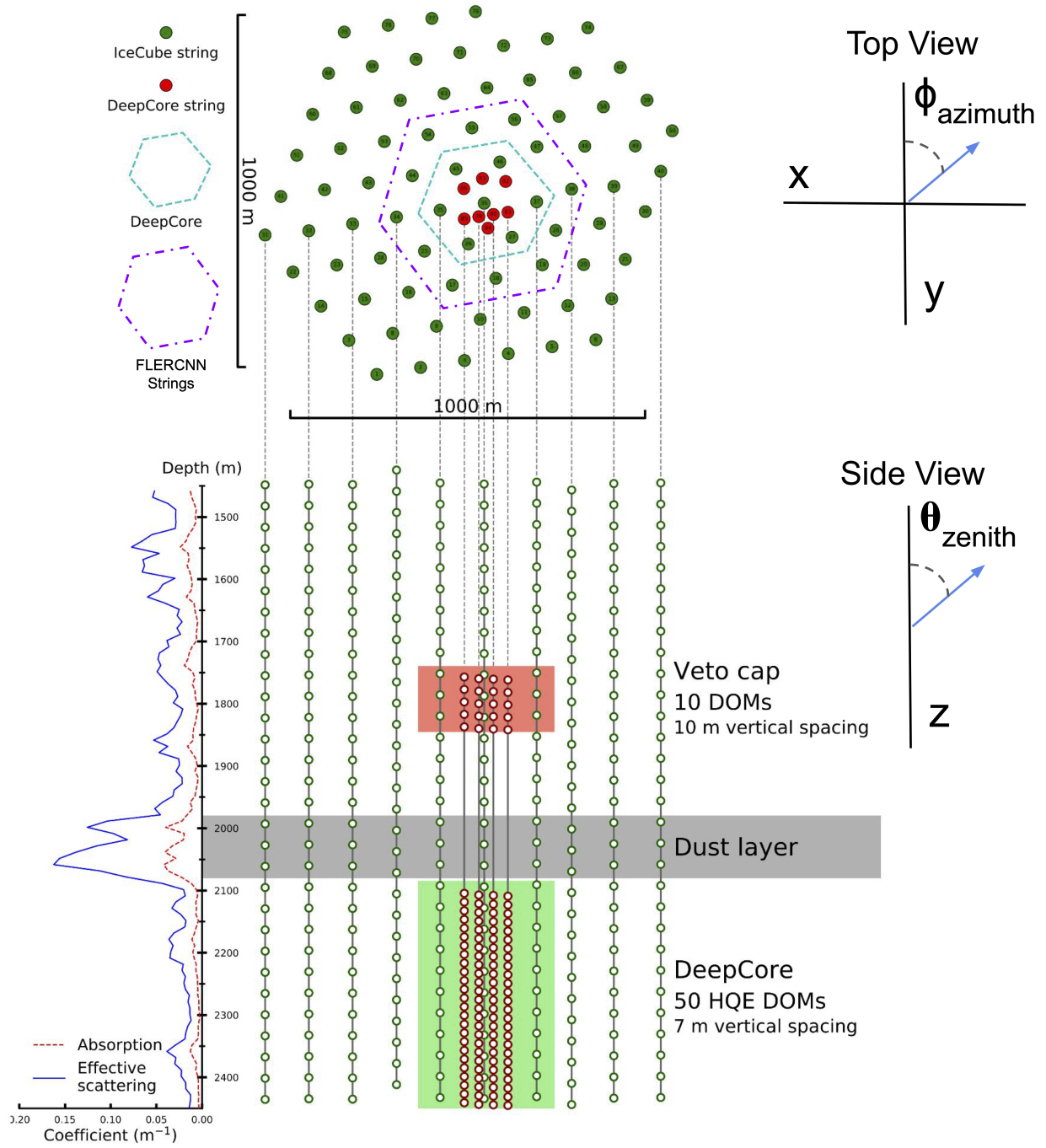


Figure 2.1: Top and side view of the IceCube detector (modified from [10]). The left side plot shows the variation in the scattering and absorption as a function of depth (Section 2.3.1). Two footprints are highlighted in the top view: the DeepCore array circled in a dotted blue line and the strings used to train the CNN reconstruction (Chapter 4) circled in purple.

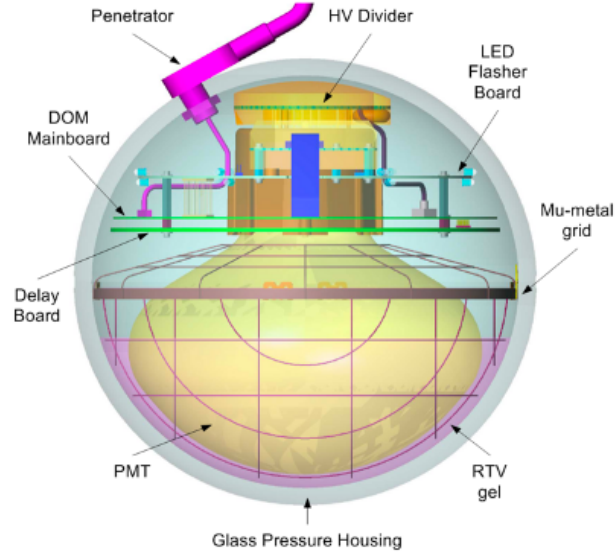


Figure 2.2: Schematic of the digital optical module (DOM).

Section 2.2.1 will first discuss when and how the secondary particles produce light. Section 2.2.2 will compare the different signatures visible in the IceCube detector and their interaction sources.

### 2.2.1 Cherenkov Radiation

Cherenkov light is radiation emitted when a charged particle travels faster than the speed of light in a medium, primarily in the blue spectrum or at wavelengths of about 400 nm [40]. For a material's refractive index ( $n$ ), charged particle's velocity in a medium ( $v$ ), and speed of light in vacuum ( $c$ ), if  $v > \frac{c}{n}$  then a cone of light will be emitted. The cone of light is such that the opening angle  $\theta_C$  relative to the path of the charge particle is

$$\cos \theta_C = \frac{c}{nv}. \quad (2.1)$$

As  $v \rightarrow c$  and using  $n = 1.32$  in ice [23],  $\theta_C \approx 41^\circ$ . Using  $v = \frac{c}{n}$ , we can find the energy threshold to see Cherenkov radiation emission. The relativistic energy must be used in this case [24]:

$$E_C \geq mc^2 \gamma \geq \frac{mc^2}{\sqrt{1 - (\frac{v}{c})^2}} \geq \frac{mc^2 n}{\sqrt{n^2 - 1}} \quad (2.2)$$

$$(2.3)$$

which is dependent on the particle mass ( $m$ ). Thus, the minimum energy an electron needs to create Cherenkov radiation in the ice is 780 keV and a muon needs at least 160 MeV. Cherenkov radiation occurs with any charged particle that has reached the threshold energy in the ice, and thus light is expected from both the charged leptonic ( $e, \mu, \tau$ ) and hadronic (made up of two or more quarks including proton, neutrino, pion, and kaon) secondaries. While they create similar cones of light, there are some differences in the signatures that these leave in the detector.

### 2.2.2 Detection Signatures in IceCube

There are three main components of the neutrino interaction that create Cherenkov light. There is the lepton itself, which is the secondary that occurs in the charged current interactions (Section 1.3). There are also electromagnetic and hadronic showers. *Electromagnetic showers* are created by electrons, which do not travel far before scattering in interactions with other particles in the ice and creating an abundance of localized light from these interactions. *Hadronic showers* are created by hadrons, such as the neutrons and protons, which are disrupted during the DIS interactions and daughter particles emitted from the nucleons (Section 1.3). Figure 2.3 shows the different expected Cherenkov light contributors from the various flavors and interaction types. In the detector, these different interaction outputs for  $< 100$  GeV neutrinos break down into two categories: cascade or track topology.

#### 2.2.2.1 Topology of Cascade-Like Events

Cascade-like events are signatures in the detector that are localized or have a spherical structure relative to the detector spacing. This structure occurs due to the hadronic and electromagnetic

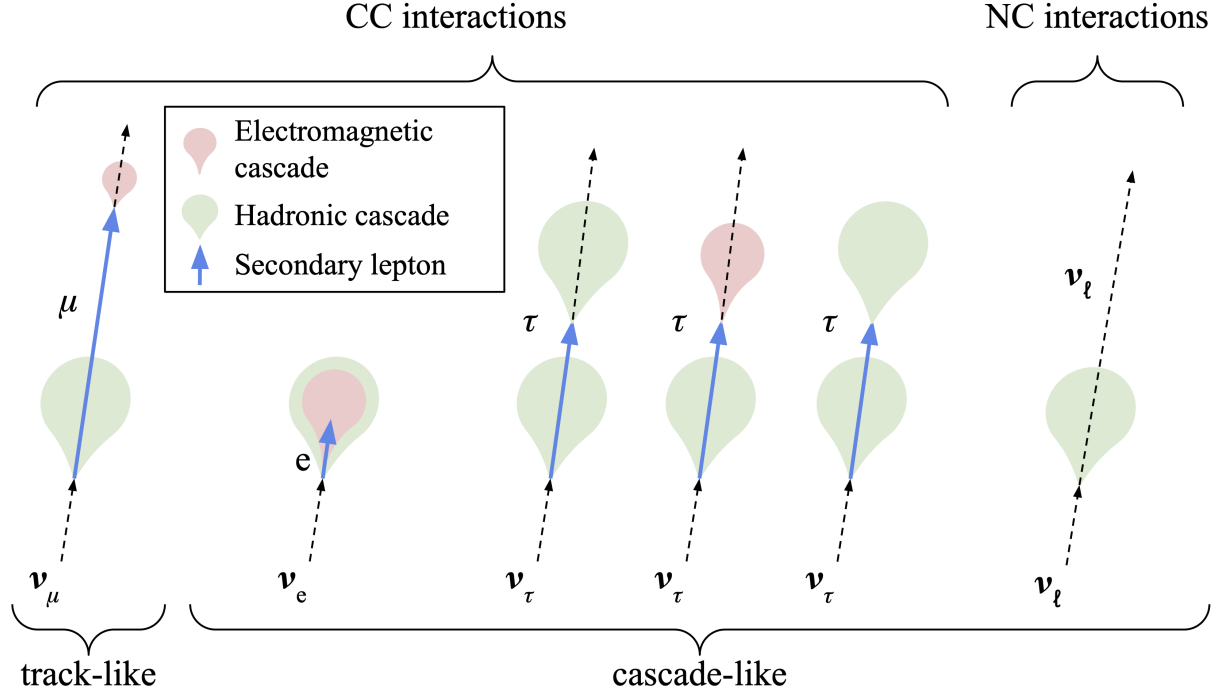


Figure 2.3: Different types of interactions, their secondaries, and the expected signatures for incident neutrino energies  $< 100$  GeV (modified from [11]).

showers creating a cascade of charged particles in all directions, with the starting vertex at the center. This makes it difficult to resolve the direction the incident neutrino is traveling at these energies due to the small length scale of the shower compared to the detector spacing.

Events that have a cascade signature are all  $\nu$  flavor NC events,  $\nu_e$  CC events, and most  $\nu_\tau$  CC events at these energies (Figure 2.3). The NC events have an invisible outgoing neutrino, thus their only signature in the detector will be the hadronic shower from the disruption of the nucleus. The  $\nu_e$  CC events have an outgoing electron, which will create the electromagnetic shower, giving the event the cascade-like signature.

For the  $\nu_\tau$  CC events, there are a few possibilities for the expected interactions. This is due to the fact that the tau lepton has heavy mass and thus has a shorter lifetime than both the muon and electron, so it is expected to decay quickly in the detector. Thus, we must examine the secondaries that the tau lepton decay will emit, which has three main expected decay modes. The largest fraction of the branching ratio is where the tau decays into various hadrons to create a hadronic shower, with

a branching ratio of 65% [23]. The next most dominate decay is where the  $\tau$  decay produces an electron ( $\tau \rightarrow e \nu_\tau \nu_e$ ) with a branching ratio of 18% [23]. This would result in an electromagnetic shower after the tau decay. At energies  $< 100$  GeV, the tau is expected to only survive on the order of millimeters. Thus, the potential *double bang* signature of the two cascades (one from the original  $\nu_\tau$  interaction and the  $\tau$  decay) is not resolved in the detector, since the vertex of these signatures would overlap and be indistinguishable from a single cascade.

The last fraction of the tau decay branching ratio is the tau decaying into a muon secondary ( $\tau \rightarrow \mu \nu_\tau \nu_\mu$ ), with a branching ratio of 17% [23]. This would lead to a track-like signature (Section 2.2.2.2), but with the expected  $\nu_\tau$  CC component in the analysis sample comprising of  $< 5\%$  of the total sample (Table 5.2), this is not expected to have a large effect in this work. Thus, the majority of the  $\nu_\tau$  CC events will have a cascade topology and will be considered as such for this analysis.

#### 2.2.2.2 Topology of Track-Like Events

Track-like signatures in the IceCube detector are events that have a deposit of light whose source travels across the detector. The primary example of this is the outgoing muon from the  $\nu_\mu$  CC interaction (Figure 2.3). The energy of the outgoing muon affects both the number of photons emitted and the distance the muon travels before decaying. From an initial neutrino energy of 100 GeV, the outgoing muon is expected to travel less than 400 m in the ice. Thus, at the lowest energies, a track-like event can appear indistinguishable from a cascade-like event. As the track length increases, these events are easier to reconstruct the direction of travel, since there is a longer signature across the detector.

### 2.3 Optical Properties of the Ice

IceCube takes advantage of the large volume of natural glacier ice at the South Pole. This is what allows IceCube to instrument such a large area, but also creates challenges to handle the natural irregularities. The key optical properties that IceCube needs to understand are the expected scattering and absorption lengths in the ice so the simulated photons can be accurately propagated

in the ice. When the photon scatters, it changes direction which increases the difficulty to trace the photon back to its origin. When a photon is absorbed, it interacts in the ice, so it is no longer detected by the DOMs. Achieving a better understanding of the ice is done by using the LED flashers on the DOMs to perform calibration [39, 41], along with using laser *dust loggers* to record back-scattering in the ice [42].

### 2.3.1 Bulk Ice Properties

The *bulk ice* refers to the unaltered ice used as the detector volume for IceCube. It is used to contrast the sections of refrozen columns of ice, the *hole ice* (Section 2.3.2), where the detector strings were inserted. The optical properties of the bulk ice are not uniform across the detector, but their variations have been explored and their dependencies accounted for in the simulation's *ice model*. The ice model is used when propagating the simulated photons in the ice, the process which is described in Section 3.3. The ice model used for this work is called *SPICE* version 3.2.1, with more details in [11]. The following sections will describe important features of the ice model. While the model is used during the simulation, the parameters for the bulk ice's scattering and absorption coefficients are also systematic parameters that are free during the analysis (Section 6.4). These apply a multiplicative factor to increase/decrease the amount of the absorption or scattering in the bulk ice, which in turn affects the number of photons expected to be detected by the DOMs.

#### 2.3.1.1 Depth Dependence

Figure 2.4 shows the variation of the scattering and absorption properties of the ice as a function of depth. The location of the DOMs on the IceCube strings are highlighted in pink, starting at 1450m, and extend down to overlap with the DeepCore DOM depths. One reason that the detector starts at depths greater than 1450 m is to utilize regions of more pure ice.

DeepCore's main volume instruments even deeper ice, to avoid the *dust layer*. The dust layer is the region with increased scattering and absorption of the photons found at depths of 2000 m to 2100 m (Figure 2.4). There are other, more subtle variations of the ice too that are  $xy$  dependent.

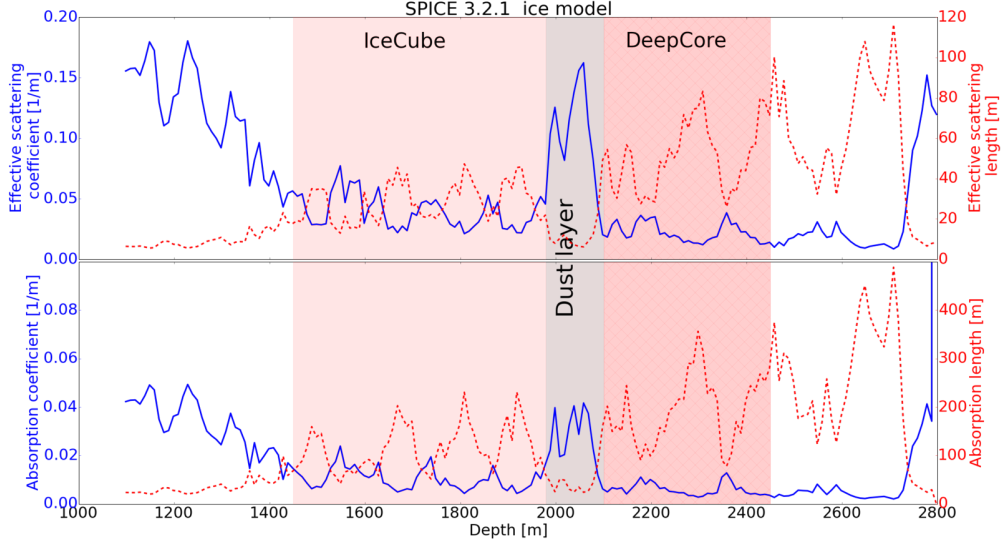


Figure 2.4: Scattering and absorption coefficients (left scale in blue line) and lengths (right scale in red) as a function of depth. Smaller values of the coefficients indicate cleaner ice [12].

### 2.3.1.2 Azimuthal Dependence

The depth dependence is not perfectly uniform across each  $z$  position, due to the structure of the bedrock. Rather, there is a *tilt* to the horizontal layer structure where some of the ice properties are shared up to  $\pm 56$  m apart [41]. This creates an azimuthal dependence, where position relative to the angle in the  $xy$  plane has an impact on the ice properties. There is also an additional azimuthal dependence seen in a preferential direction for the scattering in the ice, due to the glacial ice flow [13]. Figure 2.5 shows the relative charge differences seen between a simulation that does not account for these differences with the data, as a function of the azimuth angle.

### 2.3.2 Hole Ice Properties

The *hole ice* refers to the refrozen cylindrical volume around the strings, which has different optical properties than the undisturbed bulk ice. Even though the same melted ice was used to refreeze in the drill holes, there is a difference observed in the region around the DOMs [43]. One potential reason for the different behavior is that the hole ice contains more air bubbles than the bulk ice [43]. During the refreezing processes, gases were released and without the years of compression

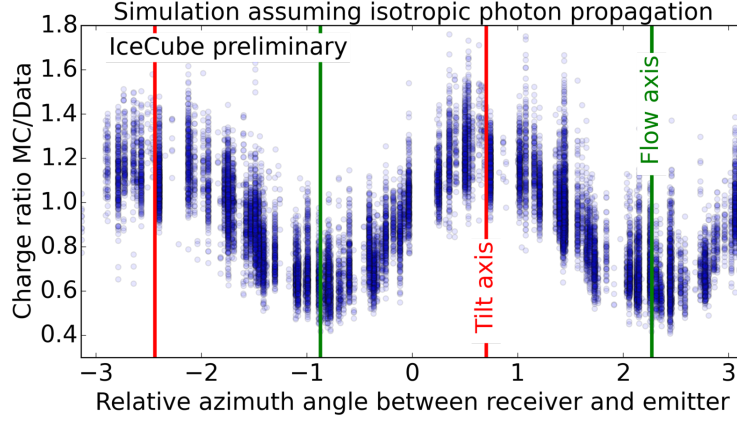


Figure 2.5: The charge difference between MC and data measured as a function of the azimuth angle, showing the impact of the tilt and flow anisotropy [13].

that the bulk ice experienced, a column of bubbles has been observed in the frozen ice using in-ice cameras [43]. This creates a local scattering effect that is independent from the observed bulk ice effects.

The hole ice is accounted for in the simulation using a function to adjust the angular acceptance of the photons on the DOM. It is parameterized using two variables derived from a principal component analysis that describe the angular acceptance expected in the hole ice,  $p_0$  and  $p_1$ . These two variables are varied and multiple sets of simulation are generated with the different values of  $p_0$  and  $p_1$  to observe their effects and systematic parameters that are accounted for during the analysis (Section 6.4).



## CHAPTER 3

### LOW ENERGY SIMULATION AND DATA SAMPLE PROCESSING

#### 3.1 Monte Carlo Samples

The Monte Carlo (MC) samples are generated to simulate a wide range of particles and interaction types. There are three main categories of simulation used in this analysis: events triggered by neutrinos, atmospheric muons, and noise. Multiple samples, or sets, were generated for the neutrino and atmospheric muon simulations, with each set having different settings for the detector systematics. For this work, there are two main purposes for the generated neutrino samples: analysis and neural network training. The different uses and goals for these samples are described below, followed by Section 3.2 which describes how each of the types of MC are generated. The last section describes the shared processing for both MC and data to create samples that are pure enough to have the final reconstruction applied (Chapter 4). A flow chart of the processing pipeline for the MC and data is in Figure 3.1.

##### 3.1.1 Neural Network Training Samples

Two sets of neutrino samples are independently generated and specifically optimized for creating neural network training samples. Unlike the typical goal of analysis MC sets, which is to simulate the expected distributions of the data, these MC sets are generated with the same physics and interaction modeling but their final distributions are intended to be flat with respect to a single variable. The aim is to create samples that are unbiased with respect to the specific variable the network is being trained to reconstruct, and that ideally extend past the target reconstruction region to avoid artificial bias at the edges of the sample.

The two special samples generated for training are intended to have the first sample uniformly distributed in energy (per GeV) and second sample uniformly distributed in zenith angle (per radian). The zenith network is trained for the zenith angle, not cosine of the angle, due to

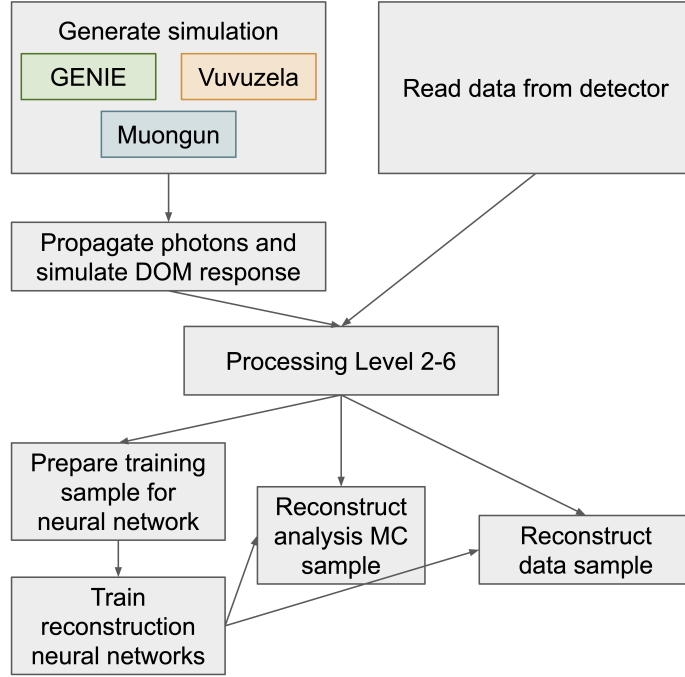


Figure 3.1: Flow chart describing the processing for simulation (left) and data (right).

training complications at the boundary of the sample, further explained in Section 4.2.5.2. The uniform energy training sample is also used to make the optimized training samples for the track classifier (Section 4.2.5.4) and muon classifier (Section 4.2.5.5). The vertex training sample uses an unoptimized sample, also derived from the energy training sample (Section 4.2.5.3).

These samples undergo the same processing as the analysis sample up until Level 6 (Section 3.4.5). After Level 6, specific training optimization cuts are made to each training sample (Chapter 4). These samples are then used to train the networks, but are not used for any further purposes in the analysis.

### 3.1.2 Analysis Sample

The goal of the analysis sample is to generate a MC sample that is as close to the expected data distribution as possible. While there is some importance sampling implemented to minimize the computational burden when simulating these events, there is close accounting of the sampling during generation so that weighting schemes can be applied to match the analysis distribution to the

expected flux. This sample, referred to hereafter as the *analysis sample*, will be used to evaluate the robustness of the analysis pipeline, project the sensitivity of our measurement of the oscillation parameters, assess the impact of the systematic parameters, and assess the goodness of fit (Chapter 6). The analysis sample is also used as the testing sample for the neural network reconstructions to gauge their performance on a data-like distribution that is completely independent of the training samples (Section 4.3).

## 3.2 Monte Carlo Generation

### 3.2.1 Neutrino Events

The neutrino MC samples are generated using GENIE (version 2.12.8), which is a neutrino MC generator with an emphasis on neutrino interaction physics at the few-GeV energy range including QE, RES, and DIS interactions (Section 1.3) [32]. The cross-sections for all flavors of neutrinos and antineutrinos are pre-calculated by GENIE from 0-10 TeV for the analysis sample. The neutrinos are generated using a given *generated spectrum*, which does not have to necessarily match the expected physical flux since it will be weighted using a given model.

The neutrino simulation is run in parallel batches split by energy to help distribute the computational load of the simulation. The neutrinos are uniformly injected in a manually-defined cylindrical volume centered on DeepCore with a defined spectral index. This is optimized such that the generation volume is adjusted with the neutrino energy range of each batch so that events are concentrated in the DeepCore region. This will also improve the chances that the events make it to the final level of processing. The ratio of generated neutrinos and antineutrinos is also defined as 70% and 30%, and kept constant between both the analysis sample and the training samples.

For the analysis sample, the optimized settings are shown in Table 3.1 [44]. The analysis sample uses a constant spectral index of  $E^{-2.0}$  for all flavors and energy ranges. The events are then weighted to match the expected model flux, so that the computational optimization does not jeopardize the physics result.

The training samples for the reconstruction training specifically avoid matching the expected,

Flavor	Energy (GeV)	Radius (m)	Length (m)	Events/file	Files
$\nu_e + \bar{\nu}_e$	1-4	250	500	450k	650
	4-12	250	500	100k	
	12-100	350	600	100k	
	100-10000	550	1000	57.5k	
$\nu_\mu + \bar{\nu}_\mu$	1-5	250	500	408k	1550
	5-80	400	900	440k	
	80-1000	450	1500	57.5k	
	1000-10000	550	1500	6.7k	
$\nu_\mu + \bar{\nu}_\mu$	1-4	250	500	1,500k	350
	4-10	400	900	300k	
	10-50	350	600	375k	
	50-1000	450	800	200k	
	1000-10000	550	1500	26k	

Table 3.1: Simulation settings used to generate neutrinos for the analysis sample.

physical model distribution since the goal of these samples is to create a flat distribution and weighting is not used when training. Thus, the  $\nu_\mu$  and  $\nu_e$  events generated for the training sample have various energy ranges and cylinder volumes used. The correlation between the energy and cylinder volume remains similar to the analysis sample (Table 3.1), but the number of events and files generated and even the spectral index vary significantly to create the artificially flat sample.

Both the analysis sample and the training samples are generated with the nominal simulation settings, with detector systematics set to their calibrated defaults. The analysis sample is also generated as various *systematics sets* that are simulated with variations in five of the detector systematic settings (relating to DOM efficiency, hole ice, and bulk ice). More details about these settings and how they are integrated into the analysis are give in Section 6.4.4.

### 3.2.2 Muon Events

The muon MC samples are generated with MuonGun, an IceCube software package that simulates atmospheric muon interactions [45]. The muons also use a generated spectrum that is optimized for computational efficiency and later weighted to an expected physical flux.

The muons are generated on a defined cylindrical surface to determine their starting position, such that the generation surface is slightly larger than the IceCube detector volume. Muongun then

propagates the muons inward toward the detector. There is also an inner cylinder defined that the muons must intersect. In the case of the oscillation sample, this inner surface is defined as the DeepCore array to ensure that only muons that go through the DeepCore detector will be simulated.

The unweighted muon flux used for generation is tuned to match some target distribution to further improve efficiency of generating time-intensive simulations that will be relevant to the final analysis. For the oscillation sample, it is optimized to generate muon events that pass cuts up to Level 5 (Section 3.4.4) and a scaling is applied to compensate for events that rarely make it through the optimization.

Three categories of muon simulation were generated. The first was produced without target volume and/or filtering applied, to tune the early (Level 3-5) cuts. The second category is the muon sample with the baseline detector settings, optimized to pass the cuts up to the Level 5 processing, so that computational time is saved when generating a nominal sample to account for the expected atmospheric muon contamination (Section 5.4.3). The last category is a set of muon samples that have various shifts in the detector systematics (Section 6.1), to explore the effect of detector systematics on the muon distribution.

### **3.2.3 Noise Hits**

Noise hits, or *dark noise*, which describes DOMs triggered by something other than a Cherenkov photon from an external source, are expected in every event. The noise can originate from various sources that causes an electron emission in the photocathode, including thermal emission, electronic noise, or radioactive decays in the DOM components [39]. Noise hits are not a separate type of event, but rather are hits added into the pulse series for the neutrino and muon events to model the expected data pulse series more realistically. These simulated noise hits are generated with Vuvuzela, an IceCube software module written to simulate the expected noise rate of the DOMs [46].

### **3.2.4 Noise Events**

Pure noise events occur in the detector, when the event triggers from noise hits alone, not a neutrino interaction or atmospheric muon. These events are simulated using the Vuvuzela software, and requires a continuous, time-intensive simulation of all detector components until a random noise coincidence triggers the DeepCore filter [11]. These events are very rare, though they are more common in the low energy IceCube samples due to the low threshold for event triggering due to targeting low energy events. Various cuts and filters are applied to specifically target these events, such that their final presence in the analysis sample is negligible (Table 5.2).

### **3.3 Propagating MC Photons and Detector Response**

Once generated, the process of propagating the different particles and accounting for the detector response is similar between the different simulation types. The interactions of the primary particles are simulated so that secondary particles and their properties are recorded. The Cherenkov photons, from the initial interaction and the secondary particles' interactions, are then propagated through the ice and recorded if they intersect a DOM. The propagation is dependent on the ice model to determine local scattering and absorption properties (Section 2.3.1). The DOM response is taken into account by converting photons from the simulation into the hits that the DOM would record. This step has dependence on the PMT efficiency in the DOM, the thermal noise, and the hole ice (Section 2.3.2). Finally, the online data filter algorithms are applied to the simulation to determine which of the filters the event would pass.

### **3.4 Data Processing**

At this stage, both data and MC undergo the same processing. The first processing level (Level 2) applies the DeepCore filter, cleans the pulse series, and performs the first fast cuts to select events in DeepCore. Level 3 focuses on atmospheric muon and noise event removal to improve data/MC agreement. Level 4 uses machine learning methods to further reduce atmospheric muon and noise events. Level 5 targets coincident and corridor events using new algorithms and tighter

machine learning cuts. Level 6 calculates variables needed for the final analysis cuts. After Level 6, the samples are ready to be prepared for the reconstruction neural network training or to have the reconstruction applied to the analysis sample. The following sections outline the processing pipeline that has been developed by various other researchers in the IceCube collaboration [44].

### 3.4.1 Level 2

For the purpose of the oscillation analysis, the DeepCore filter trigger is the first requirement, to ensure that events happen in the DeepCore region of the detector. The DeepCore filter requires at least 3 HLC hits (defined below) in the DeepCore array within a sliding time window of  $2.5 \mu\text{s}$ . Level 2 cuts any events that do not pass the DeepCore filter.

Level 2 also applies an initial cleaning algorithm to the pulse series, to remove noise pulses that are not correlated with the neutrino event interaction. This algorithm uses the two categories of hits, the Soft Local Coincidence (SLC) and the Hard Local Coincidence (HLC). When a DOM records a hit, the PMT waveform is digitized and read out if the signal surpasses the threshold. An HLC hit is when there is a nearby coincidence with a neighboring hit DOM on the string (within 2 DOMs above or below) within  $1 \mu\text{s}$ . In contrast, an isolated hit is called an SLC hit, with this naming convention used for historical reasons. The SLC hits are useful for reconstructing low energy events, which produce less light for the detector to record, but are also more likely to be dark noise. Thus, since SLC hits are included in low energy events, additional algorithms are necessary to clean the pulse series. The Seeded-RT (Radius and  $\Delta\text{Time}$ ) cleaning algorithm uses the HLC hits as its starting list to seed the radius and time cleaning procedure. It loops through all of the HLC hits and searches for nearby SLC hits, checking if they occur within a certain radius and time window relative to the HLC hit. For the DeepCore array, it must be within 75 m and 500 ns and for the IceCube array, it uses a looser 150 m and 1000 ns to account for the larger distance between DOMs. These SLC hits within the radial and time window are added to the kept series, and the procedure is re-run to see if any remaining pulses are now near enough with the new series. This is repeated until no new pulses are added to the series, and all remaining pulses outside of the clean

pulse series.

In addition to the Seeded-RT algorithm, a static time window cleaning is applied to the pulse series. This removes all pulses that are more than  $4\ \mu\text{s}$  before or  $6\ \mu\text{s}$  after the DeepCore trigger. These two algorithms, the Seeded-RT and time window cleaning, define the cleaned pulse series that is used as input for many Level 3-6 algorithms as well as the neural network reconstruction (Chapter 4).

Level 2 is also used to run another fast event selection algorithm that calculates the center of gravity (COG) using hits in the DeepCore fiducial region in the Seeded-RT cleaned pulse series. The COG calculation is then used to explore any remaining hits in the veto region and remove these events if their derived speed relative to the COG indicates the hit could be causally connected to a muon crossing the detector.

### 3.4.2 Level 3

The goal of Level 3 is to apply fast algorithms that remove background muons, pure noise, and event types that are not simulated in the MC such as muon bundles and coincident events. Bundles are multiple muons that trigger in the detector in the same time window. MuonGun injects single muons, not multiples (i.e. bundles), so these are not accounted for in the simulation sample. Muon bundles are easy to remove using veto cuts, so there is not a concern of them in the final sample once these cuts are applied here. Coincident events are events that have two independent particles that interact in the detector within the same triggered time window, such as a neutrino and an atmospheric muon. Since these particles are simulated separately, they are also not accounted for in the MC sample.

To remove noise, events are cut if they do not meet the minimum number of hit DOMs ( $< 6$  cleaned hit DOMs), the minimum number of hits in a sliding time window (to check for local coincidence), or the minimum number of hits in the fiducial region ( $< 2$  hits). To target muons, the cuts focus on the position of the hit DOMs. This includes cutting events if there are too many hits in the veto regions ( $> 7$  hits the veto region or an overall ratio of 1.5 times the veto hits compared to



fiducial), cutting on a quick estimate of the starting vertex  $z$  position to loosely constrain it within DeepCore ( $< -120$  m) which indicates particle production at the neutrino interaction vertex, and demanding that most of the hits are concentrated in the first 600 ns of the event (cut if the ratio is  $< 0.37$ ). To remove coincident events, the time between the first and last pulse is calculated and constrained. This is applied both to the uncleaned ( $< 13000$  ns) and cleaned pulse ( $< 5000$  ns) series, with a tighter constraint applied to the cleaned pulses. These cuts remove 95% of the background atmospheric muon events and 99% of the pure noise events from the Level 2 sample [44].

### 3.4.3 Level 4

Now that the obvious background events are removed, machine learning algorithms can be applied to further identify and remove background events. There are two classifiers, the first has been trained to classify pure noise events and the second to classify atmospheric muon events at this level.

A noise classifier is trained using Level 4 MC to distinguish pure noise events from neutrino events. It uses a Boosted Decision Tree (BDT) that takes in variables calculated at Level 3 and 4 to distinguish the probability that an event is neutrino-like rather than noise. The cut that is applied on the BDT score ( $> 0.7$ ) reduces the noise rate by two orders of magnitude (Figure 5.5) while keeping 96% of neutrino events [44] between Level 3 and 4.

The muon classifier also uses a BDT, which was trained to distinguish atmospheric muons from neutrino events. The training sample that is used is a subset of real detector data processed up to Level 4 for the atmospheric muons, since the Level 4 sample is expected to be composed of 99% atmospheric muon events. The GENIE MC is still used as the training sample for the neutrino events, also processed up to Level 4. The chosen cut on the BDT score ( $> 0.65$ ) removes 94% of atmospheric muon events and keeps 87% of neutrino events between Level 3 and 4.

### 3.4.4 Level 5

Level 5 focuses on removing the atmospheric muon events that have eluded the previous cuts. The signatures of the remaining muons are distinct from the clear muon-like signatures already cut. Level 5 applies two categories of cuts to remove the remaining muon events: starting containment and *corridor* cuts.

Starting containment is a check to ensure the event begins inside of the DeepCore fiducial volume. At Level 5, three fast estimates of the vertex are used to to apply a cut on the radial ( $< 150$  m) and vertical ( $-490 \text{ m} < z < -225 \text{ m}$ ) containment.

Corridor muons are events that sneak through the gaps between the hexagonal structure without hitting any strings until the DeepCore region. An algorithm is run at Level 5 that searches for DOMs hit along the corridor regions, using a COG calculation to determine the closest string associated with it. It then searches for DOMs along that corridor using a cylindrical spatial and time window to determine if light was detected in the region associated with a corridor event. If two or more DOMs are found to have light associated with the corridor region, the event is removed.

The last cuts applied at Level 5 use the Level 4 classifiers again, further tightening their cuts (noise  $> 0.85$  and muon  $> 0.9$ ).

### 3.4.5 Level 6

For the purposes of this analysis, Level 6 includes some fast algorithms that are run to determine variables that are later used in final analysis cuts to improve data/MC agreement. No cuts are applied at this level, only preparation for the analysis cuts after the reconstruction has been applied. This includes the variables calculated to further cut coincident and corridor events. Section 5.2.1.1 and 5.2.2 contain more information about which variables are used and their effect on data/MC agreement. This is where my work begins, creating selections for each of the training samples and analysis samples after Level 6.

## CHAPTER 4

### FAST LOW ENERGY RECONSTRUCTION USING CONVOLUTIONAL NEURAL NETWORKS

Previous low energy reconstruction methods for IceCube use likelihood methods to compare the observed pulse distributions in the DOMs to pre-calculated tables containing the number of expected photons. One such algorithm, Retro Reconstruction [10], compares the observed and expected distributions for the DOMs for various scenarios of neutrino energy, interaction vertex, and direction. The algorithm then calculates the likelihood of each hypothesis and finds the best match. This method takes 40 seconds on average to reconstruct a low energy neutrino event.

IceCube has a large data sample of atmospheric neutrinos to reconstruct, with about 100,000 events expected per year, even after undergoing the processing cuts up to Level 6. The low energy reconstruction must handle an even greater sample of Monte Carlo (MC) generated to explore effect of detector systematics and backgrounds, reconstructing  $O(10^8)$  MC events. The Retro Reconstruction method takes months to accomplish this task, even while leveraging a computing cluster. A solution was needed that would speed up the reconstruction time, without losing accuracy in the difficult-to-reconstruct low energy regime.

This is the motivation behind the Fast Low Energy Reconstruction using Convolutional Neural Networks (FLERCNN). FLERCNN is written with the intention to reconstruct neutrino events for a  $\nu_\mu$  disappearance oscillation analysis and informs the choices of variables that FLERCNN is optimized to reconstruct. To complete the oscillation analysis, one needs the  $L/E$  and flavor of the incoming neutrino. FLERCNN reconstructs zenith angle ( $\cos(\theta_{zenith}) \approx L$ ), energy ( $E$ ), and scores how confident it is that an event is track-like (predominately  $\nu_\mu$  CC at these energies) (Section 1.2.2). In addition, previous oscillation analyses on IceCube have benefited from cuts that ensure the neutrino events start in the DeepCore region (since its instrumentation is optimized for lower light events) and a cut to further remove background atmospheric muons [17, 47]. FLERCNN is trained with the capability to apply these two purity cuts in addition to reconstructing the necessary

analysis variables.

The following sections will discuss how FLERCNN was specifically optimized to handle the low light events at the 10-GeV scale, the training procedure for each of the 5 FLERCNN regression and classification networks, and the performances. The network is trained and run using Keras software with a Tensorflow backend [48, 49].

## 4.1 Architecture

Convolutional Neural Networks (CNNs) are known for their capabilities in image recognition [50, 51, 52]. They function by using a kernel, or window, to span nearby pixel inputs and having them share their weights while training. CNNs have translational equivariance, which means they can identify objects regardless of their overall location in the image [53, 50]. This is useful for identifying similar events in different parts of the detector. The IceCube detector is a fairly regular array of optical modules, equating the DOMs to the pixels in an image. The input from the DOMs equate to the RGB (red, green, blue) input from typical images. This has already been proven successful, as IceCube has used CNNs to reconstruct the energy and direction of TeV to PeV events [20]. That high energy CNN loses resolution at the 100s of GeV scale, though, and resolution down to the 10-GeV scale is needed to resolve oscillation signatures. The architecture of the high energy CNN was used as a starting point to motivate the design for FLERCNN's architecture and input features.

To focus the CNN on the well-resolved low energy events, FLERCNN only uses strings in and around the DeepCore array. The CNN is given input from all 8 DeepCore strings and the 19 center-most IceCube strings (Figure 4.1). This accounts for the first dimension in the input array, which corresponds to an unordered string index in either the 8 DeepCore strings or 19 IceCube strings. The second dimension in the input array is the relative DOMs index, ordered by their position on the string. The last dimension is five variables that are chosen to summarize information of all photons that hit a given DOM in the event time window:

- Sum of the charge

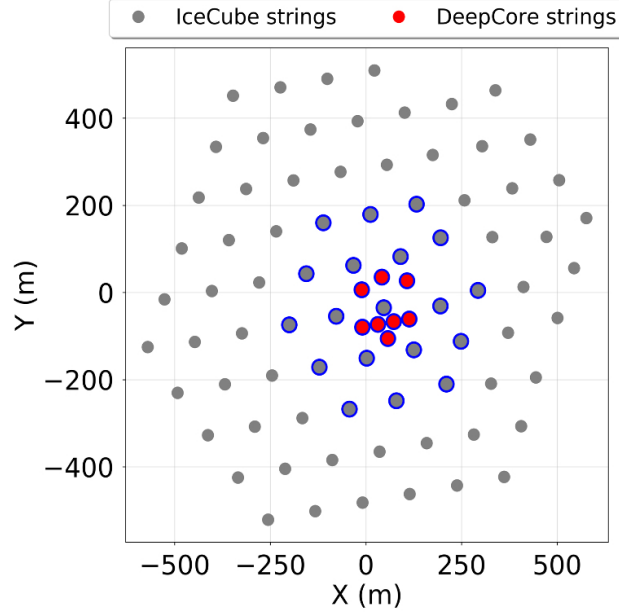


Figure 4.1: Top view of the IceCube detector, with DeepCore strings in red, IceCube strings in gray, and a blue highlight on all strings used for the low-energy CNN.

- Time of the first hit
- Time of the last hit
- Charge-weighted mean of the hits
- Charge-weighted standard deviation of the hits

More details on these input features, how they are calculated, and their re-scaling is in Section 4.2.2.

The inputs from the DeepCore and IceCube strings are recorded in separate arrays, and given separately to the CNN since they have different  $z$ -spacing between DOMs, resulting in a two-branched network architecture for the convolutional layers (Figure 4.2). The CNN kernel spans only the vertical, or  $z$ -depth, of the strings. Since the DeepCore strings are deployed in an irregular array, no convolution is applied in the  $xy$ -plane. The same convolutional kernel applied to the IceCube strings gave satisfactory results, and thus both branches of the CNN use the same kernel applied in  $z$ -depth only.

The network structure in Figure 4.2 shows the 8 convolutional layers in green, with their kernel sizes specified each layer. They vary between 1-7 DOMs included in the kernel, always using an odd number so the same number of DOMs above and below the target DOM are taken into account. Every layer also included a batch normalization and dropout layer. The batch normalization is used to improve optimization and stability of the training [54]. The dropout layer randomly masks 20% of the nodes so that they do not contribute to the training during that forward pass. The nodes that are masked are changed every iteration through the training to help make the network more stable and more generalized. Two max pooling layers are included at layer 1 and 4, since max pooling is known to improve CNN recognition if the object is skewed or rotated slightly within the image [53]. After the 8th convolutional layer, both branches of the network are flattened into a single layer and concatenated together into a single, fully connected dense layer. The batch normalization and dropout is applied once more on the entire input before the network gives its prediction of the output.

The network is trained by using the events in the training sample to input into the network. The forward pass occurs by taking the event's inputs and running it through the existing network's configuration of the nodes and their weights. After the last layer, an *activation function* is used to transform the output. Then, a chosen *loss function* is used to score how accurate the CNN's reconstruction is. The backpropagation step then occurs, updating the weights using the gradient of the loss. The next event is then fed into the network for another forward pass, and this process continues until there is a full pass through the training data, or an *epoch*. Note that this process can also be done in batches, where all FLERCNN architectures use a batch size of 256 events for training.

The chosen loss for each network is in Table 4.1. The energy FLERCNN network uses the mean absolute percentage error loss, such that

$$\text{Loss} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{t_i - r_i}{t_i} \right| \quad (4.1)$$

where  $n$  is the number of events,  $t_i$  is the true value and  $r_i$  is the reconstructed (or estimated)

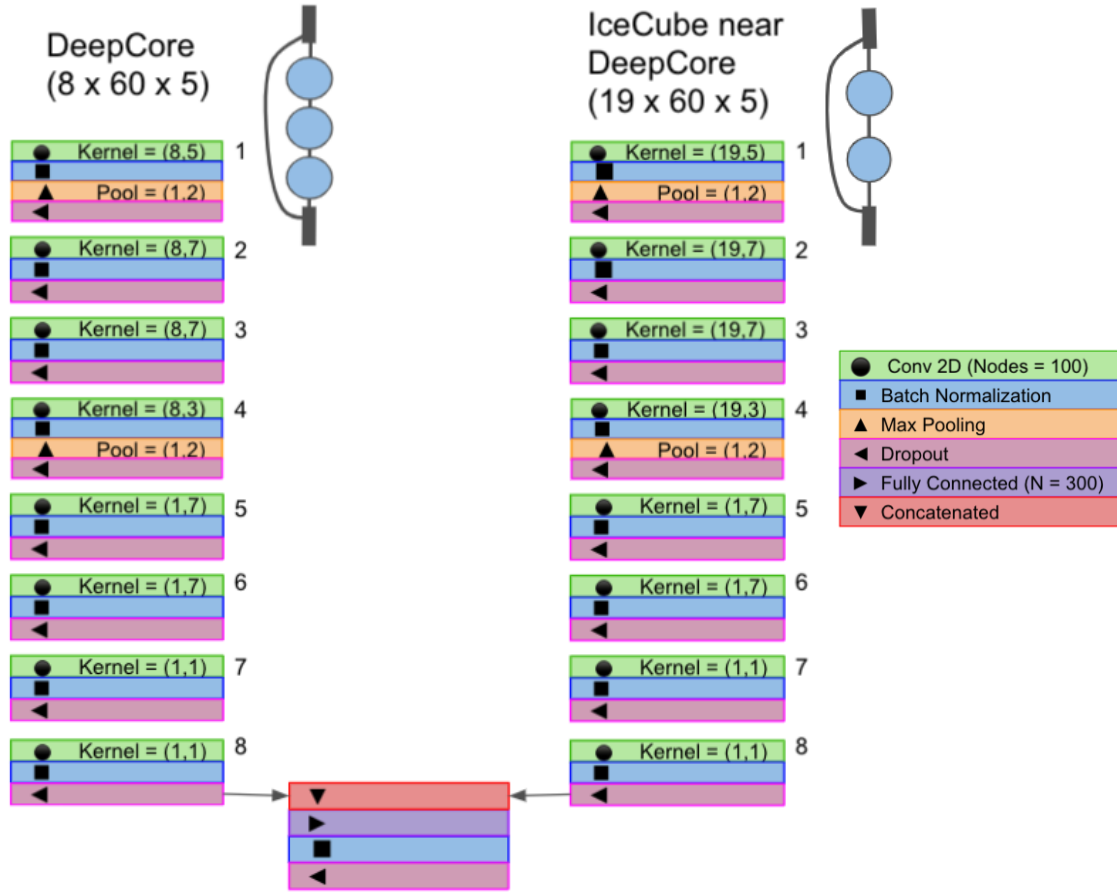


Figure 4.2: Detailed network architecture for FLERCNN with the legend for colored layers on the right. The notation at the top indicates that for each event, there are 8 DeepCore strings with 60 DOMs per string and 5 summary variables per DOM fed into one branch of the network. The other branch of the network sees the 19 IceCube strings with their 60 DOMs per string and 5 summary variables per DOM.

variable by the CNN. This was chosen carefully such that the energy network would give equal importance to reconstructing the low energy events (at a few GeV) as the high energy events (at 100 GeV). Since the  $O(10\text{GeV})$  events are the region where oscillation is expected, the percentage error ensures that the 10 GeV events are well-reconstructed without over-emphasis on the  $O(100\text{GeV})$  events.

For the vertex and zenith FLERCNN networks, these use a mean squared error loss:

$$\text{Loss} = \frac{1}{n} \sum_{i=1}^n (t_i - r_i)^2 \quad (4.2)$$

This uses the difference between the true ( $t$ ) and reconstructed ( $r$ ) values, taking the square so that larger differences are emphasized more in the loss.

Lastly, the classification networks use a binary crossentropy loss function for the output score:

$$\text{Loss} = -(t \ln(r) + (1 - t) \ln(1 - r)) \quad (4.3)$$

where  $t$  is still the true value, but it is now either a 1 or 0, and  $r$  is the classifier score.

CNN	Loss Function	Activation
Energy	Mean Absolute Percentage Error	Linear
Zenith	Mean Squared Error	Linear
Vertex	Sum(Mean Squared Error) for X, Y, Z	Linear
Muon	Binary Crossentropy	Sigmoid
Track	Binary Crossentropy	Sigmoid

Table 4.1: Describes the loss function and activation used for each network.

## 4.2 Preparing the Training Samples

The training samples account for the biggest difference between these networks. These were specially generated MC and their distributions were optimized to perform for each network's purpose. The generation of the optimized MC training samples is outlined in Section 3.1.1. These samples undergo the typical MC processing described in Section 3.4, but there are additional cuts and optimizations used to prepare the samples for training.

### 4.2.1 Data Agreement

There are a few data agreement measures that must be addressed at the level of charges. The first potential disagreement is that some of the IceCube data seasons only record quantized charge into 0.05 PE (photo-electron) steps. To make all data and MC appear uniform, all pulse charges are quantized (in 0.05 PE steps) before being used to calculate the variables for the CNN's input features.



A study performed by IceCube's Spencer Axani showed that data/MC disagreement is evident in charges recorded below 0.25 PE [14]. A plot from their technical report is included in here to show the expected mismatch between data and MC below 0.25 PE (Figure 4.3). The data and MC input into FLERCNN uses the "default discriminator" settings. Thus, all pulses below the  $< 0.25$  PE threshold are cut from both training and testing samples for FLERCNN to remove potential data/MC disagreement.

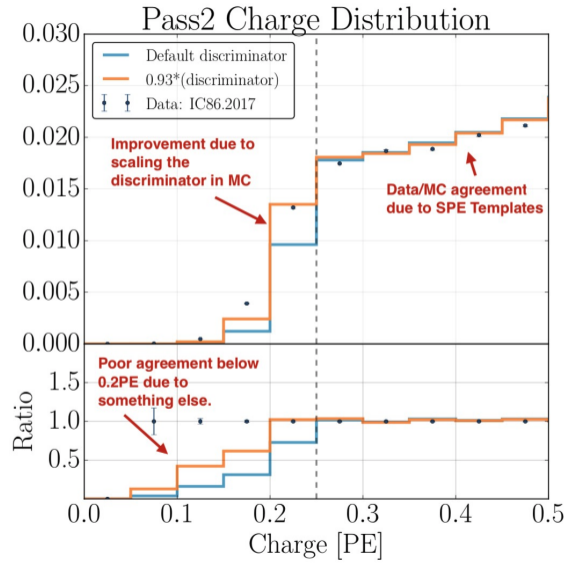


Figure 4.3: Shows the rate of the default discriminator in blue (from the SPE templates used to generate FLERCNN data and MC) compared to data as a function of charge. Agreement is seen above 0.25 PE [14].

#### 4.2.2 Preparing Input Features

In addition to using the cleaned pulse series (Section 3.4.1), all input features undergo a further selection criteria unique to the FLERCNN processing. This selection criteria helps to ensure that the pulses given to the CNN have good data/MC agreement and that the sample has minimal extraneous noise hits.

The first step done in this process is that all pulse times are shifted to be relative to the DeepCore trigger, so that the trigger time is at 0 ns. Any events with no DeepCore trigger or greater than 1 DeepCore triggers are then removed, since it is unclear how to define 0 (trigger time) for the event

<b>Input Variable</b>	<b>Shift</b>	<b>Division Factor</b>	<b>Requirement</b>
Sum Charge	None	25	$> 0.25$ PE; rounded to nearest 0.05 step
Time First Pulse	By trigger time	4000	Uses times in $[-500, 4000]$ ns
Time Last Pulse	By trigger time	4000	Uses times in $[-500, 4000]$ ns
Charge Weighted Mean	Uses times shifted by trigger time	4000	Uses times in $[-500, 4000]$ ns; pulses $> 0.25$ in 0.05 PE steps
Charge Weighted Standard Deviation	Uses times shifted by trigger time	2000	Uses times in $[-500, 4000]$ ns; pulses $> 0.25$ in 0.05 PE steps

Table 4.2: Describes the transformations that the input features undergo during preprocessing.

in these cases. After this shift is applied, only the pulses in  $[-500, 4000]$  ns window around the trigger are used to train the CNN. This corresponds to about 8 scattering lengths, and thus it is assumed that most pulses outside this window are noise. If the pulse times do not fall into the time window, they are not used for any of the five summary variables, including the sum of the charge for that DOM.

Some re-scaling is applied to the input variables to help the CNN train faster and make sure it treats all inputs proportionally using a uniform range [55]. Thus, the input variables are divided by static, near maximum numbers determined early on from histograms of the raw input so that the input features all fall near the  $[-1, 1]$  range. Each of the five summary variables are divided by a factor listed in the third column of Table 4.2.

### 4.2.3 Sample Cleaning Differences between Training Sets

As these networks were developed over the course of three years, along with developing unique samples optimized for specific training variables, there are some differences in the pre-processing for the FLERCNN networks. Table 4.3 summarizes which networks use the pre-processing methods described below.

The cuts in Table 4.3 are only applied to the described training samples. The cuts that will be used for testing and analysis samples are described separately in Tables 4.7 and 5.1.

CNN	ProbNu > 0.95	# Pulses ≥ 8	nDOM CNN strings cut	Containment
Energy	No	No	≥ 7	No
Zenith	No	No	No	Yes
Vertex	Yes	Yes	No	No
Track	Yes	Yes	No	No
Muon	Yes	Yes	≥ 4	No

Table 4.3: Describes the additional training sample cuts. For ease of viewing, cells highlighted in green indicate that the cut in that column was applied to the training sample for the network in that row.

#### 4.2.3.1 Neutrino vs. Noise BDT

To ensure that the events appear as "neutrino-like" as possible, a cut is applied on the BDT trained at Level 4 to distinguish noise events (Section 3.4.3). A tight cut is applied here at ProbNu > 0.95, to make sure the neutrino event used for training looks very neutrino-like. Even with this tight cut, this typically only removes a few percent neutrino events per sample, as earlier cuts up through Level 6 have already removed any suspicious-looking events. It should be noted that none of the CNN training samples include pure noise events, but all do include pulses in the neutrino (or muon) event that originate from noise, as described in Sections 3.4.1, 4.2.1, and 4.2.2. Note that this tighter noise BDT cut is only applied to the training sample, but is not applied during testing or analysis (see Table 4.7 for cuts used during testing and Table 5.1 for cuts applied for the analysis sample).

#### 4.2.4 Number of Pulses and DOMs Hit

In accordance with previous low energy IceCube reconstruction methods, a cut is applied to some CNN training samples to check that there are at least 8 hits in the pulse series. Newer models of the CNN optimized this cut by focusing on the CNN input itself, and thus a cut on number of DOMs in the CNN strings (8 DeepCore and center-most 19 IceCube strings) is introduced for some training samples. Some training samples apply one or both of these cuts, though it is suggested that future training samples for the CNN use just the number of DOMs in the CNN strings. Note that the

number of pulses cut is only applied to the training sample, but is not applied during testing. The number of DOMs hit is applied to all samples, from the training, testing, and final analysis sample (see Table 4.7 for cuts used during testing and Table 5.1 for cuts applied for the analysis sample).

#### 4.2.4.1 Containment

Lastly, most training samples relied on early processing (Level 2-5 described in Sections 3.4) to eliminate events with starting vertices outside of the DeepCore region. Training the zenith network, however, is particularly sensitive to seeing examples of full track signatures. Thus, a cut was applied so that all events started and ended in the CNN strings, using the true vertex, direction, and track length to determine. This cut is called "containment" in Table 4.3. Note that the containment cut is only applied to the training sample, but is not applied during testing or analysis (see Table 4.7 for cuts used during testing and Table 5.1 for cuts applied for the analysis sample).

#### 4.2.5 Training Each CNN

Each network has its own training sample, after applying the unique combination of cuts described in Section 4.2.3. Table 4.4 describes the content of each network's training sample, including the particles,  $\nu$  interactions, number of events, and energy range they cover.

To confirm the network does not overfit to the given training sample, 20% of each training sample is set aside to validate the network's progress after each training pass. At the end of each pass or evaluation of a single training file, the network is evaluated on both the training and validation sample using a specified *loss function*, which is defined by the user and chosen per network. Note that all FLERCNN variables were trained using multiple files containing a random subset of the training sample, so the loss function was evaluated at the end of each of these training files on a subset of the validation sample. The model is chosen at an *epoch*, or full pass through all the training files, that shows a near-minimum result in the loss along with no signs of overtraining. Overtraining would be indicated by the training loss continuing to decrease while

the validation loss stagnates or even increases, diverging from each other. Table 4.5 contains the information of the chosen epoch for each model, where epoch = raw training pass / number of files.

The *learning rate*, or step size of changing the gradient of the weights while training, is updated throughout the training process. The learning rate is decreased as the loss decreases so that it can settle at a minimum loss. The equation used to update the learning rate is  $R_n = 0.001^{\text{floor}(n/E)*D}$  where  $n$  is number of raw training passes through the files,  $E$  = number of passes between drops (LR Drop),  $D$  = drop factor (LR Drop Factor). The floor function rounds down to the nearest integer, so that the rate only changes every  $E$  passes. Table 4.6 has the settings used for the learning rate progression for each network, which varies  $E$  (LR Drop) and  $D$  (LR Drop Factor) specific to each network. The table has the final learning rate at the chosen model, which is determined by setting  $n$  to the raw training pass at the chosen epoch.

The differences between the CNN training sample size, composition, and step size (*learning rate*) progression are summarized in Tables 4.4, 4.5, 4.6 with more depth in the following sections.

CNN	Particle(s)	$\nu$ Interactions	# Training Events	Energy Range (GeV)
Energy	$\nu_\mu$	CC	9 million	1-500
Zenith	$\nu_\mu$	CC	5 million	1-300
Vertex	$\nu_\mu$	CC	5 million	1-500
Muon	$\nu_\mu, \nu_e, \mu$	CC, NC	7 million	$\nu$ 5-200, $\mu$ 150-5000
Track	$\nu_\mu, \nu_e$	CC, NC	5 million	5-200

Table 4.4: Describes the training samples for each network, which more details expanded in the following sections.

CNN	# Files	Raw Training Pass	Epoch
Energy	18	594	33
Zenith	10	700	70
Vertex	8	232	29
Track	12	192	16
Muon	12	240	20

Table 4.5: Describes the number of files used for training, the model chosen in terms of the raw training pass and epoch.

CNN	LR Start	LR Drop	LR Drop Factor	LR at chosen model
Energy	0.001	200	0.5	$2.5 \times 10^{-4}$
Zenith	0.001	80	0.8	$1.7 \times 10^{-4}$
Vertex	0.001	50	0.1	$1 \times 10^{-7}$
Track	0.001	50	0.1	$1 \times 10^{-6}$
Muon	0.001	-	-	0.001

Table 4.6: Describes the training samples learning rate and changes.

#### 4.2.5.1 Energy

A sample of more than 9 million unweighted,  $\nu_\mu$  CC events is generated to train FLERCNN for energy. The goal is to have a flat, unbiased energy distribution for training, between 1-500 GeV, since the target reconstruction region is 5-100 GeV for FLERCNN to use for the  $\nu_\mu$  disappearance analysis. Most energy bins (1 GeV per bin) have 20,000 events per bin, creating the "flat" sample distribution seen in Figure 4.4a.

Events with a low number of hit DOMs are especially difficult for the network to resolve, and thus, these events are removed both when training and evaluating the network (Tables 4.3 and 5.1). The cut requires that the CNN has at least seven DOMs with input information, requiring that there are seven DOMs hit between the 8 DeepCore strings and 19 inner-most IceCube strings for the event to be reconstructed with FLERCNN.

The total sample size is 9,026,280 events with 80% of the sample used for training and 20% for validation. To cope with the large file size that comes with storing the necessary pulse series information for training, the sample is split into 18 files and each file is loaded into the CNN during training in succession. The progression of the loss function with respect to each pass through the files is shown in Figure 4.4b. Since the training and validation are split into 18 files, there are 18 evaluations for the loss over the course of each epoch.

The settings for the learning rate progression are  $R_n = 0.001^{\text{floor}(n/E)*D}$  where  $E = 200$  and  $D = 0.5$  for the energy network. This combination of settings produces a fairly stable loss decrease over training progression (Figure 4.4b). Since each training pass is only through 1/18 of the data, the drop occurs at the equivalent of 11.11 epochs. This occurs since  $E = 200$ , meaning the learning

rate is set to drop after 200 training passes. Each training pass scans one of the 18 files, therefore it changes at the equivalent of  $200/18 = 11.11$  epochs or passes through the full data set. The model used was after 594 training passes, or 33 epochs. Figure 4.4b shows that the validation loss (in cyan) vs. the training loss (in blue) does not diverge at epoch 33, so no overtraining is evident. The performance of the energy CNN will be evaluated in Section 4.3.2.

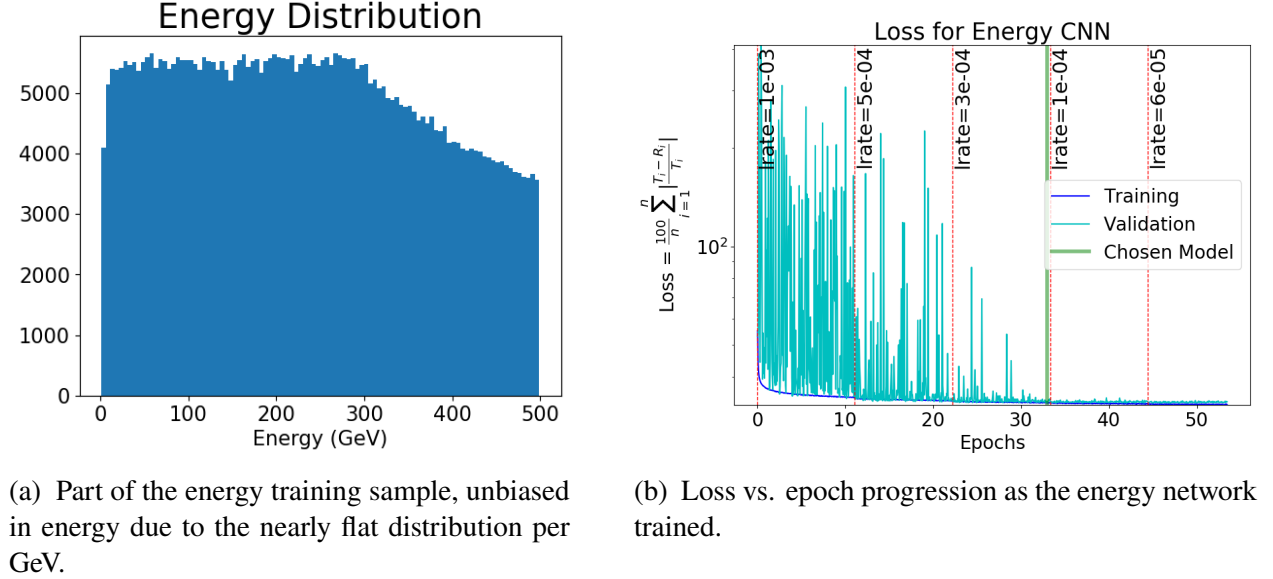


Figure 4.4: Energy training sample and loss progression.

#### 4.2.5.2 Zenith Angle

The direction that the neutrino is traveling is used to determine the distance that it has traveled ( $L$ ) after originating from cosmic ray interactions in the atmosphere. The zenith angle, relative to the  $z$ -vertical axis of the IceCube detector, denotes if the neutrino is traveling from the South Pole atmosphere toward the Earth ( $\theta_z = 0^\circ$ ) ranging to if it is traveling through the Earth's core toward the atmosphere above the detector ( $\theta_z = 180^\circ$ ). Determining the exact angle of travel at the GeV-scales, where only a few DOMs could be hit, is challenging for all reconstructions due to the limited information. Furthermore, track-like events with their hits distributed across multiple DOMs are typically easier for reconstructions to resolve. Thus, the training sample for training the zenith CNN is a  $\nu_\mu$  CC only sample with about 5 million events.

Optimizing the training sample is particularly tricky for the zenith sample as the sample cannot be generated with extended bounds in the same way that the energy network is. Even though the training bounds can not be extended in zenith angle, the energy range is still extended past the target analysis region, so that it has 1-300 GeV events. The distribution is flattened relative to the zenith angle (in radians) to train for the angle while minimizing bias (Figure 4.5a). This flat zenith training distribution was generated and tested by Dr. Shiqi Yu. A containment cut is applied to both the starting and ending locations of the muon from the  $\nu_\mu$  CC interaction that requires it to be within  $-505 \text{ m} < z < 500 \text{ m}$  and  $\rho_{36} < 260 \text{ m}$ , where  $\rho_{36}$  describes the radius from the center IceCube string (#36).

After these cuts, there are a total of 5,024,876 events with 80% of the sample used for training and 20% for validation. To cope with the large file size that comes with storing the necessary pulse series information for training, the sample was split into 10 files.

The settings for the learning rate progression are  $R_n = 0.001^{\text{floor}(n/E)D}$ , where  $E = 80$  and  $D = 0.8$  for the zenith network (Figure 4.5b). Since each training pass is only through 1/10 of the data, the drop occurs at the equivalent of 8 epochs. The model used was trained for 700 passes, or 70 epochs. The performance of the zenith CNN will be evaluated in Section 4.3.3.

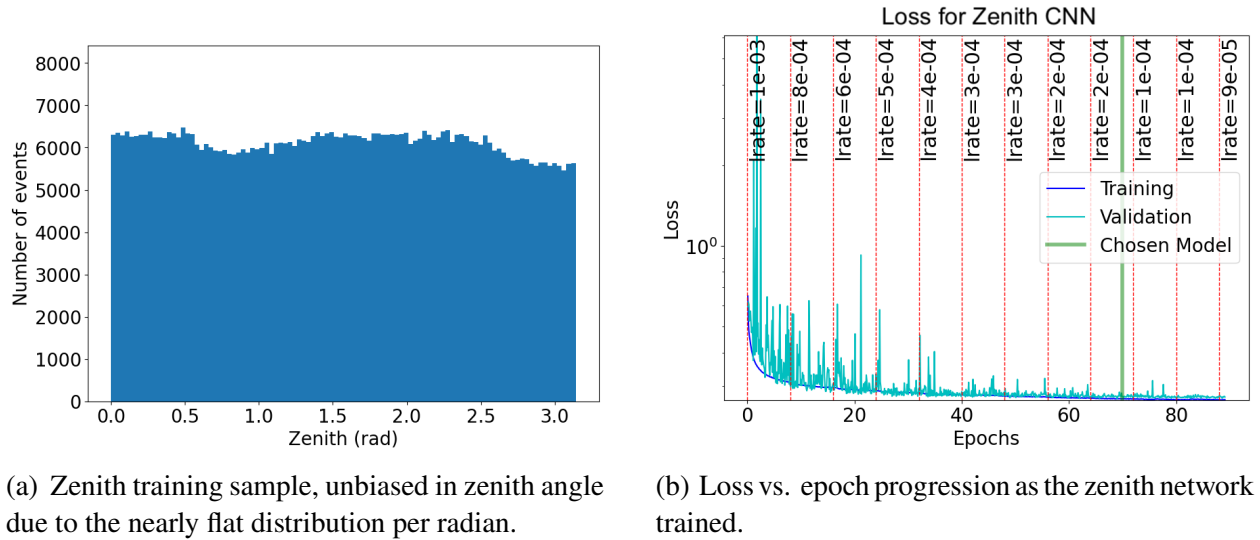


Figure 4.5: Zenith training sample and loss progression.

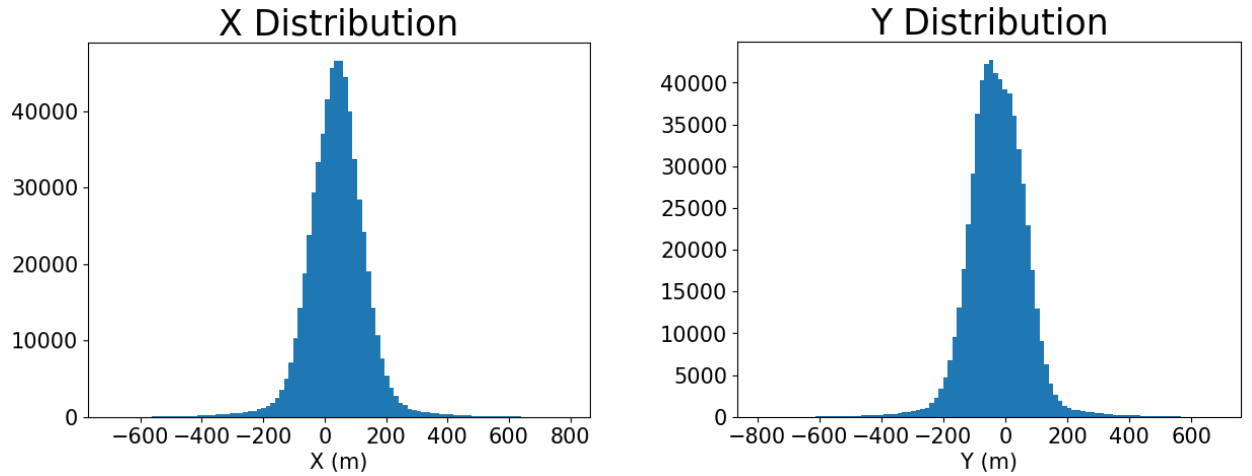


### 4.2.5.3 Vertex

The vertex CNN was trained by undergraduate researcher Julia Willison using an early version of a training sample initially created to train the energy CNN. As with the final energy training sample, it includes only  $\nu_\mu$  CC events and is flat in energy. Instead of the number of DOMs in CNN strings cut, it uses the number of pulses cut (required at least 8 hits). This training sample was not specifically optimized for vertex reconstruction, but it was found that it had a passable resolution and thus no further optimizations were applied (Figures 4.6a, 4.6b, 4.7a).

The total sample size is 4,982,335 events with 80% of the sample used for training and 20% for validation. To cope with the large file size that comes with storing the necessary pulse series information for training, the sample was split into 8 files.

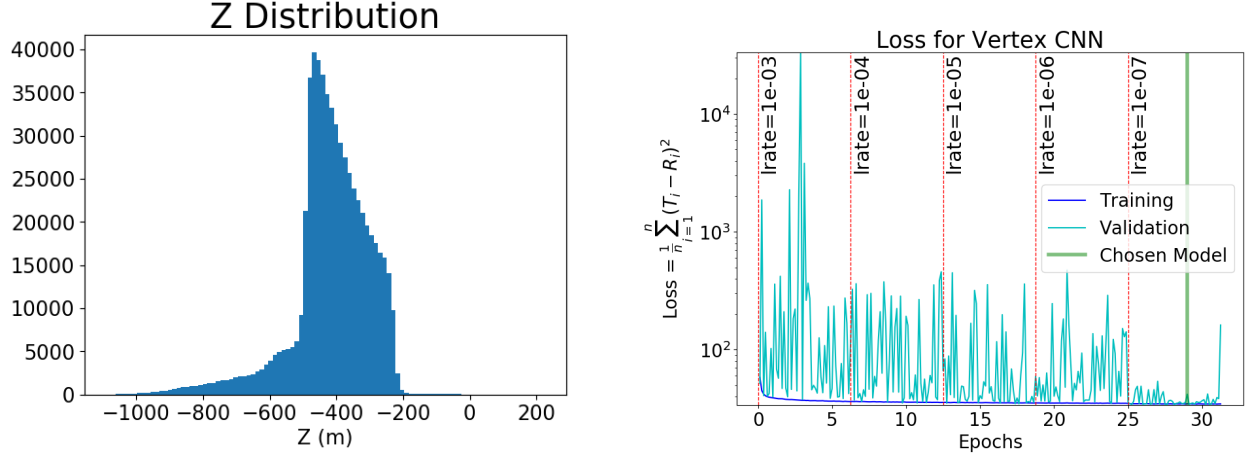
The settings for the learning rate progression are  $R_n = 0.001^{\text{floor}(n/E)D}$  where  $E = 50$  and  $D = 0.1$  for the vertex network (Figure 4.7b). Since each training pass is only through 1/8 of the data, the drop occurs at the equivalent of 6.25 epochs. The model used was trained for 232 passes, or 29 epochs. The performance of the vertex CNN will be evaluated in Section 4.3.4.



(a)  $x$  distribution for part of the vertex CNN training sample, optimized to be unbiased in energy but repurposed to train the vertex classifier.

(b)  $y$  distribution for part of the vertex CNN training sample, optimized to be unbiased in energy but repurposed to train the vertex classifier.

Figure 4.6: X position training sample and Y position training sample.



(a)  $z$  distribution for part of the vertex CNN training sample, optimized to be unbiased in energy but repurposed to train the vertex classifier.

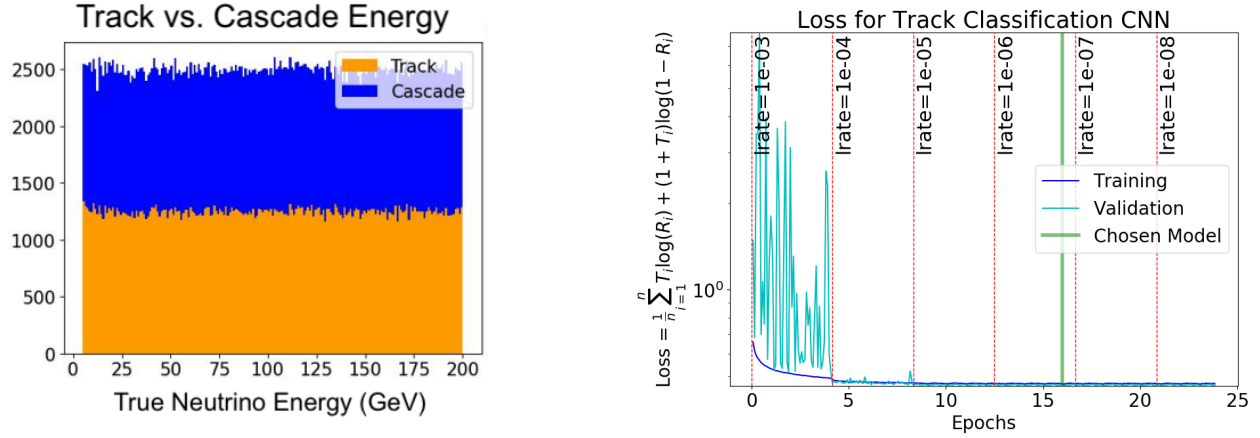
(b) Loss vs. epoch progression as the vertex network trained, where the loss is the sum of the mean squared error calculation for  $x$ ,  $y$ , and  $z$ .

Figure 4.7: Z training sample and vertex loss progression.

#### 4.2.5.4 Track Classifier

The track classifier aims to distinguish the track topology from the cascade topology (Section 2.2.2). It is sometimes referred as the PID (particle identification) classifier, since distinguishing the topology leads to the separation of the  $\nu_\mu$  CC events, thus identifying one of the particles. The training sample was created from the flat-energy generated training sample, cut to have a 50:50 distribution of track:cascade-like events. It uses  $\nu_\mu$  CC events for track-like signatures and uses  $\nu_\mu$  NC and  $\nu_e$  CC & NC events for cascade-like signatures. This was enforced per energy bin, so that across the energy range (5-200 GeV), there are  $\sim 15,000$  track events and  $\sim 15,000$  cascade events per 1 GeV bin. In total, it contained 5,263,514 events, with 80% used for training and 20% for validation. To cope with the large file size that comes with storing the necessary pulse series information for training, the sample was split into 12 files.

The settings for the learning rate progression are  $R_n = 0.001^{\text{floor}(n/E)*D}$  where  $E = 50$  and  $D = 0.1$  for the track classifier. Since each training pass was only through 1/12 of the data, the drop occurs at the equivalent of 4.17 epochs. The model used was trained for 192 passes, or 16 epochs. The performance of the track classifier will be evaluated in Section 4.3.5.



(a) Energy distribution for the track classifier training sample, illustrating that the sample was balance both in track/cascade composition along with energy distribution.

(b) Loss vs. epoch progression as the track classification network trained, using the binary crossentropy loss function.

Figure 4.8: PID training sample and loss progression.

#### 4.2.5.5 Muon Background Classifier

The muon classifier aims to distinguish background atmospheric muons that start outside of the detector from the neutrino signatures that interact inside of the detector. The neutrino portion of the track classifier training sample was re-purposed to train the muon classifier, with the addition of the existing oscNext Level 4 muon MC. The neutrino sample was cut down in size to match the muon sample, to give a 20:40:40 ratio between  $\nu_e:\nu_\mu:\mu$ . The  $\nu_e$  and  $\nu_\mu$  events contain both CC and NC interactions in the sample, and include true neutrino energies up to 200 GeV and all zenith angles (Figure 4.9). The Level 4 muon simulation sample is used rather than the Level 6 muon sample due to lack of statistics at Level 6, which can be seen by the muon rate in Figure 5.5. The muon energy in Figure 4.9 is the energy at generation, rather than the neutrino energy shown at its interaction point, which is why it is always much greater than the neutrino energies. This training sample also has a cut applied enforcing that the number of DOMs in the CNN strings  $\geq 4$ .

The total sample size is 6,923,001 events, with about 2.8 million atmospheric  $\mu$ , 2.8 million  $\nu_\mu$  events, and 1.8 million  $\nu_e$  events. Again, 80% of the sample was used for training and 20% for validation. To cope with the large file size that comes with storing the necessary pulse series information for training, the sample was split into 12 files.

To train, the learning rate started at 0.001 and was not changed through the duration of training. The model used was trained for 240 training, or 24 epochs since there are 12 files total in the training sample. The performance of the background muon classifier will be evaluated in Section 4.3.6.

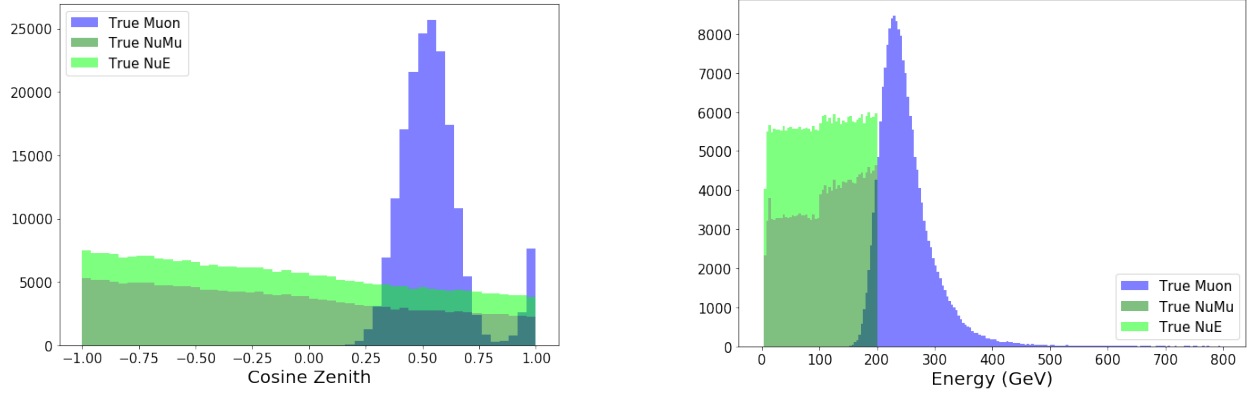


Figure 4.9: True cosine zenith and energy distribution used to train the muon classifier. The true energy for the muon simulation is the energy of the muon at generation, which is why it is different than the energy of the neutrino, which is shown at the interaction point. The  $\nu_\mu$  and  $\nu_e$  histograms are stacked, in comparison with the muon histogram (not stacked).

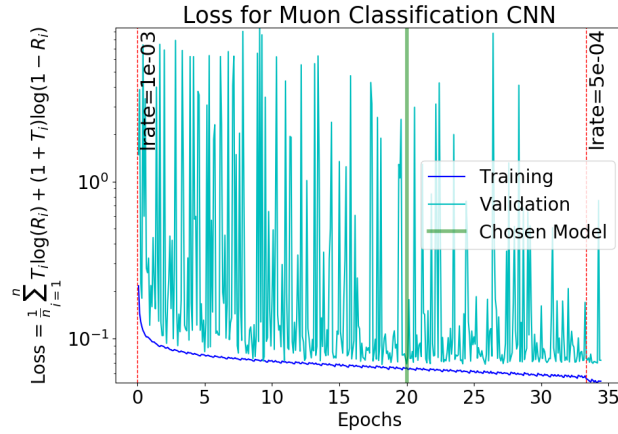


Figure 4.10: Loss vs. epoch progression as the muon classification network trained, using the binary crossentropy loss function.

### 4.3 Performance

Testing samples are independently generated neutrino MC: the analysis sample described in 3.1.2. These distributions are weighted with atmospheric flux and oscillation weights, making them

look as close as possible to the expected data distribution. The analysis cuts are therefore applied to this sample, which are independent of the training sample cuts. The motivation for these cuts are described in Chapter 5, but a table summarizing the cuts is included here (Table 4.7). Similar cuts are made on the likelihood reconstructed sample, using its own reconstructed variables to apply their cuts, along with other cuts optimized for a  $\nu_\mu$  disappearance analysis using this method.

Cut Name	CNN Cut	High Stats Cut
Reconstructed radius	$\rho_{\text{CNN}} < 200 \text{ m}$	$\rho_{\text{Retro}} < 300 \text{ m}$
Reconstructed $z$	$-495 \text{ m} < Z_{\text{CNN}} < -225 \text{ m}$	$-500 \text{ m} < Z_{\text{Retro}} < -200 \text{ m}$
Reconstructed energy	$5 \text{ GeV} \leq E_{\text{CNN}} \leq 100 \text{ GeV}$	$5 \text{ GeV} \leq E_{\text{Retro}} \leq 300 \text{ GeV}$
Reconstructed cosine zenith	$\cos \theta_{\text{CNN}} \leq 0.3$	$\cos \theta_{\text{Retro}} \leq 0.3$
PID using reconstructed input	$PID_{\text{CNN}}$ boundary: [0.25, 0.55]	$PID_{\text{Retro}}$ BDT boundary: [0.5, 0.85]
Muon BDT using reconstructed input	$BDT_{\text{CNN}} \geq 0.8$	$BDT_{\text{Retro}} \geq 0.8$
# DeepCore Triggers	only 1	N/A
# CNN DOMs	$\geq 7$	N/A
# Pulses	N/A	$\geq 8$
Reconstructed time	N/A	$t_{\text{Retro}} < 14500 \text{ ns}$
Passed Retro	N/A	True
# Retro Iterations	N/A	$> 10$
DOMs with direct pulses	$> 2.5$	$> 2.5$
# hits in top 15 DOMs	$< 0.5$	$< 0.5$
# hits in outer strings	$< 7.5$	$< 7.5$

Table 4.7: FLERCNN and Retro Reconstruction sample cut comparison after Level 5. All analysis cuts that apply to the FLERCNN sample are described in more detail in Sections 5.1 and 5.2.

The differences between the training samples and testing sample is helpful to evaluate the adaptability of the CNN and make sure it is not overtrained, or over-optimized, to only perform well on distributions that look like the training sample. Figure 4.11 shows the distribution of energy and zenith for the testing sample, with the CNN analysis cuts applied (described in Table 5.1). The shape of the distribution is similar for the likelihood-based sample's testing distribution, with its respective cuts applied.

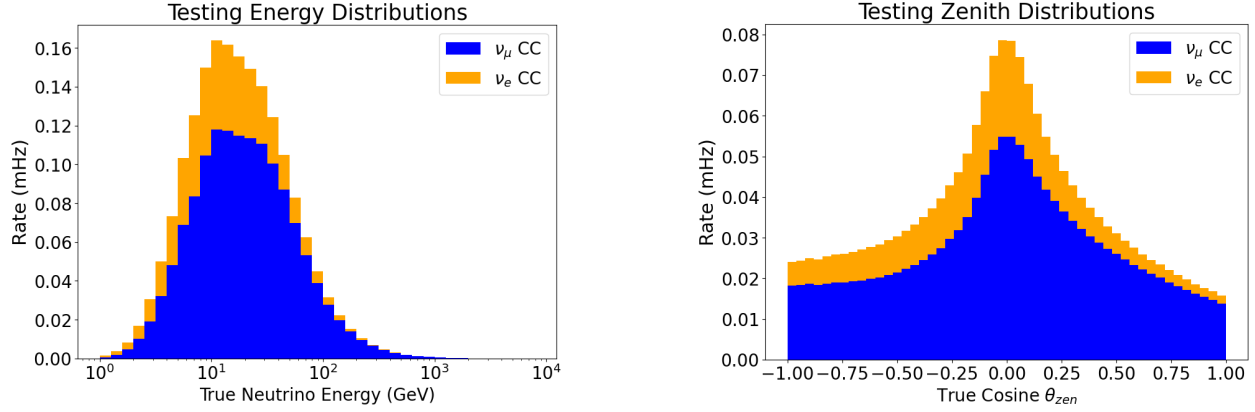


Figure 4.11: Testing distributions for the  $\nu_\mu$  CC and  $\nu_e$  CC events. The energy and zenith cuts are not applied in this case, since they will be applied based on the individual reconstructions' outputs during testing. All other cuts applied are the CNN Cuts described in Table 5.1.

#### 4.3.1 Reconstruction Runtime

One of the main motivating factors for pursuing the CNN reconstruction method is for its evaluation speed. Table 4.8 shows the times for running all 5 FLERCNN reconstructions and classifications (currently done in serial) vs. the likelihood-based reconstruction (which returns 8 variables, not including the PID and muon classification scores).

Reconstruction Method	Avg Time per Event	Events per day (single core)	Time for $10^8$ events (1000 CPU cores or GPU nodes)
5 CNNs on GPU	0.0011 s	80,000,000	2 min
5 CNNs on CPU	0.29 s	300,000	8 hours
Likelihood-based method	40 s [10]	2,000	46 days

Table 4.8: Runtime and implications of applying the different reconstructions to the MC sample.

The GPU timing for the CNN is determined using an NVIDIA v100 GPU node. The CPU timing for the CNN is benchmarked using a single core on a 2.40 GHz Intel Xeon E5-2680v4, though Tensorflow does have the capability to use multithreaded operations if multiple CPU cores are available. The likelihood-base method was also benchmarked using a single core [10] and only has CPU capability. There is a difference in the number of output variables for the two methods. FLERCNN uses five networks run in succession to produce seven output values: true neutrino energy, zenith direction, (x,y,z) starting vertex, PID classification score, and Muon classification

score. In contrast, the likelihood-based method is evaluated once per event to output eight variables: track energy, cascade energy, (x,y,z,t) starting vertex, zenith direction, azimuth direction. The benchmark numbers are thus the total times for the CNN to output all seven variables and the likelihood-based method to output all eight of its variables.

Note that the final column of the table indicates the differences with some generous assumptions. While IceCube has access to multiple computing clusters, completely reserving 1000 cores for the low energy reconstruction alone is not typically feasible, especially if they are reserved for days. There are also more CPU cores available than GPU, so the number of reasonably expected cores would be less for the GPUs, though they would only be required for a shorter duration due to the faster runtime. These number also do not factor in the read and write time of the events, which are stored as a special IceCube-specific file type in  $O(10^3)$  event batches. Thus, this increases the runtime for both the CNN and the likelihood-based reconstruction in the true application. FLERCNN has successfully reconstructed the full MC and data samples within a week. The likelihood-based reconstruction takes months to reconstruct the full MC and data samples.

### 4.3.2 Energy

Figure 4.12 shows the fractional energy resolution as a function of energy. The resolution is evaluated separately for four samples, first  $\nu_\mu$  CC and  $\nu_e$  CC events which have their deposited energy in the detector approximately equal to the true energy of the incoming neutrino. For the  $\nu_\tau$  CC and all flavor NC events, these have secondary neutrinos from their interactions, and thus will have part of their energy not deposited in the detector (carried away from the outgoing neutrino). For the latter case, these resolutions are plotted against the deposited energy, which is a more accurate representation of what the reconstructions can hope to estimate. This is a reasonable representation though since the analysis is performed in the reconstructed energy space, so the deposited energy is a good reflection of how these flavors will behave in the analysis.

Comparing to the likelihood-based reconstruction, the CNN performs well at low energies for  $\nu_\mu$  events, where the CNN's median is near the 0 line (indicating good agreement) while the

likelihood-based plot overestimates the energies in this region. The CNN was specifically trained and optimized on the  $\nu_\mu$  CC events, with the intention of using it for the  $\nu_\mu$  disappearance analysis. This is why the behavior is best for the CNN on that flavor. The CNN still has a comparable resolution for the other flavors compared to the likelihood-based method. The spread for the CNN continues to be smaller than the likelihood-based reconstruction at low energies. There is some offset for the median, but since the same reconstruction is applied for all flavors and our detection methods are unable to distinguish beyond the two PID topology (track vs. cascade), adjustments cannot be made for each flavor accordingly. Overall, the CNN energy reconstruction behaviors similarly to the likelihood-based reconstruction, including its trends between flavors and the resolution is comparable.

Note that for all these energy resolution plots, the energy cut is not applied but all other final level analysis cuts are (described in Chapter 5). This is because it would cause an artificial bias on the edges of the true energy resolution range.

### 4.3.3 Zenith

Figure 4.13 shows the absolute resolution for the CNN and likelihood-based reconstructions as a function of energy for track events (left plot) and cascade events (right plot). The high energy tracks are the best resolved events for both reconstructions due to the long lever arm that the reconstructions have to work with. The slight difference in the 68% region for the CNN at high energies is likely due to the track leaving the CNN string region (19 IceCube strings + 8 DeepCore strings). Future improvements to resolution could consider adding additional strings or training a network to apply an ending position cut on the track to make sure it is contained in the CNN input region, but this would only offer a slight improvement in a region that is already better resolved than (for example) the cascade zenith angle resolution at any energy.

Cascades prove to be more challenging for both reconstructions, as both methods have similar resolutions across all energies. Even though the CNN is trained on  $\nu_\mu$  CC events only, it manages to reconstruct reasonable zenith positions on average for the sample with the median resolution



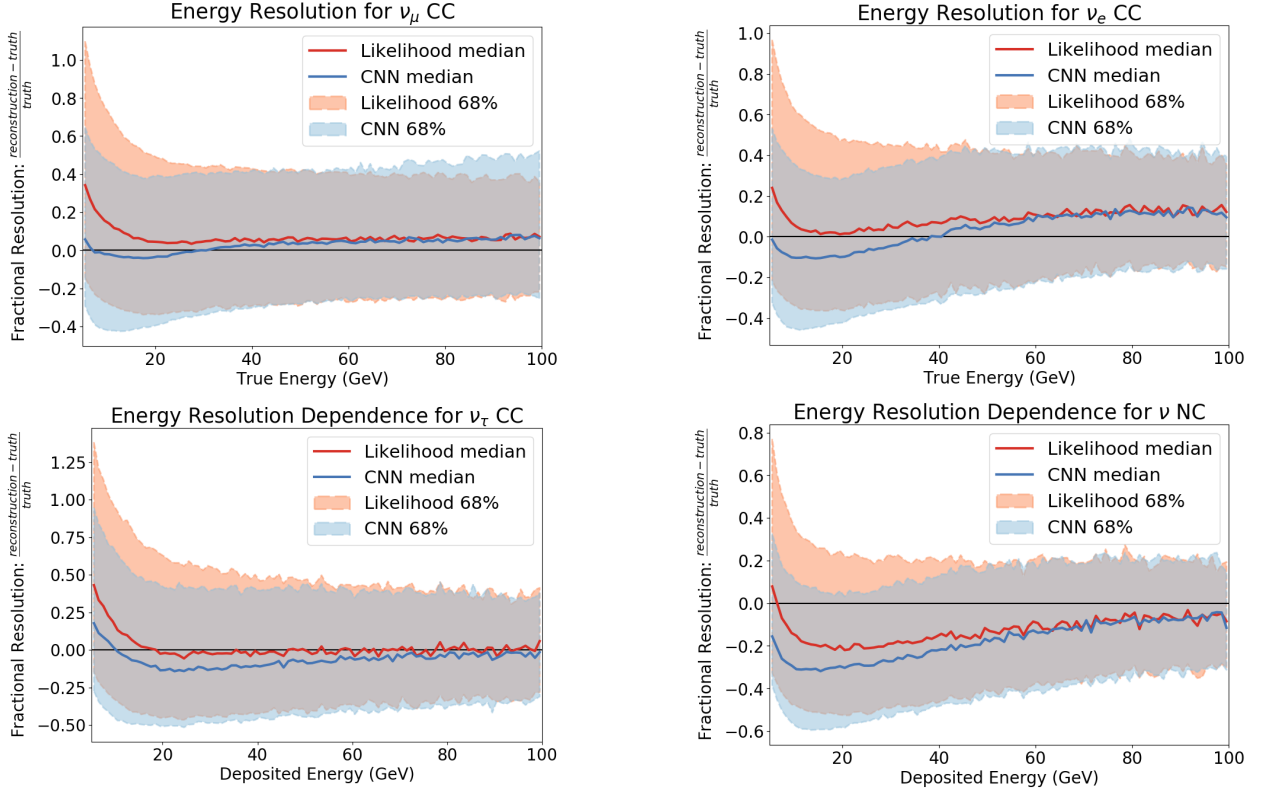


Figure 4.12: Resolution shown for true energy for the  $\nu_\mu$  CC and  $\nu_e$  CC events. With the invisible energy carried off by the outgoing neutrinos from NC interactions and some  $\nu_\tau$  CC interactions, the resolution is shown vs. deposited energy for these selections.

near the zero line and the 68% spread comparable to the likelihood-based method. Note that the artifact in both the reconstruction methods at low energy (seen in both the track and cascade plots) is likely due to the events being weighted with the expected flux and oscillation weights—the latter of which is likely the cause of this artifact.

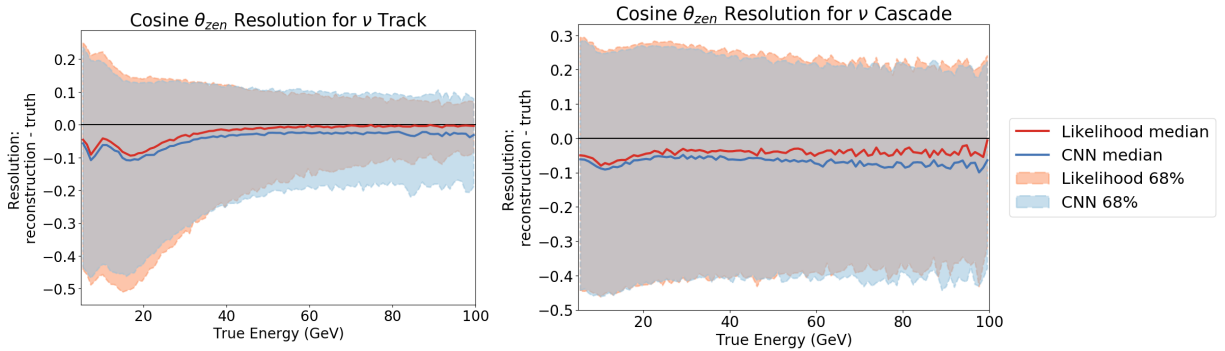


Figure 4.13: Resolution shown for reconstructed cosine zenith for the  $\nu_\mu$  CC tracks and remaining cascade events.

#### 4.3.4 Vertex

The CNN resolution is fairly constant across all energies, for both the track and cascade radial reconstructions (Figure 4.14). The median is comparable with the likelihood-based reconstruction, but unlike the likelihood-based reconstruction, the 68% spread does not improve with increased energy. This could be due to the fact that the CNN was trained only using the center-most strings, so that as events go up higher in energy, some of their hits could be outside of the region the CNN uses for input. Note that the difference between the two 68% range is at maximum 10 m. This is much smaller than the distance between strings, and as the intention of this reconstruction is to use it to apply a containment cut to make sure the events start in the DeepCore region, improving the resolution further is superfluous.

The results for the  $z$  reconstruction in Figure 4.15 are similar to the radial reconstruction in that the CNN has a larger 68% spread on the order of 10 m maximum difference. The median reconstructed  $z$  per energy bin is very comparable to the likelihood-based reconstruction. Overall, the spread for the  $z$ -position is smaller than the radial reconstruction, evident for both reconstruction methods, which is not surprising since the DOM  $z$ -spacing is much finer than the  $xy$ -spacing (Section 2.1.1).

The performance of using this reconstruction as a classifier to cut events inside the DeepCore volume vs. outside the contained volume is in Figure 4.16. Note that it keeps most of the events inside of the desired detector region (95.65%) and successfully eliminates 52.03% of the events outside of the region.

For the vertex reconstruction, there is a known artifact in the CNN reconstruction where the string positions are favored in the  $x$  and  $y$  reconstruction (contributing to the radial reconstruction) and where the DOM positions are favored in  $z$ . Since this reconstruction is used for a containment cut, the precise reconstruction position is not necessary to make a reasonable containment cut. The favoring of string positions is evident in Figure 4.14.

Future work optimizing the vertex classifier could be to generate a training sample that is uniform in the  $(x,y,z)$  starting vertex position, or at least a sample that has more events outside of

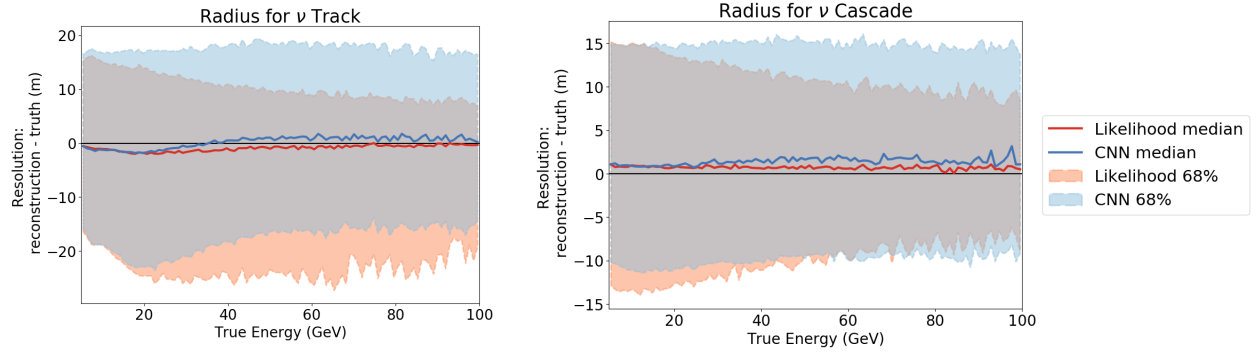


Figure 4.14: Resolution shown for reconstructed radius from the center string (string 36) of the starting vertex for the  $\nu_\mu$  CC tracks and remaining cascade events.

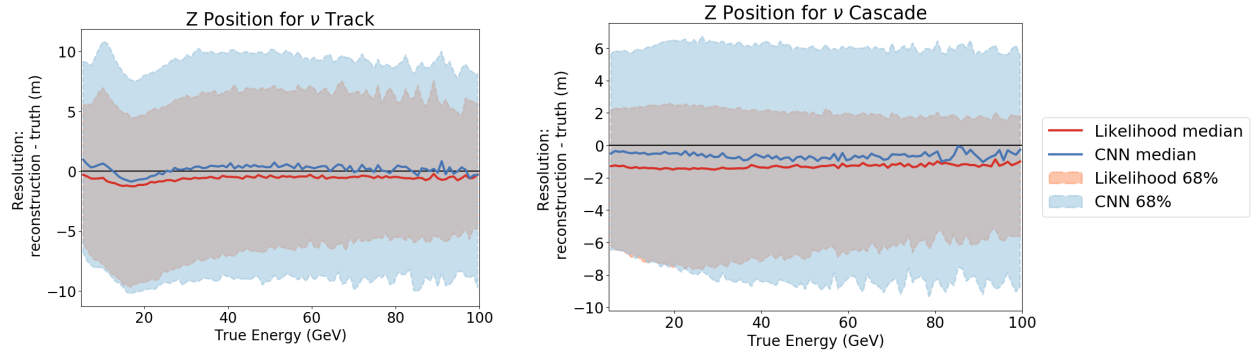


Figure 4.15: Resolution shown for reconstructed z position of the starting vertex for the  $\nu_\mu$  CC tracks and remaining cascade events.

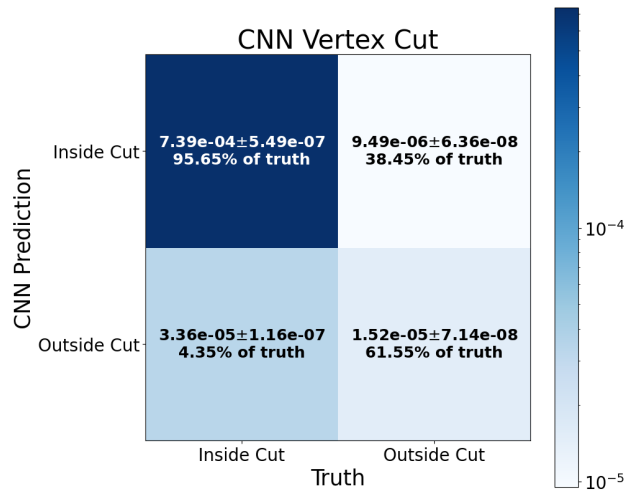


Figure 4.16: Confusion matrix for vertex reconstruction, exploring efficiency of cutting at 200m, proposed for oscillation analysis.

the desired detector region. CNNs are known for their ability to interpolate, not extrapolate [56], so it is expected that the CNN performs better on classifying the event types it has seen (inside of the desired cut volume) rather than events that were more rare in the training sample (outside of the desired cut volume). As is, this cut still has the necessary effect to improve the sample resolution by ensuring the sample predominantly contains events starting inside of DeepCore; it successfully removes 52.0% of events outside of DeepCore while keeping 95.7% of the events inside of DeepCore.

#### 4.3.5 Track Classifier

Assessing the performance of binary classifiers uses different metrics. One way is to plot the classifier's score (0-1) on events from each binary class. The left plot in Figure 4.17 shows this representation, with the true track events ( $\nu_\mu$  CC) in red and true cascade (all other neutrino flavors) in blue. Ideally, there would be a peak for the CNN track score at 1.0 for the true tracks and another peak at 0 for the true cascades.

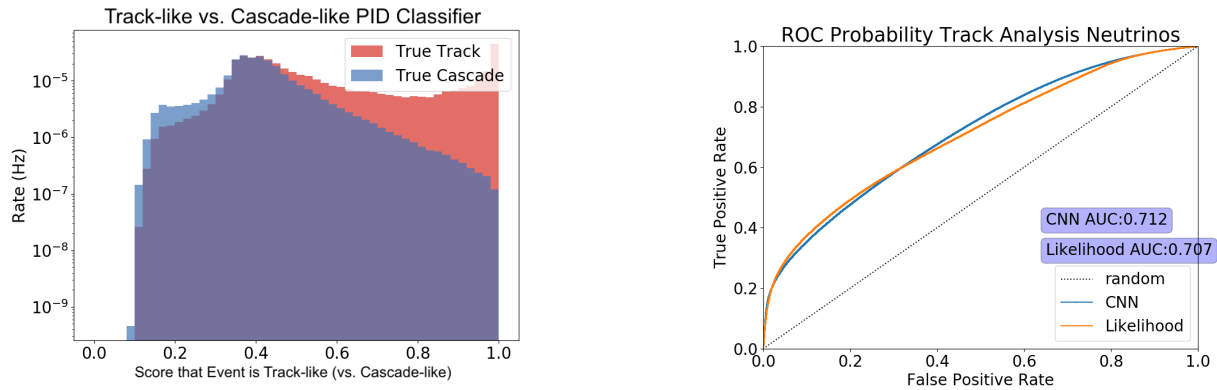


Figure 4.17: Performance of Track classifier with binary classification histogram on the left and the ROC on the right.

The FLERCNN result in Figure 4.17 shows the true track distribution has a peak near 1.0, which are the higher energy tracks that the CNN is more confident in classifying (note the log y-axis). There is not a large true cascade bump near the score of zero, which would be ideal, though there are more true cascade events classified in the region  $< 0.4$  than true track events.

Another plotting metric for binary classification is the Receiver Operator Curve (ROC), which compares the "true positive rate" (true track, classified as track) to the "false positive rate" (true cascade, classified as track). A completely random classification would get the binary classification result correct 50% of the time, which is represented by the diagonal 1:1 line on the ROC plot where the true positive and false positive rates are equal. To reduce this plot to a single number, the Area Under the Curve (AUC) can be calculated. The closer that the AUC is to 1.0, or the further away from a perfectly random result (at 0.5), the better the classification power.

The right plot in Figure 4.17 compares the ROC for both the FLERCNN classifier and the BDT trained on the likelihood-based reconstruction variables used in the traditional IceCube low energy analysis. The AUC for both are comparable, greater than 0.5 but less than 1.0. This is expected due to the classifiers' difficulty distinguishing low energy track events. The GeV-scale events often have only a handful of hit DOMs, making it difficult to distinguish the track-like structure from the cascade due to short track lengths.

#### **4.3.6 Atmospheric Muon Classifier**

To assess the atmospheric muon classifier, the same histogram and ROC method are used as above. In the classification histogram (left plot in Figure 4.18), the true neutrino events (all flavors) have a peak at the 1.0 classification score. This indicates the CNN's capability to distinguish neutrinos from the sample. The true muon sample, however, appears mostly flat. Recall that the muon classifier is trained on Level 4 muon events (Section 4.2.5.5). The true muon histogram has a peak near 0 when tested on Level 4 muons, but this is not the target sample. Even so, the ROC (right plot in Figure 4.18) does show the CNN muon classifier's ability to distinguish the neutrinos. The CNN's AUC is slightly improved compared to the BDT trained on the likelihood-based reconstruction variables. A harsh cut could be used to remove most of the true muons and keep the true neutrinos, since the neutrinos are extremely peaked at the CNN score near 1.0. Instead, the output of this CNN is used to feed into a BDT that is used as a final analysis cut to distinguish atmospheric muons, which will be discussed further in Section 5.2.1.2.

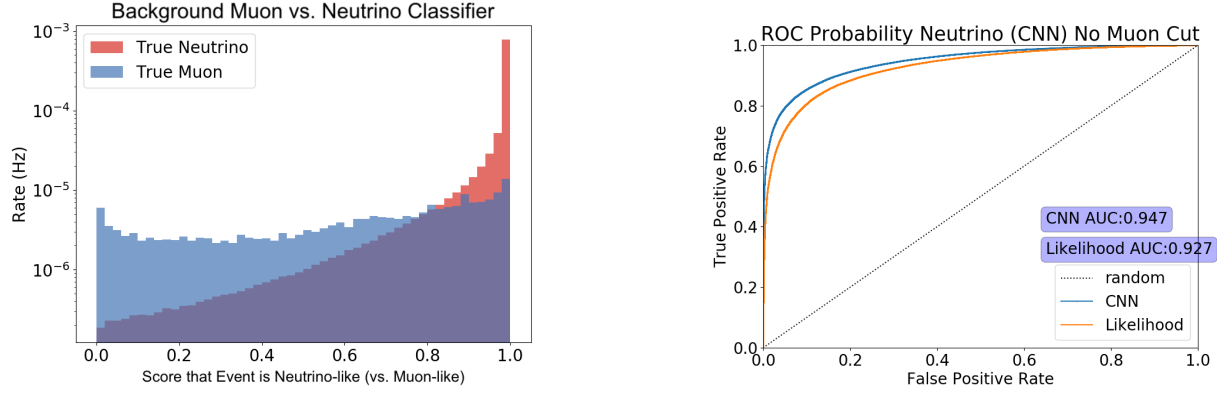


Figure 4.18: Performance of muon classifier with binary classification histogram on the left and the ROC on the right.

#### 4.4 Stability Against Detector Systematics

The neutrino simulation sets used for the tests in the previous sections are the sample set to the nominal analysis parameters for the physics and systematic expectations (Table 6.2). There also exists neutrino simulation sets that have their detector systematic parameters purposely set to off-nominal values (generally 5-10% past expectations). Some of these sets are used here to test the response that the five FLERCNN networks have to the changes in the detector settings. Note that changes are expected, since changes in these parameters will lead to more or less photons being detected by the DOMs. Thus, this check is to confirm the overall trend of these changes align with expectations and do not significantly deviate from the baseline results.

The variables that are explored in this section are the energy, zenith, and PID which contribute directly to the analysis measurement (Section 6). Figures 4.19, 4.20, and 4.21 focus only on  $\nu_\mu$  CC track-like events. The test was also performed on  $\nu_e$  CC cascades and found to have similar results. The top two plots in each figure show the FLERCNN energy reconstruction as a function of true neutrino energy. The bottom two plots in each figure show the FLERCNN zenith angle reconstruction as a function of true cosine zenith. The bias plots look at the overall shift of the median of the distribution for the calculated resolution (fractional resolution for energy and absolute resolution for zenith). The 68% spread plots give the width of the 68% contained resolution. In essence, this breaks the plots in Figures 4.12 and 4.13 into two numbers per bin: the position of the

CNN median line and the range of the CNN 68% shaded region. The PID will be discussed as a fractional change between the rates for the track, mixed, and cascade classification (Section 5.4.1): the rate of the systematic set over the rate of the baseline set.

#### 4.4.1 Reconstructing Different DOM Efficiencies

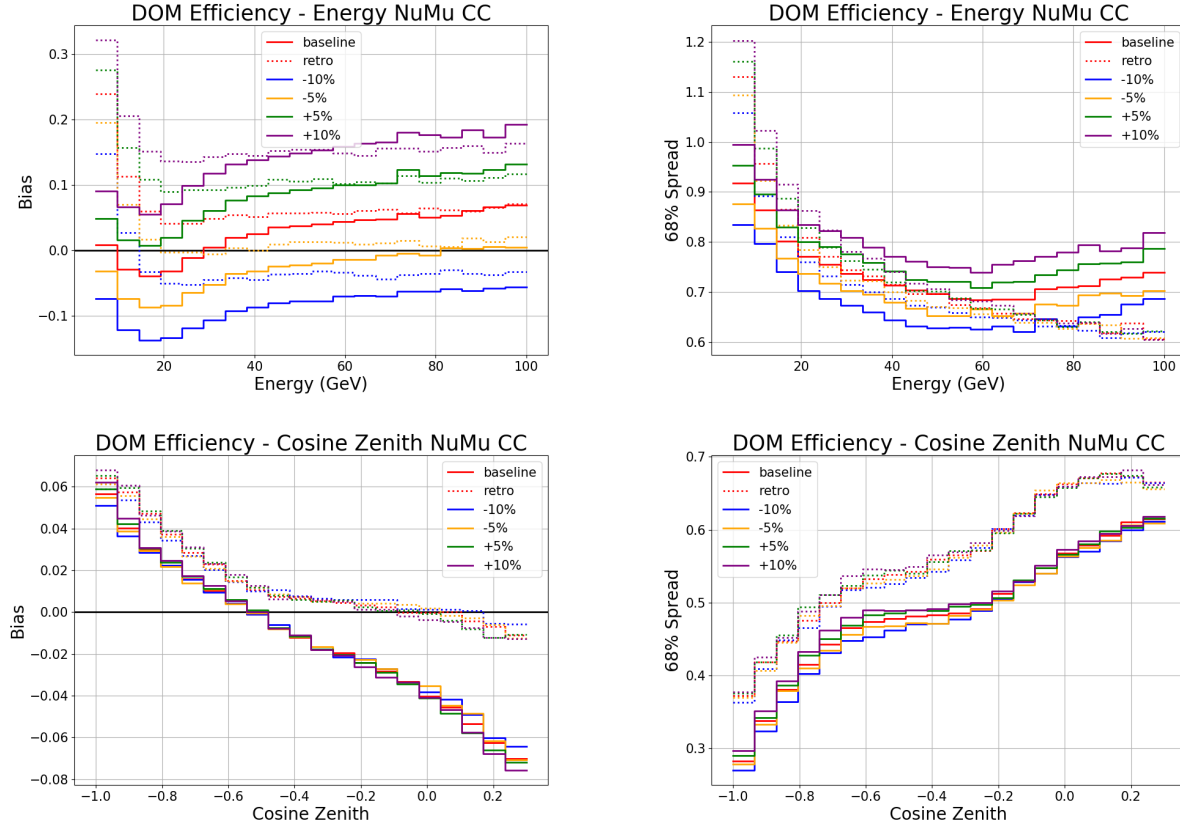


Figure 4.19: Bias and spread for the FLERCNN performance on various detector systematic sets for DOM efficiency. The FLERCNN reconstruction is shown in the solid lines and the likelihood-based reconstruction (retro) is shown in the dotted lines.

The energy reconstruction scales with increased DOM efficiency; this is as expected since increasing the DOM efficiency would increase the number of photons that the PMT detects. The shift in the bias also mirrors the change in the DOM efficiency, where the energy reconstruction bias difference between the baseline and +10% increase DOM efficiency changes by a factor of 0.10. The changed DOM efficiency should not have an affect on the angular reconstruction, which

is found to be true in Figure 4.19 as the FLERCNN zenith reconstruction shows a very consistent bias and 68% spread between these sets.

The comparison to the likelihood-based reconstruction (retro, shown in the dotted lines) shows similar behavior to the CNN. The energy effects are of similar magnitude, with a slightly larger impact for the CNN in bias and spread. The likelihood-based reconstruction also follows the trend that the spread continues to decrease as the energy increases, whereas the CNN spread levels off. These differences are small, but could be due to the fact that the CNN only uses the centermost IceCube strings, so as events get up to higher energy, there is a larger chance for underestimating the energy due to hits being outside of the CNN input region.

The zenith bias and spread show no effect for both reconstructions. Note that the likelihood-based reconstruction bias is near 0 at the horizontal angles ( $\cos \theta_z = 0$ ), better than the CNN which has a more linear trend. In contrast, the CNN shows a smaller spread than the likelihood-based reconstruction across all angles. Overall, the differences between these two reconstructions are small and the trends overall show similar behavior.

The PID rates do change with the DOM efficiency, with the cascade bin most sensitive (change in the rate by  $1.00 \pm 0.15$  corresponding to the  $\pm 10\%$  sets). This is due to the lowest energy events being classified in the cascade bin (Section 5.4.1) so that increasing the DOM efficiency means that less events are cut from the sample (creating an increase in rate) and decreasing the DOM efficiency means that more events are cut from the sample, as compared to the baseline set. Most other changes to the PID are within  $1.00 \pm 0.05$ .

#### 4.4.2 Reconstructing Different Hole Ice Parameterizations

The different hole ice sets vary the  $p_0$  and  $p_1$  hole ice parameterization of the DOM's angular acceptance (Section 2.3.2). The numbers in the legend correspond to changes in the parameterization, and their magnitude of change from the baseline settings is the most telling difference—rather than the exact values. In this case, the energy FLERCNN results are less effected by changes in the hole ice—seen in both the bias and 68% spread plots. The angular FLERCNN results do have a slight



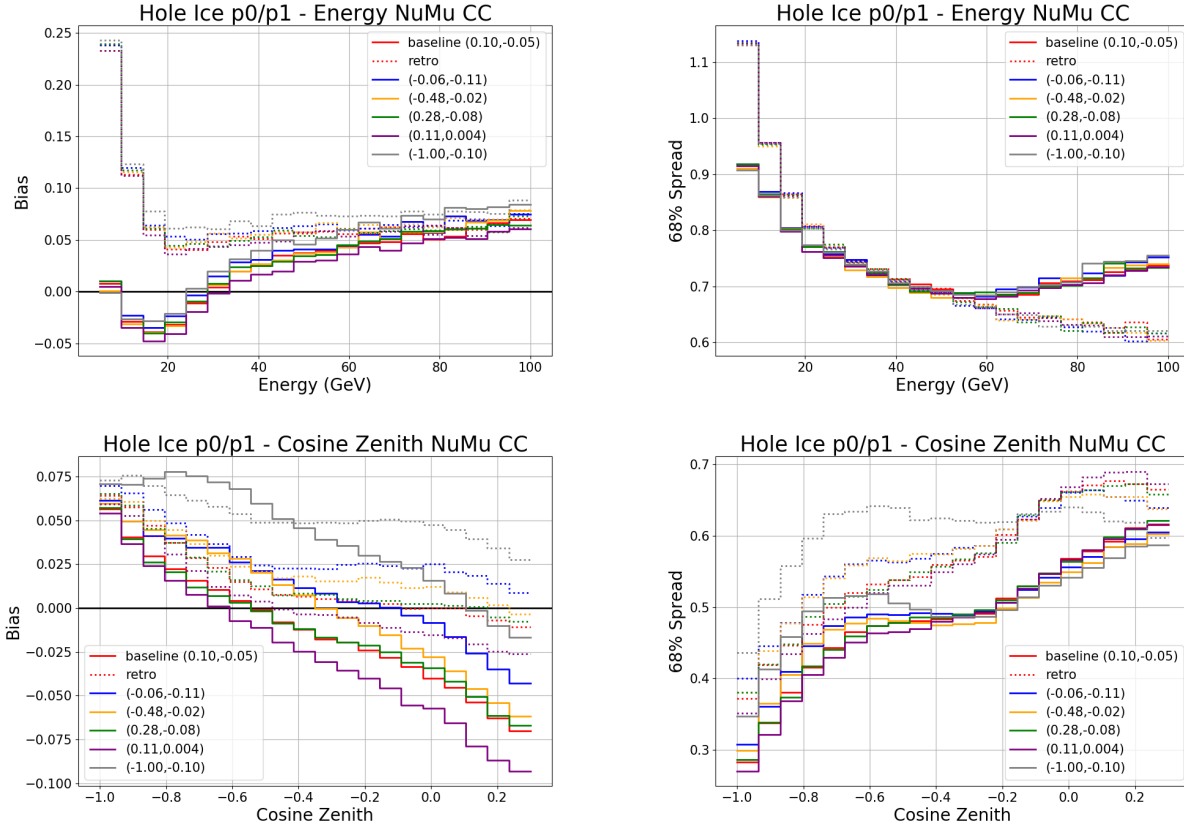


Figure 4.20: Bias and spread for the FLERCNN performance on various detector systematic sets for hole ice. The FLERCNN reconstruction is shown in the solid lines and the likelihood-based reconstruction (retro) is shown in the dotted lines.

bias with the different hole ice sets, which is expected due to the fact that the hole ice properties adjust the angular acceptance for the DOM. Thus, the CNN has likely learned the baseline angular acceptance of the DOM, and changing the acceptance could alter the angular reconstruction. Note that the gray line has the most severe hole ice changes as compared to the baseline, and the zenith CNN responds accordingly.

The comparison to the likelihood-based reconstruction (retro, shown in the dotted lines) shows similar behavior to the CNN. The energy bias and spread show little change due to the various hole ice sets. The zenith effects are of similar magnitude, with a slightly larger impact for the CNN in bias and for the likelihood-based reconstruction in spread. Overall, the direction of impact from the sets and trends are very similar between the two reconstructions.

The PID rates across all hole ice sets are very stable, changing by less than  $1.00 \pm 0.05$ , except

for a slightly larger effect on the severely pulled hole ice sample (set shown in gray in Figure 4.20). The change to the mixed and cascade bin rates is near 0.90, which could be due to the bias seen in the zenith CNN reconstruction causing more events in this sample to fail the zenith cut (Table 5.1).

#### 4.4.3 Reconstructing Different Bulk Ice Properties

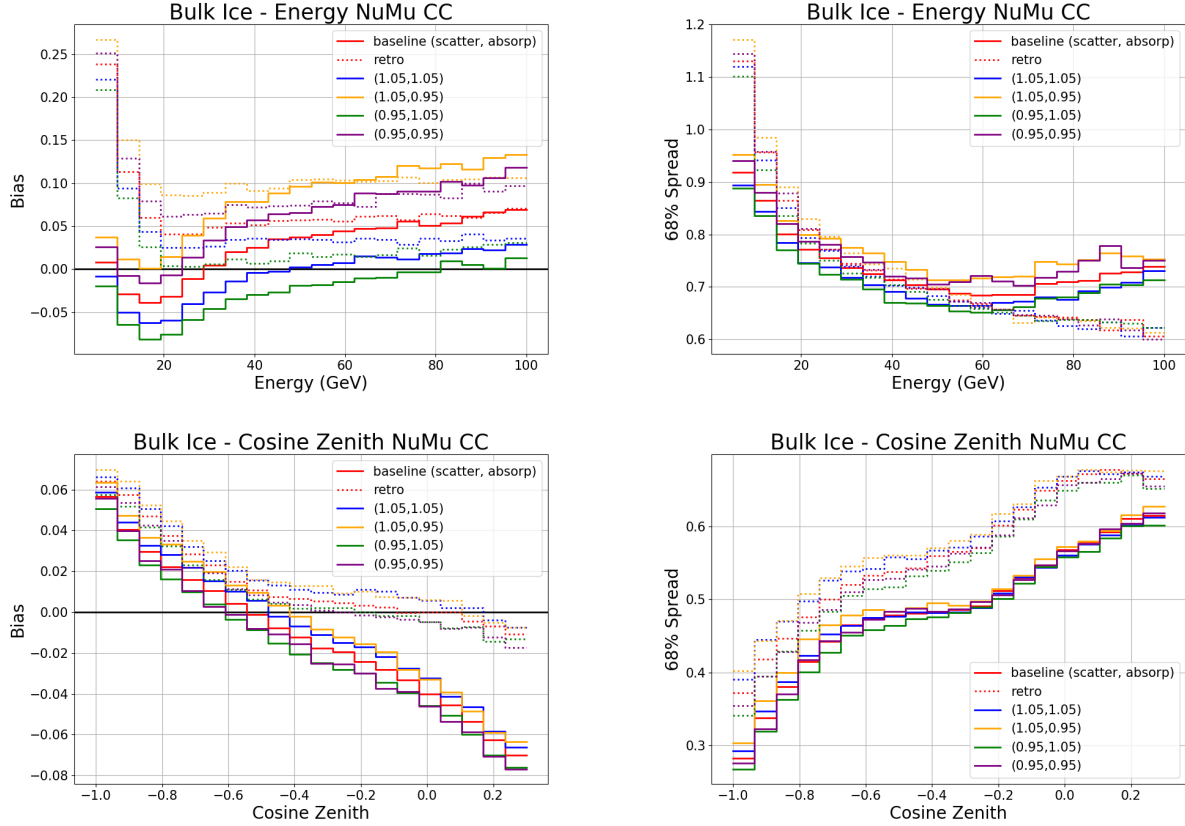


Figure 4.21: Bias and spread for the FLERCNN performance on various detector systematic sets for bulk ice. The FLERCNN reconstruction is shown in the solid lines and the likelihood-based reconstruction (retro) is shown in the dotted lines.

The scattering and absorption systematic sets change the scattering and absorption coefficients uniformly for the entire bulk ice (recall these coefficients vary with depth and are modeled by the ice model used during the simulation photon propagation - Section 2.3.1). For scattering  $> 1$ , it indicates more photon scattering in the ice and has a different influence depending on the orientation between the PMT face and the photon. If the photon is head on to the DOM, more scattering would likely divert the photon away from being detected. On the other hand, it can improve the probability

of the photon from the back of the PMT being detected due to scattering into a position with better angular acceptance. For absorption  $> 1$ , it means a higher chance the photon is absorbed in the ice, and thus is not detected by the PMT.

As with the DOM efficiency, the energy CNN reconstruction's bias scales with high energies predicted when the systematic set allows for more photons to reach the DOM along with the inverse being true. Thus, the trend of the change is reasonable and the magnitude does not exceed that of the changed DOM efficiency. The zenith CNN reconstruction does not show much sensitivity to the changes in the bulk ice, with only a slight change that scales with the scattering differences. With a change in scattering, the angular reconstruction becomes difficult since the path of the photon has been altered (whether more or less than the nominal that the CNN expects). Thus, these changes are minimal, only changing the bias by  $\sim 0.1$ , but the trend is expected.

The comparison to the likelihood-based reconstruction (retro, shown in the dotted lines) shows similar behavior to the CNN. The energy effects are of similar magnitude, with a slightly larger impact for the CNN in bias and spread. The zenith bias and spread show little change, with much of the same behavior as seen before. Overall, the CNN seems to have a slightly larger sensitivity in bias with the various systematic sets, but the overall magnitude is very close to the changes seen in the likelihood-base reconstruction. The trends are also consistent for the changes in the systematic sets themselves, so that the reconstructions behave as expected with these detector settings altered.

The PID rates across all sets are very stable, less than  $1.00 \pm 0.05$ . This is reasonable since the bulk ice systematic sets have less of a bias shift for the energy than the DOM efficiency sets, so the change in energy is not enough to cause a similar rate change seen in the  $\pm 10\%$  DOM efficiency.

#### **4.4.4 Conclusion of Robustness to Systematics**

Overall, the energy and zenith CNNs perform well on the the various systematic sets with their detector settings altered. The changes in the bias scale with the expected behavior due to the changes in number of expected photons. The notable variation in bias indicates that FLERCNN is sensitive to the detector systematics, and therefore will be able to account for these when fitting the

detector systematic parameters during the analysis. The sensitivity is well-behaved, such that the trends match expectations and the shifts are not extreme, so the reconstruction should respond well during the fit. The 68% spread is fairly consistent across all systematic sets, without concerning trends. The PID rates are very stable between the systematic sets and the baseline sets, with minor fluctuations likely due to events being added/removed from the sample due to the cuts. It is important to note that the cuts applied to the analysis sample were optimized with the nominal systematic set in mind, so this behavior is not surprising and only occurs with systematic sets at the boundary or even more severe than the expected shifts in these parameters.

## CHAPTER 5

### FLERCNN ANALYSIS SAMPLE SELECTION

After reconstruction, a series of cuts must be applied to remove remaining backgrounds, such as atmospheric muon and noise events. The data distribution is then separated into the analysis binning, by energy, zenith angle, and PID.

#### 5.1 Resolution Motivated Cuts

The first set of cuts are motivated by improving the angular and energy resolution of the remaining events. These include ensuring the events start in the DeepCore array, have a minimum number of DOMs hit for the reconstruction to perform well, and fall into the energy range where the FLERCNN energy reconstruction performs best.

##### 5.1.1 Starting Vertex

Low energy events benefit from the improved detector density that DeepCore provides. Thus, we want to make sure events are in the DeepCore array, so that they can be reconstructed with the best possible resolution. Cuts are applied using the FLERCNN reconstructed radius relative to the origin at the center string 36 ( $\rho_{36}$ ) and  $z$ -position of the starting vertex. For the radius, a cut is applied so all events must have  $\rho_{36} < 200$  m. For the  $z$  position, we use the the depth of the DeepCore strings, so that events with a  $z$  vertex position of  $-495 \text{ m} < z < -225 \text{ m}$  are kept.

##### 5.1.2 Number of hit DOMs

During the FLERCNN studies, we found that cutting on  $n_{\text{DOM}}$  gives better energy resolution in the low energy region. Thus, a cut is applied requiring at least 7 DOMs are hit in the 19 IceCube strings and 8 DeepCore strings that the CNN uses to reconstruct the events for training/testing. This reduces the total neutrino sample by only 1.05%, after all other analysis cuts are applied.

### 5.1.3 Energy Range

The  $\nu_\mu$  disappearance oscillation bands are expected in the region  $< 100$  GeV (Figure 1.7). The FLERCNN energy network was therefore optimized to reconstruct this region, and thus a cut is applied on the reconstructed energy so that only with  $E_{\nu,\text{reco}} < 100$  GeV are included. Due to the very small number of pulses recorded by DeepCore at less than 5 GeV, training the network much lower than this proved difficult. To avoid poor resolution at the lowest energy events, a cut is made to ensure the reconstructed energy of the event is  $E_{\nu,\text{reco}} > 5$  GeV.

## 5.2 Data/MC Agreement Motivated Cuts

The second category of cuts are motivated by data/MC agreement, ensuring the shape agreement between the 1D data distribution and MC prediction for each variable. Only the shape matters because there are neutrino and muon normalization terms included in the systematic parameters (Section 6.4.1.3), so that their final values will be determined during the analysis. These include variables that limit the noise events, atmospheric muon background events that sneak into corridors between the IceCube strings, and other events that are not well simulated or accounted for in the MC.

### 5.2.1 Identifying Atmospheric Muons

After Level 6 (Section 3.4.5), the muon rates are comparable with the neutrinos. Most of the remaining muons are ones that have not been identified as clearly "muon-like" in previous cuts. One category of this are coincident muon events, which occur when a muon and neutrino interact in the detector within the same triggered time window. These are not simulated due to technical limits in the MC, though they do occur in real data, and thus must be removed through cuts to ensure data/MC agreement. A few cuts are applied at the analysis level to remove any remaining muons, likely with signatures like coincident events.

### 5.2.1.1 Straight Veto Cut for Muons

There are two counting variables used to apply veto cuts to remove muons at the final level. The first is variable is the number of hits in the top 15 layers of IceCube DOMs, with the cut at  $< 0.5$ . The second variable is the number of hits in the outermost IceCube strings, and is cut at  $< 7.5$  hits. These cuts only remove a small portion of the final sample ( $< 0.05\%$ ). Figures 5.1 and 5.2 show the distributions of these variables with all analysis cuts applied, except for the one in question.

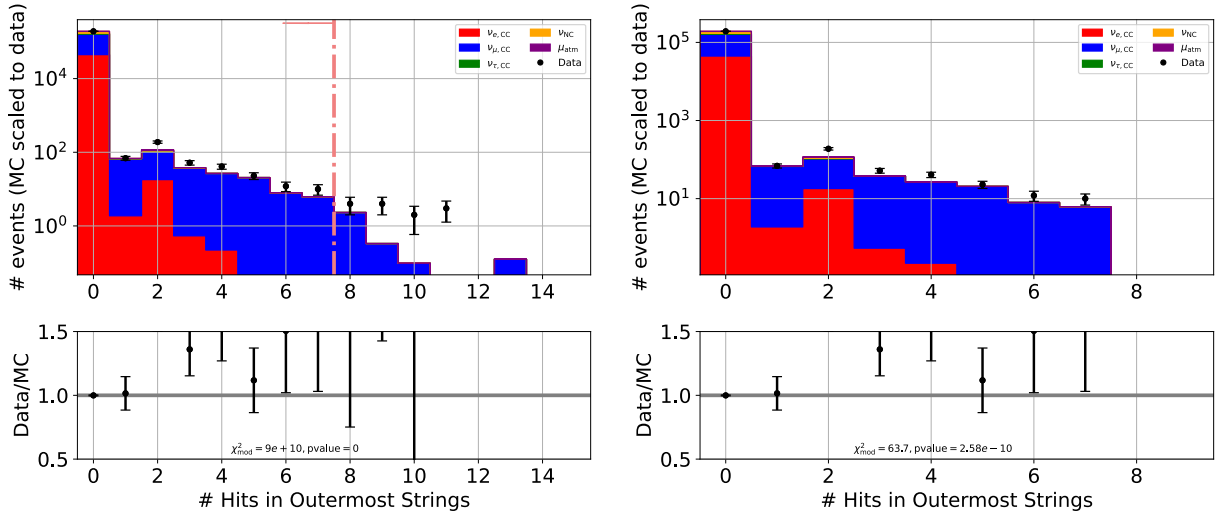


Figure 5.1: The data/MC disagreement for the number of hits in the top 15 layers of IceCube DOMs. The left plot shows before the cut (keeping events  $< 0.5$  hits) and the right plot shows the adjusted agreement after these events are removed.

### 5.2.1.2 Muon BDT

A more sophisticated cut is also employed to remove remaining atmospheric muons at the analysis level using machine learning. Rather than using the FLERCNN muon classifier directly, a BDT was trained using the FLERCNN muon classifier score as one of five training inputs. This BDT was optimized by Dr. Shiqi Yu to improve the performance of the muon identification. The training and performance of the muon BDT will be described briefly below as it relies heavily on FLERCNN variables.

The Muon BDT uses 5 variables as inputs:

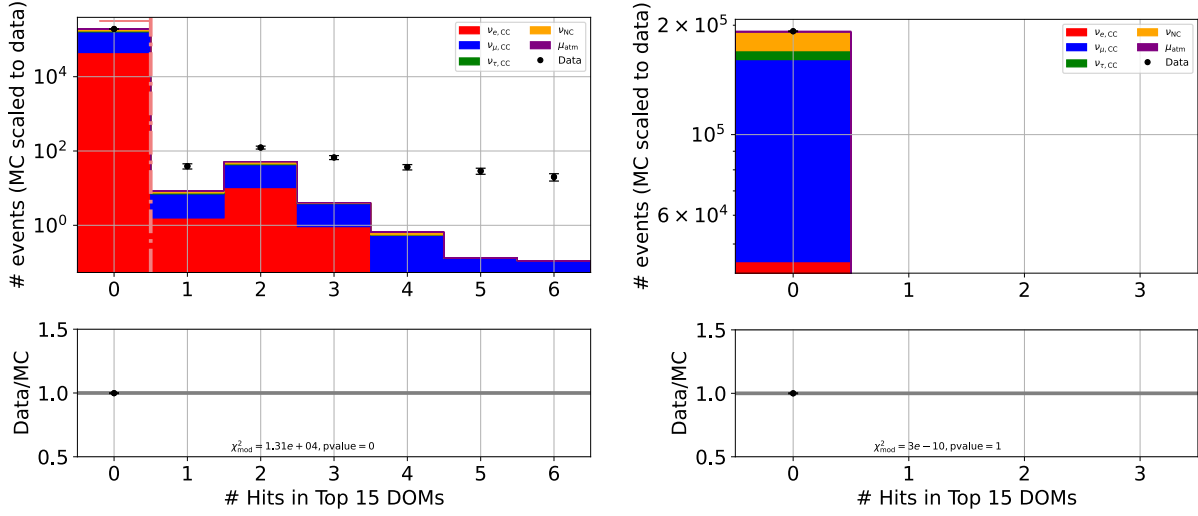


Figure 5.2: The data/MC disagreement for the number of hits in the outermost IceCube strings. The left plot shows before the cut (keeping events with  $< 7.5$  hits) and the right plot shows the adjusted agreement after these events are removed.

- FLERCNN vertex  $z$  -  $z$  depth position of the interaction vertex, reconstructed by the CNN
- FLERCNN vertex  $\rho_{36}$  - radial position of the interaction vertex, reconstructed by the CNN
- FLERCNN Muon Classification - score of how muon-like the event is according to the CNN classifier
- Probability Neutrino from Data BDT - probability the event is a neutrino, according to a BDT trained at Level 4 on real data to distinguish muons (Section 3.4.3)
- Corridor Cut's  $z$  Minimum - Lowest  $z$  position recorded that matches the Corridor Cut algorithm's hypothesis (Section 3.4.4)

The reconstructed vertex  $z$  and  $\rho_{36}$  are used because atmospheric muons are generated outside of the detector whereas the neutrinos we are detecting would have muon signatures starting inside of the detector. The muon from a  $\nu_\mu$  CC event should have its starting position at the neutrino interaction vertex, so providing the reconstructed vertex to the BDT should help it distinguish background muons that come from outside of the detector vs. the signature we want to see with the track starting inside of the detector.



The training sample for the BDT uses the MC analysis sample at Level 6 with the FLERCNN reconstruction applied (Section 3.1.2). There is only one additional cut applied, requiring  $\cos(\theta) < 0.3$ . This is applied to both the training and testing samples, where 50% of the neutrino analysis sample is used for training (50% for testing) and 70% of the muon analysis sample is used for training (30% for testing), with the exception of the 10% of the muon sample used for training the FLERCNN muon classifier. This is to prevent any bias being introduced in the training or testing of the BDT from the CNN already being optimized over this sub-sample of events.

The left plot in Figure 5.3 shows clear distinguishing power between the true atmospheric muons, identified with a peak near the score of 0 for how neutrino-like these events are, and the true neutrinos, identified near 1.0 neutrino-like score with a strong peak. To compare directly with the FLERCNN muon classifier's capability alone, the Muon BDT has an AUC of 0.961 from the ROC plot (not shown here), which is improved from the FLERCNN muon classifier's AUC of 0.947.

The right plot in Figure 5.3 illustrates the rate remaining between the neutrino and muon samples for various proposed cut thresholds using the muon BDT. The cut used for this analysis is at a 0.8 neutrino-like score, which has a strong effect of reducing the muon events down to a rate  $O(10^2)$  mHz. This cut does remove 23% of neutrinos when added to the final sample, but the background muon reduction is deemed worth this loss. In total, the total neutrino rate is still at acceptable levels on the  $O(0.1)$  mHz, and the muon is reduced to a fraction of a percent of the sample (Section 5.3).

Both the training and testing samples have similar behavior in both plots in Figure 5.3, which indicates the BDT is not overtrained and the training sample can be used in the final analysis. Thus, 100% of the muon analysis sample is used in the final analysis.

### 5.2.2 Direct Pulses

There are some remaining coincident noise events in the detector, that are not captured by the coincident cut described in Section 5.2.1.1. These can be identified by an algorithm that identifies "direct" pulses, from photons that do not undergo scattering. The SANTA algorithm accomplishes this by searching for photons that are causally connected to the reconstructed vertex [10]. The

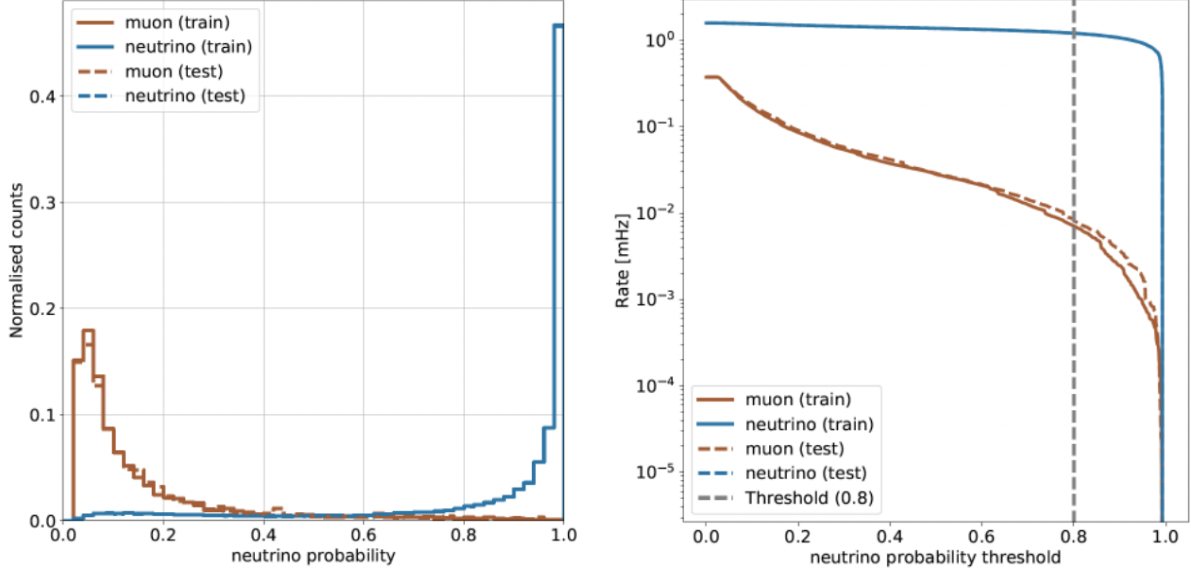


Figure 5.3: Performance of Muon BDT, with the BDT predicted neutrino probability on the left and the rate with various proposed threshold cuts on the right (plots created by Dr. Shiqi Yu).

SANTA-determined number of hit DOMs with direct pulses, shown in Figure 5.4, is used to apply a cut on the number of DOMs hit with direct pulses, requiring at least 3 pulses. This cut results in a total neutrino sample reduction of 10.50%, after all other analysis cuts are applied (Table 5.1). Figure 5.4 shows this distribution with all cuts applied except for the number of DOMs hit with direct pulses, showing a large data/MC disagreement below 3 directly hit DOMs. Note that the MC is scaled to the data, so once the events below 3 direct hit DOMs are removed, the remaining data/MC disagreement shifts to hover around the 1.0 ideal line.

### 5.2.3 Cosine Zenith Range

The zenith angle range has a cut applied to one end of the boundary. This is to avoid self-veto atmospheric effect. The self-veto atmospheric effect is where a muon and neutrino from a single air shower both interact in the detector, which would be likely removed due to the selection process. Since the muon and neutrino simulation is generated independently, this loss of neutrinos would not be accounted for in the MC sample. Thus, applying the cosine zenith cut, requiring events to have a reconstructed  $\cos(\theta) < 0.3$ , helps to remove the events where the self-veto effect is most likely.

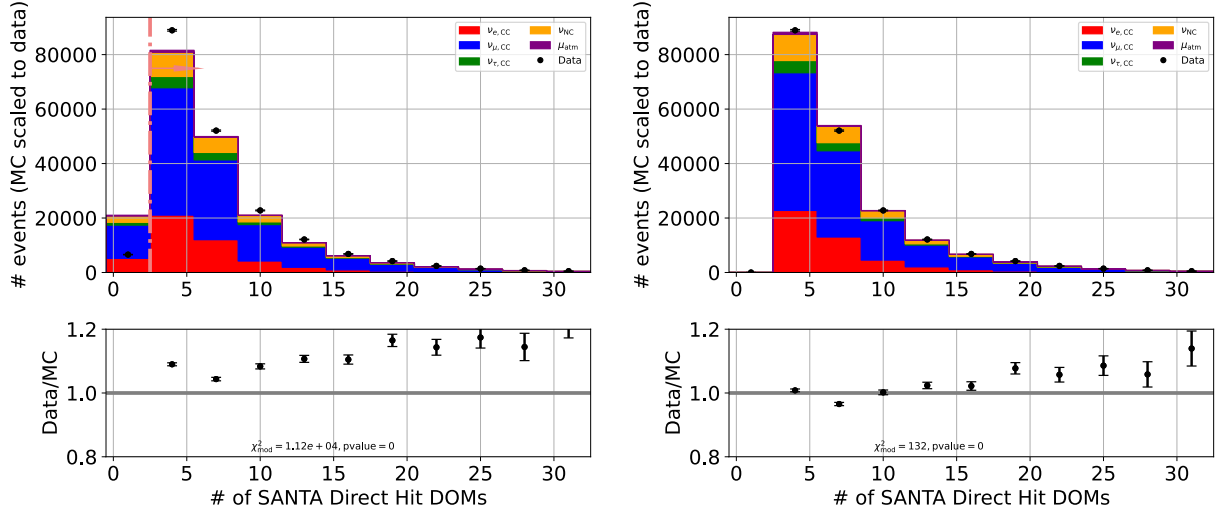


Figure 5.4: Number of DOMs hit with direct pulses, identified using the SANTA algorithm. The left plot shows before the cut (keeping events with  $< 3$  direct hits) and the right plot shows the adjusted agreement after these events are removed. The MC scaling to the data means that removing the mismatched bin improves the overall normalization for the data/MC agreement.

In addition, this cut helps to remove any downgoing muons that sneak through other muon-tuned cuts.

### 5.3 Application of Cuts

Table 5.1 summarized the impact of the cuts described on an individual basis for the neutrino sample. The percentage in the last column is the effect of the specified cut when all other cuts are already applied to the Level 5 sample, to show the impact that this last cut alone has on the final analysis sample. The energy and zenith cuts are separated in the table to notate the fact that these cuts are used in the analysis sample selection, so that some fraction of the events cut are removed due to the fact that they are not in the analysis region (rather than attempting to remove backgrounds). Thus, the fact that these percentages are larger for the energy and zenith cuts are not a concern, due to the fact that these cuts are aiming to constrain the sample to the necessary analysis region. Note that even with the individual contribution percentages, only 34.83% of the neutrinos are removed in total between the Level 5 cuts and the analysis cuts. Thus, these cuts are not mutually exclusive. The muon reduction is much larger, with the rate reduced by nearly a factor

of 200 between Level 5 and the final FLERCNN analysis cuts (Figure 5.5).

Cut Name	Value	% Removed
# CNN DOMs	$\geq 7$	1.05%
CNN Radius	$< 200$ m	0.09%
CNN Z	$-495 \text{ m} < z < -225 \text{ m}$	5.48%
CNN Vertex	R and Z cut	6.12%
DOMs w/ direct pulse	$> 2.5$	10.50%
NTop15	$< 0.5$	0.03%
NOuter	$< 7.5$	0.001%
Muon BDT	$\geq 0.8$	23.90%
CNN Energy	$5 \text{ GeV} < E < 100 \text{ GeV}$	20.70%
CNN Cosine Zenith	$< 0.3$	19.66%
Total	all above cuts	34.83%

Table 5.1: Describes cuts on the neutrino MC analysis sample and their individual impact when added to the final sample.

The final event rates after the FLERCNN cuts are listed in Table 5.2. Figure 5.5 shows the reduction of noise and atmospheric muon background in relation to the neutrino samples. Rates for Levels 2-5 are extracted from previous work done by IceCube collaborators who optimized the early level processing [44], which is contrasted to the final FLERCNN analysis sample cuts at the last data points.

At the final level, the atmospheric muon rate is an order of magnitude smaller than the lowest neutrino rate ( $\nu_\tau$ ) and the noise simulation sample has been completely eliminated. The noise rate in Table 5.2 has an upper limit of the noise rate with the limited statistics simulated. The FLERCNN selection rates do take into account the re-weighting from the neutrino hypersurfaces and smoothing from the muon KDE, which are described in Sections 6.4.4.1 and 5.4.3.

#### 5.4 Analysis Distribution Template

Chapter 4 and Section 5.3 have established that the reconstruction is completed for the necessary analysis parameters (energy, zenith, and PID) and the background has been reduced. The sample is now ready to be prepared for the  $\nu_\mu$  disappearance analysis, choosing the binning to define the analysis sample. This separation into the (energy, zenith, and PID) bins will be referred to as the

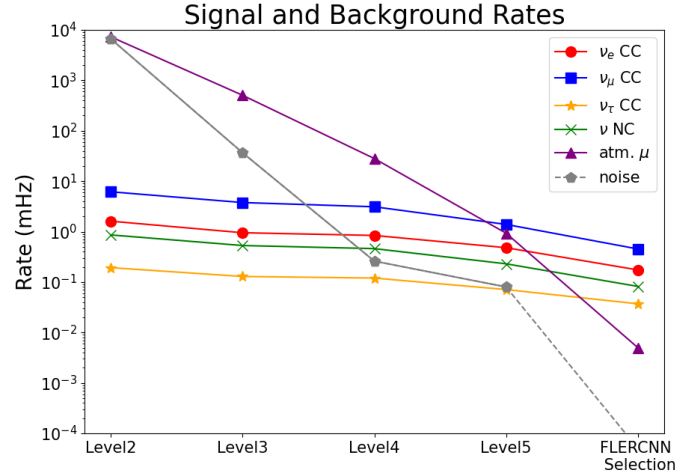


Figure 5.5: Rates of the different selections comparison from the earlier levels to the final level FLERCNN sample after analysis cuts, including re-weighting with interpolated hypersurfaces and muon KDE smoothing described in Sections 6.4.4.1 and 5.4.3.

Selection	Rate (mHz)	# of Counts (9.3 yr)	% of Sample
$\nu_e$ CC	0.175	$51,096 \pm 77$	23.2%
$\nu_\mu$ CC	0.453	$132,674 \pm 132$	60.3%
$\nu_\tau$ CC	0.037	$10,747 \pm 23$	4.9%
$\nu$ NC	0.082	$24,053 \pm 53$	10.9%
$\mu_{atm}$	0.005	$1,392 \pm 95$	0.6%
noise	< 0.0006	~0	< 0.08%
total	0.750	$219,964 \pm 190$	-

Table 5.2: Describes the final rates, including re-weighting with interpolated hypersurfaces and muon KDE smoothing described in Sections 6.4.4.1 and 5.4.3, predicted by the MC simulation.

"analysis binning," where the event count for each bin will be utilized for comparison in the final analysis (Section 6.1).

### 5.4.1 PID

Following the precedent set by the likelihood-based reconstruction IceCube 9.28 year analysis, a three PID bin approach is taken for the PID. For the three bins, two are cut so that they create the purest possible track-like and cascade-like bins after classification, with similar statistics in each. The remaining events are put into the "mixed" PID bin category, with a score somewhere between

the purest track-like and cascade-like events.

Figure 5.6 shows the classifier score for various true neutrino (and muon) flavors, with finer binning in the top plot and separated into the three analysis bins on the bottom plot. The cut for the track bin is at  $\geq 0.55$ , to create the largest sample of high purity track-like events. For the cascade bin, it is impossible to completely avoid  $\nu_\mu$  CC events since low energy events are often mis-classified due to the very short muon tracks (see Figure 5.8). The cascade cut is applied at  $\leq 0.25$  to avoid the peak of the  $\nu_\mu$  CC events near 0.35, and to keep similar statistics to the pure track-like bin.

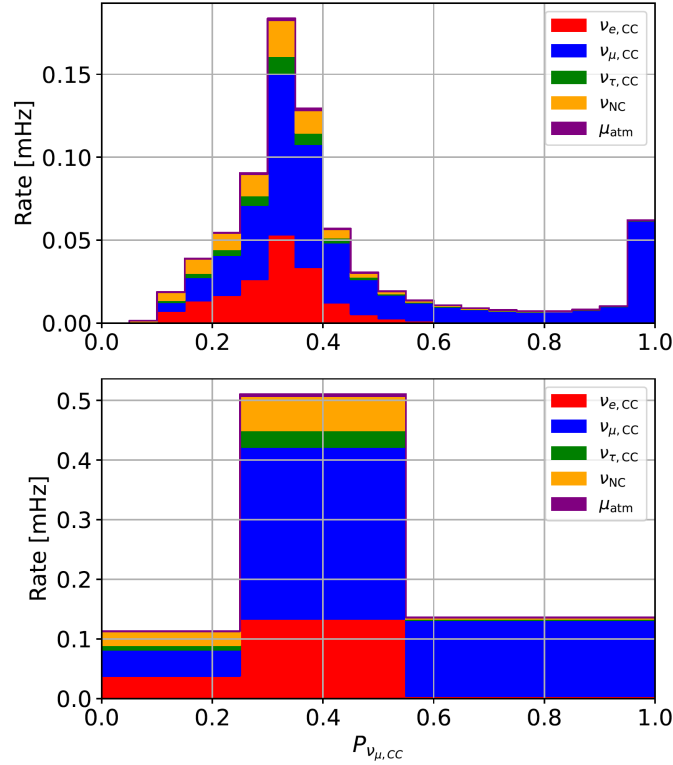


Figure 5.6: Result of the CNN track-like classifier on the nominal MC, binned more finely in the top plot and binned at the proposed 3-bin boundaries for analysis in the bottom plot. Note that the neutrino (and muon) flavors are all stacked.

After the cut has been applied, we explore the relative effectiveness of the resulting separation. Figure 5.7 shows that the high energy  $\nu_\mu$  CC events are correctly classified as tracks, whereas the low energy  $\nu_\mu$  CC events are typically classified in the mixed PID bin. Figure 5.8 shows that the track PID bin is dominated by  $\nu_\mu$  CC events and the mixed PID bin is still mostly tracks, but much

less pure. This shows that the chosen separation accomplishes the intended goal to have a high purity track-like bin.

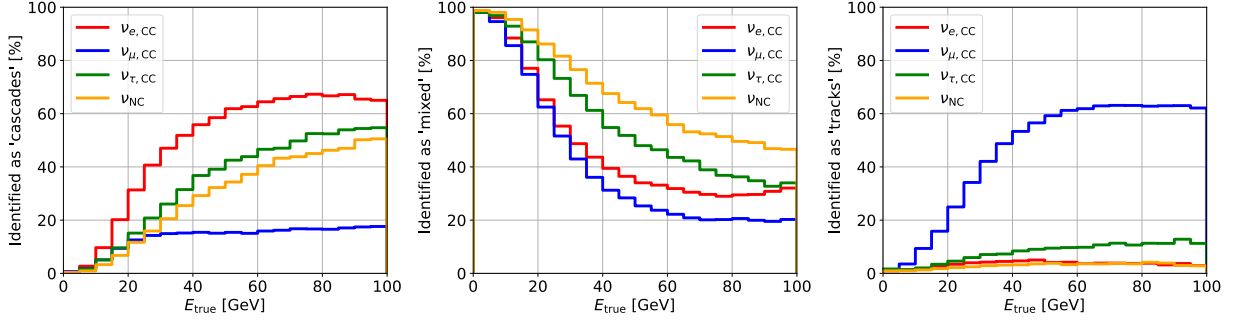


Figure 5.7: Percentages for each flavor's fraction of events located in that PID bin as a function of energy. For example, the  $\nu_\mu$  CC sample shows that 80% of the  $\nu_\mu$  CC events  $> 60$  GeV are classified as "tracks" with the chosen PID binning.

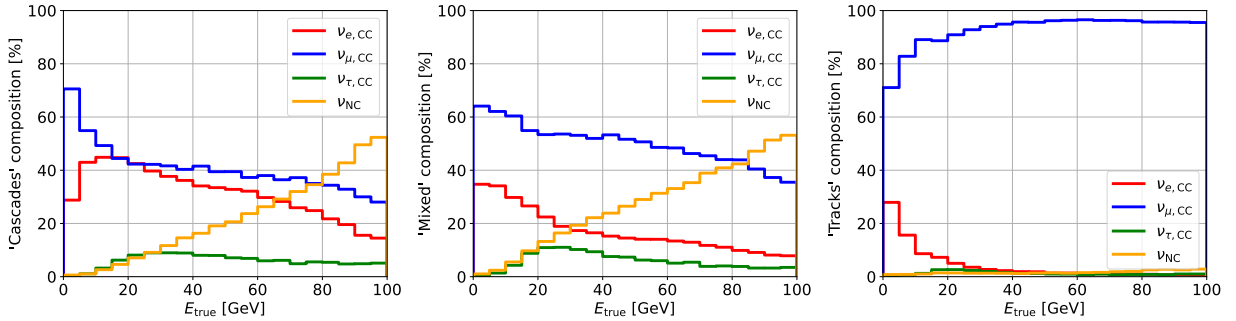


Figure 5.8: Composition of each PID bin as a function of true energy, with the relative percent of the each flavor's contribution to the events in each PID bin.

## 5.4.2 Binning

To see the oscillation signatures, the template is binned in reconstructed  $\cos(\theta_z)$ , energy, and PID. The PID binning is described in Section 5.4.1. The chosen range for the energy and zenith are described in Sections 5.1.3 and 5.2.3. The energy uses logarithmic binning to account for the spectral index favoring neutrino flux low energies and to focus on the expected oscillation signature region. The zenith axis is binned linearly in the cosine of the angle. The number of bins in energy and zenith is chosen to balance an ideal finest binning with ensuring there are enough statistics

expected per bin in 9.28 years of livetime. There are a few bins that do result in extremely low expected counts, in the cascade PID bin, and Section 5.4.4 discusses how this is dealt with in the analysis. The binning is summarized in Table 5.3 and the resulting analysis template distribution is in Figure 5.9.

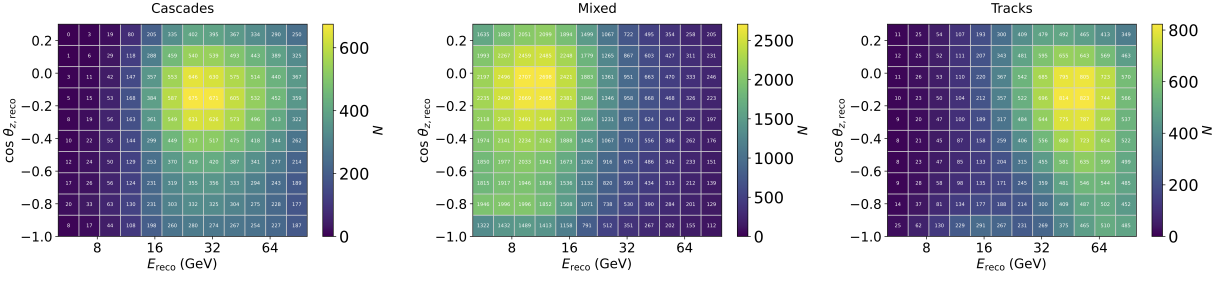


Figure 5.9: Number of expected events in each bin over the given 9.28 years of livetime, with neutrino weights adjusted by interpolated hypersurfaces to account for detector systematics and the expected muon distribution smoothed by KDE.

Variable	Number Bins	Bounds	step
PID	3	[0, 0.25, 0.55, 1.00]	linear
Energy	12	[5, 100]	log
Zenith	10	[-1.0, 0.3]	linear

Table 5.3: Describes the analysis binning using the CNN reconstructed and classified variables.

### 5.4.3 Muon Sample KDE Template

The muon sample is reduced to  $\sim 1000$  events after the analysis selection. Thus, there are not enough statistics after applying the analysis binning to produce hypersurfaces for the muon distribution.

Due to the low statistics of muons that survive to the final analysis sample, the distribution has artificially sharp features with large differences between neighboring bins (see first column in Figure 5.10). Thus, a Kernel Density Estimator (KDE) is applied on the distribution to smooth the muon template. The sample is smoothed without the analysis boundaries on energy and zenith angle applied, so that muon events on the edges of the analysis region can influence the generated KDE distribution. The result of the muon distribution after the KDE smoothing is applied is shown in the center column of Figure 5.10, with the final column showing the relative pulls between the



original template and the new smoothed version. Note that most bins are under a  $5\sigma$  pull, and in particular the signal region in the track PID bin is minimally affected.

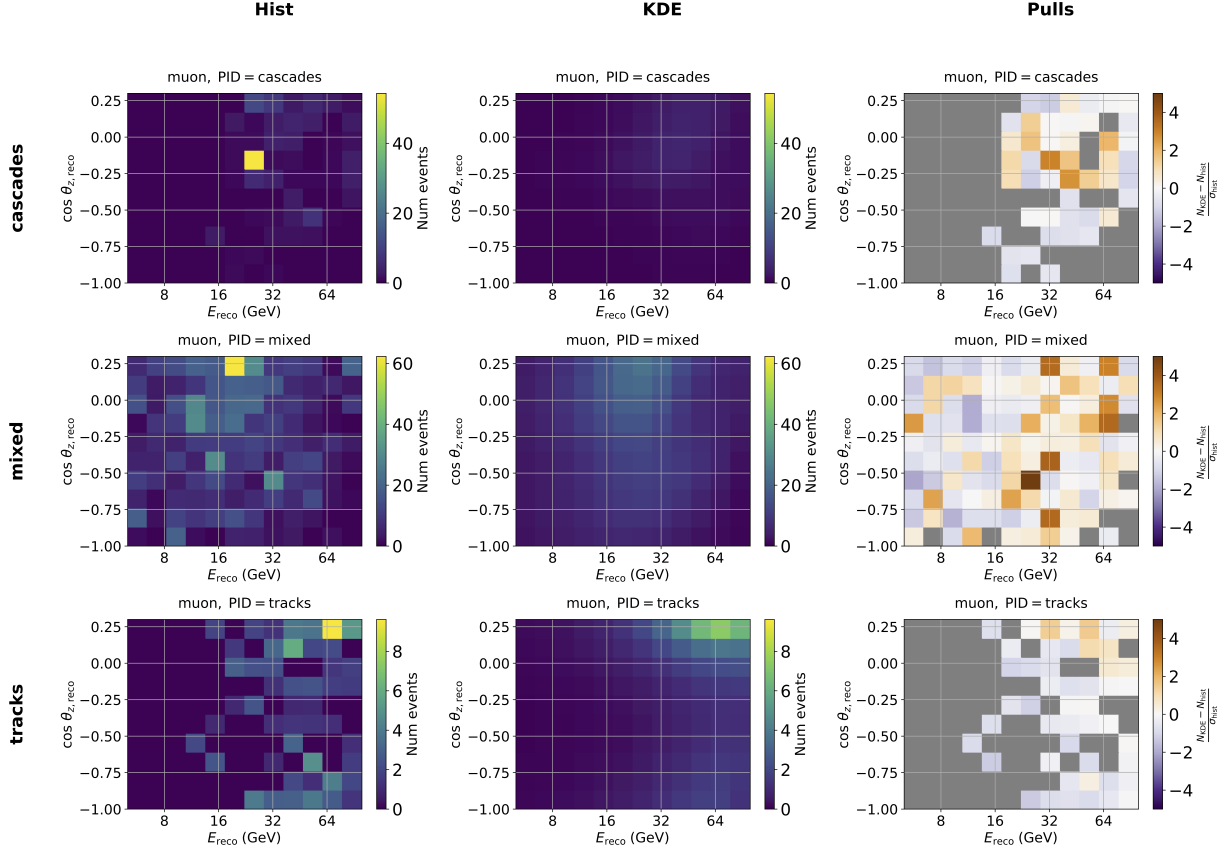


Figure 5.10: Muon distribution in the analysis binning before the KDE smoothing (left column), with sharp features due to low statistics, and after KDE smoothing applied (middle column) along with the pull in each bin (right column). The gray bins are where  $\sigma_{\text{hist}} = 0$ .

Exploring the effect of the remaining muon sample after the KDE smoothing, we look at the composition of the muons per analysis bin. Figure 5.11 shows the percentage of muons expected in the sample per analysis bin. Most bins contain less than 3% muons, except for the extremely low statistics bins in the cascade PID. This reinforces the need to mask out the analysis bins with low event counts, such as those in the upper left corner of the cascade PID histogram.

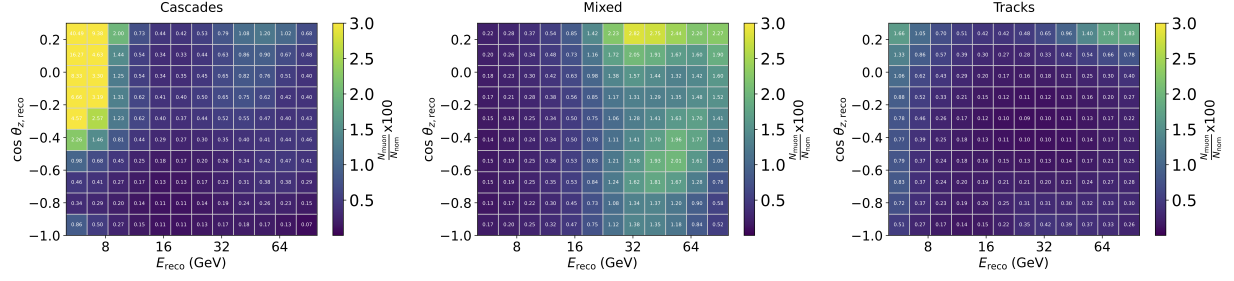


Figure 5.11: Percentage of muons expected in the sample per analysis bin after muon KDE smoothing.

#### 5.4.4 Masking

The bins with very low statistics can lead to large influences from background in those select analysis bins, as seen in the muon KDE. In addition, bins with low statistics can lead to large uncertainties in the hypersurface fits (Section 6.4.4.1). Furthermore, these bins can be the source of unwanted behavior in the oscillations fit if there are a few data events in a bin with  $< 1$  event predicted, creating a large pull due to a small statistical change.

To avoid complications from these bins, they are masked out during the analysis (Figure 5.12). In total, there are 10 bins masked in the analysis, all in the cascade PID bin. These bins each have fewer than 15 events predicted in the 9.3 year livetime, so masking these bins does not cause any appreciable change in the total expected rate for any of the flavors (as seen in Table 5.2).

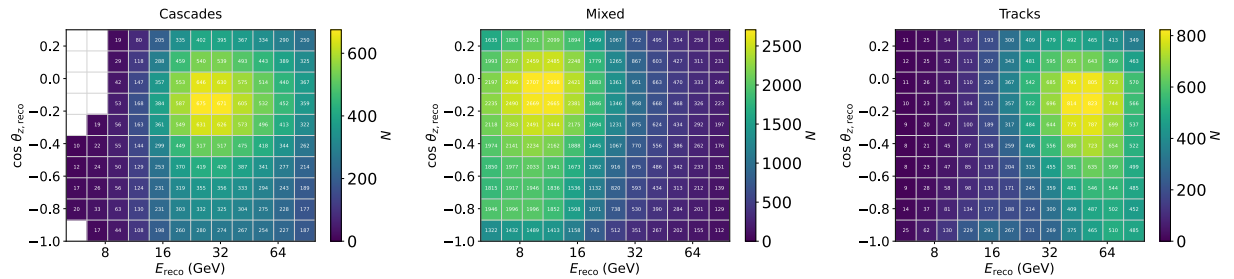


Figure 5.12: Analysis binning after masking, with the masked bins shown in white.

## CHAPTER 6

### ANALYSIS OF MUON NEUTRINO DISAPPEARANCE WITH 9.28 YEARS OF DATA

#### 6.1 Overview of Analysis Procedure

Now that the analysis sample has been established, we can perform the  $\nu_\mu$  disappearance analysis considering the full 3-flavor neutrino oscillation and matter effects. The analysis framework used is the PISA software, which was written by IceCube collaborators to use in high statistics neutrino oscillation experiments, optimized to analyze small signals [57, 58]. The goal is to use PISA to constrain the mass splitting  $\Delta m_{32}^2$  and the mixing angle  $\sin^2 \theta_{23}$  oscillation parameters. To do this, the expected and observed counts per bin are compared (Section 5.4.2).

The data undergoes the processing up to Level 6 (Section 3.4), FLERCNN reconstruction (Chapter 4) and muon BDT (Section 5.2.1.2), and application of the analysis cuts (Chapter 5). The data is then used to provide the observed counts. The MC analysis sample outlined in Chapter 5 provides the expected counts per bin (Figure 5.12), weighted to match the data livetime. The systematic uncertainties and physics parameters for the MC are then varied to create a new realization of the expected counts per bin, and a test statistic (Section 6.2) is calculated to quantify the comparison of the new bin-by-bin prediction to the data. The test statistic is minimized, finding the optimal fitted values for each of the free systematic parameters (Section 6.4) and the two physics parameters ( $\Delta m_{32}^2$  and  $\sin^2 \theta_{23}$ ).

Chapter 6 outlines the analysis using 9.28 years of livetime. This analysis is done in conjunction with Dr. Shiqi Yu and Finn Mayhew, and is still undergoing final checks in the review by the IceCube collaboration. Due to ongoing studies being performed for the 9.28 year analysis, a proof of concept analysis is run using the same procedure (with minor modifications) on 0.92 years of livetime (described in Chapter 7). This uses 10% of the full data sample to confirm the analysis procedure without biasing the analysers to the result of the full sample, since the smaller sample will be dominated by statistical uncertainties.

## 6.2 Test Statistic

The test statistic (TS) is the function that is calculated at each realization of systematic and physics parameters. The scan continues until the test statistic converges on a minimum. The function used for this analysis is a binned Poisson likelihood, evaluating the likelihood of observing the data count ( $d$ ) in each bin ( $i$ ) given the expected number of events ( $n$ ).

$$\mathcal{L}(d|n(\vec{\eta}, \vec{s})) = \prod_{i \in \text{bins}} \frac{n_i^{d_i} e^{-n_i}}{d_i!} \quad (6.1)$$

$$\ln(\mathcal{L}(d|n(\vec{\eta}, \vec{s}))) = \sum_{i \in \text{bins}} \ln \left( \frac{n_i^{d_i} e^{-n_i}}{d_i!} \right) \quad (6.2)$$

The expected number of events is dependent on  $\vec{\eta}$  which represents the physics parameters and  $\vec{s}$  which represents the systematic parameters. These systematic parameters are described in detail in Section 6.4. The natural logarithm is taken of the likelihood (abbreviated *LLH*) and the negative *LLH* is the value that is minimized. To account for the systematic parameters, a penalty term is added to use the priors on our systematics to influence the final LLH metric.

$$\ln(\mathcal{L}(d|n(\vec{\eta}, \vec{s}))) = \sum_{i \in \text{bins}} \ln \left( \frac{n_i^{d_i} e^{-n_i}}{d_i!} \right) + \sum_{j \in \text{syst}} \frac{(s_j - \hat{s}_j)^2}{\sigma_{s_j}^2} \quad (6.3)$$

where  $j$  accounts for all 16 systematic parameters with the fitted value ( $s_j$ ) compared to the nominal value ( $\hat{s}_j$ ).

This process of minimizing the  $-LLH$  over the physics and systematic parameters is called the "fit". "Pre-fit" means the MC distribution using the nominal systematic parameter values and the physics parameters set to the previous best-fit point from the IceCube  $\nu_\mu$  disappearance analysis using the 7.5 year livetime "golden event" sample (values in Table 6.2). "Post-fit" indicates the systematic and physics parameters at the best-fit values for the FLERCNN analysis sample.

### 6.3 Verifying Data Consistency

Data from IceCube seasons 2012 - 2021 is used for this analysis, for a total livetime of 9.28 years. IceCube data-taking seasons typically start in April or May of the year and do not necessarily correspond to one year of livetime. Note that only part of the 2021 season is complete at the time of this analysis, so only the 2021 season up to Jan 2022 is used for this study.

#### 6.3.1 Data Quality Check Between Years

The data quality is checked to verify that the analysis variables are consistent between different years of IceCube data. All data is used for the quality check, with a Kolmogorov–Smirnov test (KS-test) performed on various 1D distributions to assess their similarity. The KS-test measures the probability that the samples are drawn for the same underlying distribution, such that a higher  $p$ -value means they are similar. The null hypothesis is that all years of data are uniform, so a low  $p$ -value indicates higher probability of the distributions being distinguishable from each other.

The reconstructed energy distributions show good agreement between all years of the data (Figure 6.1). The left histogram shows mostly large  $p$ -values, with no concerning patterns in the distribution. The right plot in Figure 6.1 compares the actual  $p$ -values observed to the expected distribution of  $p$ -values, to assess if there are a reasonable number of low  $p$ -value bins since some are expected due to minor fluctuations over the 10 years.

The zenith angle test also shows good agreement between all years of data, with mostly large  $p$ -values spread across the 2D histogram (Figure 6.2). The two sided  $p$ -value is  $< 5\%$  in the right plot, but it pulls in the positive direction, where there are fewer low  $p$ -values actually seen than predicted. Thus, there is no concern with the quality of the zenith angle data agreement between the years.

The PID classifier data quality check shows good agreement between all years (Figure 6.3). The two-sided  $p$ -value also is above 5%, and favoring higher consistency with fewer actual low  $p$ -values than expected.

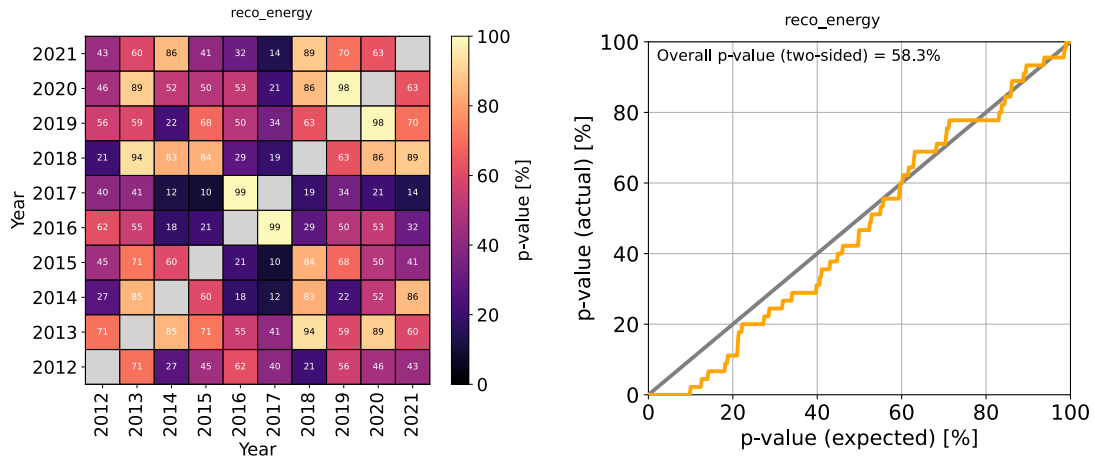


Figure 6.1: Energy distribution compared between years of data.

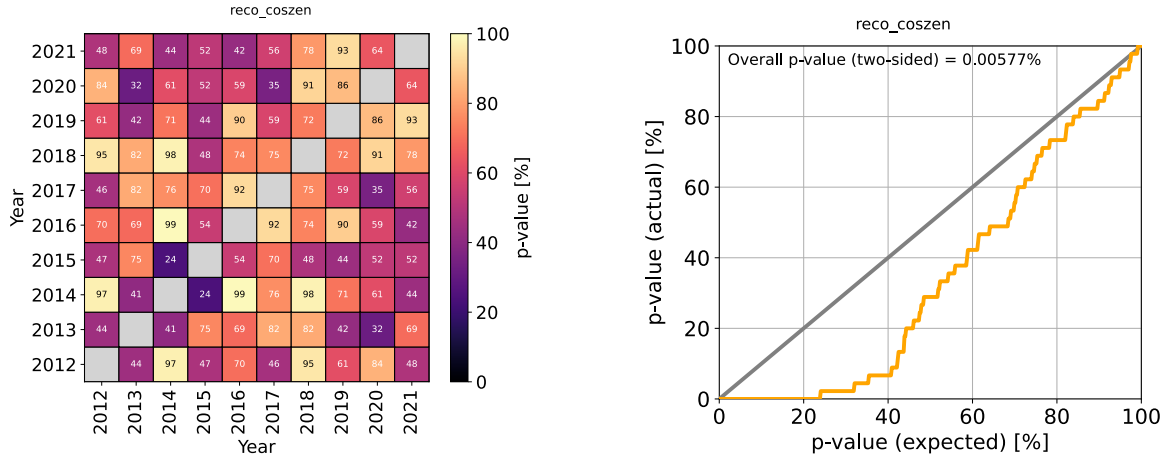


Figure 6.2: Zenith angle distribution compared between years of data.

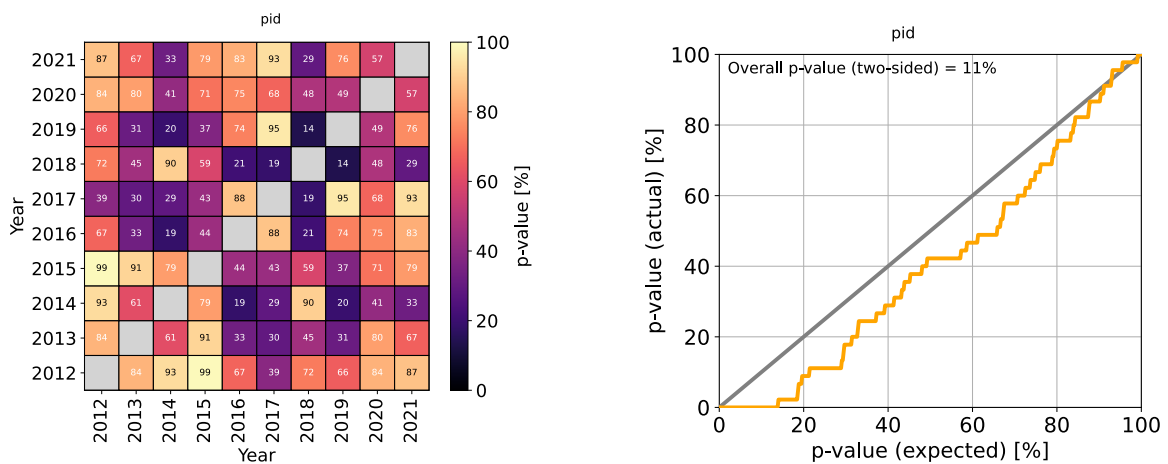


Figure 6.3: PID distribution compared between years of data.

The muon BDT data quality check shows a trend of slightly lower  $p$ -values for 2019 and 2020 (Figure 6.4). Overall, these differences are not too harsh (generally above 5%) and the two sided  $p$ -value (right plot) is well above the 5% threshold of concern. The likelihood-based 9.28 year analysis sample also saw similar behavior for their final level muon classifier's data quality check, and traced its behavior back to the Level 4 muon BDT (Section 3.4.3) that their final level BDT uses as input. Their study revealed that one of the Level 4 muon BDT inputs, an estimate of  $\rho_{36}$ , noticeably changes due to a slight change in the IceCube calibration file used for the data in the 2017 season and beyond, where a rounding error shifts the position of some DOMs by less than a mm. They theorize that the Level 4 Muon BDT distinguishes this change, creating the different behavior after 2017. The same Level 4 muon BDT is used as input for the FLERCNN final level muon BDT (Section 5.2.1.2). To fix this would require adjusting the calibration files and re-training the Level 4 BDT. This may not be the only cause, as the final level muon BDT shows larger changes in the 2019-2020 years (and reverts back in 2021). But as the effect is small and the  $p$ -values for these years do not exceed the 5% threshold set for concern, it was judged acceptable to move forward without re-optimizing for the small variation in the muon veto efficiency in those years.

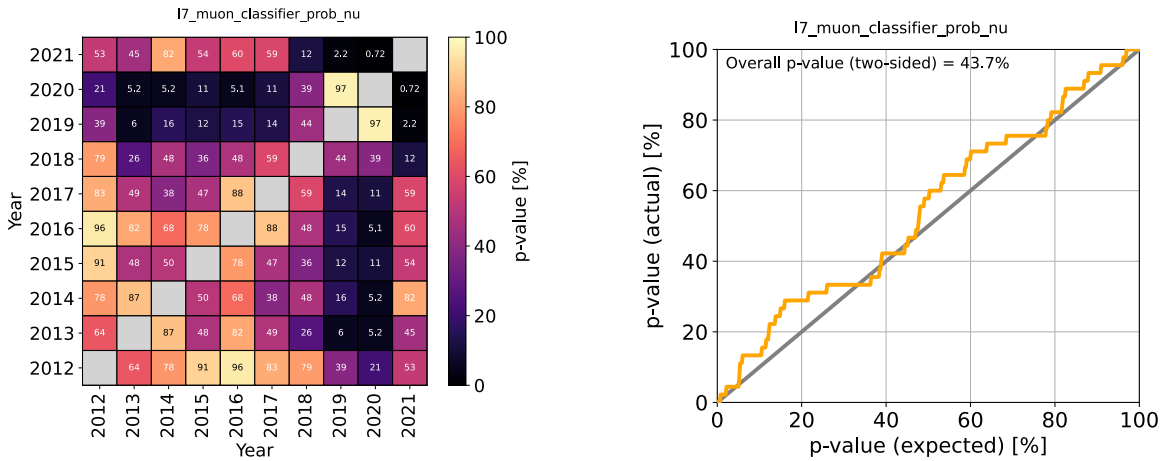


Figure 6.4: Muon BDT distribution compared between years of data.

Lastly, there is an expected variation in between the months in each year due to the seasonal atmospheric changes [37]. Nonetheless, there is a consistency expected in the months between each year, to a degree. Figure 6.5 shows the variation of the rate over each month, with the years

split by color. The results show consistency across the years, with most years being well within the error bars from each other.

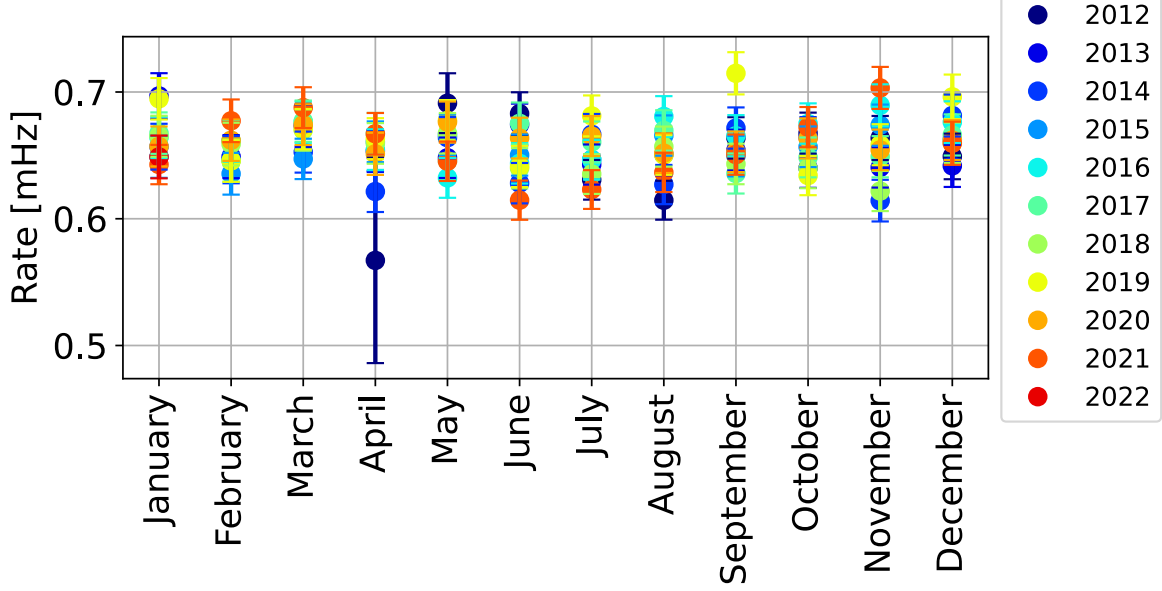


Figure 6.5: Rates compared between years of data, split up by month.

### 6.3.2 Data/MC Agreement

Data/MC agreement is evaluated by comparing the shape of the 1D distributions of the final level sample. The pre-fit agreement uses the analysis template with all physics and nuisance parameters set to nominal values (Table 6.2) with interpolated hypersurfaces applied. Before the fit is performed, we expect there to be some disagreement between the data and nominal MC comparison. Preliminary checks are still conducted to look for large discrepancies. The total MC rate is scaled to match the data rate, so this is a shape comparison only.

Figure 6.6 has the FLERCNN energy distribution on the left and FLERCNN zenith angle distribution on the right. The top histogram in each figure stacks the separated interactions in the MC indicated by the colored regions, scaled to compare with the data in the black data points. The bottom part of the figure has the data/MC ratio from each bin, bounded to look at the region around the 1.0 line representing perfect agreement.



The energy does well across most the range, with expected improvement once the MC uses the data's best fit for the parameters. The zenith angle agreement has a sharper disagreement, but to some extent this is expected since the zenith distribution is highly dependent on the flux parameters which are not yet fit to match the data. Neither of these plots show concerning, pre-fit disagreement and their trends are expected to be resolved after the fit parameters are applied to the MC (and is resolved post-fit in the one year analysis – see Section 7.4.1).

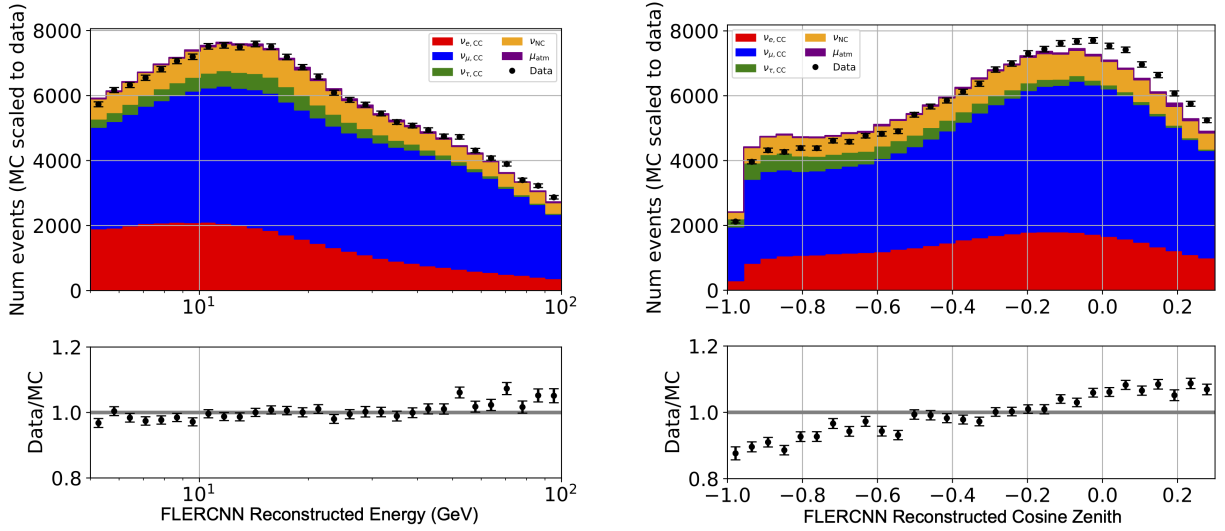


Figure 6.6: Data/MC agreement for reconstructed energy and zenith angle.

Figure 6.7 shows the data/MC agreement for the PID (left) and the atmospheric muon classifier (right). Both plots show good agreement across their included ranges (0-1 for PID and 0.8-1.0 for muon classifier).

The vertex  $z$  MC prediction in Figure 6.8 (left) shows good agreement with data across the included  $z$  region from -495 to -225 m. For the radial reconstruction, there is a known artifact in the CNN reconstruction where the string positions are favored in the  $x$  and  $y$  reconstruction, contributing to the two-peak shape seen in the  $\rho_{36}$  histogram (right plot in Figure 6.8). Since this reconstruction is used for the containment cut, the precise reconstructed position is not necessary to make a reasonable containment cut.

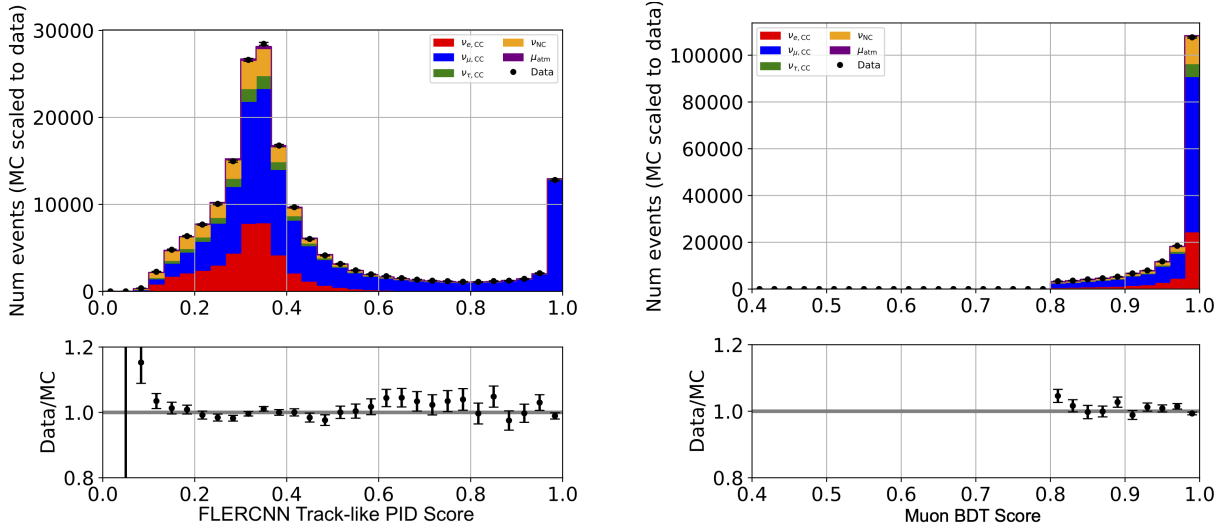


Figure 6.7: Data/MC agreement for PID and muon BDT classifiers. Note that the muon BDT classifier has a cut applied at 0.8, which accounts for the behavior from 0.4-0.8.

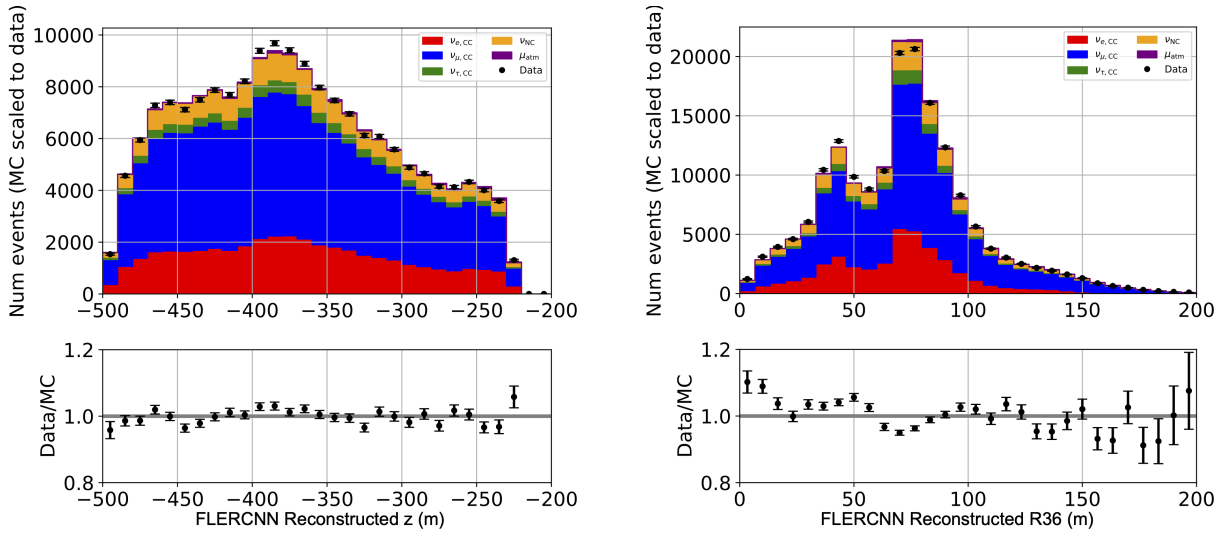


Figure 6.8: Data/MC agreement for starting vertex  $z$  and  $\rho_{36}$  reconstructed variables (radial position relative to the center IceCube string #36).

## 6.4 Systematic Uncertainties

Various systematic influences are assessed for this analysis. There are 16 systematic parameters which are free during the analysis fit, while others have been determined to not have a large effect on the data and thus are fixed. Selecting only the systematics that impact the analysis is important to help the minimizer converge, which becomes increasingly complex as additional parameters are added. Broadly, the systematics considered fall into four categories: atmospheric flux (Section 6.4.1), oscillation modeling (Section 6.4.2), cross sections (Section 6.4.3), and detector systematics (Section 6.4.4). Section 6.4.5 shows how the impact of the systematic parameters are determined, to justify fixing the other oscillation parameters to their nominal values.

### 6.4.1 Atmospheric Flux

The atmospheric flux model has a significant influence on the expected MC counts. The uncertainties for the flux models vary over energy and zenith angle, so that they must be accounted for carefully. The flux is dependent on the choice of the cosmic ray model, meson production yield, atmospheric model density, and hadronic interaction model. The FLERCNN analysis uses the MCEq toolkit to calculate the atmospheric flux [59]. Instead of enforcing the oscillation analysis to rely on the result of a single flux model, the introduction of the systematic parameters below allows the expected neutrino flux to vary and span the possible models and their uncertainties. The flux can be updated iteratively during the analysis due to the MCEq framework, allowing the counts to be re-weighted to reflect the flux estimation.

#### 6.4.1.1 Cosmic Ray Uncertainty

Shifts in the spectral index of the neutrino flux ( $\Delta\gamma_\nu$ ) and the energy pivot are due to the cosmic ray flux uncertainties, such that

$$\Delta\Phi = \left( \frac{E}{E_{\text{pivot}}} \right)^{\Delta\gamma_\nu} \quad (6.4)$$

The energy pivot is fixed during the fit at 24.1 GeV. The neutrino flux spectral index is free, and Figure 6.11 indicates that this systematic is ranked as one of the leading impacts on the analysis, along with the detector systematics.

#### 6.4.1.2 Uncertainty on Meson Production

As discussed in Section 1.4, atmospheric neutrinos are produced through pion and kaon decay in the atmosphere. Thus, the uncertainty on the neutrino production from the meson flux is a factor that can affect the neutrino expectation. The meson flux uncertainties are energy and zenith angle dependent, outlined by Barr *et al.* in [15], and hence the naming convention in Table 6.2. Letters a-i refer to the pion uncertainty and w-z the kaon uncertainty, with the letter indicating the energy and  $x_{\text{lab}}$  region that the Barr parameter accounts for [15]. The MCEq framework allows variation for each the systematic parameters individually within their own region corresponding to the Barr term.

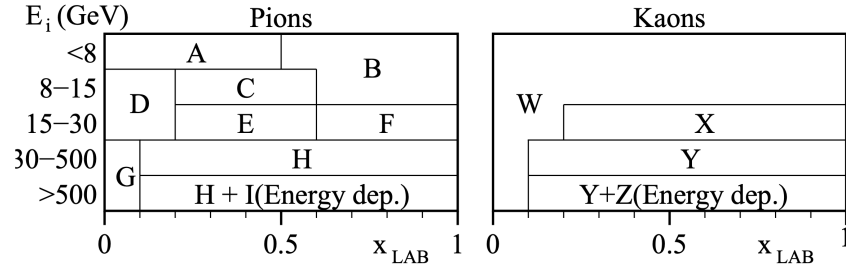


Figure 6.9: Barr parameters labels, chart from [15].

The Barr parameters mostly show a weak influence on the fit (Figure 6.11), with only one parameter having a  $\Delta LLH > 0.1\sigma$ . To keep consistency with the IceCube likelihood-based 9.28 year analysis, a total of five of the Barr parameters are chosen to remain free during the fit even though their expected impact is low for the FLERCNN analysis [9].

### 6.4.1.3 Normalization Terms

The normalization terms ( $N_x$  in Table 6.2) uniformly scales the rate for the specified particle in the subscript across the entire energy range. The only normalization terms that are free during the fit are the neutrino normalization ( $N_\nu$ ) and muon normalization ( $N_\mu$ ). The neutrino normalization scaling helps to take into account things like the neutrinos removed during the veto cuts, such as coincident events. The muon normalization can account for things such as count variations due to the detector systematics, which are not directly factored into the muon estimated counts (how these are handled for the neutrinos in Section 6.4.4). Thus, both of these normalization terms free during the fit to allow additional flexibility for the fit to account for possible uniform scaling in the expected counts.

### 6.4.2 Oscillation Modeling and Parameters

The oscillation modeling script must take into account parameters such as the earth density, depth of the detector, and height above the earth at which the neutrino is generated in the atmosphere (prop height). The earth density affects the matter effects—this analysis uses the 12 layer estimate of the preliminary reference Earth model [60] to account for the different density of from the core to the crust. There are three terms that account for the electron density in the inner core (YeI), outer core (YeO), and mantle (YeM). These values are fixed at the inputs used for the oscillation modeling computation (using Prob3++ [61].), along with the detector depth and neutrino generation height.

The  $\nu_\mu$  disappearance analysis aims to fit  $\Delta m_{31}^2$  and  $\sin^2 \theta_{23}$ , but since we are using the 3-flavor paradigm, the probability of oscillation also depends on the other mixing angles, other mass squared differences, and the  $\delta_{CP}$  violating term. This analysis is not sensitive to these variables, since they are subdominant to the  $\nu_\mu \rightarrow \nu_\tau$  oscillation (Section 1) so their values are fixed during the fit.

### 6.4.3 Cross Sections

The uncertainties on the cross sections for DIS, resonant, and QE scattering must be considered as systematic parameters. Contributions to all three cross sections are free during the fit.

The DIS systematic parameter in this analysis, simply referred to as "DIS" in Table 6.2, is a term that accounts for how "CSMS-like" the cross section is vs. "GENIE-like" (Section 1.3.3). This is done by finding the ratio of the GENIE and CSMS cross section and adjusting this ratio as a constant scale. With its current implementation, the DIS term is not applied with any energy dependence, but scales the total DIS cross section with an energy independent interpolation.

As described in Sections 1.3.2 and 1.3.1, the cross section for resonance and quasi-elastic interactions are determined by the form factor, which is dependent on the axial mass ( $M_A$ ) that has been determined from experimental measurements. Thus, the axial mass terms are separated for QE scattering  $M_{A,QE}$  and resonant scattering  $M_{A,res}$ , as they are implemented separately in GENIE [31]. Both of these axial mass terms are free systematic parameters that are fitted during the analysis. Note that a higher  $M_A$  implies higher cross section and thus higher neutrino yield—so including these terms in the systematics is critical to account for potential differences in the observed neutrino yield.

### 6.4.4 Detector Systematics

The detector systematic parameters accounted for in this analysis are the DOM efficiency, two variables used to parameterize the hole ice effects ( $p_0$  and  $p_1$ ), and the scattering and absorption of the bulk ice. Various MC systematic sets are generated with altered settings for these five detector systematics. These simulations provide discrete MC distributions that include the effect of these changed parameters, which are both energy and zenith dependent.

Table 6.1 describes the systematic sets' detector settings in detail, highlighting the changes from the nominal baseline values (first row of table). For each systematic set (corresponding to a row), there are 1-3 parameters changed from their nominal setting. The DOM efficiency indicates the PMT's sensitivity to photons, with a value of 1.0 indicating an efficiency aligned with calibration

measurements (Sections 2.1.3 and 4.4.1) and values  $< 1$  scale the efficiency so that less light per event is detected. The hole ice systematics describe the parameterize of the angular acceptance, where the magnitude of the changes indicate shifts in the upgoing ( $p_0$ ) and downgoing ( $p_1$ ) photon acceptances (Sections 2.3.2 and 4.4.2). The scattering and absorption values indicate scaling of the bulk ice scattering and absorption coefficients (Section 2.3.1). Thus, if the scattering systematic parameter is  $> 1$ , it indicates more photon scattering in the ice, which can decrease the detection rate of photons initially from the head on direction but increase the rate from more downgoing photons from being scattered toward the PMT face (Section 4.4.3). A absorption systematic parameter of  $> 1$  indicates more absorption in the ice and a lower likelihood of the photon being detected. In total, 30 systematic sets are used to generate *hypersurfaces* that interpolate between the discrete settings used to create the 30 systematic sets.

#### 6.4.4.1 Hypersurfaces

The hypersurfaces are used to interpolate between the discrete systematic sets and return a continuous, smooth space for the five detector systematics in each analysis bin (interaction type, reconstructed energy, reconstructed zenith). The sample is first split by PID, separated into three true categories due to their topology (Section 2.2.2):  $\nu_\mu$  CC (including  $\bar{\nu}_\mu$  CC),  $\nu_\tau$  CC (including  $\bar{\nu}_\tau$  CC), and  $\nu_e$  CC and all NC flavors (and their antineutrinos). Then, for each flavor they are calculated per analysis bin using a linear function where a slope and intercept are fit for each of the five detector systematic parameters.

Figure 6.10 illustrates the hypersurface fit projected on the DOM efficiency axis for the  $\nu_\mu$  CC (including  $\bar{\nu}_\mu$  CC) sample. Each hypersurface is fit on one [PID, energy, cosine zenith] bin, with the example in Figure 6.10 on the [track, 28.7 - 38.6 GeV, -1.0 - -0.87] analysis bin using only the  $\nu_\mu$  CC sample. The uncertainty on the hypersurface fit (shown in the red shading around the line) is less than the statistical uncertainty on the counts in the bin in the baseline simulation, so they are also used to re-weight the expected nominal simulated count in each MC bin (y-axis of Figure 6.10). Notice that the re-weighting factor for this example has the nominal (1.0) DOM efficiency

DOM Efficiency	Hole Ice $p_0$	Hole Ice $p_1$	Scattering	Absorption
1.00	0.101569	-0.049344	1.00	1.00
0.90	0.101569	-0.049344	1.00	1.00
0.95	0.101569	-0.049344	1.00	1.00
1.05	0.101569	-0.049344	1.00	1.00
1.10	0.101569	-0.049344	1.00	1.00
1.20	0.101569	-0.049344	1.00	1.00
1.00	-0.064805	-0.108849	1.00	1.00
1.00	-0.483907	-0.01712	1.00	1.00
1.00	0.280321	-0.075404	1.00	1.00
1.00	0.112231	0.003502	1.00	1.00
0.88	-0.049899	-0.054374	1.00	1.00
0.93	-0.372924	0.034933	1.00	1.00
0.97	0.296566	-0.036394	1.00	1.00
1.03	0.124492	-0.11328	1.00	1.00
1.12	-0.314209	-0.077367	1.00	1.00
1.00	-0.5	-0.049344	1.00	1.00
1.00	0.1	-0.049344	1.00	1.00
1.00	0.3	-0.049344	1.00	1.00
1.00	0.101569	-0.1	1.00	1.00
1.00	0.101569	0	1.00	1.00
1.00	0.101569	0.05	1.00	1.00
1.00	0.101569	0.10	1.00	1.00
1.00	-0.2	-0.049344	1.00	1.00
1.00	0.101569	-0.049344	1.05	1.05
1.00	0.101569	-0.049344	1.05	0.95
1.00	0.101569	-0.049344	0.95	1.05
1.00	0.101569	-0.049344	0.95	0.95
1.00	0.101569	-0.049344	1.00	1.10
1.00	0.101569	-0.049344	1.00	0.90
1.00	0.101569	-0.049344	1.10	1.00
1.00	0.101569	-0.049344	0.90	1.00

Table 6.1: Settings for the systematic sets used to generate the hypersurfaces, with each row describing a new set. Cells highlighted in green indicate settings that vary from their baseline values (row 1). Rows highlighted in gray indicate systematic sets that are currently under study with potential mismatched simulation (Section 7.2).



offset to  $\sim 1.1$ .

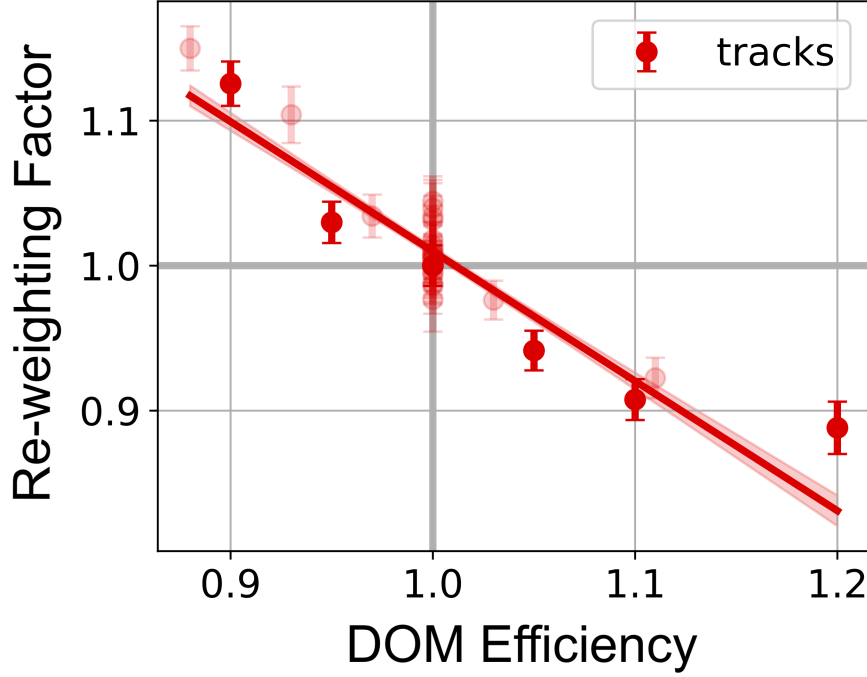


Figure 6.10: Example of hypersurface fit in one analysis bin projected on the DOM efficiency axes, using the discrete sets to create a continuous fit (plot from [8]). The lighter data points indicate systematic sets that have additional detector parameters pulled in addition to the DOM efficiency. The darker data points are systematic sets with only DOM efficiency changed.

Since the hypersurfaces are calculated using binned simulation, they are subject to dependence on non-detector parameters. To reduce this bias, the hypersurface fit for the five systematics is then interpolated over the oscillation and flux parameters since these are most significant non-detector parameters for this analysis. Thus, fitted hypersurfaces for the detector systematics are also interpolated in the  $\Delta m_{32}^2$ ,  $\theta_{23}$ , and the neutrino flux spectral index ( $\Delta\gamma_\nu$ ) parameter space.

#### 6.4.5 Assessment of Systematic Parameters' Impact

The systematic impact test aims to quantify the impact on the measured physics parameters if a given systematic parameter is fixed at a nominal, but the nominal is the incorrect value. Overall, this test should help determine which systematic parameters can safely be fixed during the analysis fit. The degree of *mismodeling* is calculated, which represents the potential that the true physics

parameters are rejected due to a specific systematic parameter being wrongly fixed and affecting the underlying data distribution. Note that no statistical fluctuations are applied in this test (referred to as an "Asimov" test). The procedure for the test is as follows:

1. Take one systematic parameter and pull by  $1\sigma$  (if Gaussian prior) or half way to its upper bound (if uniform prior).
2. Generate pseudo-data with this setting, and a choice for the true physics parameters.
3. Return parameter to nominal value and fix it there for the fit.
4. Fit the pseudo-data for the physics and systematic parameters, without fitting the fixed systematic parameter in question.
5. Calculate the difference in the TS between the fit with these settings vs. point representing the true physics parameters.
6. Repeat for different true values of the physics parameters, scanning 9 different combinations of the physics parameter settings.
7. Take the largest difference in TS for Figure 6.11.
8. Repeat for each systematic parameter in question.

To determine the relative  $\sigma$  values for the  $-2\Delta LLH$ , Wilks' theorem is used such that the  $LLH$  is assumed to be  $\chi^2$  distributed. The  $\sigma$  values were used as a metric to determine the systematic parameters that should be included in the fit, due to the mismodeling potential they have if they are fixed to an incorrect assumption. Thus, all parameters with a  $-2\Delta LLH > 0.1\sigma$  are left free during the analysis (indicated by orange color bar). In addition, there are some variables that are also left free, even with the expected mismodeling impact low.

The  $N_\mu$  normalization is free during the fit, so that there is a parameter to account for the scaling of the expected muon count. This is the only systematic parameter included to account for differences in the muon distribution, so it is not fixed in this analysis. The  $M_{A,QE}$  falls just below

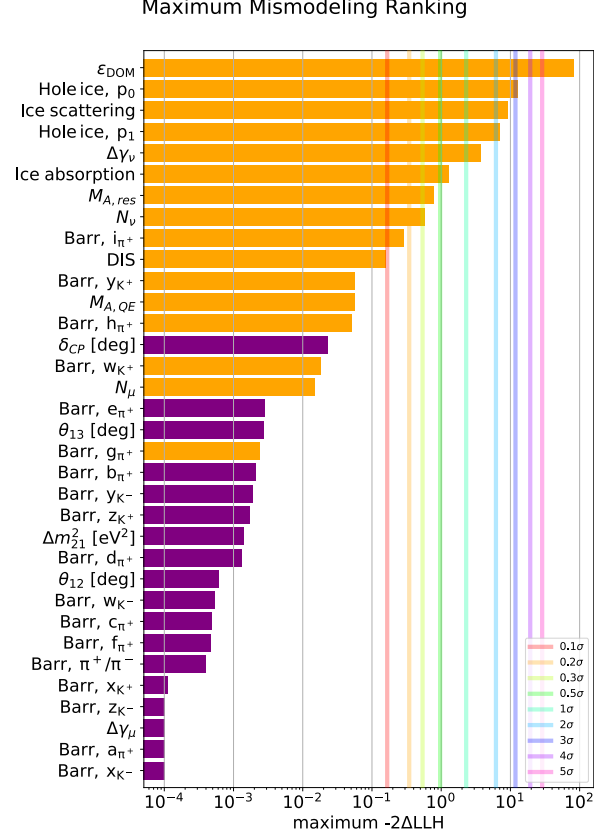


Figure 6.11: Relative effect of each systematic parameter on the analysis, determined by pulling the parameter from nominal to generate pseudo-data but fixing it to nominal when fitting. Bars in orange are systematic parameters that are free for the proposed analysis fit, while the bars in purple are proposed to be fixed.

the  $1\sigma$  threshold, but due to the fact that  $M_{A, QE}$  and  $M_{A, rest}$  are implemented separately in the GENIE simulation, both of these variables are left free for the fit. Lastly, four of the Barr parameters ( $y_{K^+}, h_{\pi^+}, w_{K^+}, g_{\pi^+}$ ) that have  $-2\Delta LLH > 0.1\sigma$  are left free for comparison purposed to the other IceCube 9.28 year analysis using the likelihood-based reconstruction.

#### 6.4.6 Summary of All Systematic Parameters

In total, there are 16 systematic parameters that are left free during the fit. These parameters are all seeded at their nominal value in Table 6.2, based on expected values for each systematic. The other systematic parameters are fixed at their nominal values.

Category	Param	Fixed/Free	Nominal	Prior	Range
flux	$\Delta\gamma_\nu$	Free	0	gaussian, $\sigma = \pm 0.1$	[ -0.5, 0.5 ]
flux	energy pivot [GeV]	Fixed	24.1	-	-
flux	Barr, $\pi^+/\pi^-$	Fixed	0	-	-
flux	Barr, a_ $\pi^+$	Fixed	0	-	-
flux	Barr, b_ $\pi^+$	Fixed	0	-	-
flux	Barr, c_ $\pi^+$	Fixed	0	-	-
flux	Barr, d_ $\pi^+$	Fixed	0	-	-
flux	Barr, e_ $\pi^+$	Fixed	0	-	-
flux	Barr, f_ $\pi^+$	Fixed	0	-	-
flux	Barr, g_ $\pi^+$	Free	0	gaussian, $\sigma = \pm 0.3$	[ -1.5, 1.5 ]
flux	Barr, h_ $\pi^+$	Free	0	gaussian, $\sigma = \pm 0.15$	[ -0.75, 0.75 ]
flux	Barr, i_ $\pi^+$	Free	0	gaussian, $\sigma = \pm 0.61$	[ -3.05, 3.05 ]
flux	Barr, w_ $K^+$	Free	0	gaussian, $\sigma = \pm 0.4$	[ -2, 2 ]
flux	Barr, x_ $K^+$	Fixed	0	-	-
flux	Barr, y_ $K^+$	Free	0	gaussian, $\sigma = \pm 0.3$	[ -1.5, 1.5 ]
flux	Barr, z_ $K^+$	Fixed	0	-	-
flux	Barr, w_ $K^-$	Fixed	0	-	-
flux	Barr, x_ $K^-$	Fixed	0	-	-
flux	Barr, y_ $K^-$	Fixed	0	-	-
flux	Barr, z_ $K^-$	Fixed	0	-	-
osc model	earth model	Fixed	osc/PREM.12layer.dat	-	-
osc model	YeI	Fixed	0.466	-	-
osc model	YeM	Fixed	0.496	-	-
osc model	YeO	Fixed	0.466	-	-
osc model	detector depth [km]	Fixed	2	-	-
osc model	prop height [km]	Fixed	20	-	-
osc param	$\theta_{12}$ [deg]	Fixed	33.8	-	-
osc param	$\theta_{13}$ [deg]	Fixed	8.61	-	-
osc param	$\theta_{23}$ [deg]	Free	45.4	uniform	[ 0, 90 ]
osc param	$\delta_{CP}$ [deg]	Fixed	0	-	-
osc param	$\Delta m_{21}^2$ [eV <sup>2</sup> ]	Fixed	7.39e-05	-	-
osc param	$\Delta m_{31}^2$ [eV <sup>2</sup> ]	Free	0.00248	uniform	[ 0.001, 0.005 ]
cross section	$M_{A,QE}$	Free	0	gaussian, $\sigma = \pm 1$	[ -2, 2 ]
cross section	$M_{A,res}$	Free	0	gaussian, $\sigma = \pm 1$	[ -2, 2 ]
cross section	DIS	Free	0	gaussian, $\sigma = \pm 1$	[ -3, 3 ]
nu weight	lifetime [common_year]	Fixed	9.28	-	-
nu weight	$N_\nu$	Free	1	uniform	[ 0.1, 2 ]
nu weight	$N_{\nu_\tau}$	Fixed	1	-	-
nu weight	$N_{\nu_\tau,CC}$	Fixed	1	-	-
nu weight	$N_{\nu,NC}$	Fixed	1	-	-
atm. muons	$\Delta\gamma_\mu$	Fixed	0	-	-
atm. muons	$N_\mu$	Free	1	gaussian, $\sigma = \pm 0.4$	[ 0.1, 3 ]
detector sys	DOM Efficiency	Free	1	gaussian, $\sigma = \pm 0.1$	[ 0.8, 1.2 ]
detector sys	Holeice, p <sub>0</sub>	Free	0.102	uniform	[ -0.6, 0.5 ]
detector sys	Holeice, p <sub>1</sub>	Free	-0.0493	uniform	[ -0.2, 0.2 ]
detector sys	Ice absorption	Free	1	gaussian, $\sigma = \pm 0.05$	[ 0.9, 1.1 ]
detector sys	Ice scattering	Free	1.05	gaussian, $\sigma = \pm 0.1$	[ 0.85, 1.25 ]

Table 6.2: Systematic parameters, including those fixed during fit, along with the priors and bounds on the parameters that are free during fit. Parameters that are fixed have their row highlighted in gray.

## 6.5 Verifying the Analysis Pipeline

Aspects of the analysis that are verified in this section are the minimizer's ability to recover true parameters, the projected sensitivity, and the reproducibility of fits with statistical fluctuations. These tests use pseudo-data generated with varied settings for the systematic and physics parameters, depending on their purpose. Some of these pseudo-data sets have no statistical fluctuations (Asimov) and some have Poisson statistical fluctuations applied dependent on the expected count in each bin. Each of the tests and results are described in the following sections.

### 6.5.1 Asimov Inject-Recover Test

The Asimov inject/recover test uses MC samples without any statistical fluctuations. This test is run to check that the minimizer and analysis pipeline can recover various injected values of the physics parameters, along with the nominal systematic parameters. The procedure for the test is as follows:

1. Generate pseudo-data with certain combinations of mixing angle ( $\theta_{23}$ ) and mass splitting ( $\Delta m_{31}^2$ ) values.
2. Run the fit on each set of pseudo-data to attempt to recover the physics and systematic parameters, with the physics parameters seeded at the best-fit point from the 7.5 year IceCube  $\nu_\mu$  disappearance analysis.
3. Compared the recovered (i.e. fitted) physics parameters with the injected values. With no statistical fluctuations, the physics parameters are expected to be recovered almost perfectly. The systematic parameters are also checked to ensure reasonable recovery.

The recovered physics parameters are compared to the injected values at 25 different combinations of mass splitting and mixing angle. Figure 6.12 shows that all injected combinations of the physics parameters were successfully recovered by the minimizer. The systematic parameters are

also recovered with with similar precision. The basis of the minimizer is Migrad from the Minuit2 public library [62, 63, 64].

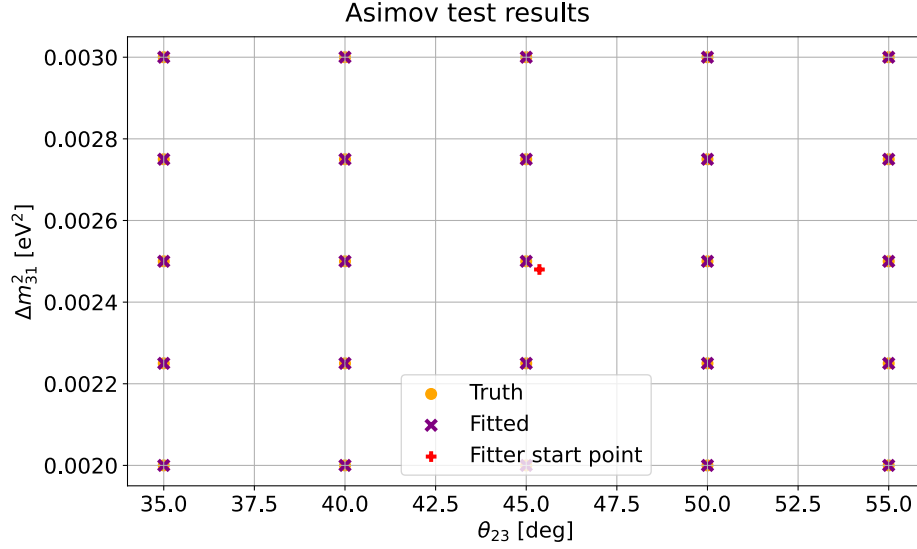


Figure 6.12: Scan over physics parameters showing the inject values for the pseudo-data in the orange dot, the fitted value in the purple cross, and the fitter starter point in red plus (the latter is the same for all scans). All fitted physics parameters recover their injected truth parameters.

### 6.5.2 Projected Sensitivity To Oscillation Parameters

The sensitivity test evaluates the likelihood of multiple combinations of  $\sin^2(\theta_{23})$  and  $\Delta m_{32}^2$ . The physics parameter space is scanned, and the  $LLH$  at each point is compared to the  $LLH$  at the best fit. The contour for the 90% confidence level is then drawn based on the a  $\chi^2$  distribution, assuming Wilks' theorem. The procedure for the sensitivity is as follows:

1. Weight the MC to nominal physics parameter values and used as pseudo-data for the fit (with no statistical fluctuations).
2. Get the  $LLH$  value at the best fit point by performing a fit to the pseudo-data.
3. Fix the physics parameters to a selected off-nominal value and perform a new fit to the pseudo-data.

4. Calculate the difference in  $LLH$  between the two fits (with off-nominal physics parameters vs. free fit to best fit point).
5. Repeat steps 3 and 4 for various combinations of the physics parameters to map the likelihood space
6. Use Wilks' theorem to determine the test statistic value that represents the desired ( $1\sigma/90\%$ ) confidence level.
7. Compare differences in  $LLH$  to draw the contour where the difference is greater than the critical value of the test statistic.

The result in Figure 6.13 shows the FLERCNN  $\nu_\mu$  disappearance analysis sensitivity in the red contour. The best fit point from the IceCube 7.5 year  $\nu_\mu$  disappearance result was assumed to be true (contour in yellow). There is a direct comparison with the 9.28 year IceCube analysis using the likelihood-based reconstruction method (in black), showing that the FLERCNN sample has comparable sensitivity. These contours are contrasted with global experimental results from other neutrino oscillation experiments, indicating that FLERCNN projected sensitivity aligns with the current global experiments [4, 5, 6, 7, 8, 9].

### 6.5.3 Parameter Ensemble Test

The parameter ensemble test looks at the ability of the fitter to recover the injected parameters, particularly for large, off-nominal values of the physics and systematic parameter. The procedure for the parameter test is as follows:

1. Use the Gaussian priors and uniform ranges to select values for the physics and systematic parameters.
2. Generate Asimov pseudo-data using the chosen values for the physics and systematic parameters.

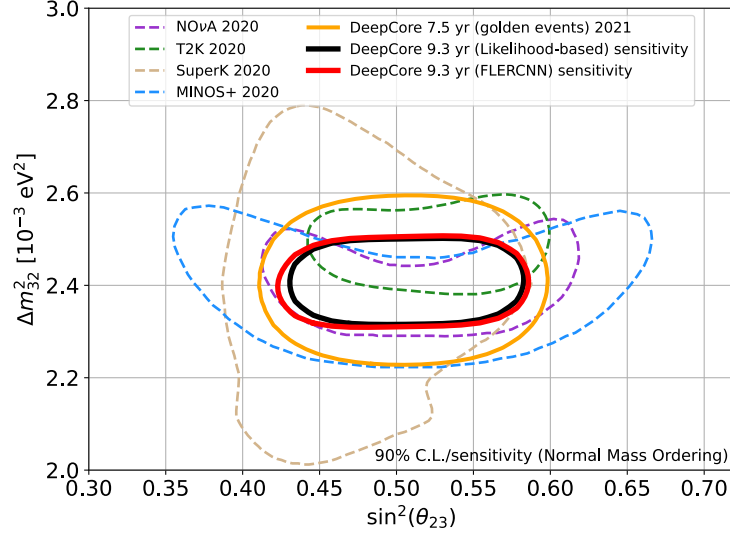


Figure 6.13: Sensitivity of the FLERCNN and *LLH*-based analysis at the verification sample best fit point, compared to global results.

3. Perform the fit, using the same initial seed for each parameter over the multiple trials.
4. Repeat with different values selected for the physics and systematic parameters.

Figure 6.14 shows the fitter’s ability to recover the injected physics and systematic parameters. Even with the same seed used for each parameter, the fit is typically able to recover the true value used to generate the pseudo-data. A perfect result would be on the diagonal 1:1 line, with parameters like  $\Delta m^2_{31}$ , hole ice, scattering, absorption, and DOM efficiency show the best capability of recovery (trials are nearest to the 1:1 line). Some of the parameters with the weakest recovery capability are the Barr parameters, which is expected from the results in Section 6.4.5 showing the systematic impact of the Barr parameters is minimal. There is also an interesting feature in the  $\theta_{23}$  histogram where the octant flip is visible (fitting  $55^\circ$  rather than the true  $35^\circ$ ), but this is unique to fitting the angle and only happens in a small fraction of the trials. For nearly all parameters, there are some fits that have a large fitted error, but these typically are correlated with a large negative *LLH* (green to yellow dots), indicating that their fit failed. This is where blind fit checks (Section 6.6.1) are useful, where the goodness of fit will be evaluated by comparing the fit’s *LLH* to the *LLH* of statistically fluctuated trial fits around the best fit point to ensure the final fit is not one of



these outliers (step 3 in Table 6.3).

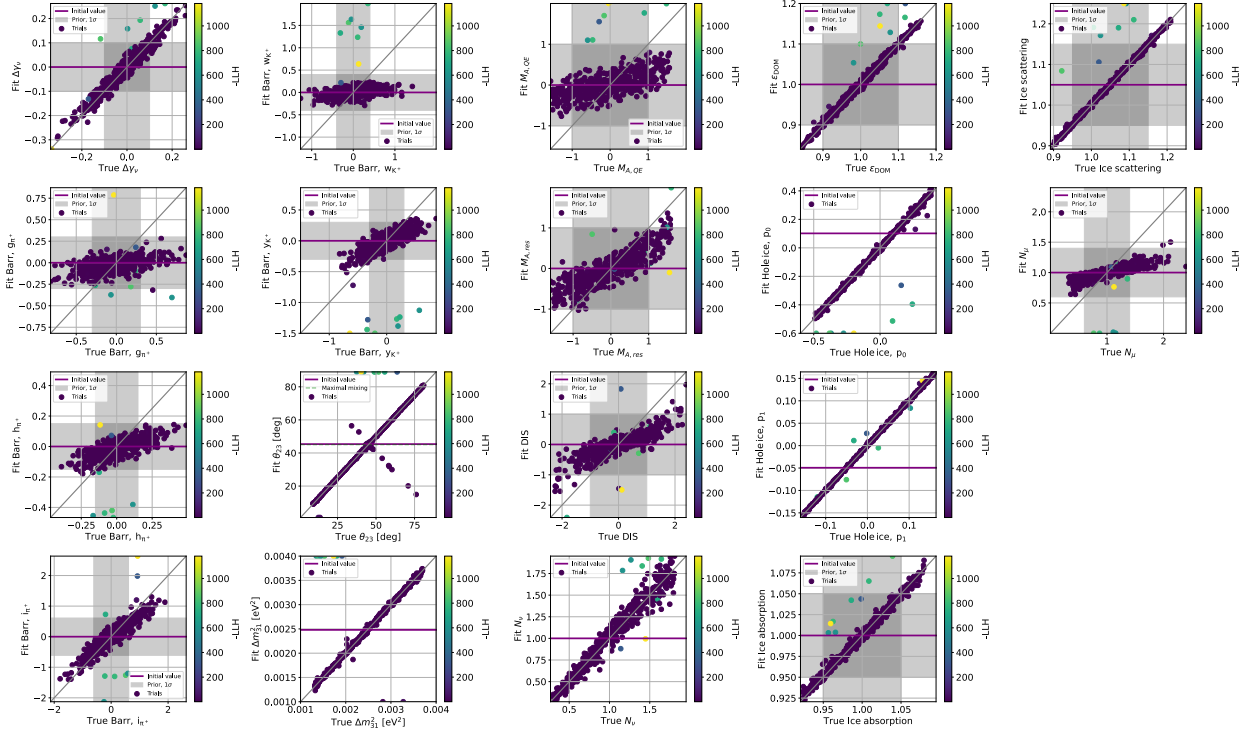


Figure 6.14: Quantifying the ability of the fitter to recover systematic parameters injected at various off-nominal values, with a perfect recovery on the diagonal 1:1 gray line and the results for various fits shown in the dots. The color of the dots corresponds to the fit’s total  $-LLH$  result, where a smaller value (purple) of  $-LLH$  indicates a better fit than a larger value (yellow). This test is run assuming normal ordering of the neutrino masses.

## 6.5.4 Ensemble Test

The ensemble test explores the effect of statistical fluctuations on the analysis, performing the fit on the MC pseudo-data with added Poisson fluctuations over many trials. The procedure for the ensemble test is as follows:

1. Generate pseudo-data by adding Poisson fluctuations to the nominal MC template.
2. Fit physics and systematic parameters for each pseudo-data trial.
3. Explore performance over  $n$  trials for the fitted physics and systematic parameters.

The results of the ensemble test for the 500 trials are in Figures 6.15 and 6.16, for both the normal and inverted orderings respectively. The histogram of the results for each parameter shows the fit values centered at the nominal parameter value. The 90% bounds for all histograms are within the prior  $1\sigma$  width, for the parameters with Gaussian priors. This means that the systematic and physics parameters are all fitting values within the expected range and centered at their expected values.

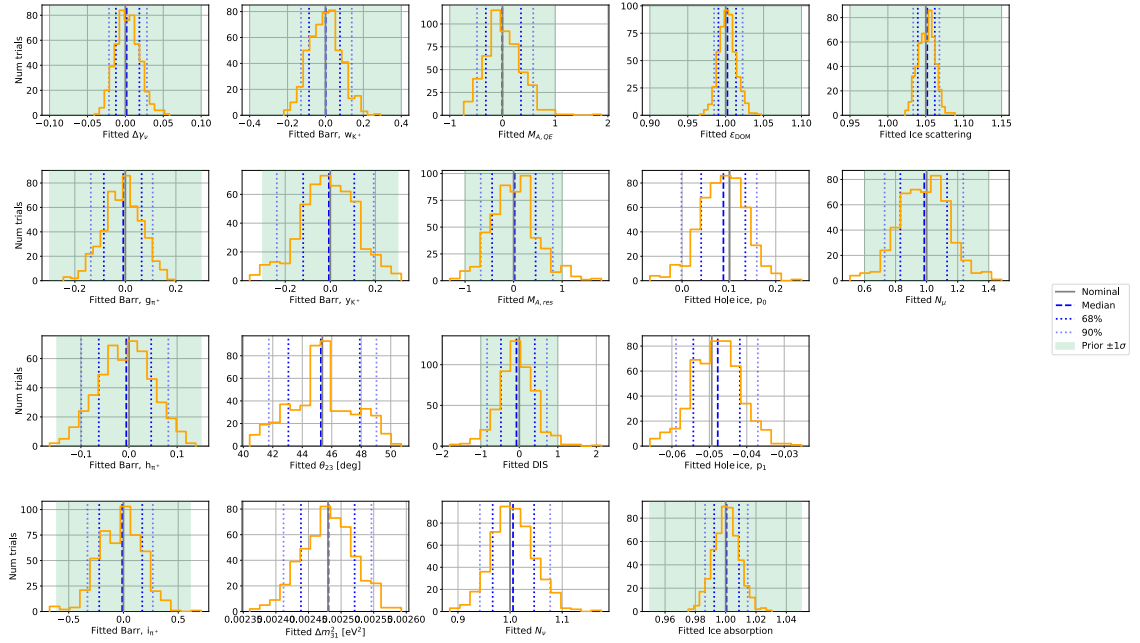


Figure 6.15: Fit results for all systematic and physics parameters after 500 ensemble test trials, assuming normal ordering. The distribution of fitted values for each parameter, including the distribution's median value which should ideally be aligned with the nominal value and the distribution's  $1\sigma$  which should be within the range of the prior.

## 6.6 Blind Fit

To reinforce objectivity, the fit is run on the data and some checks are performed without looking directly at the best fit point result. These checks include ensuring that each systematic parameter is fit at a value that is not outside of its expected range, each analysis bin is not pulled far from its expectation, and there is a reasonable goodness of fit. The blind fit procedure is described in Section 6.6.1 and the outcome in Section 6.6.2.

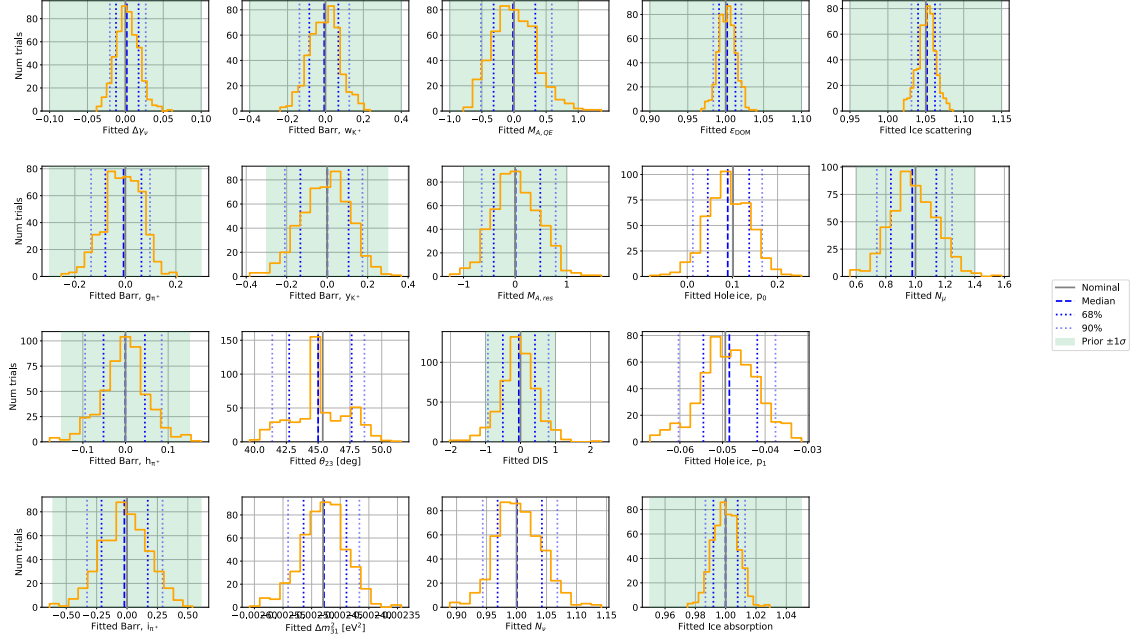


Figure 6.16: Fit results for all systematic and physics parameters after 500 ensemble test trials, assuming inverted ordering. The distribution of fitted values for each parameter, including the distribution’s median value which should ideally be aligned with the nominal value and the distribution’s  $1\sigma$  which should be within the range of the prior.

### 6.6.1 Blind Fit Procedure

The following is the blind fit procedure to evaluate the behavior of the fit without looking directly at the result. The conditions under which the analysis will be paused to assess concerns (called *stop conditions*) along with the items that are reported are in Table 6.3. These terms were set to by the IceCube collaboration’s oscillations working group, in alignment with previous analyses and their precedents.

1. Fit both normal and inverted hierarchy. IceCube is not expected to be sensitive to the mass ordering in this analysis, but this is still done to ensure blindness. The one with the better Test Statistic (TS) is used for the rest of the blind fit tests, but it is not reported or known by the analyzer.
2. Check pulls on all systematic parameters. Allow one (out of 16 systematic parameters) to have a larger pull: stop if one parameter is pulled  $> 3\sigma$  or more than one parameter hits

listed stop conditions (Table 6.3).

3. Check pulls in each analysis bin between the data and MC prediction for the Best Fit Point (BFP) and stop if  $> 5\sigma$ . Run an ensemble test at blind fit BFP to get expected TS distribution using 500 trials of pseudo-data. Check goodness of fit: stop if TS from data is above the 95% percentile (5%  $p$ -value) compared to the pseudo-data trials.
4. Check 1D data/MC distributions for energy, zenith, and PID to make sure they agree with  $p$ -value  $> 5\%$ .

Test #	Test	Stop Condition	Report
2	Parameter pulls: Non-Detector	one parameter $> 3\sigma$ ; two or more parameters $> 2\sigma$ or at boundary for non-Gaussian prior	Distance to boundary and absolute pull in $\sigma$
2	Parameter pulls: DOM efficiency	Fit outside of $[-10\%, +10\%]$	Distance to boundary and absolute pull in $\sigma$
2	Parameter pulls: Hole Ice	$p_0$ fit outside of $[-0.60, +0.40]$ $p_1$ fit outside of $[-0.13, +0.10]$	Distance to boundary and absolute pull in $\sigma$
2	Parameter pulls: Bulk Ice	Scattering or absorption fit outside of $[-10\%, +10\%]$	Distance to boundary and absolute pull in $\sigma$
3	Goodness of fit: Test Statistic	On sided $p$ -value $< 5\%$	Plot of $p$ -values
3	Goodness of fit: bin-wise pulls	Any bin $> 5\sigma$	bin-wise plot of pulls
4	1D Data/MC agreement	One sided $p$ -value $< 5\%$	1D data/MC plots

Table 6.3: Stop conditions for the blind fit plan.

### 6.6.2 Blind Fit Result

The fit was successfully run for normal hierarchy and inverted hierarchy, and a preference chosen blindly based on the better test statistic result. All systematic parameter pulls passed their stop conditions, fit within  $1\sigma$  for the Gaussian priors and not at the boundary for uniform priors (Figure 6.17). All fit counts per analysis bin pulled less than  $4\sigma$  from the expected distribution (Figure

6.18). This passes tests 1, 2, and part of 3. The second part of blind fit test 3 is checking the  $p$ -value of the test statistic. This required running the ensemble test at the best fit point, and comparing the data fit's test statistic with the pseudo-data trials test statistics to check the goodness of fit. This is the blind fit test that the fit failed, with a one sided  $p$ -value at 0.6%, less than the 5% stop condition (Figure 6.19).

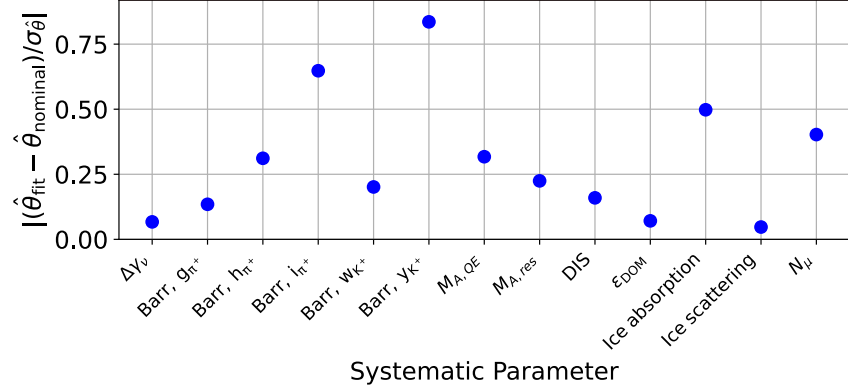


Figure 6.17: Systematic pulls for the nine year FLERCNN sample blind fit shown in terms of  $\sigma$  for parameters with a Gaussian prior. None of the parameters hit their stop condition, with all Gaussian pulls under  $1\sigma$ .

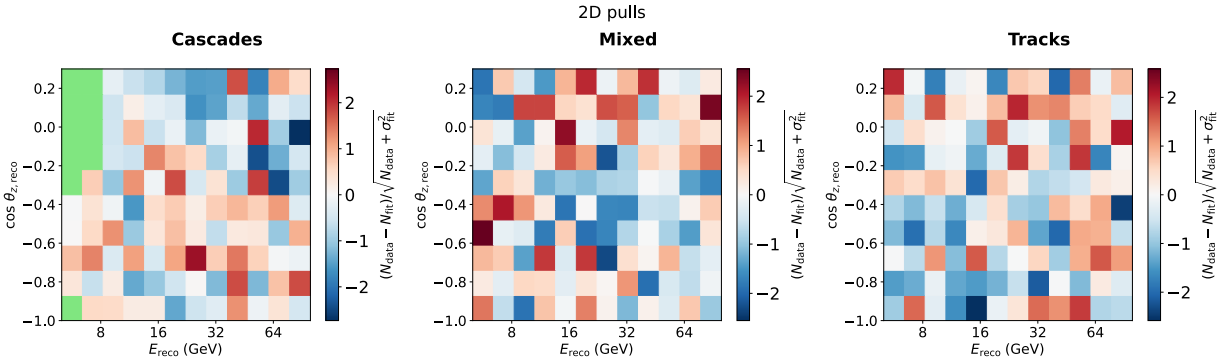


Figure 6.18: Pulls in  $\sigma$  for each of the analysis bins between the data and the MC at the fitted physics and systematic parameters for the nine year FLERCNN sample. None of the bins hit the stop conditions for the pull, with all bins pulling less than  $3\sigma$ .

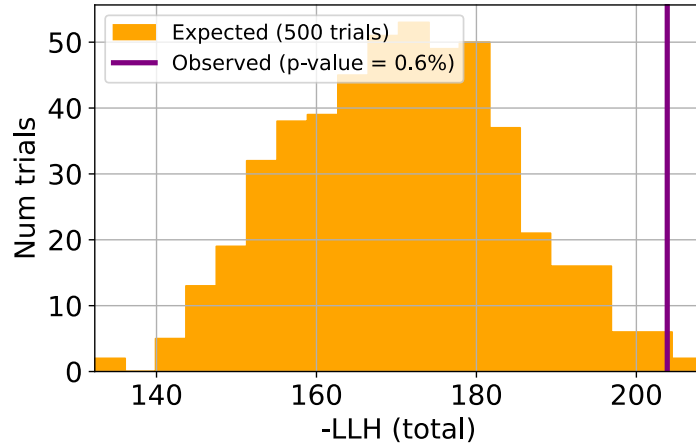


Figure 6.19: Test statistic result for nine years of data fit shown in purple line with a distribution of ensemble test TS results in yellow showing the expected TS spread for 499 trials at the data best fit point. The data TS falls within the ensemble distribution of TS results with a p-value of 0.6% which hits the stop condition requiring it to be  $> 5\sigma$ .

### 6.6.3 Potential Contributions to the Blind Fit Result

The goodness of fit result at 0.6% means that the data fit is not outside the prediction of the pseudo-data trials, but does have a higher  $LLH$  than 99.4% of the trials. This does not necessarily indicate a major problem in the fit, but there could be potential issues leading to the weak goodness of fit. Thus, the analysis has been paused while a few areas of potential contribution to this result are explored.

The post-fit 1D data/MC agreement was checked for some of the reconstructed and cut variables, in particular for variables that are unique to the FLERCNN analysis such as the FLERCNN reconstructed variables and muon BDT classifier. The energy and zenith distributions were not revealed so as to keep the analysis blind to the variables that directly influence the measurement of the oscillation parameters. There were no anomalous disagreements between the explored post-fit data/MC distributions.

Due to the small sample size, no muon statistical errors or systematic changes (beyond the constant scaling muon normalization parameter) were factored into the analysis. The justification for this was that the muons only contributed to 0.6% of the total sample (Table 5.2), but if there

are underestimations in the muon flux, then this fraction could be larger. The muon normalization term in the systematic parameters is intended to account for this, but as Figure 6.11 indicates, the analysis is not very sensitive to this systematic parameter. While this simultaneously indicates that mismodeling impact for  $N_\mu$  should not have a large effect on the resulting  $LLH$ , it could have enough of an effect to cause the slightly lower than expected goodness of fit. Furthermore,  $N_\mu$  only accounts for a constant scaling across all the analysis bins, such that energy or zenith dependent muon count fluctuations, such as those caused by detector systematics, could also potentially be contributing to the pulls.

A first attempt was made to account for the error on the muon counts by adding a bootstrap error estimation on the KDE smoothed muon distribution. It was also noted that the muon KDE smoothing was implemented over linear binning for energy, when the analysis binning uses log binning for energy. Thus, this was also updated so that the KDE smooths over the log binned energy parameter for muons. These changes did result in a slight increase in the goodness of fit  $p$ -value up to 2%. This is likely due to better modeling of the muon distribution along with the bootstrap error producing a larger variance for the pseudo-data trials, more in line with the muon statistical error. This did not increase the test statistic above the stop condition threshold, but indicates that there could be contribution from the lack of muon statistics in the sample and properly accounting for their error, including the variation due to the detector systematics.

Another potential source of disagreement is in the ice model, which is used to propagate photons in the simulation. A new, more detailed ice model has been recently developed on IceCube that accounts for the *of* polycrystals in the ice [12]. The birefringence leads to increased diffusion of photons along the flow axis and a small deflection toward the flow axis. Accounting for this in the ice model has shown improved modeling of the photon arrival time at a basic level, and its implementation to high energy IceCube reconstructions shows improved results at a higher level [65].

Using the new ice model would require regenerating all of the MC simulation, since the ice model is used during the early, time intensive photon propagation step (Section 3.3). This would

take months to re-simulate the  $10^8$  events used in this analysis, including the necessary suite of systematic sets to create a robust fit for the neutrino hypersurfaces (Section 6.4.4.1). Since FLERCNN was trained on the simulation using a different ice model, it would also need to be verified that it can estimate the reconstructed and classification variables well with the new ice model simulation. If FLERCNN needs to be re-trained using the new ice model, then another independently generated  $\sim 10^7$  MC events would need to be simulated for the FLERCNN training samples. Given the robustness of FLERCNN to other ice variations (Section 4.4), it seems likely that it would not require re-training. In that case, the pre-processing and reconstruction should only take a week from this point—which is much improved from the months it takes the likelihood-base reconstruction.

It should be noted that the other IceCube 9.28 year  $\nu_\mu$  disappearance analysis using the likelihood-based reconstruction sees similar issues, with an even smaller goodness of fit. Since these analyses use the same base data and MC up to Level 6 (diverging at the reconstruction and final analysis cuts) and the same analysis software, these are the areas of focus for searching for potential issues. This was reinforced after the post-fit data/MC distributions for the FLERCNN sample did not show any concerning results, indicating that potential issues could be in the underlying assumptions such as the muon background or ice model.

The 9.28 year  $\nu_\mu$  disappearance analysis using the likelihood-based reconstruction saw strongest pulls in the DIS cross section scaling and one of the Barr parameters (Section 6.4.1). As mentioned in Section 6.4.3, the DIS cross section adjustment is energy independent and may benefit from a treatment allowing flexibility that scales with the energy. There are currently studies being pursued on adding this implementation. Similarly, there is work being done to apply more advance handling of the Barr terms for the atmospheric flux. Unlike the likelihood-based reconstruction analysis sample, the FLERCNN analysis did not see a pull  $> 2\sigma$  in any systematic parameters. The DIS and Barr parameters could still be contributing to the weaker goodness of fit, but there is no direct indication of any single systematic parameter pulling strongly in the FLERCNN analysis.

As these studies are time-consuming and potentially open-ended, a small subset of the data will



instead be used as a proof of concept for the FLERCNN analysis without unblinding the full result. Only 10% of the analysis sample will be used, so that statistical uncertainties in the measured physics parameters will be large enough to preserve blindness in the 9.28 year analysis. Using a smaller analysis sample also requires less precise agreement of the data and MC. This will help to validate the analysis pipeline and show a result using the FLERCNN sample. Due to the low statistics, the analysis on 10% of the data will not create competitive bounds on the oscillation parameters, but should still fit in a region that aligns with other experiments.

## CHAPTER 7

### ANALYSIS OF MUON NEUTRINO DISAPPEARANCE WITH 0.92 YEARS OF DATA

To demonstrate that the new FLERCNN reconstruction and analysis sample yields reasonable results, while the 9.28 year FLERCNN analysis remains under collaboration review, the analysis is run with one year of livetime. The analysis with only one year of livetime does not expect to have competitive oscillation constraints due to the limited statistics, but it will determine if the analysis is in agreement with other results and provide a proof of concept for the FLERCNN sample.

#### 7.1 One Year Sample

Only one year of data is used henceforth from the 2013 data taking season, which ran from April 2013 to May 2014. The total livetime for the 2013 season is 0.92 years, which is 10% of the full sample livetime. This sample will be referred to as the "one year sample," as compared to the FLERCNN "nine year sample." The one year sample is used to show a proof of concept that the new FLERCNN reconstruction and analysis procedure produce results that are in line with previous expectations. This is used to undergo the analysis fit and create the  $\nu_\mu$  disappearance oscillation parameter contours.

#### 7.2 Changes in Analysis with One Year Sample

The same MC sample is used to predict the expected analysis distribution (Section 5), with the number of expected counts re-weighted to the new livetime (Table 7.1). There is a minor update to the muon KDE for the one year analysis, which now has the smoothing applied with log binning in energy so that it better matches the analysis binning. The one year sample also updated the systematic set list, as some systematic sets were identified with potential concerns with the uniformity (rows highlighted in gray in Table 6.4). These sets are ones that have a more severe pull in the systematic parameters (one for DOM efficiency and eight for hole ice), so the detector systematic behavior is still well-defined within the expected systematic ranges. This did require

making a new version of the hypersurfaces, using 21 of the originally identified 30 systematic sets, which results in a new re-weighting for the nominal analysis sample. This re-weighting only changed the bin count by  $\pm 1$  event per bin, for a total count difference of  $O(10)$  for the entire sample. All of these changes are minor, indicated by the fact that there is no change in the total rate or fractions expected from each of the flavors (Table 7.1).

Selection	Rate (mHz)	# of Counts (0.92 yr)	% of Sample
$\nu_e$ CC	0.175	$5,070 \pm 7$	23.2%
$\nu_\mu$ CC	0.453	$13,166 \pm 13$	60.3%
$\nu_\tau$ CC	0.037	$1,066 \pm 2$	4.9%
$\nu$ NC	0.082	$2,387 \pm 5$	10.9%
$\mu_{atm}$	0.005	$138 \pm 6$	0.6%
noise	$< 0.0006$	$\sim 0$	$< 0.08\%$
total	0.750	$21,829 \pm 17$	-

Table 7.1: Total counts with 0.92 years of livetime predicted by the updated MC simulation.

Due to the decrease in statistics, some of the analysis bins have near-zero expected counts. As described in Section 5.4.4, these low statistics bins should be masked during analysis. Thus, the mask is extended for the one year sample to cover any bin with less than eight events. This occurs in the low energy cascade and track PID bins for the one year sample (Figure 7.1).

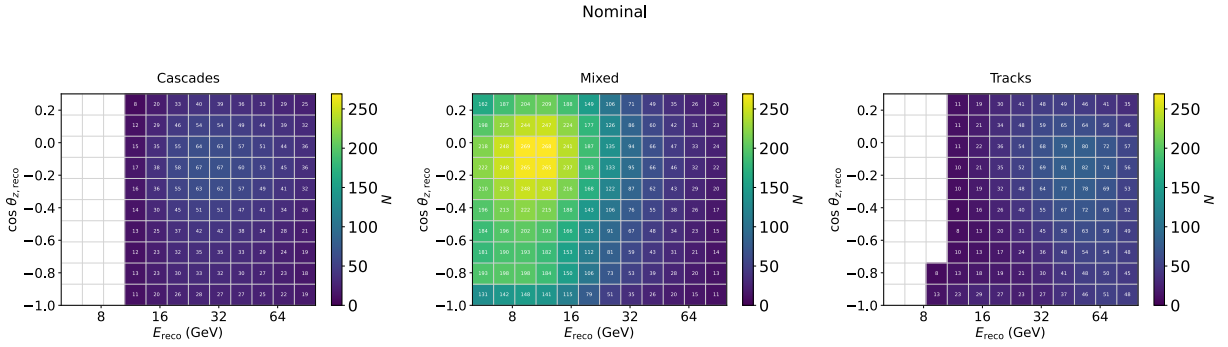


Figure 7.1: Analysis binning after masking, with the masked bins shown in white, for the one year of livetime.

### 7.2.1 Adjusting the Test Statistic

Some of the exploration for the nine year sample's goodness of fit was looking at the muon error, which now includes an estimate of the muon uncertainty using the newly implemented bootstrap method. To make sure this is incorporated directly into the test statistic, the one year analysis will use a  $\chi^2_{\text{mod}}$  that was used as the test statistic by the IceCube 7.5 year golden sample analysis. This  $\chi^2_{\text{mod}}$  contains an additional term in the denominator ( $(\sigma_i^n)^2$ ) that directly accounts for the MC error. Thus, this will be the new test statistic for the one year analysis:

$$\chi^2_{\text{mod}} = \sum_{i \in \text{bins}} \frac{(n_i - d_i)^2}{n_i - (\sigma_i^n)^2} + \sum_{j \in \text{syst}} \frac{(s_j - \hat{s}_j)^2}{\sigma_{s_j}^2} \quad (7.1)$$

The expected counts per bin ( $n_i$ ) are determined by the sum of the MC weights per bin. The statistical uncertainty on the simulation is included as  $(\sigma_i^n)^2$  which is the sum of the squared MC weights. The second term of the  $\chi^2_{\text{mod}}$  is a penalty term on the systematics, where  $j$  accounts for the systematic parameters with the fitted value ( $s_j$ ) compared to the nominal value ( $\hat{s}_j$ ). This is the same penalty term included for the *LLH* test statistic 6.2.

### 7.2.2 Summary of Changes

Table 7.2 summarizes the changes for the one year analysis. Note that the last three changes have been tested on the nine year sample, but the goodness of fit remains <5% still with these changes. All other aspects of the one year sample remains the same as the proposed nine year analysis, since this one year sample is being used as a proof of concept. The changes described are not expected to affect most of the analysis checks previously performed, which were re-run on the one year sample to confirm as much. The only analysis checks where the outcomes change are for the projected sensitivity and the systematic impact test, as expected due to the change in statistics. Thus, these checks on the one year sample are discussed in Sections 7.3.1 and 7.3.2.

Change in Analysis	One Year Sample	Nine Year Sample
Livetime (yrs)	0.92	9.28
Bins masked	58	10
Test statistic	$\chi^2_{\text{mod}}$	LLH
Muon Error	bootstrap estimate	None
Muon KDE energy	log bins	linear bins
Detector systematic sets	21	30

Table 7.2: Changes made for the one year analysis to handle updates and reduced in statistics.

### 7.3 Adjusted Expectations with One Year of Livetime

#### 7.3.1 Sensitivity Projection for One Year

The sensitivity projection is run using the new livetime of 0.92 years. Figure 7.2 shows the change in sensitivity between the one year (red) and full nine year (black) FLERCNN analysis samples. The main difference is due to the changed sample size, as the analysis pipeline has not changed significantly otherwise. The results from other IceCube samples and other experiments are included for comparison. It is expected that only one year of IceCube livetime would not be competitive with previous results; this is necessary to avoid biasing the result from the nine year analysis sample.

#### 7.3.2 Impact of Systematics on One Year Analysis

The other expected difference is the mismodeling impact of the systematic parameters on the one year analysis. It is now dominated by the statistical uncertainty, so each of the systematic parameters generally have less mismodeling potential. Figure 7.3 shows the comparison between the mismodeling for one year (left) vs. the full nine year FLERCNN sample (right). Note that the test statistic has changed between these two tests (Table 7.2), but both TS are assumed to follow  $\chi^2$  distributions so that their  $\sigma$  critical values remain the same.

As expected, the magnitude of the mismodeling for each systematic parameter is less for the one year sample than that seen with the nine year sample. The one year sample has a maximum mismodeling impact of  $< 3\sigma$  for DOM efficiency, whereas the nine year sample's DOM efficiency mismodeling potential surpasses  $5\sigma$ . The only exception to this trend in the difference of magnitude

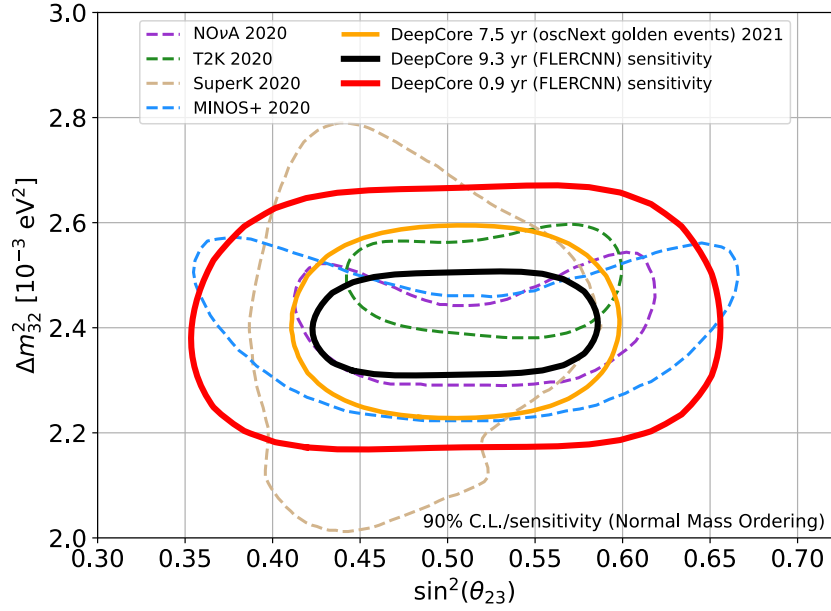


Figure 7.2: Projected sensitivity for the one year analysis (red) vs. the nine year analysis (black), where the expected contour is wider with less statistics.

is  $N_\nu$ , which has a slightly larger overall impact ( $> 0.5\sigma$ ) in the one year sample than this same parameter in the nine year sample ( $> 0.3\sigma$ ). The difference is minor, and could be due to the fact that the nine year sample is sensitive to more systematics so that other systematic parameters could adjust to compensate for a mismodeled neutrino normalization.

The same 16 systematic parameters will be kept free for the one year fit despite the decrease in magnitude of the mismodeling impact. This is motivated by the fact that the one year analysis is a proof of concept for the nine year analysis, so it uses the same systematics list. Thus, Table 6.2 still accurately describes the settings for the systematic parameters used in the one year analysis.

### 7.3.3 Pre-Fit Data/MC Agreement for One Year

With the decrease in statistics, the error bars on the data/MC agreement comparison are larger. The same trends are evident here that were seen with the nine year analysis data/MC agreement plots, discussed in Section 4.2.1. This is expected since Section 6.3.1 shows consistency between most variables over the total ten years. Figures 7.4, 7.5, and 7.6 are included for completeness.

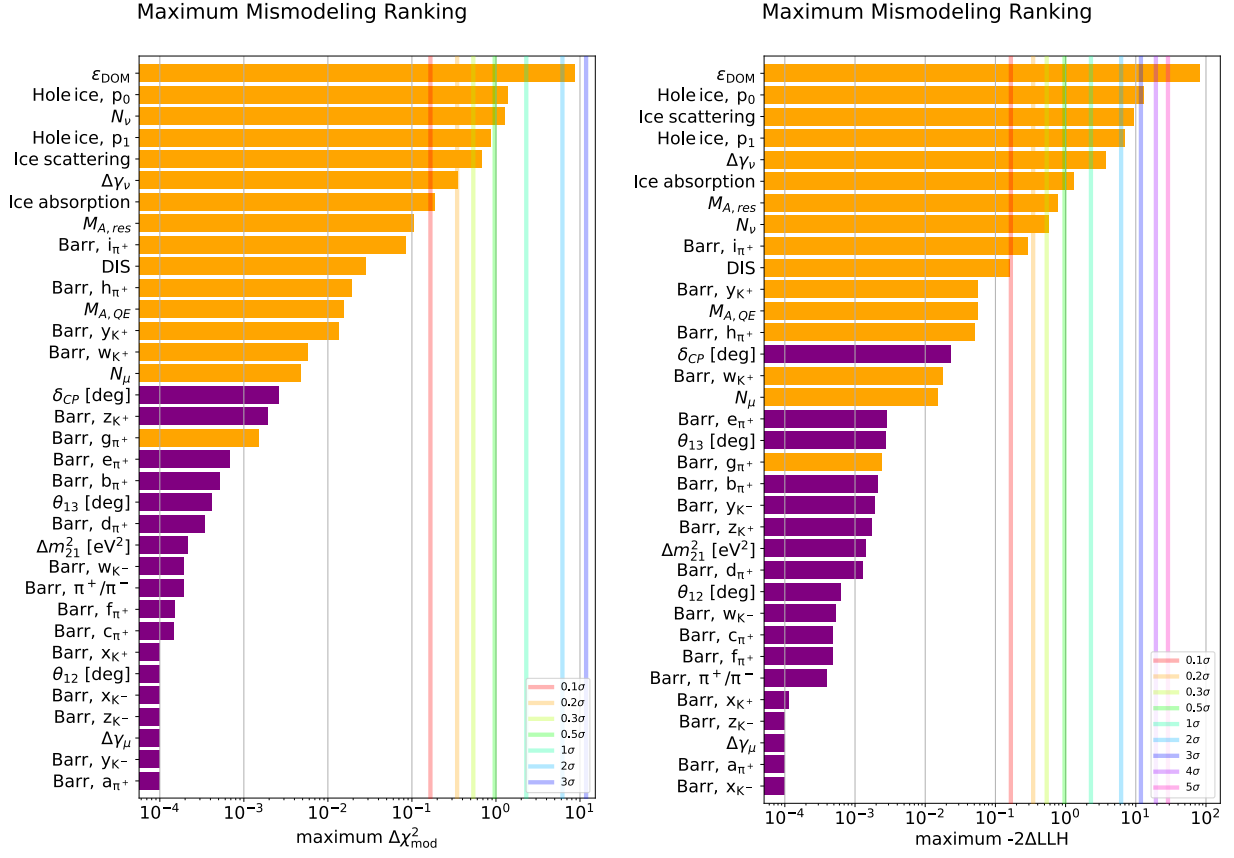


Figure 7.3: New systematic mismodeling plot is on the left for the one year sample, with comparison to the right plot (from Figure 6.11) for the nine year sample.

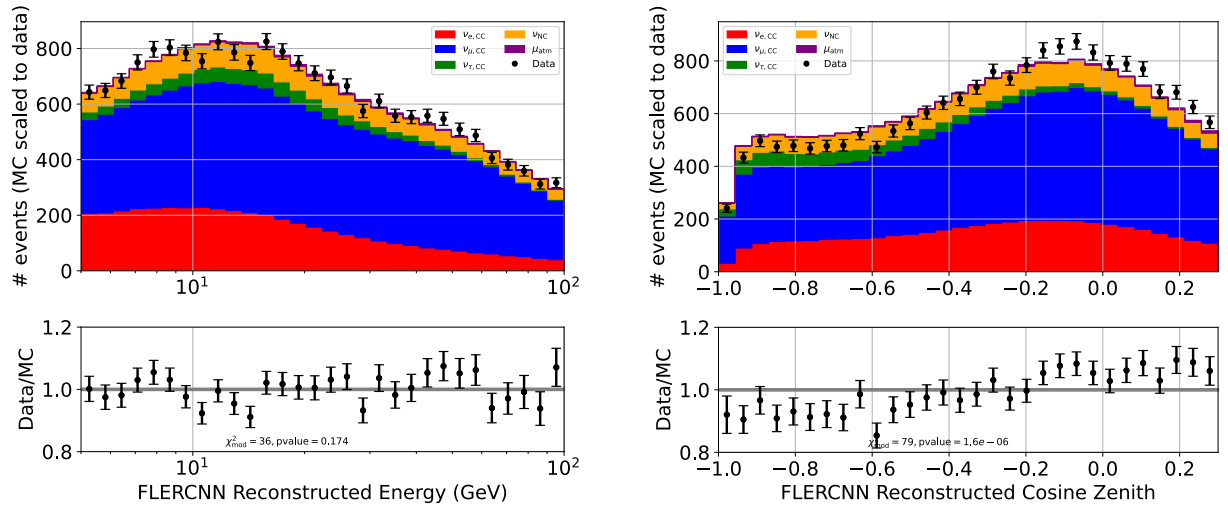


Figure 7.4: Data/MC pre-fit agreement for one year of livetime with the reconstructed energy (left) and reconstructed zenith angle (right).

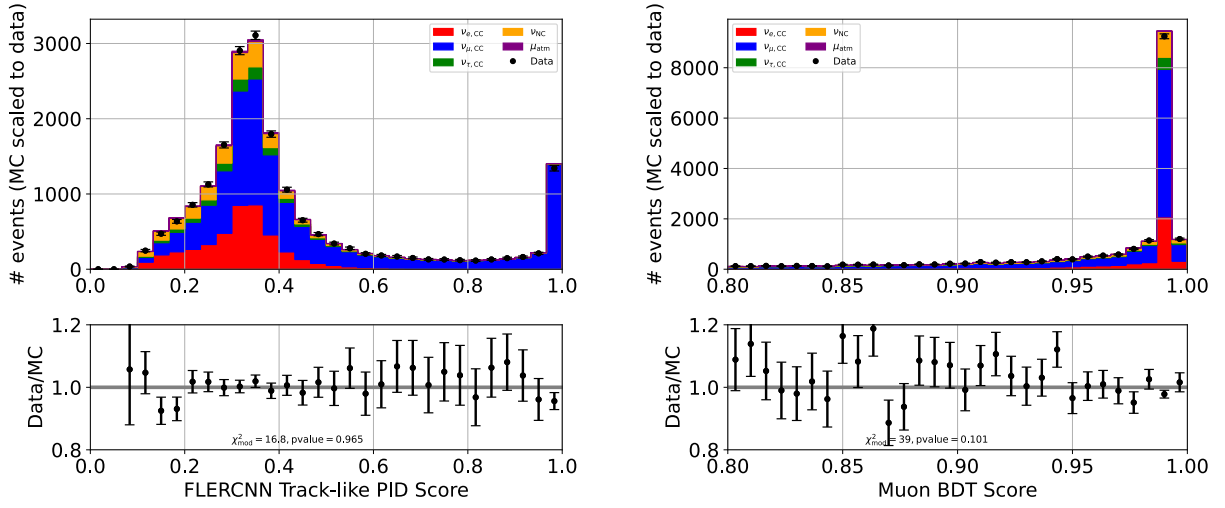


Figure 7.5: Data/MC pre-fit agreement for one year of livetime with the CNN classified PID (left) and BDT Muon classification (right).

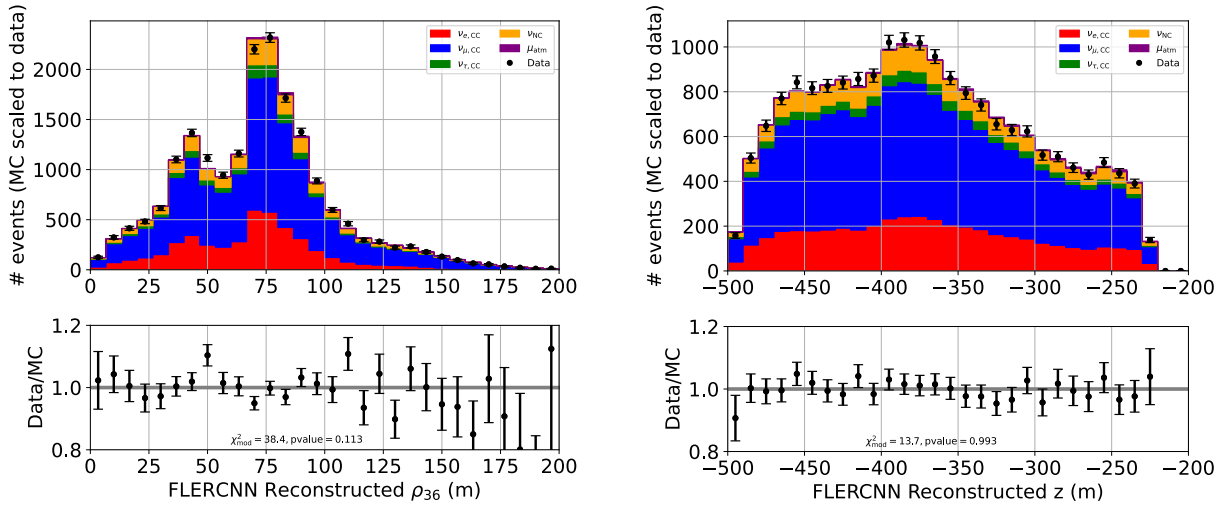


Figure 7.6: Data/MC pre-fit agreement for one year of livetime with the starting vertex reconstructed  $\rho_{36}$  (left) and reconstructed  $z$  positions (right).



## 7.4 Blind Fit

The blind fit procedure and stop conditions remain the same for the one year analysis (Section 6.6.1). The one year analysis passes the blind fit checks:

- ✓ Successful fit of normal and inverted hierarchy
- ✓ Systematic pulls fall within stop condition (Figure 7.7 for Gaussian pulls, Figure 7.13 for a complete picture)
- ✓ Pulls in analysis bins fall within stop conditions (Figure 7.8)
- ✓ Goodness of fit TS  $p$ -value  $> 5\%$  (Figure 7.9)
- ✓ Data/MC distributions for energy, zenith, and PID agree with  $p$ -value  $> 5\%$  (Section 7.4.1)

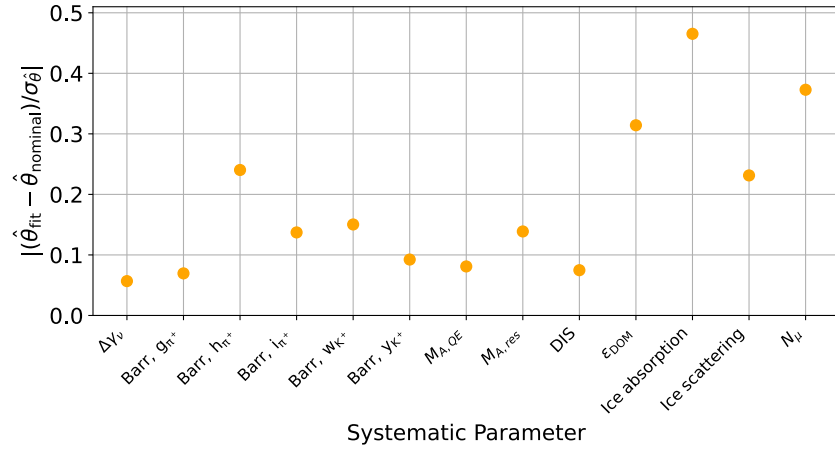


Figure 7.7: Systematic pulls for the one year blind fit shown in terms of  $\sigma$  for parameters with a Gaussian prior.

### 7.4.1 Data/MC Agreement using Best Fit Values

The data/MC agreement is compared after the fit in Figures 7.10, 7.11, 7.12. Note that the comparison now relies on the fitted  $N_{\nu}$  and  $N_{\mu}$  systematic parameters to scale the MC to the data, rather than an artificial scaling applied during plotting with the pre-fit distributions.

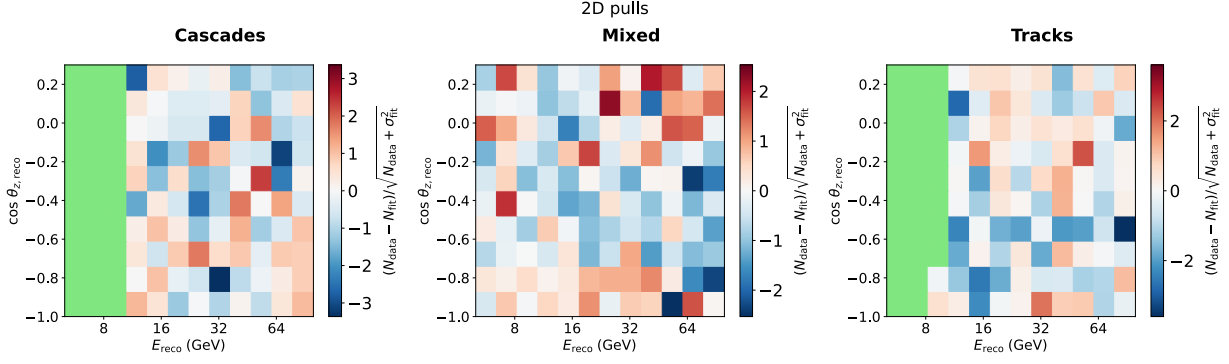


Figure 7.8: Pulls in  $\sigma$  for each of the analysis bins between the data and the MC at the fitted physics and systematic parameters.

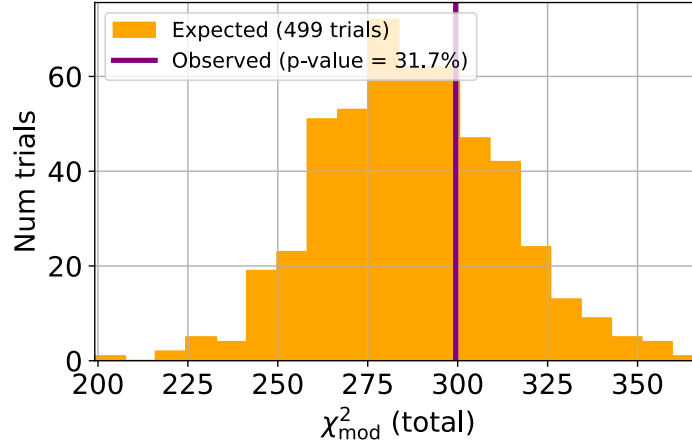


Figure 7.9: Test statistic result for the data fit shown in purple line with a distribution of ensemble test TS results in yellow showing the expected TS spread for 499 trials at the data best fit point. The data TS falls within the ensemble distribution of TS results with a  $p$ -value of 31.7%.

First, we will focus on the three FLERCNN variables that contribute directly to the analysis: energy, zenith, and PID. The energy has good data/MC agreement across the 5-100 GeV range, with no bias or trend evident. The  $p$ -value comparing the data and MC distribution is 0.141, which is greater than the required 5% threshold. The zenith distribution bias also no longer exists after the fit, with a large improvement in the  $p$ -value up to 0.773 post-fit. The PID continues to have good agreement after the fit too, with a  $p$ -value remaining near 1.0.

All post-fit  $p$ -values for the core reconstructed and classified variables using FLERCNN (or FLERCNN as input for the BDT) show good post-fit agreement via this metric. The post-fit

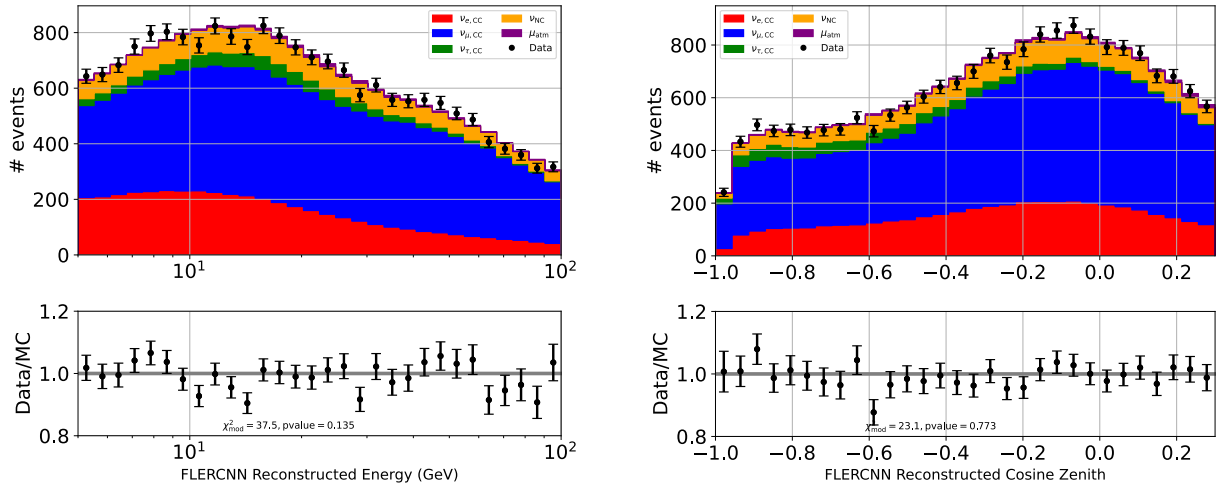


Figure 7.10: Post-fit data/MC agreement for the energy (left) and zenith angle (right).

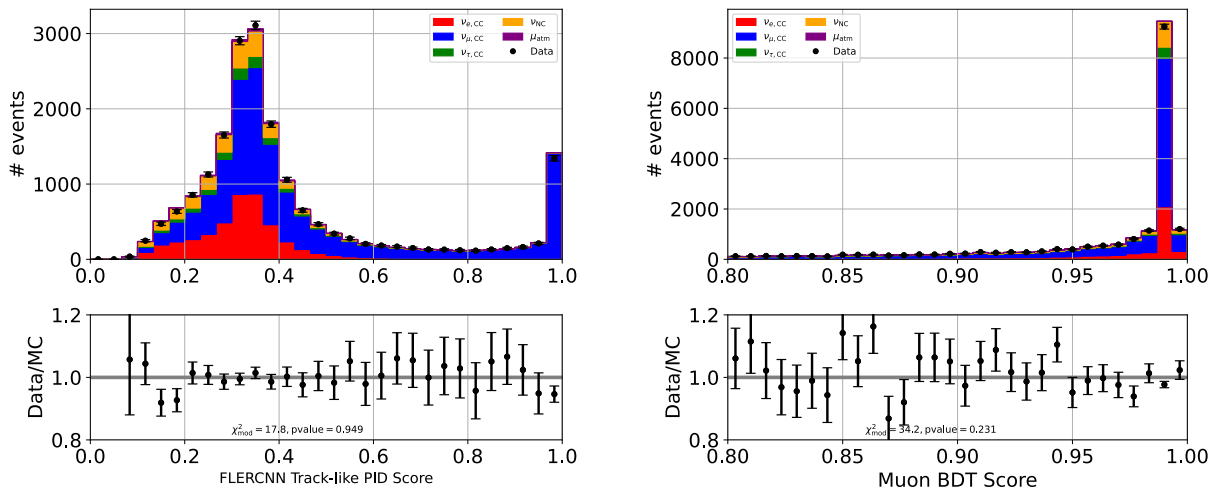


Figure 7.11: Post-fit data/MC agreement for the PID score (left) and background muon BDT score (right).

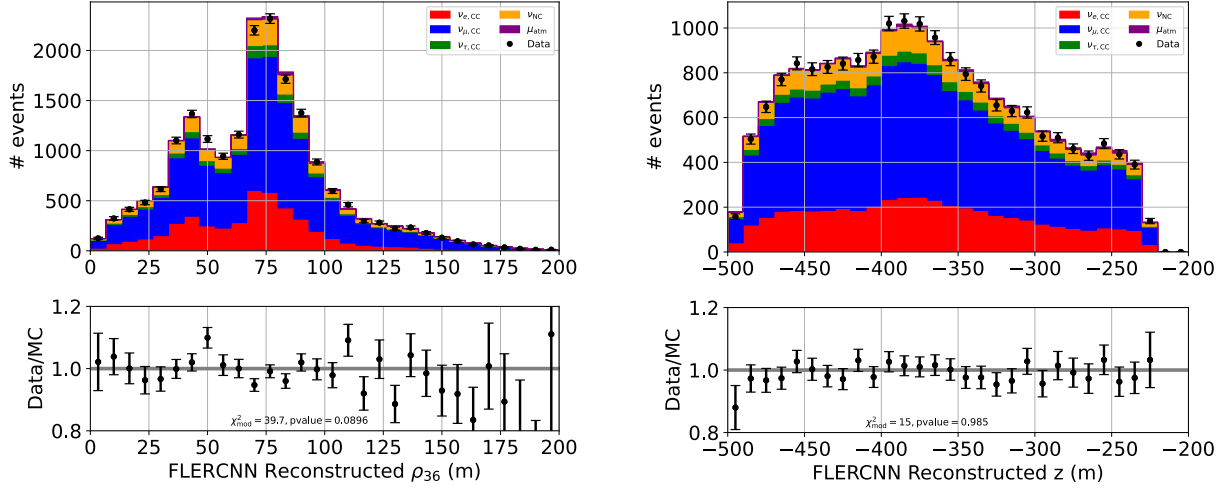


Figure 7.12: Post-fit data/MC agreement for the  $\rho_{36}$  (left) and  $z$  (right).

$p$ -values either large improvements or minimal change for the data/MC comparison (Table 7.3). Notably, the zenith distribution's  $p$ -value improved by five orders of magnitude and the muon BDT score distribution's  $p$ -value is more than doubled. Variables that saw a decreased  $p$ -value only changed by  $< 0.05$ , none of which drop below the 5% threshold.

Reconstructed	Post-Fit $p$ -value	Pre-Fit $p$ -value
energy	0.135	0.174
$\cos \theta_z$	0.773	$1.6 \times 10^{-6}$
PID Score	0.949	0.965
Muon BDT Score	0.231	0.102
$\rho_{36}$	0.090	0.113
$z$	0.985	0.993

Table 7.3: Agreement between the data and post-fit MC using the best fit values for the physics and systematic parameters, quantified by the  $p$ -value pre- vs. post-fit.

## 7.5 Best Fit Result

The result for the best fit is  $\sin^2(\theta_{23}) = 0.634^{+0.049}_{-0.300}$  and  $\Delta m_{32}^2 = 2.41^{+0.22}_{-0.21} \times 10^{-3} \text{ GeV}^2$ . Table 7.4 contains the best fit results for all the physics and systematic parameters after the FLERCNN 0.92 year fit, including the pulls from their priors when applicable. Figure 7.13 has the full picture of this table, including the boundaries for the stop conditions where applicable. The notable result is that the mixing angle is fit off-maximal, closer to  $52^\circ$  rather than  $45^\circ$ , but is still consistent

with the fit at maximal at the  $1\sigma$  level. Since this analysis is performed with low statistics, the uncertainty and resulting contours are quite large—as intended so as not to bias the results of the nine year FLERCNN analysis. None of the physics or systematic parameters hit concerning boundaries though, as confirmed during the blind fit checks, indicating that they are consistent with expectations (Figure 7.13).

Parameter	Fitted Value	Pull ( $\sigma$ )	Nominal Value
$\sin^2(\theta_{23})$	0.626	-	0.506
$\Delta m_{32}^2$	0.00241	-	0.00241
DOM eff	1.0314	0.314	1.0
Hole ice $p_0$	-0.446		0.101569
Hole ice $p_1$	-0.0294	-	-0.049344
Ice scattering	1.07	0.231	1.05
Ice absorption	0.977	-0.465	1.0
$\Delta\gamma_\nu$	0.00567	0.0567	0
$M_{A,QE}$	0.0810	0.0810	0
$M_{A,res}$	-0.139	-0.139	0
DIS	0.0748	0.0748	0
$N_\nu$	0.862	-	1.0
$N_\mu$	1.15	0.372	1.0
Barr, $g_\pi^+$	-0.0209	-0.0695	0
Barr, $h_\pi^+$	-0.0360	-0.240	0
Barr, $i_\pi^+$	-0.0837	-0.137	0
Barr, $w_K^+$	0.0601	0.150	0
Barr, $y_K^+$	0.0277	0.0924	0

Table 7.4: Best fit values for the physics and systematic parameters using one year of livetime with FLERCNN sample. The pull describes the difference in  $\sigma$  of the parameters with Gaussian priors.

### 7.5.1 Best Fit Values and Contours

The contour is determined by scanning values of  $\Delta m_{32}^2$  and  $\sin^2(\theta_{23})$  and performing the fit again. The  $\chi_{\text{mod}}^2$  result is subtracted from the  $\chi_{\text{mod}}^2$  at the best fit point, and the difference is compared to the  $\chi^2$  distribution's critical values to determine 90% and  $1\sigma$  confidence intervals. Figure 7.14 focuses on the result of the one year sample, including its contours and best fit point.

Figure 7.15 shows the FLERCNN 0.92 year sample in red, compared to previous IceCube results (left plot) and the most recent results from other neutrino experiments. The best fit is indicated

in the cross, and the comparison to IceCube results shows that not only does the best fit align closely to the other 90% contours, but also agrees very well with the other  $\Delta m_{32}^2$  results from the IceCube 7.5 year golden neutrino sample. The off-maximal fit for  $\sin^2(\theta_{23})$  is not unique, with the IceCube 2019 result seeing a similar  $\sin^2(\theta_{23})$  fit, but the  $1\sigma$  contours still do not eliminate the possibility of maximal mixing. It is important to note that the contours for the one year fit are wide, as expected, and the best fit point has large uncertainty on it due to the low statistics. This comparison is reassuring that the FLERCNN sample has results that agree with past IceCube and neutrino experiment results, and should have much more precise conclusions once the 9.28 year sample is unblinded.

Since the one year sensitivity project was performed using the IceCube 7.5 year best fit point at maximal mixing, there is a shape mismatch for the sensitivity projection (Figure 7.2) and the actual contour around the off-maximal best fit. To confirm that the result contour is aligned with the analysis expectation, the sensitivity projection re-run using the FLERCNN result's best fit point. Figure 7.16 confirms the sensitivity and result are consistent when using an off-maximal best fit.

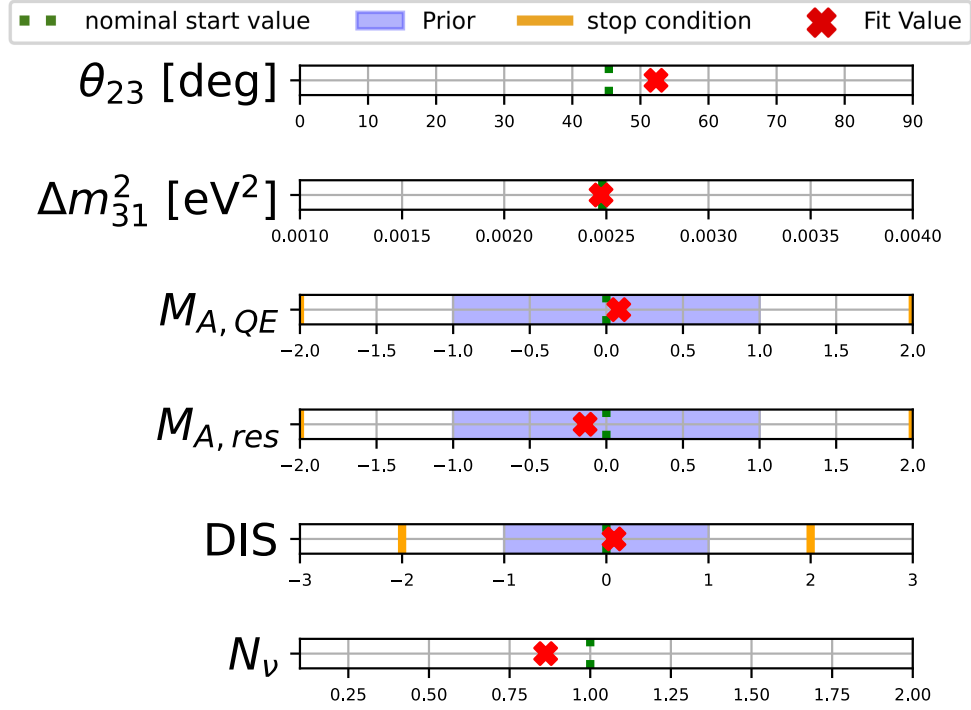
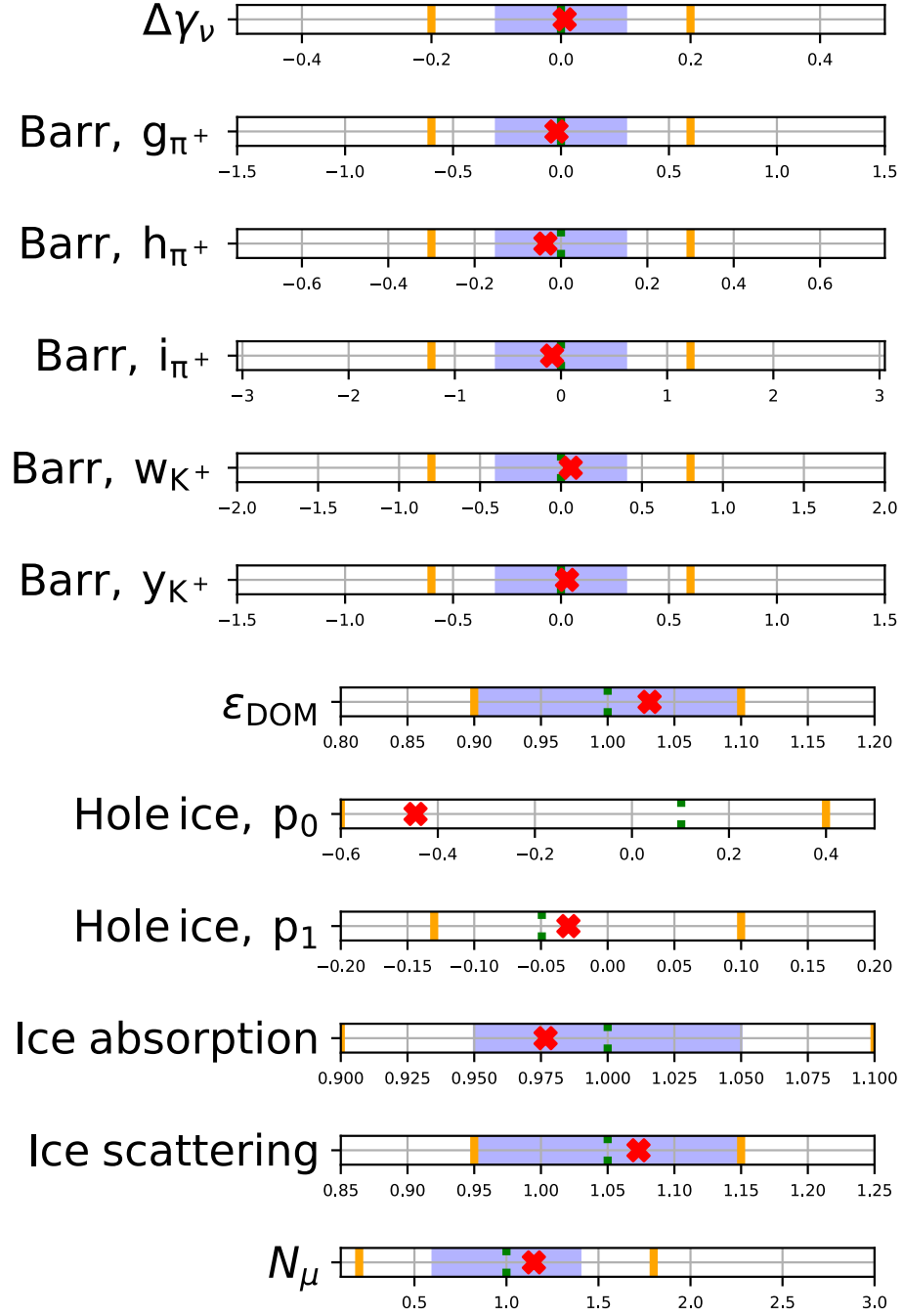


Figure 7.13: Result FLERCNN one year fit shown on the explored range (range of number line plot), Gaussian prior if relevant (blue shaded region), nominal value (green dashed line), stop condition boundary (orange lines), and fitted value (red cross). For historical reasons, the analysis software fits for  $\Delta m^2_{31}$ , solving for  $\Delta m^2_{32} = \Delta m^2_{31} - \Delta m^2_{21}$ .

Figure 7.13: (Cont'd)





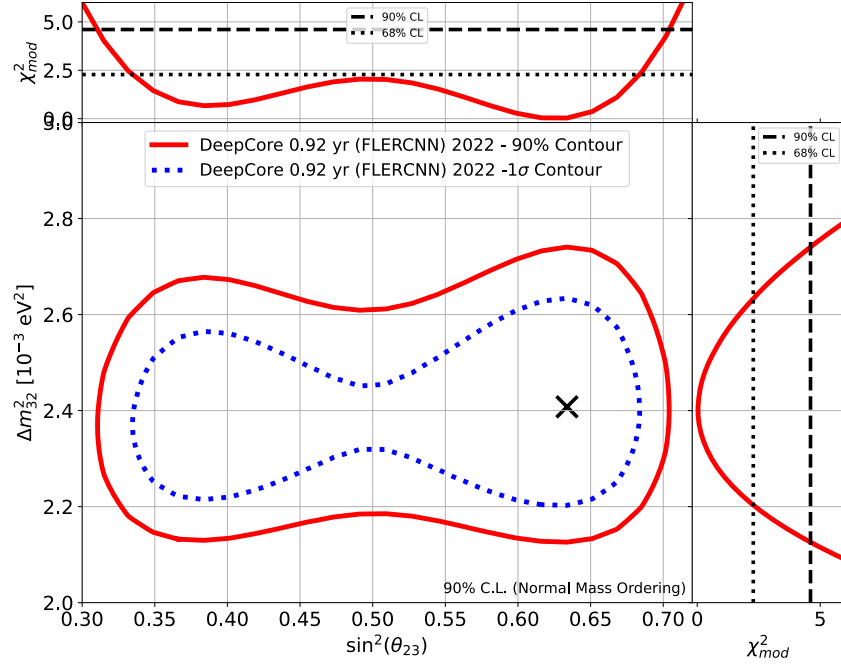


Figure 7.14: Result of the one year analysis, with 90% contour in red,  $1\sigma$  contour in blue, and the best fit point in the black X. The 1D projections for the physics parameters are shown on top of and to the right of the main plot, with the  $\chi^2_{\text{mod}}$  value shown for each physics parameter.

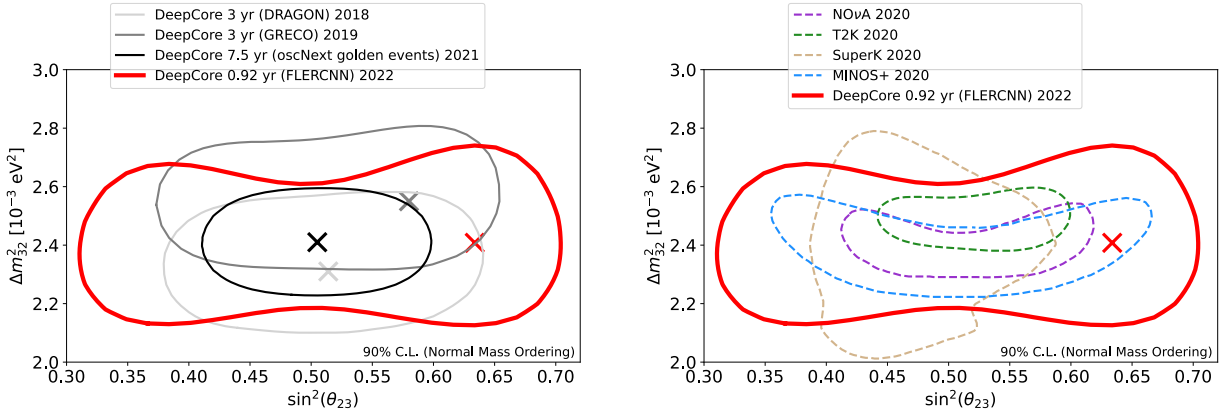


Figure 7.15: Result FLERCNN one year contours compared to other IceCube  $\nu_\mu$  disappearance analyses (left) [16, 17, 8] and global experimental results (right) [4, 5, 6, 7].

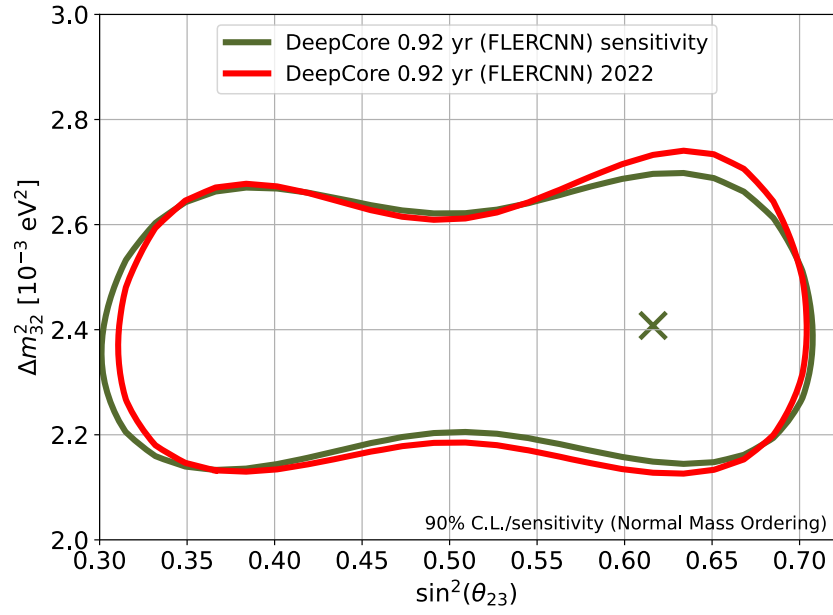


Figure 7.16: Projected sensitivity of the one year analysis using the one year result BFP.

## CHAPTER 8

### CONCLUSIONS

This work has introduced the Fast Low Energy Reconstruction using Convolutional Neural Networks as an alternative, fast reconstruction method for low energy neutrino events in the IceCube detector. The FLERCNN method provides five reconstructed variables (neutrino energy, zenith angle,  $x$ ,  $y$ ,  $z$  positions) and the score for the track-like and muon-like classification for each event. The performance of these networks is comparable with the likelihood-based reconstruction method (Section 4.3), and notably improve the reconstruction speed by a factor of 6000. This permits IceCube’s large, atmospheric neutrino MC and data samples to be reconstructed fully in days rather than months. This speedup opens the potential for faster analysis pipelines and to expanding the MC systematic sets for even more sophisticated handling of systematic parameters, where the simulation generation will be the bottleneck rather than the reconstruction itself.

The 9.28 year FLERCNN analysis is a collaborative effort by the IceCube group, lead by the author and Dr. Shiqi Yu. Extensive checks have been performed to prove the viability of using an analysis sample defined by the FLERCNN reconstruction and cuts. The FLERCNN sample selection has shown reasonable pre-fit data/MC agreement and reconstructed parameters in the data samples have similar distributions between years (Section 6.3). Numerous systematic parameters are accounted for and their potential impact on the analysis is quantified, including parameters affecting the neutrino flux, cross sections, and detector systematics (Section 6.4). The analysis pipeline itself is confirmed using multiple tests to explore the recovery potential of the fit for both the physics and systematic parameters in the pseudodata (Section 3.1.2). While all of these tests pass the minimum requirements agreed upon by the IceCube collaborators, there are some small anomalies that were judged to not have significant impact on the final results.

The blind fit performed on the 9.28 year sample resulted in small pulls per systematic parameter and per bin, when compared with the MC expectations, that were below the predetermined threshold. The goodness of fit yielded a p-value at 2%, failing the  $>5\%$  threshold pre-determined to pause

the analysis. While several potential sources of mismodeling have been put forward to explain the goodness of fit result (Section 6.6.3), comprehensive evaluation of these issues will be time consuming. Thus, 10% of the total sample was taken to use as a confirmatory proof of concept for a  $\nu_\mu$  disappearance analysis using the FLERCNN sample.

The results of the one year analysis show good agreement with both previous IceCube analyses and global experimental measurements (Section 7.5.1). The small statistics on the sample do not lead to constraining contours as intended, but the predicted sensitivity for the analysis aligns with the results. This indicates that the predicted sensitivity for the 9.28 year sample should hold as expected, and thus be comparable to the global constraints from accelerator oscillation experiments. This is particularly exciting since IceCube probes a different source of neutrinos over multiple baselines compared to single-baseline accelerator experiments. This will lead to similarly constraining contours on the oscillation parameters using multiple sources, energies, and baselines. Constraining these parameters leads us one step closer to understand the many open questions about neutrinos and their behavior in the Standard Model and beyond.

## **8.1 Potential for Future Work on FLERCNN**

The FLERCNN framework was developed over three years with multiple contributions from undergraduate students, graduate students, and one postdoctoral researcher. The author developed the framework and documentation, along with training other collaborators on the tools. While FLERCNN has reached a stage enabling it to be used as a robust, fast reconstruction for an IceCube oscillation analysis, there are future developments that could further improve the existing and future capability of the tools. Some potential suggestions are outlined in the following sections.

### **8.1.1 Optimize Current Architecture**

To improve the current FLERCNN regression and classification networks, metaparameter optimization could be explored in future studies. Some initial studies of the neural network's number of layers and nodes has been preformed by undergraduate Brandon Pries, under the guidance of the

author. These did not reveal any clear improvements on a first pass, but this could be revisited now that the training samples have been optimized. In addition, there are other metaparameters that received some limited study (such as the loss functions and learning rate changes), that lead to the current chosen settings but could be conducted more systematically for each individual FLERCNN network.

One theory for the jagged verification loss plots (cyan line Figure 4.4b for example) is that they are due to the use of the batch normalization, which normalizes the layers over each subset of data (or batch) used during training and is typically known to lead to smoother, more stable behavior of the loss gradients [54]. Part of the architecture optimization could be to explore other options, such as instance normalization [66] to see if this normalization handles the sparse nature of the input data better to lead to more stable training. One could also explore potential optimization specific for sparse input arrays. Other neutrino physics experiments have even developed their own sparse CNNs [67], which could be potentially adapted to IceCube.

Another addition to the current FLRECNN setup would be to add input from the outer IceCube strings. While the low energy samples themselves are tuned to only save events with interactions in the DeepCore array, the inclusion of summarized information from the outer rings of IceCube strings has shown some promise in improving resolutions, as seen in a study by undergraduate student Emma Hettinger. This summary input allows the network to use information from the outer strings without overemphasizing them and without excessively increasing the burden of the training file sizes, which are already quite large (Table 4.5). This could lead to improved resolution for some of the FLERCNN reconstructions, with the zenith and vertex reconstructions particularly benefiting from additional information from DOMs outside the currently included region.

### **8.1.2 Expand FLERCNN Reconstruction Output**

The FLERCNN architecture and training samples could be used to train for additional variables, such as the ending position of the muon track from  $\nu_\mu$  CC interactions or predictions of its own uncertainty on its reconstruction. Both of these variables could be used to improve the purity of

sample by cutting these lower quality events. Ensuring that the ending position of the track is within DeepCore is a cut that has been used for past IceCube oscillation analyses [17] and also showed improvements in training the zenith angle CNN (Section 4.5a), so training for this variable would help to apply a cut to only keep well-reconstructed events. Simultaneously training the networks for their own uncertainty has shown promise on other IceCube projects [20], and this could also be used as a purity selection for the sample if implemented. Adding the uncertainty prediction to the energy FLERCNN reconstruction is currently being pursued by graduate student Finn Mayhew.

### 8.1.3 Broaden FLERCNN Applications

FLERCNN opens the potential for future analyses on the current IceCube geometry. FLERCNN could be used in its current state to explore other oscillation studies, such as the  $\nu_\tau$  appearance [16]. Additional training would be necessary for FLERCNN to be adapted to other studies, such as a GeV dark matter search where azimuth angle would be necessary to reconstruct [68], but these are also potential applications for the networks.

FLERCNN could also be developed to be versatile for other detector geometries. The future IceCube Upgrade intends to deploy seven densely instrumented strings near the center of DeepCore to improve the low energy neutrino detection [69]. Another branch in the FLERCNN architecture could be added and developed to reconstruct the next generation of IceCube low energy atmospheric neutrino data. This network would also need to be adapted to the new DOM designs, which employ multi-PMT modules as opposed to the single, downward facing PMT design for the IceCube and DeepCore DOMs. This work is being pursued by graduate student Finn Mayhew.

## 8.2 Potential for Future Work on IceCube Neutrino Oscillations

As discussed previously, the 9.28 year FLERCNN analysis is pursuing follow up studies to explore potential contributions to the goodness of fit result. Both this analysis and the 9.28 year  $\nu_\mu$  disappearance analysis using the likelihood-based reconstruction are working to pursue the potential issues under study in Section 6.6.3. Looking beyond the current  $\nu_\mu$  disappearance study,

FLERCNN provides a fast reconstruction solution that is adaptable to optimizing samples for various neutrino oscillation analyses and beyond. With the improved reconstruction speed, even higher statistics of simulations could be generated. This could further help to parameterize the detector systematics, improving the hypersurfaces used in the oscillation studies (Section 6.4.4.1). As mentioned in Section 8.1.3, other studies that are performed with IceCube’s previous low energy samples include  $\nu_\tau$  appearance [16], neutrino mass ordering [70], non-standard interactions [71], and sterile neutrino searches [72]. The FLERCNN analysis cuts could be tuned to optimize samples for these specific explorations.

## **BIBLIOGRAPHY**



## BIBLIOGRAPHY

- [1] M G Aartsen et al., “PINGU: a vision for neutrino and particle physics at the South Pole”, *Journal of Physics G: Nuclear and Particle Physics* vol. 44 (5) p. 054006 (2017), doi:10.1088/1361-6471/44/5/054006, URL <https://doi.org/10.1088%2F1361-6471%2F44%2F5%2F054006>
- [2] J. A. Formaggio and G. P. Zeller, “From eV to EeV: Neutrino cross sections across energy scales”, *Reviews of Modern Physics* vol. 84 (3) pp. 1307–1341 (2012), doi:10.1103/revmodphys.84.1307, URL <https://doi.org/10.1103%2Frevmodphys.84.1307>
- [3] M. Honda et al., “Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model”, *Physical Review D* vol. 92 (2) (2015), doi:10.1103/physrevd.92.023004, URL <https://doi.org/10.1103%2Fphysrevd.92.023004>
- [4] M. A. Acero et al., “An Improved Measurement of Neutrino Oscillation Parameters by the NOvA Experiment”, (2021), doi:10.48550/ARXIV.2108.08219, URL <https://arxiv.org/abs/2108.08219>
- [5] K. Abe et al., “Improved constraints on neutrino mixing from the T2K experiment with  $3.3 \times 10^{21}$  protons on target”, *Physical Review D* vol. 103 (11) (2021), doi:10.1103/physrevd.103.112008, URL <https://doi.org/10.1103%2Fphysrevd.103.112008>
- [6] Volodymyr Takhistov, “Atmospheric Neutrino Oscillations with Super-Kamiokand”, URL: [https://indico.cern.ch/event/868940/contributions/3817174/attachments/2082629/3498300/ICHEP2020\\_SK\\_Takhistov.pdf](https://indico.cern.ch/event/868940/contributions/3817174/attachments/2082629/3498300/ICHEP2020_SK_Takhistov.pdf) (2020)
- [7] P. Adamson et al., “Precision Constraints for Three-Flavor Neutrino Oscillations from the Full MINOS+ and MINOS Dataset”, *Physical Review Letters* vol. 125 (13) (2020), doi:10.1103/physrevlett.125.131802, URL <https://doi.org/10.1103%2Fphysrevlett.125.131802>
- [8] Summer Blot et al., “Oscnext Verification Sample Analysis (internal)”, Tech. rep. (2022)
- [9] Étienne Bourbeau et al., “oscNext High Stats Sample (internal)”, Tech. rep. (2022)
- [10] R. Abbasi et al. (IceCube), “Low Energy Event Reconstruction in IceCube DeepCore”, (2022), doi:10.48550/ARXIV.2203.02303, URL <https://arxiv.org/abs/2203.02303>
- [11] Martin Leuermann, Testing the Neutrino Mass Ordering with IceCube DeepCore, Ph.D. thesis, RWTH Aachen U. (2018), doi:10.18154/RWTH-2018-231554
- [12] Dmitry Chirkin and Martin Rongen (ICECUBE), “Light diffusion in birefringent polycrystals and the IceCube ice anisotropy”, *PoS* vol. ICRC2019 p. 854 (2019), doi:10.22323/1.358.0854
- [13] Dmitry Chirkin (ICECUBE), “Evidence of optical anisotropy of the South Pole ice”, in “33rd International Cosmic Ray Conference”, p. 0580 (2013)

- [14] Spencer N. Axani, “SPE Templates V2 Technote (internal)”, Tech. rep. (2020)
- [15] G. D. Barr, S. Robbins, T. K. Gaisser, and T. Stanev, “Uncertainties in atmospheric neutrino fluxes”, *Physical Review D* vol. 74 (9) (2006), doi:10.1103/physrevd.74.094009, URL <https://doi.org/10.1103%2Fphysrevd.74.094009>
- [16] M. G. Aartsen et al., “Measurement of atmospheric tau neutrino appearance with IceCube DeepCore”, *Physical Review D* vol. 99 (3) (2019), doi:10.1103/physrevd.99.032007, URL <https://doi.org/10.1103%2Fphysrevd.99.032007>
- [17] M. G. Aartsen et al. (IceCube), “Measurement of Atmospheric Neutrino Oscillations at 6–56 GeV with IceCube DeepCore”, *Phys. Rev. Lett.* vol. 120 p. 071801 (2018), doi:10.1103/PhysRevLett.120.071801, URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.071801>
- [18] H. Watanabe et al., “First study of neutron tagging with a water Cherenkov detector”, *Astroparticle Physics* vol. 31 (4) pp. 320–328 (2009), doi:<https://doi.org/10.1016/j.astropartphys.2009.03.002>, URL <https://www.sciencedirect.com/science/article/pii/S0927650509000401>
- [19] Alexander Radovic et al., “Machine learning at the energy and intensity frontiers of particle physics”, *Nature* vol. 560 (7716) pp. 41–48 (2018), doi:10.1038/s41586-018-0361-2, URL <https://doi.org/10.1038/s41586-018-0361-2>
- [20] R. Abbasi et al., “A convolutional neural network based cascade reconstruction for the IceCube Neutrino Observatory”, *Journal of Instrumentation* vol. 16 (07) p. P07041 (2021), doi:10.1088/1748-0221/16/07/p07041, URL <https://doi.org/10.1088/1748-0221/16/07/p07041>
- [21] Y. Fukuda et al. (Super-Kamiokande Collaboration), “Evidence for Oscillation of Atmospheric Neutrinos”, *Phys. Rev. Lett.* vol. 81 pp. 1562–1567 (1998), doi:10.1103/PhysRevLett.81.1562, URL <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>
- [22] Q. R. Ahmad et al. (SNO Collaboration), “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* vol. 89 p. 011301 (2002), doi:10.1103/PhysRevLett.89.011301, URL <https://link.aps.org/doi/10.1103/PhysRevLett.89.011301>
- [23] P.A. Zyla et al. (Particle Data Group), “Review of Particle Physics”, *PTEP* vol. 2020 (8) p. 083C01 (2020), doi:10.1093/ptep/ptaa104
- [24] D. Griffiths, *Introduction to Elementary Particles*, Wiley-VCH, second, revised edn. (2008)
- [25] K. Hirata et al., “Observation of a neutrino burst from the supernova SN1987A”, vol. 58 (14) pp. 1490–1493 (1987), doi:10.1103/PhysRevLett.58.1490, URL <https://ui.adsabs.harvard.edu/abs/1987PhRvL..58.1490H>
- [26] B. Pontecorvo, “Inverse beta processes and nonconservation of lepton charge”, *Zh. Eksp. Teor. Fiz.* vol. 34 p. 247 (1957)

- [27] Ziro Maki, Masami Nakagawa, and Shoichi Sakata, “Remarks on the Unified Model of Elementary Particles”, *Progress of Theoretical Physics* vol. 28 (5) pp. 870–880 (1962), doi: 10.1143/PTP.28.870, <https://academic.oup.com/ptp/article-pdf/28/5/870/5258750/28-5-870.pdf>, URL <https://doi.org/10.1143/PTP.28.870>
- [28] Mark Thomson, *Neutrinos and Neutrino Oscillations*, Cambridge University Press (2013)
- [29] M.C. Gonzalez-Garcia and Michele Maltoni, “Phenomenology with massive neutrinos”, *Physics Reports* vol. 460 (1-3) pp. 1–129 (2008), doi:10.1016/j.physrep.2007.12.004, URL <https://doi.org/10.1016%2Fj.physrep.2007.12.004>
- [30] Kaoru Hagiwara, Naotoshi Okamura, and Ken ichi Senda, “The earth matter effects in neutrino oscillation experiments from Tokai to Kamioka and Korea”, *Journal of High Energy Physics* vol. 2011 (9) (2011), doi:10.1007/jhep09(2011)082, URL <https://doi.org/10.1007%2Fjhep09%282011%29082>
- [31] Costas Andreopoulos et al., “The GENIE Neutrino Monte Carlo Generator: Physics and User Manual”, (2015), doi:10.48550/ARXIV.1510.05494, URL <https://arxiv.org/abs/1510.05494>
- [32] C. Andreopoulos et al., “The GENIE neutrino Monte Carlo generator”, *Nucl. Instrum. Methods Phys. Res. A* vol. 614 (1) pp. 87–104 (2010), doi:<https://doi.org/10.1016/j.nima.2009.12.009>, URL <https://www.sciencedirect.com/science/article/pii/S0168900209023043>
- [33] Júlia Tena-Vidal et al. (GENIE Collaboration), “Neutrino-nucleon cross-section model tuning in GENIE v3”, *Phys. Rev. D* vol. 104 p. 072009 (2021), doi:10.1103/PhysRevD.104.072009, URL <https://link.aps.org/doi/10.1103/PhysRevD.104.072009>
- [34] Amanda Cooper-Sarkar, Philipp Mertsch, and Subir Sarkar, “The high energy neutrino cross-section in the Standard Model and its uncertainty”, *Journal of High Energy Physics* vol. 2011 (8) (2011), doi:10.1007/jhep08(2011)042, URL <https://doi.org/10.1007%2Fjhep08%282011%29042>
- [35] T. K. Gaisser and M. Honda, “Flux of Atmospheric Neutrinos”, *Annual Review of Nuclear and Particle Science* vol. 52 (1) pp. 153–199 (2002), doi:10.1146/annurev.nucl.52.050102.090645, URL <https://doi.org/10.1146%2Fannurev.nucl.52.050102.090645>
- [36] Sebastian Euler, *Observation of oscillations of atmospheric neutrinos with the IceCube Neutrino Observatory*, Ph.D. thesis, RWTH Aachen Univ. (Germany) (2013), ()
- [37] M. Honda, T. Kajita, K. Kasahara, and S. Midorikawa, “New calculation of the atmospheric neutrino flux in a three-dimensional scheme”, *Phys. Rev. D* vol. 70 p. 043008 (2004), doi: 10.1103/PhysRevD.70.043008, URL <https://link.aps.org/doi/10.1103/PhysRevD.70.043008>
- [38] K. Abe et al., “Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV”, *Physical Review D* vol. 97 (7) (2018), doi:10.1103/physrevd.97.072001, URL <https://doi.org/10.1103%2Fphysrevd.97.072001>
- [39] M. G. Aartsen et al. (IceCube), “The IceCube Neutrino Observatory: Instrumentation and On-line Systems”, *JINST* vol. 12 (03) p. P03012 (2017), doi:10.1088/1748-0221/12/03/P03012, 1612.05093

- [40] P. A. Čerenkov, “Visible Radiation Produced by Electrons Moving in a Medium with Velocities Exceeding that of Light”, *Phys. Rev.* vol. 52 pp. 378–379 (1937), doi:10.1103/PhysRev.52.378, URL <https://link.aps.org/doi/10.1103/PhysRev.52.378>
- [41] M.G. Aartsen et al., “Measurement of South Pole ice transparency with the IceCube LED calibration system”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* vol. 711 pp. 73–89 (2013), doi:10.1016/j.nima.2013.01.054, URL <https://doi.org/10.1016/j.nima.2013.01.054>
- [42] S. Blot M. Rongen, R.C. Bay, “Observation of an optical anisotropy in the deep glacial ice at the geographic South Pole using a laser dust logger”, *The Cryosphere* vol. 14 p. 2537–2543 (2020), doi:"<https://doi.org/10.5194/tc-14-2537-2020>"
- [43] Rongen, Martin, “Measuring the optical properties of IceCube drill holes”, *EPJ Web of Conferences* vol. 116 p. 06011 (2016), doi:10.1051/epjconf/201611606011, URL <https://doi.org/10.1051/epjconf/201611606011>
- [44] Summer Blot et al., “oscNext (internal)”, Tech. rep. (2022)
- [45] Yee Lam Elim Elim Cheung, Neutrino Interactions in IceCube above 1 TeV: Constraints on Atmospheric Charmed-Meson Production and Investigation of the Astrophysical Neutrino Flux with 2 Years of IceCube Data taken 2010–2012, Ph.D. thesis, University of Wisconsin-Madison (2014), (<https://docushare.icecube.wisc.edu/dsweb/Get/Document-72536/thesis.pdf>, page 130)
- [46] Michael J. Larson, Simulation and identification of non-Poissonian noise triggers in the IceCube neutrino detector, Ph.D. thesis, University of Alabama (2014), (<https://ir.ua.edu/handle/123456789/1927>)
- [47] Yee Lam Elim Elim Cheung, Measurement of Atmospheric Neutrino Oscillation Parameters Using Three Years of IceCube-DeepCore Data, Ph.D. thesis, University of Maryland (2018), ([https://drum.lib.umd.edu/bitstream/handle/1903/21383/Cheung\\_umd\\_0117E\\_19177.pdf](https://drum.lib.umd.edu/bitstream/handle/1903/21383/Cheung_umd_0117E_19177.pdf))
- [48] François Chollet et al., “Keras”, <https://keras.io> (2015)
- [49] Martín Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”, (2015), software available from [tensorflow.org](https://www.tensorflow.org/), URL <https://www.tensorflow.org/>
- [50] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, *Commun. ACM* vol. 60 p. 84 (2017), doi:10.1145/3065386
- [51] Christian Szegedy et al., “Going Deeper with Convolutions”, *CoRR* vol. abs/1409.4842 (2014), 1409.4842, URL <http://arxiv.org/abs/1409.4842>
- [52] Yann LeCun et al., “Handwritten Digit Recognition with a Back-Propagation Network”, in D. Touretzky (Ed.), “Advances in Neural Information Processing Systems”, vol. 2, Morgan-Kaufmann (1989), URL <https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf>

- [53] Arohan Ajit, Koustav Acharya, and Abhishek Samanta, “A Review of Convolutional Neural Networks”, in “2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)”, pp. 1–5 (2020), doi:10.1109/ic-ETITE47903.2020.049
- [54] Shibani Santurkar, Dimitris Tsipras, Andrew Ilyas, and Aleksander Madry, “How Does Batch Normalization Help Optimization?”, in S. Bengio et al. (Eds.), “Advances in Neural Information Processing Systems”, vol. 31, Curran Associates, Inc. (2018), URL <https://proceedings.neurips.cc/paper/2018/file/905056c1ac1dad141560467e0a99e1cf-Paper.pdf>
- [55] Baijayanta Roy, “All about Feature Scaling”, (2020), URL <https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35>
- [56] Jiefeng Chen et al., “Robust Out-of-distribution Detection for Neural Networks”, (2020), doi:10.48550/ARXIV.2003.09711, URL <https://arxiv.org/abs/2003.09711>
- [57] IceCube collaboration, “PISA”, [https://user-web.icecube.wisc.edu/~peller/pisa\\_docs/index.html](https://user-web.icecube.wisc.edu/~peller/pisa_docs/index.html) (2021)
- [58] M. G. Aartsen et al. (IceCube), “Computational Techniques for the Analysis of Small Signals in High-Statistics Neutrino Oscillation Experiments”, (2019), 1803.05390
- [59] Anatoli Fedynitch, “mceq”, <https://mceq.readthedocs.io/en/latest/> (2022)
- [60] Adam M. Dziewonski and Don L. Anderson, “Preliminary reference Earth model”, *Physics of the Earth and Planetary Interiors* vol. 25 (4) pp. 297–356 (1981), doi:[https://doi.org/10.1016/0031-9201\(81\)90046-7](https://doi.org/10.1016/0031-9201(81)90046-7), URL <https://www.sciencedirect.com/science/article/pii/0031920181900467>
- [61] Super-Kamiokande collaboration, “Prob3++”, <https://github.com/rogerwendell/Prob3plusplus> (2018-2022)
- [62] Piti Ongmongkolkul and the iminuit team, “iminuit”, <https://iminuit.readthedocs.io/en/stable/index.html> (2020)
- [63] F. James and M. Roos, “Minuit - a system for function minimization and analysis of the parameter errors and correlations”, *Computer Physics Communications* vol. 10 (6) pp. 343–367 (1975), doi:[https://doi.org/10.1016/0010-4655\(75\)90039-9](https://doi.org/10.1016/0010-4655(75)90039-9), URL <https://www.sciencedirect.com/science/article/pii/0010465575900399>
- [64] M. Winkler and F. James, “Minuit2”, [https://root.cern.ch/doc/master/md\\_math\\_minuit2\\_doc\\_Minuit2.html](https://root.cern.ch/doc/master/md_math_minuit2_doc_Minuit2.html) (2022)
- [65] Rasha Abbasi et al., “A novel microstructure based model to explain the IceCube ice anisotropy”, in “Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)”, vol. 395, p. 1119 (2021), doi:10.22323/1.395.1119
- [66] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, “Instance Normalization: The Missing Ingredient for Fast Stylization”, (2016), doi:10.48550/ARXIV.1607.08022, URL <https://arxiv.org/abs/1607.08022>

- [67] P. Abratenko et al., “Semantic segmentation with a sparse convolutional neural network for event reconstruction in MicroBooNE”, *Physical Review D* vol. 103 (5) (2021), doi: 10.1103/physrevd.103.052012, URL <https://doi.org/10.1103%2Fphysrevd.103.052012>
- [68] R. Abbasi et al. (IceCube Collaboration), “Search for GeV-scale dark matter annihilation in the Sun with IceCube DeepCore”, *Phys. Rev. D* vol. 105 p. 062004 (2022), doi:10.1103/PhysRevD.105.062004, URL <https://link.aps.org/doi/10.1103/PhysRevD.105.062004>
- [69] Aya Ishihara (IceCube Collaboration), “The IceCube Upgrade – Design and Science Goals”, (2019), doi:10.48550/ARXIV.1908.09441, URL <https://arxiv.org/abs/1908.09441>
- [70] M. G. Aartsen et al., “Development of an analysis to probe the neutrino mass ordering with atmospheric neutrinos using three years of IceCube DeepCore data”, *The European Physical Journal C* vol. 80 (1) (2020), doi:10.1140/epjc/s10052-019-7555-0, URL <https://doi.org/10.1140%2Fepjc%2Fs10052-019-7555-0>
- [71] R. Abbasi et al., “All-flavor constraints on nonstandard neutrino interactions and generalized matter potential with three years of IceCube DeepCore data”, *Physical Review D* vol. 104 (7) (2021), doi:10.1103/physrevd.104.072006, URL <https://doi.org/10.1103%2Fphysrevd.104.072006>
- [72] M. G. Aartsen et al., “Search for sterile neutrino mixing using three years of IceCube DeepCore data”, *Physical Review D* vol. 95 (11) (2017), doi:10.1103/physrevd.95.112002, URL <https://doi.org/10.1103%2Fphysrevd.95.112002>