

Department of Physics and Astronomy

University of Heidelberg

Master Thesis in Physics

submitted by

Leander Fischer

born in Heidelberg

2019

A New Method for Low Energy Event Classification in IceCube DeepCore

This Master thesis has been carried out by Leander Fischer at the
Deutsches Elektronen-Synchrotron, DESY
under the supervision of
Prof. Dr. Marek Kowalski (DESY, Zeuthen)
Prof. Dr. Ulrich Uwer (Heidelberg University)

Zusammenfassung

Die Beobachtung, dass Neutrinos ihr Lepton-Flavor nach Ausbreitung durch den Weltraum verändern können, ist ein Beweis für ihre von Null verschiedenen Massen und damit ein Indiz für neue Physik außerhalb des Standardmodells. IceCube ist ein quadratkilometer großer Cherenkov Neutrino Detektor, der sich unterhalb des geografischen Südpols im antarktischen Gletschereis befindet. DeepCore, das dichter instrumentierte Subarray von IceCube, ist in der Lage Neutrinos bis herab zu Energien von einigen GeV zu detektieren.

Diese Arbeit steht in enger Verbindung mit der Messung des Verschwindens von atmosphärischen Myon-Neutrinos als einer der möglichen Detektionskanäle für Neutrinooszillationen. Die Identifizierung des Lepton-Flavor der gemessenen Neutrinos ist ein unerlässlicher Bestandteil um die Neutrinooszillations-Parameter Δm_{32}^2 und θ_{23} zu bestimmen. Der Kern dieser Arbeit ist die Entwicklung eines neuartigen Verfahrens zur Unterscheidung von Lichtspuren, die von Myon-Neutrinos in geladenen Strömen der schwachen Wechselwirkung erzeugt werden, und Lichtkaskaden, die sowohl von Elektron- und Tau-Neutrinos in geladenen Strömen als auch in allen neutralen Strömen der schwachen Wechselwirkung erzeugt werden. Das Verfahren verwendet eine mehrdimensionale Methode des Maschinellen Lernens um die Trennung der Event-Klassen gegenüber traditionellen, eindimensionalen Techniken zu verbessern. Die Anwendung dieser Methode in der Bestimmung der Neutrinooszillations-Parameter führt zu einer Verbesserung der Genauigkeit von Δm_{32}^2 um 13,0 % und $\sin^2(\theta_{23})$ um 7,2 %.

Abstract

Neutrino oscillations, the phenomenon that neutrinos can change their flavor after propagation through space, is proof of their non-zero masses and, therefore, a sign of new physics beyond the Standard Model. IceCube is a cubic kilometer Cherenkov neutrino detector buried in the Antarctic glacial ice at the geographic South Pole. DeepCore is a more densely instrumented sub-array located at the center of IceCube. It can detect neutrinos down to energies as low as a few GeV.

This work is closely related to measurements of atmospheric muon neutrino disappearance as one of the possible detection channels of neutrino oscillations. Identifying the flavor of detected neutrinos is essential to determine the neutrino oscillation parameters Δm_{32}^2 and θ_{23} . The core of this thesis is the development of a novel method to distinguish tracks, caused by muon neutrino charged-current interactions, from cascades, caused by both neutral-current interactions of all flavors and charged-current interactions of electron and tau neutrinos. The method utilizes a Gradient Boosting Machine to enhance the separation between these event classes over the traditional, univariate techniques. Applying this method to DeepCore data leads to an improvement in the sensitivities to Δm_{32}^2 and $\sin^2(\theta_{23})$ of 13.0 % and 7.2 %, respectively.

Contents

Abstract	v
Contents	vii
1 Introduction	3
2 Neutrinos in the Standard Model and Beyond	5
2.1 General Properties	5
2.2 Neutrino Interactions	6
2.3 Atmospheric Neutrinos	7
2.4 Neutrino Oscillations	9
3 Detection and Reconstruction of Neutrinos in Ice	13
3.1 The IceCube Neutrino Observatory	13
3.2 Cherenkov Effect	15
3.3 Energy Losses	16
3.4 Event Reconstruction	18
3.4.1 Hit Selection	18
3.4.2 Directional Reconstruction	19
3.4.3 Energy Reconstruction	21
3.5 Event Topologies	22
4 Improving Particle Identification	25
4.1 Reconstructed Variables	26
4.2 Gradient Boosting Classifier	27
4.2.1 Feature Selection	29
4.2.2 Data Preprocessing	30
4.3 Classifier Results	31
4.3.1 Classifier Evaluation	33
4.3.2 Hyper-Parameter Tuning	35

CONTENTS

5 Sensitivity to Atmospheric Neutrinos Oscillations	37
5.1 Analysis Strategy	37
5.1.1 Binned Poisson Likelihood	37
5.1.2 Sensitivity Estimation	39
5.1.3 Systematic Uncertainties	39
5.2 Analysis Binning	40
5.3 Sensitivity Improvement	45
6 Summary and Conclusions	47
Bibliography	I
Appendices	V
A Data/MC Checks	VII
B Expected Event Distributions	XI
List of Figures	XV
List of Tables	XVII

CONTENTS

T

Chapter 1

Introduction

The neutrino was postulated by Wolfgang Pauli [1] in 1930 to explain the continuous energy spectrum of electrons originating from beta decay. Cowan and Reines confirmed this prediction of a light, neutral particle in 1956 when they discovered the electron neutrino using inverse beta decay [2]. Two additional neutrino flavors were found in the following years, and with the discovery of the muon neutrino in 1962 [3] and the tau neutrino in 2001 [4], the current theory of neutrinos in the standard model (SM) was established.

Although neutrinos were first believed to be massless, experimental evidence showing the existence of mixed neutrino states started to appear in the 1960s [5]. Mixing between different physical representations of neutrinos is proof for differences in their masses. The resulting phenomenon of neutrino oscillations can be incorporated into the standard model by extending it to include massive neutrinos. How massive they are and how strong is the mixing between neutrino states has to be obtained from measurement. Today there are a variety of precision oscillation experiments using solar, reactor and atmospheric neutrinos to tighten the constraints on the neutrino oscillation parameters. IceCube is one of those leading experiments probing the oscillation theory with atmospheric neutrinos.

The IceCube Neutrino Observatory [6] was constructed between 2004 and 2010 at the geographic South Pole. It is the first cubic kilometer Cherenkov neutrino detector and consists of 5160 optical sensors attached to 86 strings, drilled down to a maximum depth of ~ 2500 m into the Antarctic ice. Neutrinos are detected by the Cherenkov light that is emitted by secondary particles produced in neutrino-nucleon scattering interactions in the ice. With DeepCore, a more densely instrumented sub-array of IceCube, the neutrino detection energy threshold can be lowered to approximately 5 GeV.

At these energies, the similarity in event signatures poses difficulties in identifying different neutrino flavor interactions. Muon neutrino charged-current interactions produce light tracks as opposed to charged-current interactions of electron and tau neutrinos as well as neutral-current interactions of all neutrinos that produce light cascades. The sparse instrumentation of IceCube makes it more challenging to separate track- and cascade-like events. In this thesis,

a novel method to distinguish those two event types is developed. In contrast to previously used univariate separation techniques, the multivariate machine learning method applied here maximizes the use of information from the detector response. Through the use of a Gradient Tree Boosting algorithm the separation of events in track and cascade is improved. As a result of the improved separation, the uncertainty to the atmospheric neutrino oscillation parameters Δm_{32}^2 and θ_{23} is significantly reduced.

This thesis is structured as follows. Chapter 2 starts with an introduction to neutrino properties and their SM interactions followed by the description of atmospheric neutrinos and the formulation of standard oscillations. Chapter 3 introduces the IceCube Neutrino Observatory, its detection principle along with the observed event signatures and the reconstruction method used for this work. Chapter 4 describes the machine learning technique that was utilized to develop a novel particle identification variable. Chapter 5 applies this method to achieve improved sensitivities in the measurement of atmospheric neutrino oscillations. Finally, Chapter 6 gives a concluding summary and a short outlook.

Chapter 2

Neutrinos in the Standard Model and Beyond

This chapter introduces the basic properties of neutrinos, their place in the Standard Model of particle physics (SM) and their peculiarities following the description of [7]. Section 2.1 and Section 2.2 state the general properties of neutrinos and the neutrino-nucleon interactions. After describing atmospheric neutrinos in Section 2.3 the phenomenon of neutrino oscillations is presented in Section 2.4.

2.1 General Properties

The neutrino is an elementary particle in the SM [7]. It belongs to the class of leptons, which itself is a subclass of elementary fermions (spin $\frac{1}{2}$ particles). The fermions - six quarks and six leptons - form the matter content of the universe. Quarks take part in all three interaction types (forces) of the SM: strong, weak, and electromagnetic (EM) [8]. The charged leptons - electron, muon, and tau - are subject to the weak and the EM interaction. Neutrinos carry neither electric charge nor color charge and therefore only take part in weak interactions. There are three distinct neutrino flavors - electron neutrinos, muon neutrinos and tau neutrinos (ν_e , ν_μ , and ν_τ) [9] - each corresponding to their charged lepton counterparts.

In the SM, weak interactions are mediated by the three massive bosons \mathbf{W}^+ , \mathbf{W}^- , and \mathbf{Z}^0 [7]. The large boson masses ($m_{\mathbf{W}} \sim 80 \text{ GeV}$, $m_{\mathbf{Z}} \sim 90 \text{ GeV}$) result in a short range of the force of about 10^{-18} m . Weak interactions carried by \mathbf{W}^\pm bosons are called charged-current (CC) interactions, because charge is transferred between the interacting particles. In CC interactions, a neutrino is converted into its corresponding charged lepton or vice versa. Neutral current (NC) interactions are those mediated by \mathbf{Z}^0 bosons. Here no charge is transferred. The Feynman diagrams for CC and NC interactions are shown in Figure 2.1.

Although neutrinos are massless in the SM, we know today that they do have a small mass. The observed phenomenon of neutrino oscillations (see Section 2.4) is based on the fact that

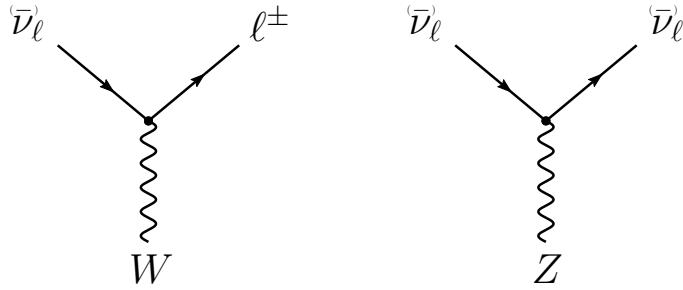


Figure 2.1: Feynman diagrams of charged-current (left) and neutral-current (right) neutrino weak interactions. Taken from [10].

there is a mass difference between the three neutrino mass eigenstates. From neutrino oscillation measurements the absolute mass scale cannot be determined, since they only depend on the mass differences, but there are upper limits on the sum of all neutrino masses from cosmological observations. These upper limits are typically between 0.3 and 1.3 eV [9].

2.2 Neutrino Interactions

To describe the neutrino detection principle of IceCube explained in Chapter 3 we need to understand the weak interaction processes that occur at the energies relevant for this work (10–100 GeV). The cross-sections are dominated by the following neutrino-nucleon interactions: quasi-elastic scattering (QE), resonant scattering (RES), and deep inelastic scattering (DIS). The relative importance of the different processes depends on energy as can be seen in Figure 2.2.

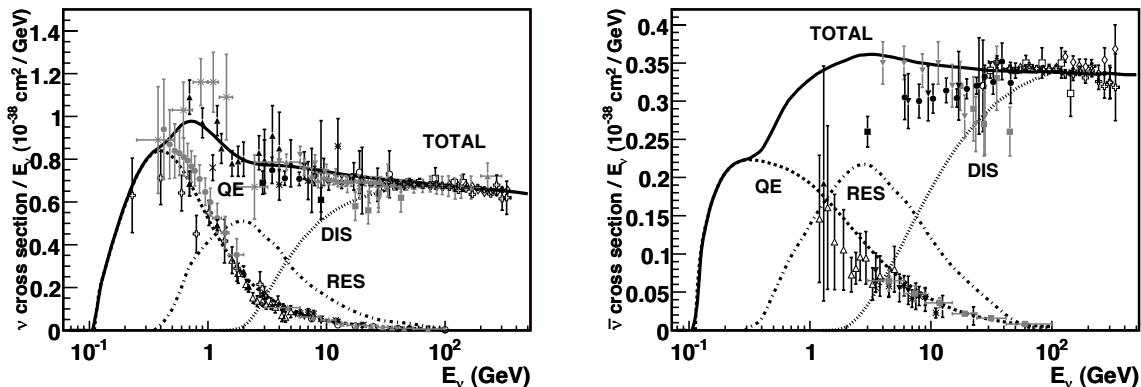


Figure 2.2: Total neutrino(left) and antineutrino(right) per nucleon cross-section divided by neutrino energy plotted against energy. The three main scattering processes quasi-elastic scattering (QE), resonant scattering (RES), and deep-inelastic scattering (DIS) are depicted. Taken from [11].

An extensive description of all the interactions and the differences between neutrino and antineutrino cross-sections can be found in [11]. At energies below 5 GeV, QE and RES occur and the neutrinos interact with approximately point-like protons and neutrons. The cross-sections of these processes are not linear in energy and the transition region to higher energies

is poorly understood. At higher energies, the interactions are dominated solely by DIS which has a linear dependence on energy above ~ 20 GeV. For a given neutrino energy, it is possible to predict the cross-section in this region. Here neutrinos interact with a single quark, breaking apart the nucleus and producing a shower of relativistic secondary particles. Neutrino DIS is the primary detection channel of IceCube. From Figure 2.2 it can be seen that the interaction cross-sections are very small of the order of 10^{-38} cm 2 . Because of the small interaction cross-section, very large volume detectors are required to capture a sufficiently large sample of neutrinos to use for precision studies of their properties. For example, the interaction length of a neutrino with $E_\nu = 10$ GeV is of $\mathcal{O}(10^{10}$ km).

Quasi-elastic scattering (QE) with nucleons is the main process below 1 GeV. Protons are converted to neutrons in antineutrino interactions and vice-versa for neutrino interactions. Additionally, a charged lepton corresponding to the neutrino/antineutrino flavor is produced.

Resonant scattering (RES) describes the process of a neutrino scattering off a nucleon producing an excited state of the nucleon in addition to a charged lepton. RES is the leading process at 1.5-5 GeV for neutrinos and 1.5-8 GeV for antineutrinos.

Deep inelastic scattering (DIS) occurs if a neutrino carries sufficient energy to resolve the underlying structure of the nucleon and interacts with one of the composing quarks. DIS is the dominant process above 10 GeV. The nucleon breaks up and a lepton accompanied by a set of hadronic final states is produced. Whether the lepton is the charged lepton corresponding to the interacting neutrino type, or the neutrino itself depends on the type of DIS interaction. DIS happens via CC as in

$$\begin{aligned} \nu_l + N &\rightarrow l^- + X, \\ \bar{\nu}_l + N &\rightarrow l^+ + X, \end{aligned} \tag{2.1}$$

or NC interactions as

$$\nu_l + N \rightarrow \nu_l + X. \tag{2.2}$$

Here, X stands for any set of final state hadrons and N for the nucleon. The Feynman diagrams for the processes in Equations (2.1) and (2.2) are shown in Figure 2.3.

2.3 Atmospheric Neutrinos

The flux of neutrinos used for this work exclusively comes from the Earth's atmosphere. The nominal flux model is calculated by [12] in the energy range of 100 MeV to 10 TeV. When highly relativistic cosmic rays (protons and heavier nuclei [9]) interact in the upper atmosphere they produce a shower of particles. Neutrinos emerge from the decays of charged pions and kaons (π

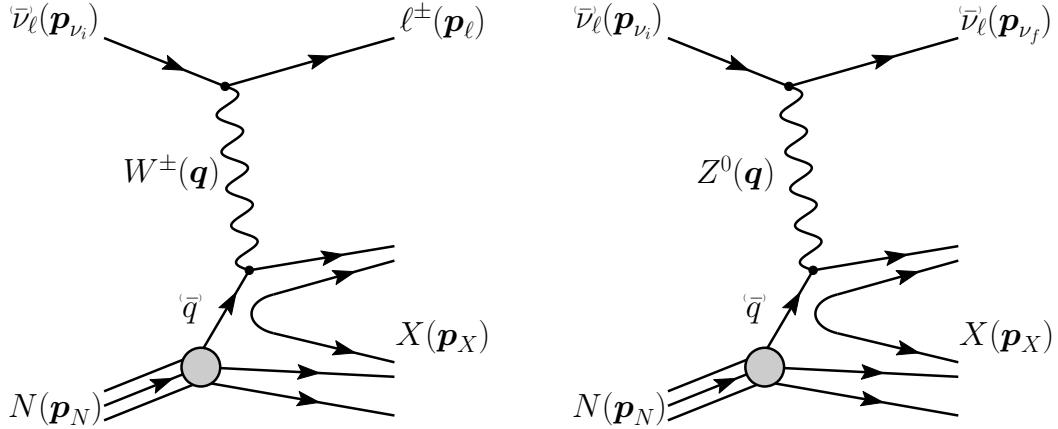


Figure 2.3: Feynman diagrams for deep inelastic scattering of a neutrino with a nucleon via charged-current (left) and neutral current (right) interactions. Taken from [10].

and K mesons) present in these showers. For energies below 100 GeV, the leading contribution comes from the pion decay chain

$$\begin{aligned} \pi^\pm &\rightarrow \mu^\pm + \nu_\mu (\bar{\nu}_\mu), \\ \mu^\pm &\rightarrow e^\pm + \bar{\nu}_\mu (\nu_\mu) + \nu_e (\bar{\nu}_e). \end{aligned} \quad (2.3)$$

The muons that also originate from this process are considered the main background source for IceCube. The left part of Figure 2.4 shows the atmospheric neutrino flux for the very broad energy spectrum in which they are produced. The flux expectations are calculated for the South Pole [12], where the IceCube detector is located. From Equation (2.3) the ratio between muon and electron neutrinos can be inferred to be $N_{\nu_\mu} : N_{\nu_e} \approx 2 : 1$. This is only the case at muon energies below 1 GeV, where all muons decay in flight. For higher energies, muons can reach earth before decaying increasing the ratio to approximately 10:1 at around 100 GeV as shown in the right part of Figure 2.4. Additionally, kaon decays start to contribute which also increases the number of muons and muon neutrinos.

In cosmic ray interactions, charged mesons or tau particles can also be produced, which leads to the formation of tau neutrinos. However, at the energy range considered for this work, the resulting tau neutrino flux is negligible as compared to the muon neutrino flux [13] and is not taken into account. It should be stated here that there is a rather large uncertainty on the normalization of the atmospheric neutrino flux on the order of 20-30% [14] in the energy region of interest. This is mainly due to uncertainties in the primary cosmic ray spectrum and modeling of the hadronic interactions.

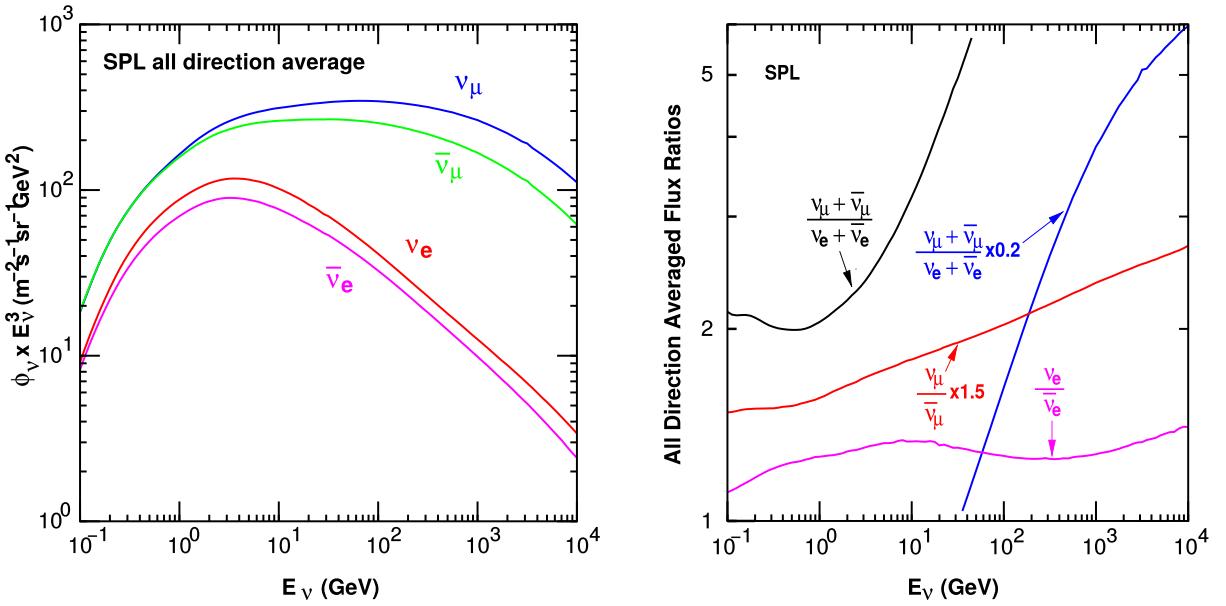


Figure 2.4: Atmospheric neutrino fluxes of the different flavors as a function of energy (left) and ratios between muon- and electron-neutrinos as well as ratios between neutrinos and antineutrinos for both flavors (right). Calculations are done for the geographic South Pole. Taken from [12].

2.4 Neutrino Oscillations

There are two ways to describe neutrino wave functions based on their Hamiltonian eigenvalues [15], as mass eigenstates or as flavor eigenstates. When applying a plane wave approach to explain the propagation of neutrinos in vacuum, their mass eigenstates evolve as

$$|\nu_k(t)\rangle = e^{-iE_k t/\hbar} |\nu_k\rangle, \quad (2.4)$$

where $E_k = \sqrt{\vec{p}^2 c^2 + m_k^2 c^4}$ is the energy of the mass eigenstate $|\nu_k\rangle$, with momentum \vec{p} and mass m_k . Alternatively, they can be described in terms of their flavor eigenstates, which relate the neutrinos to the charged leptons they interact with in weak CC interactions. The flavor eigenstates are ν_e, ν_μ , and ν_τ , whereas the mass eigenstates are called ν_1, ν_2 , and ν_3 in the standard three-neutrino model. To understand the propagation of distinct neutrino flavors in time (in vacuum) we need to relate the flavor eigenstates to the mass eigenstates. For massive neutrinos, each flavor eigenstate is a superposition of mass eigenstates [9]

$$|\nu_\alpha\rangle = \sum_k U_{\alpha k}^* |\nu_k\rangle, \quad (2.5)$$

where $|\nu_\alpha\rangle$ are the weak flavor states with $\alpha = e, \mu, \tau$ and $|\nu_k\rangle$ the mass states with $k = 1, 2, 3$. $U_{\alpha k}$ is the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix defining the mixing between mass

and flavor eigenstates. The mixing matrix can be parameterized as [9]

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{diag}(e^{i\rho_1}, e^{i\rho_2}, 1), \quad (2.6)$$

where $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$ are cosine and sine of the mixing angle θ_{ij} , that defines the strength of the mixing between the mass eigenstates i and j and δ_{CP} is the neutrino CP-violating phase. Nonzero, non-equal neutrino masses and the neutrino mixing relation in Equation (2.5) lead to the observed phenomenon of neutrino oscillations. Oscillation means that a neutrino changes from its initial flavor to another flavor and back after traveling a certain distance. A produced flavor eigenstate $|\nu_\alpha\rangle$ propagates through space as a superposition of mass eigenstates. To find the probability that the initial flavor state $|\nu_\alpha\rangle$ ends up as the final flavor state $|\nu_\beta\rangle$ after the time t we calculate

$$P_{\nu_\alpha \rightarrow \nu_\beta}(t) = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2, \quad (2.7)$$

where P is the probability calculated by applying Fermi's Golden Rule [16]. Fermi's Golden Rule explains the transition rate from one energy eigenstate to another depending on the strength of the coupling between the two. The strength of the coupling is described by the square of the matrix element. Using the unitarity of the mixing matrix $U^{-1} = U^\dagger$ to reverse the relation (2.5) and then time evolve the mass eigenstates with Equation (2.4) we get the time evolution of the flavor state $|\nu_\alpha(t)\rangle$. Inserting this result into Equation 2.7 yields

$$P_{\nu_\alpha \rightarrow \nu_\beta}(t) = \sum_{j,k} U_{\beta j}^* U_{\alpha j} U_{\beta k} U_{\alpha k}^* e^{-i(E_k - E_j)t/\hbar}, \quad (2.8)$$

where the indices j and k run over the mass eigenstates. For small neutrino masses compared to their kinetic energy, we can approximate the energy as

$$E_k \approx E + \frac{c^4 m_k^2}{2E} \rightarrow E_k - E_j \approx \frac{c^4 \Delta m_{kj}^2}{2E}, \quad (2.9)$$

where $\Delta m_{kj}^2 = m_k^2 - m_j^2$ is the mass-squared splitting between states k and j . If we now replace the time in Equation (2.8) by the distance traveled by the relativistic neutrinos $t \approx L/c$ we get

$$\begin{aligned} P_{\nu_\alpha \rightarrow \nu_\beta}(t) &= \delta_{\alpha\beta} - 4 \sum_{j>k} \mathbf{Re}(U_{\beta j}^* U_{\alpha j} U_{\beta k} U_{\alpha k}^*) \sin^2\left(\frac{c^3 \Delta m_{kj}^2}{4E\hbar} L\right) \\ &\quad + 2 \sum_{j>k} \mathbf{Im}(U_{\beta j}^* U_{\alpha j} U_{\beta k} U_{\alpha k}^*) \sin^2\left(\frac{c^3 \Delta m_{kj}^2}{4E\hbar} L\right), \end{aligned} \quad (2.10)$$

which is referred to as the survival probability if $\alpha = \beta$ and the transition probability if $\alpha \neq \beta$. The probability in Equation 2.10 is only nonzero if there are neutrino mass eigenstates with masses greater than zero. Additionally, there must be a mass-squared difference Δm^2 and

nonzero mixing between the states. Since we assumed propagation in vacuum in Equation 2.4, the transition and survival probabilities correspond to vacuum mixing.

Two-neutrino Approximation Looking at Equation (2.10) it can be noted, that the oscillation frequency for a neutrino with a given energy depends on the mass-squared splitting. From experiments, we know that the mass splittings Δm_{21}^2 and Δm_{23}^2 differ by two orders of magnitude [9]. As a result, neutrino oscillation experiments are usually designed to measure at the scale of one of the two mass splittings. When observing one of the two oscillation frequencies, the other will either be too fast or too slow to be observed. The oscillations seen in an experiment can, therefore, be described in a simplified two-neutrino scheme, where we only assume two mass eigenstates ν_1, ν_2 that are connected to two flavor eigenstates ν_α, ν_β by the mixing matrix

$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad (2.11)$$

which is a two-dimensional rotation matrix that depends only on one mixing angle θ between the two states. Consequently, the resulting survival/transition probability from Equation (2.10) only depends on one mass-squared splitting Δm^2 and one mixing angle.

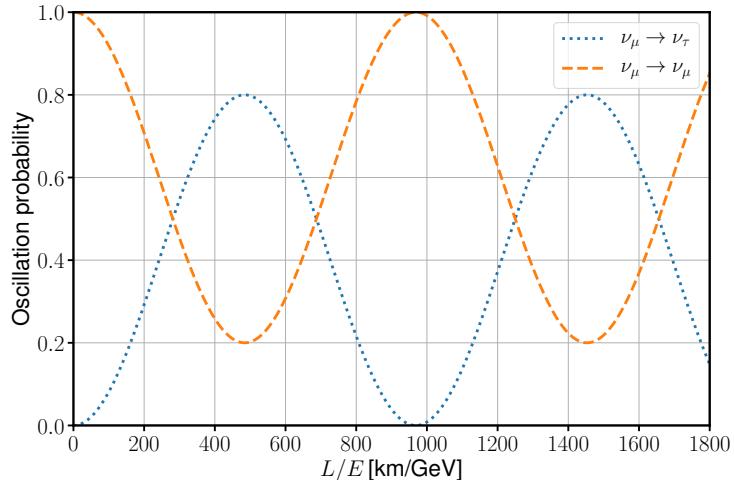


Figure 2.5: Transition and survival probability for muon neutrinos in the two-neutrino approximation in vacuum (equations (2.13) and (2.14)) plotted as a function of L/E for $\Delta m_{32}^2 = 2.54 \cdot 10^{-3} \text{ eV}^2$ and $\sin^2(\theta_{23}) = 0.8$.

DeepCore is sensitive to the fast atmospheric oscillations caused by the bigger mass splitting Δm_{23}^2 , because it is optimized for neutrinos at energies of $\mathcal{O}(25 \text{ GeV})$, which have an oscillation length related to Δm_{23}^2 on the order of the Earth's diameter. The current global best fit value is $\Delta m_{23}^2 = 2.54 \cdot 10^{-3} \text{ eV}^2$ [9]. Here the dominant oscillation happens between the flavor states ν_μ and ν_τ . Formulating the neutrino oscillation phase of Equation (2.10) in appropriate units

used by the experiment, we get

$$\frac{c^3 \Delta m^2 L}{4E\hbar} \approx 1.267 \frac{\Delta m^2 [\text{eV}^2] \cdot L [\text{km}]}{E [\text{GeV}]}.$$
 (2.12)

Inserting this for the two-flavor approximation into the transition probability, we end up with

$$P_{\nu_\mu \rightarrow \nu_\tau}(L) = \sin^2(2\theta) \sin^2\left(1.267 \cdot \frac{\Delta m^2 [\text{eV}^2] \cdot L [\text{km}]}{E [\text{GeV}]}\right),$$
 (2.13)

whereas the survival probability becomes

$$P_{\nu_\mu \rightarrow \nu_\mu}(L) = 1 - P_{\nu_\mu \rightarrow \nu_\tau}(L) = 1 - \sin^2(2\theta) \sin^2\left(1.267 \cdot \frac{\Delta m^2 [\text{eV}^2] \cdot L [\text{km}]}{E [\text{GeV}]}\right).$$
 (2.14)

The transition and survival probability in the two-neutrino approximation are plotted against the traveled neutrino distance divided by the energy in Figure 2.5. The mixing angle in $\sin^2(\theta_{23})$ dictates the amplitude of the oscillation while the oscillation length (distance between extrema) is governed by the mass splitting Δm_{32}^2 .

Chapter 3

Detection and Reconstruction of Neutrinos in Ice

This chapter explains the detection principle of the IceCube Neutrino Observatory as well as one of the reconstruction algorithms applied to extract information from the raw detector signals. Section 3.1 first presents the layout of IceCube and DeepCore before Section 3.2 and 3.3 explain the Cherenkov effect and the propagation of particles through ice. Finally, Section 3.4 outlines the reconstruction algorithm itself in detail.

3.1 The IceCube Neutrino Observatory

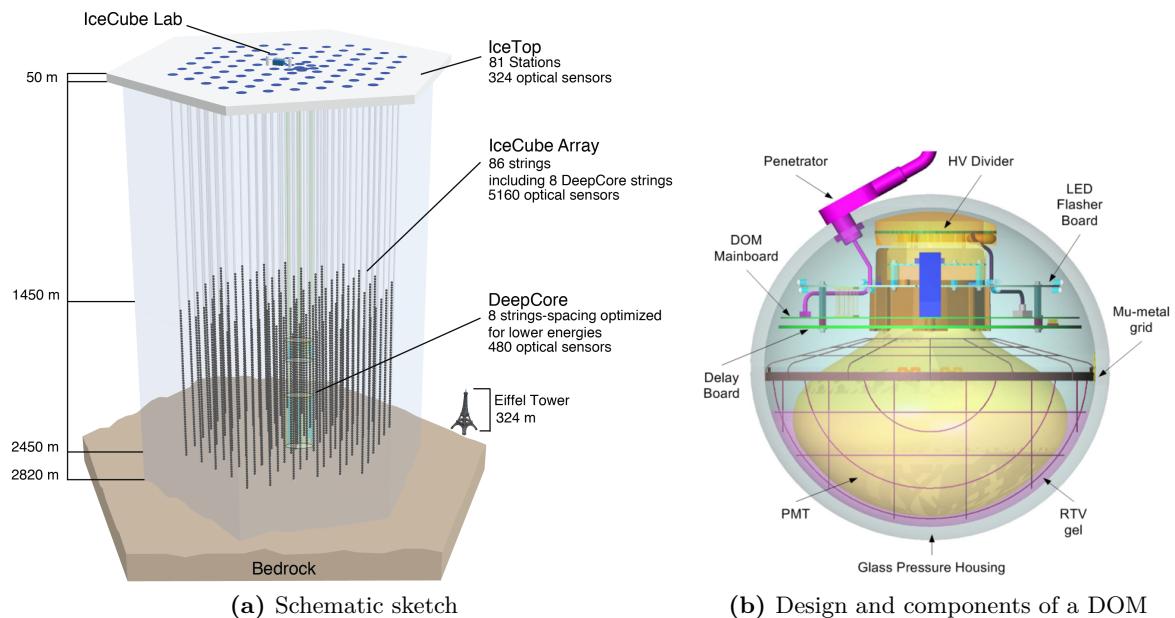


Figure 3.1: The IceCube Neutrino Observatory and a schematic view of the Digital Optical Module (DOM). More detailed information can be found in [6].

The IceCube Neutrino Observatory [6] is a cubic-kilometer particle detector located at the geographic south pole. The detector consists of 86 vertical strings with 5160 optical sensors called digital optical modules (DOMs) [17]. Figure 3.1a shows a schematic overview of the detector layout with its main components. The strings were drilled down into the Antarctic ice in a hexagonal arrangement shown in Figure 3.2. 60 optical sensors are placed on each string at depths of 1450-2450 m below the ice. IceCube is designed to detect neutrinos in the energy range $\mathcal{O}(\text{GeV})$ - $\mathcal{O}(\text{PeV})$. Neutrinos can only be detected by the secondary charged particles created when they interact in the ice. The optical sensors detect Cherenkov light produced by these charged particles (see Section 3.2). Using the observed light pattern, the neutrino energy, direction, and flavor can be reconstructed as described in Section 3.4. The DOMs are made of a spherical glass housing containing a 10 " photomultiplier tube (PMT) in addition to the readout and processing electronics. The modules are made to collect light, digitize the resulting PMT signal and send the data to the central data acquisition at the surface if a trigger condition is met. The design of a DOM and its components is shown in Figure 3.1b.

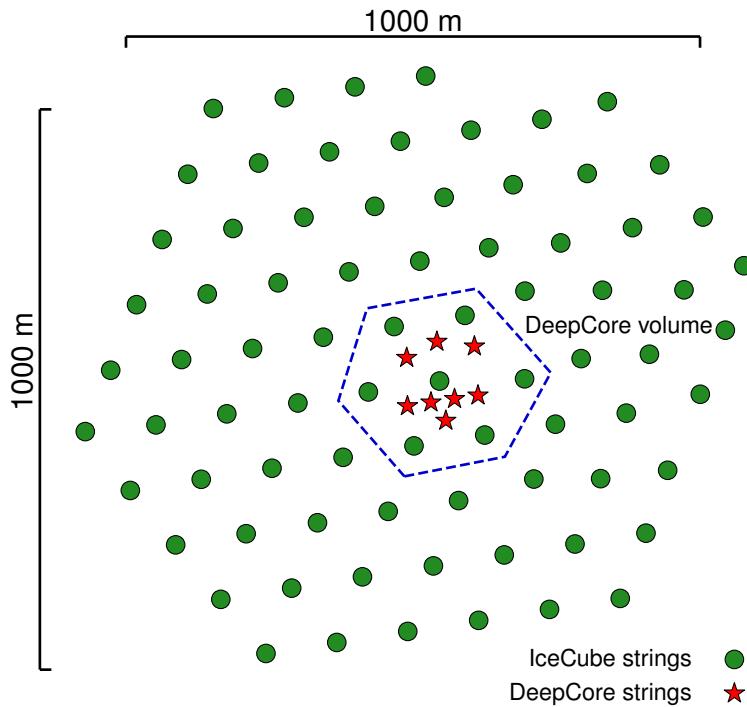


Figure 3.2: Top view of the IceCube array.

The main part of IceCube is formed by 78 strings with 125 m horizontal spacing and 17 m vertical spacing between DOMs. This results in a neutrino detection energy threshold of 100 GeV. There are 8 additional strings that form a denser sub-array of IceCube called DeepCore [18]. DeepCore is at the bottom-center of the detector and includes these 8 strings as well as the 7 surrounding IceCube strings in the fiducial volume shown in Figure 3.2. The strings in this

volume have an average distance of about 70 m. The majority of the DeepCore DOMs are placed between 2100 m and 2450 m below the ice and the vertical spacing is 7 m. Around 2050 m there is a dust layer with bad optical properties. Above the dust layer, there are a few additional DeepCore modules used as a veto cap. The denser spacing and the application of high quantum efficiency DOMs in DeepCore as well as the fact that the ice between 2100 m and 2450 m has the best optical properties leads to a lowered neutrino detection energy threshold of around 5 GeV. DeepCore is optimized for the observation of neutrinos in the energy range of 10 – 100 GeV. This is the energy range where the oscillation signal from atmospheric neutrinos can be observed and studied. The surrounding IceCube strings and the veto cap are used to reject atmospheric muons that are the main background in neutrino oscillation measurements.

The coordinate system [6] used in IceCube is centered at 46500'E, 52200'N at an elevation of 883.9 m. It is defined as a right-handed coordinate system with the y-axis pointing along the Prime Meridian (Grid North) towards Greenwich, UK and the x-axis pointing 90 degrees clockwise from the y-axis (Grid East). The z-axis is normal to the Earth's surface, pointing away from the surface. Additionally, the depth in IceCube is defined as the vertical distance from the ice surface, set to be at an elevation of 2832 m.

3.2 Cherenkov Effect

Cherenkov radiation is emitted when a charged particle moves through a medium with a velocity that is greater than the speed of light in that medium. The continuous energy loss due to the emission of Cherenkov radiation is small, at the order of $\mathcal{O}(10^{-4})$, as compared to the main energy losses that will be described in Section 3.3. The observation of this radiation in IceCube and DeepCore, however, is fundamental for the detection of the charged particles originating from the neutrino interactions that were outlined in Section 2.2. The Cherenkov effect was first observed by Pavel Cherenkov in 1934 [19]. As can be derived from trigonometry, the Cherenkov light is emitted in a parallel wavefront at the Cherenkov angle

$$\theta_c = \arccos\left(\frac{1}{\beta n}\right), \quad (3.1)$$

with n being the refractive index of the medium and β the speed of the particle in units of the speed of light. A sketch of the wavefront is shown in Figure 3.3, where the black circles depict spherically emitted light and the blue line is the formed Cherenkov light front. Typical values for the Antarctic ice are $n \approx 1.3$ and as a result $\theta_c \approx 41^\circ$ [20]. Additionally, one can calculate the number of photons produced by a Cherenkov emitter based on the description in [21]. For a wavelength λ with $(300 \text{ nm} < \lambda < 500 \text{ nm})$ 250 photons per cm are emitted assuming a very relativistic particle with $\beta \simeq 1$ [22].

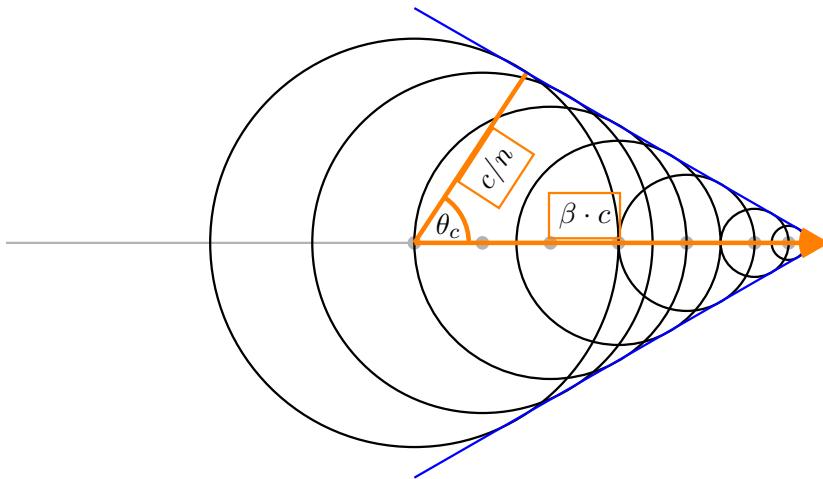


Figure 3.3: Schematic formation of the Cherenkov light front (blue) produced by a charged particle traveling faster than the speed of light in the medium. The black circles are spherically emitted light and the orange arrow shows the direction of the particle.

3.3 Energy Losses

When charged particles travel through matter they interact and lose energy by several interaction processes. The Cherenkov light emitted by the particles as described in Section 3.2 only contributes a small amount to the total energy loss. The dominant processes depend on the type of Cherenkov light source, which we can broadly categorize into the three groups: quasi-continuous energy loss by muons, electromagnetic cascades, and hadronic cascades.

Muons lose their energy mainly by *ionization*, *bremsstrahlung*, *pair production* and the *photoneuclear interaction*. Considering that ionization only has a weak energy dependence for muons above 1 GeV and combining the other three components into one, the total energy loss is given by

$$-\frac{dE}{dx} = a_I(E) + b_R(E) \cdot E, \quad (3.2)$$

where E is the energy and $a_I(E)$ and $b_R(E) \cdot E$ are the energy loss by ionization and the combined radiative losses, respectively. For the energy range of interest for this work, the parameters $a_I(E)$ and $b_R(E)$ only have a weak energy dependence and equation 3.2 reduces to

$$-\frac{dE}{dx} = a + b \cdot E. \quad (3.3)$$

This description results in a critical energy $E_{crit} = a/b$ separating the two energy regimes where ionization or radiative losses are dominant. Typical values are $a \approx 2.59 \text{ MeV/cm}$ and $b \approx 3.63 \cdot 10^{-6} \text{ cm}^{-1}$ [23] leading to a critical energy of $\sim 770 \text{ GeV}$. Since the considered energy range for atmospheric neutrino oscillations is below the critical energy we only consider ionization

losses by setting $b = 0$ which easily relates the range of a muon R_μ to its initial energy by

$$R = \frac{E_0}{a}. \quad (3.4)$$

With equation 3.4 it is clear that by measuring the length of a muon track, its energy can be estimated if the full track is contained in IceCube. This treatment is only an approximation and does not take into account the stochastic nature of some of the energy losses. Especially bremsstrahlung and photo-nuclear interactions occur rarely, but when they happen, they deposit a large amount of energy. More detailed information is found in [24].

Electromagnetic cascades are induced by electrons and positrons or photons. All of them are either produced directly in the neutrino interactions or in interactions of secondary particles. Photons lose energy via pair production whereas for electrons and positrons the dominant energy loss is due to bremsstrahlung. For both cases, the interaction process happens repeatedly and an electromagnetic shower is formed when pair production and bremsstrahlung take place in turn. Every time one of the interactions takes place, more electrons and positrons or photons are produced with smaller energy. This proceeds until the energy of the particles falls below the critical energy E_c and the remaining energy is quickly lost. For electrons and positrons, this happens through ionization and excitation of the surrounding atoms. For photons, the Compton effect and the photoelectric effect become the dominant energy losses. Electromagnetic showers are characterized by the radiation length X_0 at which the energy of electrons or positrons is reduced to $1/e$ of their initial energy. For photons, X_0 is $7/9$ of the mean free path for pair production. For ice, the critical energy is $E_c \approx 78$ MeV and the radiation length is $X_0 \approx 39.3$ cm [9].

Hadronic cascades are always produced in the neutrino interactions described in Section 2.2, either from the breaking nucleus or as decay products. Similar to an electromagnetic cascade, a hadronic cascade forms as a result of the production of secondary particles from the strong interactions of hadrons with the traversed matter. Hadronic cascades also contain an electromagnetic component, for example through the decay of neutral pions into two photons. The shower profile and the light emission are very dependent on the produced particle type, which leads to larger fluctuations between individual showers with the same energy. The observed energy also varies, because energy gets lost in the hadronic binding process and muons and neutral particles produce less light and no light, respectively. On top of that, hadrons have a higher energy threshold for Cherenkov light production, due to their higher mass. The relative brightness of hadronic showers as compared to electromagnetic showers is given by [24]

$$F(E) = \frac{T_{\text{hadron}}}{T_{\text{EM}}}, \quad (3.5)$$

where $T_{\text{hadron/EM}}$ is the total track length of a hadronic/electromagnetic shower with the same energy. The ratio $F(E)$ is always smaller than 1, but increases with energy as the electromagnetic fraction of the hadronic cascade becomes larger. A more detailed parameterization of $F(E)$, as well as fitted values for several particle types, can be found in [24].

3.4 Event Reconstruction

There are several methods to select and reconstruct events in IceCube. At energies around 10-40 GeV, where we expect the oscillation signal, the events are faint and only a few DOMs detect light. One approach for the reconstruction of such events is described in this section. The reconstruction uses only photons that traveled along a straight line - called *direct* photons. Using direct photons has the benefit of reducing the systematic biases caused by the large variations of the bulk ice properties; scattering and absorption. With an average distance of 70 m between strings in DeepCore and an effective scattering length of about 50 m [25], there will always be a fraction of direct photons arriving at the DOMs. The used method applies a stepwise procedure, where first a cleaning routine selects events with direct photons as described in Section 3.4.1. Afterward, the direction of the particle is reconstructed and finally the energy is determined as outlined in Sections 3.4.2 and 3.4.3.

3.4.1 Hit Selection

To select hits that originated from direct photons, a procedure closely related to the one described in [26] is applied. The cleaning is based on removing hits from DOMs that could have originated from light emitted by any of the other hit DOMs on the same string. The selection solely uses the time of arrival (TOA) of the pulses. It is carried out for every detected event in the following steps:

- (i) Select strings where at least 3 DOMs have seen light.
- (ii) Every hit DOM is characterized by the time of the earliest pulse (above a threshold of 0.1 photoelectron (PE)) and the integrated charge of all pulses.
- (iii) For every string passing these criteria the following steps are performed:
 - (a) Remove DOMs with hit outside of a time window of [-250 ns, +2000 ns] around the median TOA of all hits on the string.
 - (b) Using the DOM with the highest charge as reference (estimate for point of closest approach), check if any of the other DOMs on the string lies in the time window

$$\left[t_r - \frac{d_{r,i}}{c_{\text{ice}}} - t_{\text{delay}}, t_r + \frac{d_{r,i}}{c_{\text{ice}}} + t_{\text{delay}} \right], \quad (3.6)$$

where t_r is the TOA of the reference DOM, $d_{r,i}$ is the absolute distance between the two DOMs considered, c_{ice} is the speed of light in ice and t_{delay} is the allowed time delay. A time delay of 20 ns is used to limit the selection to photons with little scattering.

- (c) For each of the selected DOMs, it is now verified that, compared to each of the other selected DOMs, none was hit after the time t_{\max}

$$t_{\max} = t_i + \frac{d_{i,j}}{c_{\text{ice}}} + t_{\text{delay}}, \quad (3.7)$$

where the subscripts i and j stand for the two DOMs in questions, and all combinations are checked.

- (d) As the last step, it is checked whether there are more than six empty modules between selected modules. Keeping the DOM with the largest charge, the other DOMs are checked going upwards and downwards along the string. Finally, only strings that still have three or more selected DOMs are kept and their hits are identified as direct pulses.

3.4.2 Directional Reconstruction

The procedure to reconstruct the direction of events follows closely the one first described in [27]. For a track-like event, it is assumed that the muon, produced in the ν_μ -CC interaction, travels at constant speed along a straight line. Additionally, knowing the angle at which Cherenkov photons are emitted and the fact that we only consider direct photons, the expected arrival time t_γ of a photon at depth z along the string is well described by

$$t_\gamma(z) = t_c + \frac{1}{c} \left((z - z_c) u_z + \sqrt{n^2 - 1} \sqrt{d_c^2 + (z - z_c)^2 (1 - u_z^2)} \right), \quad (3.8)$$

where t_c , d_c , and z_c are time, distance and depth of the point of closest approach of the muon track to the string and $u_c \equiv \cos \theta_z$ as in [26, 27]. Without loss of generality the string position was set to coincide with the z -axis and the interaction vertex was assumed to be at the point of closest approach to derive this formula. Schematics of these typical, hyperbolic shapes can be seen in Figure 3.4.

Modeling cascade-like events as isotropic light sources, the arrival time can be obtained by modifying equation 3.8 and assuming light spreading from the interaction vertex only. The result also has a hyperbolic shape of the arriving light in time and depth, as described by

$$t_\gamma(z) = t_0 + \frac{n}{c} \sqrt{d_c^2 + (z - z_c)^2}, \quad (3.9)$$

where d_c and z_c are depth and distance to the cascade source, respectively.

The equations 3.8 and 3.9 hold for individual strings, although there is no preferred x-y

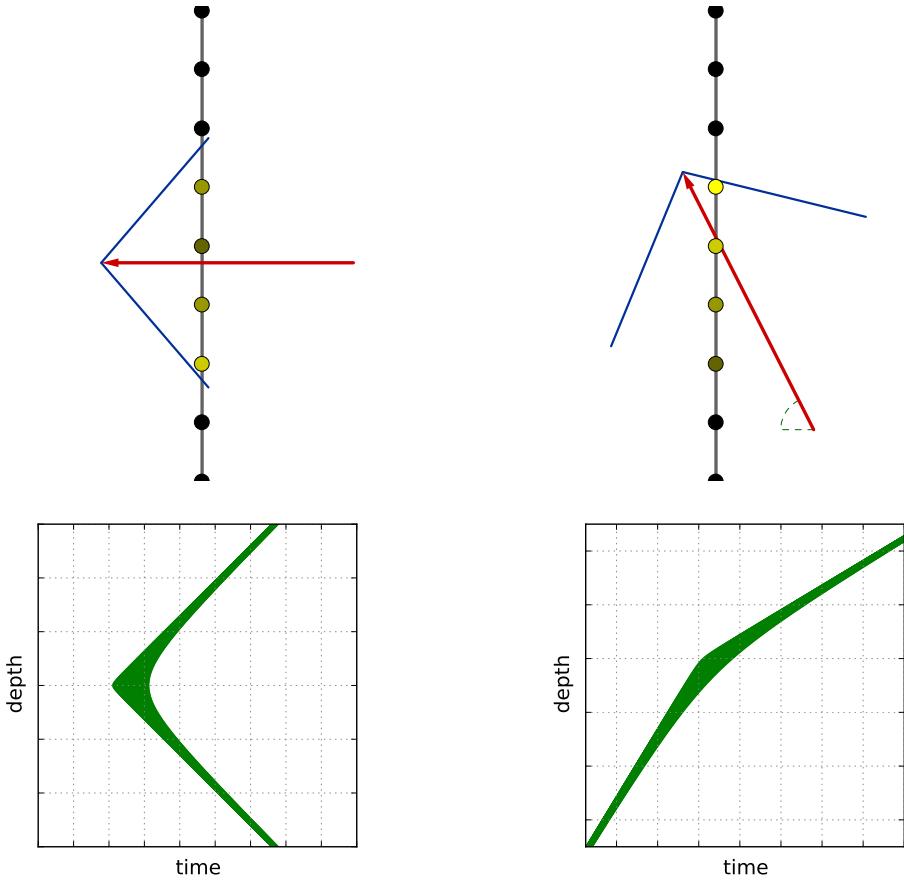


Figure 3.4: Top: Illustrated arrival time differences (dark yellow-early, bright yellow-late) and muon track including Cherenkov light cone (red arrow, blue line). Bottom: Hyperbolas formed in time and depth by the direct photons for the two orientations shown in the top. Taken from [26].

direction if only one string passed the cleaning procedure. For those so-called *single-string* events only the zenith angle can be reconstructed and the azimuth has to be obtained by a different reconstruction. For *multi-string* events, where two or more strings passed the cleaning step, the azimuth angle can also be reconstructed.

The observed time of arrivals are fitted using the modified χ^2 function

$$\chi_{\text{mod}}^2 = \sum_i \left[\frac{(t_\gamma(z_i)^{\text{exp}} - t_\gamma(z_i)^{\text{obs}})^2}{\sigma_\gamma^2} + \frac{2 \cdot q(z_i) \cdot \sqrt{r_{\text{DOM}}^2 + d_\gamma^2}}{(\cos \phi_\gamma \cdot \bar{q} \cdot d_0)} \right], \quad (3.10)$$

where $t_\gamma^{\text{exp/obs}}$ are the observed and expected arrival times in the i th DOM at depth z_i and $\sigma_\gamma = 3 \text{ ns}$ is the effective time resolution of the DOM. The additional term in equation 3.10 penalizes long photon light paths, since photons traveling short distances are expected to produce the largest signal. Observing a charge $q(z_i)$ will be disfavored depending on the photon distance d_γ traveled to the DOM, with \bar{q} being the mean charge over all selected DOMs, r_{DOM} the module

radius and $\cos \phi_\gamma$ the projected photon arrival angle on the DOM. $d_0 = 10\text{ m}$ is a distance that quantifies the strengths of the penalization and it can be understood as the typical distance at which 1 PE signal is anticipated.

3.4.3 Energy Reconstruction

The neutrino energy is reconstructed in two steps as first described in [28]. The steps are based on the previously determined direction of the event. The position of vertex and endpoint of the track will be re-iteratively modified until the best fit is found. In the first step, the most probable position, where the muon track has ended, is found by minimizing the logarithm of the likelihood ratio

$$\text{LLHR} = \log \left(\frac{\prod_{i=1}^{\text{DOMs}} p_i(\text{noHit}|\text{Track})}{\prod_{i=1}^{\text{DOMs}} p_i(\text{noHit}|\text{noTrack})} \right), \quad (3.11)$$

with $p_i(\text{noHit}|\text{Track})$ being the probability to see no light if we assume that the muon traveled along an infinite track and $p_i(\text{noHit}|\text{noTrack})$ the probability to see no light assuming a track ending at a certain point. The products in equation 3.11 extend over all DOMs in a certain radius around the track that do not have seen any light. This procedure is executed as in [20]. An illustration of the process can be seen in Figure 3.5. In a second step, the interaction vertex

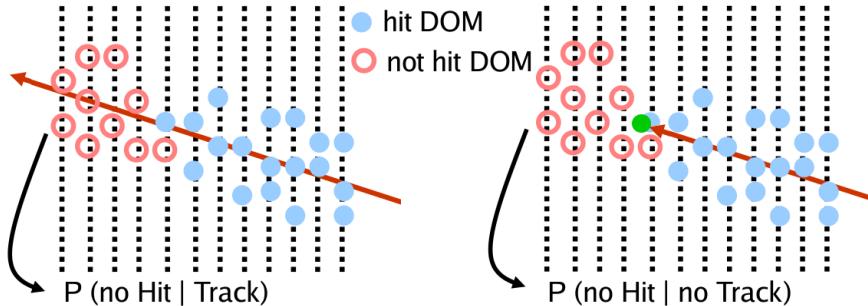


Figure 3.5: Visualization showing the track endpoint reconstruction, with the track in red, the endpoint in green, and the DOMs with signal and without a signal in blue and red, respectively. $P_i(\text{noHit}|\text{Track})$ is the probability to see no light for an infinite track and $P_i(\text{noHit}|\text{noTrack})$ the probability to see no light for a finite track. Taken from [29].

is found using the *hit-no-hit* likelihood with the method presented in [28] and [29]. This is done by minimizing the negative logarithmic likelihood

$$\text{LLH} = - \sum_i^{\text{DOMs}} \log p_i(\lambda_i), \quad (3.12)$$

where $p_i(\lambda_i)$ is the probability for a module i to have seen or not to have seen a signal, depending on whether it actually had a signal or not. The probabilities are calculated with Poisson statistics

using

$$\begin{aligned} p_{\text{no-hit}}(\lambda) &= P_0(\lambda) = e^{-\lambda} \text{ and} \\ p_{\text{hit}}(\lambda) &= 1 - P_0(\lambda) = 1 - e^{-\lambda}, \end{aligned} \quad (3.13)$$

where λ denotes the light expectation. The total light expectation can be estimated with

$$\lambda = \lambda_{\text{track}} + \lambda_{\text{cascade}}^{\text{1GeV}} \cdot E_{\text{cascade}}^{\text{EM}} + \nu \cdot \Delta T_{\text{event}}, \quad (3.14)$$

containing the expected number of photons λ_{track} and $\lambda_{\text{cascade}}^{\text{1GeV}}$ for a track and for a 1 GeV cascade, respectively, the EM equivalent energy $E_{\text{cascade}}^{\text{EM}}$ of the hadronic cascade and the noise rate ν of the DOM. The noise rate is scaled with the time duration T_{event} of the pulse for the considered event. The expected number of photons are taken from tables that are based on simulated tracks and electromagnetic cascades at various positions in the detector. For a given DOM the light expectation of an event located at any point in the detector can be read out from these *photonic* lookup tables [25].

The total energy of the neutrino is calculated based on the underlying assumption of the occurrence of a hadronic cascade at the interaction point and a muon track originating from this point. By minimizing the likelihood in equation 3.12 we find the energy $E_{\text{cascade}}^{\text{EM}}$ of the hadronic cascade in EM equivalent and the energy of the muon defined by the total length L_μ of the muon track. With these quantities and the description of muon energy loss from [30], the total neutrino energy is given by

$$E_\nu = \frac{b}{a}(e^{b \cdot L_\mu} - 1) + F^{-1}(E_{\text{cascade}}^{\text{EM}}), \quad (3.15)$$

where $a = 0.266 \text{ GeV/m}$ and $b = 4.6 \cdot 10^{-4} \text{ m}^{-1}$. The factor F^{-1} accounts for the different amount of light produced by hadronic cascades as compared to electromagnetic cascades.

3.5 Event Topologies

The signals that IceCube detects vary depending on the neutrino flavor and interaction type of the event. The two main signatures that can be observed are track-like and cascade-like events. The observed Cherenkov light is produced by the secondary particles originating from the neutrino interactions described in Section 2.2. Table 3.1 shows an overview of the possible event signatures. Minimum ionizing muons can travel for long distances and are seen as extended light signatures called tracks. Muons can come from ν_μ -CC interactions or from ν_τ -CC followed by the decay of the τ to a muon. However, the τ only decays to a muon with a branching ratio of BR=17%.

Cascades are the light signal produced by the EM/hadronic showers described in Section 3.3. They come from ν_e -CC and most of the ν_τ -CC interactions because the electron and the tau lose

all their energy quickly and only travel a short distance. They are also produced in all ν -NC interactions since only the hadronic shower is observable and the produced neutrino escapes unseen. The cascades at the energies considered in this work have a smaller radius than the spacing of the DOMs and are therefore seen as point-like light emitters.

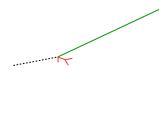
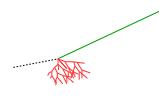
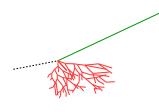
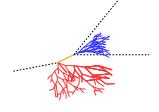
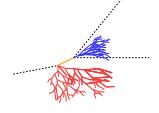
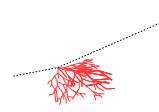
Interaction	Secondary particles	Signature
CC $\nu_\mu^{(-)}$	 μ^\pm track	Track-only
	 μ^\pm track and hadrons	Track with cascade
CC $\nu_\tau^{(-)}$	 τ^\pm decaying into μ^\pm ($\sim 17\%$ BR), hadrons	
	 τ^\pm decaying into e^\pm or hadrons ($\sim 83\%$ BR)	
CC $\nu_e^{(-)}$	 e^\pm , hadrons	Cascade-only
NC $\nu_\ell^{(-)}$	 hadrons	

Table 3.1: IceCube event signatures, their underlying interaction type and the particles that produce them. Also shown are the secondary particles produced in the interactions. Black dashed lines represent neutrinos, green lines muons, and blue and red lines are particles in electromagnetic and hadronic cascades, respectively. Taken from [10].

The existence of the two types of event topologies and their origins imply that by identifying track-like events we can identify events coming (mainly) from ν_μ -CC interactions and therefore obtain a flavor identification. This is a crucial part of performing an oscillation analysis as will be further discussed in Section 5.1.

Chapter 4

Improving Particle Identification

As already indicated in Section 3.5, particle identification (PID) is needed to perform an oscillation analysis, and improving the PID leads to increased precision when determining oscillation parameters. The fundamental difficulty is that tracks at low energies are very hard to distinguish from cascades because the muon produced in a ν_μ -CC interaction can have low energy and will therefore not produce a long visible track. This causes the detector response for the two event types to look very similar which can be seen in Figure 4.1.

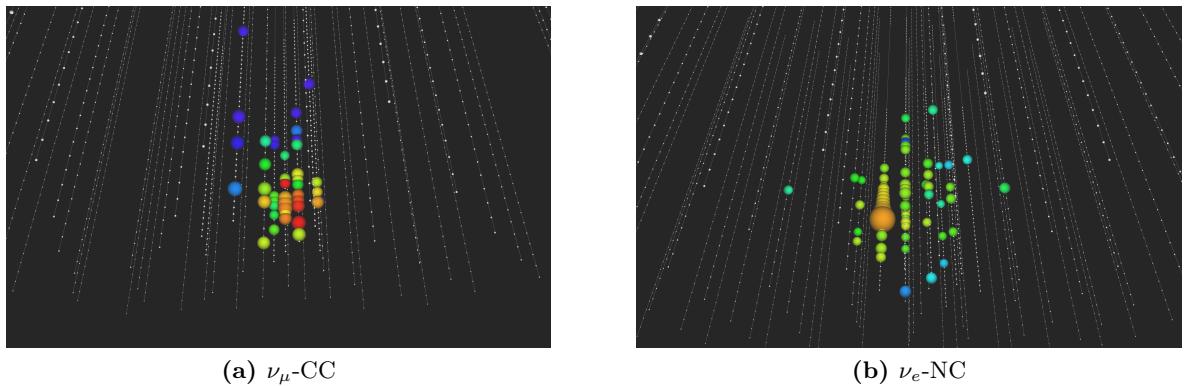


Figure 4.1: Examples of low energy track and cascade events. The color of the spheres indicates the signal time, where red is early and blue is late. Their size is proportional to the integrated charge in each DOM. The depicted events have a neutrino energy of $\mathcal{O}(\text{GeV})$.

In previous works, a single reconstructed variable was used as a PID discriminator. Here instead several variables are used to train a multivariate machine learning algorithm to find a more powerful PID discriminator. Section 4.1 introduces a selection of relevant reconstructed variables. Section 4.2 presents the multivariate method and Section 4.3 shows the results of the classification.

4.1 Reconstructed Variables

The main motivation for the work done in this thesis is the number of reconstructed quantities that might be useful in discriminating event signatures between tracks and cascades. All variables are taken from the event reconstruction outlined in Section 3.4. Figure 4.2 shows the distributions of selected variables, split in ν_μ -CC (track) and ν_e -CC+ ν_e -NC+ ν_μ -NC (cascade).

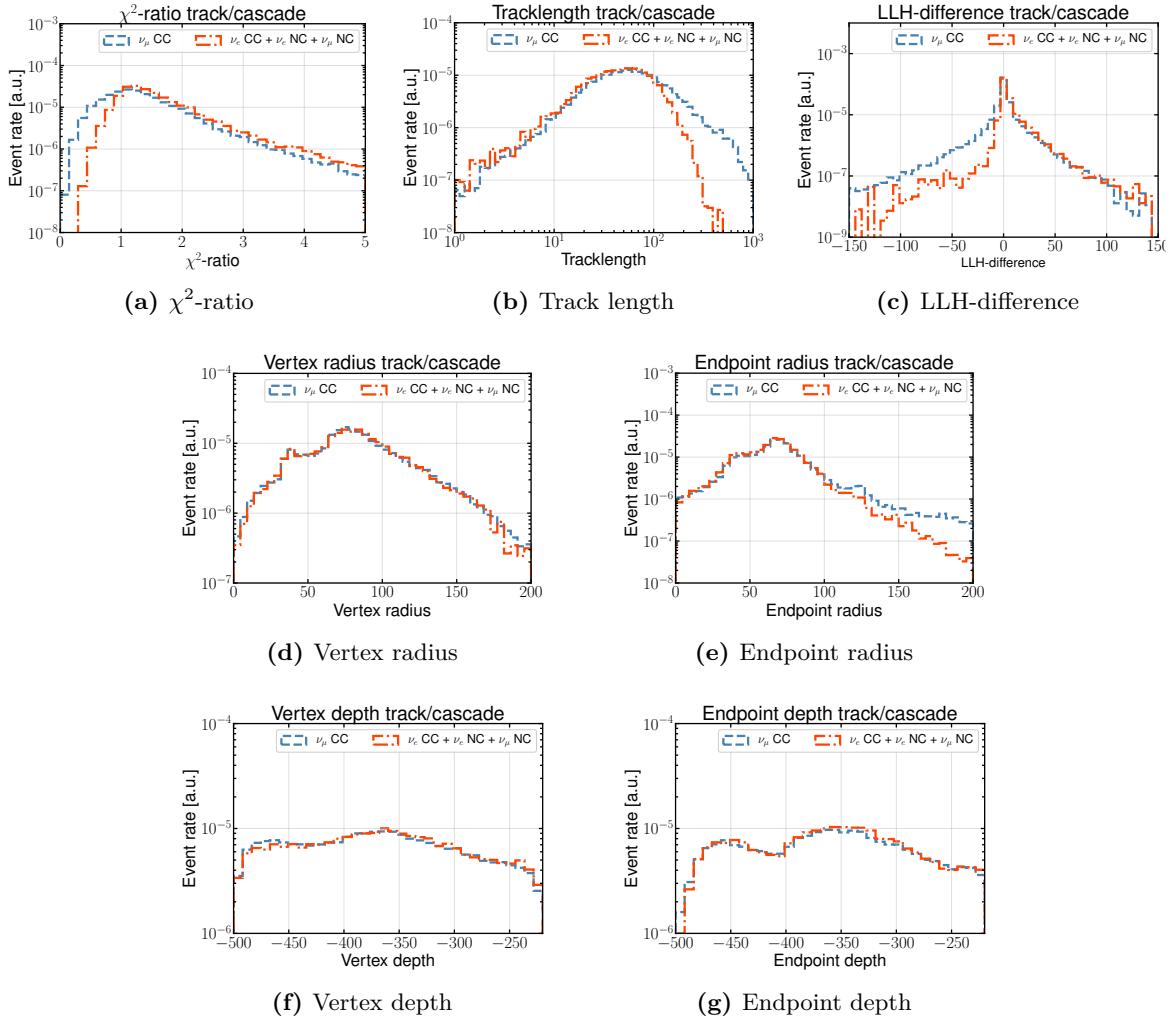


Figure 4.2: Distributions of the selected input features split in ν_μ -CC (track) and ν_e -CC+ ν_e -NC+ ν_μ -NC (cascade) events. The rates are scaled to the same area so that the shapes can be compared.

Two of these variables are related to the fit performance of the event reconstruction, while the rest of them are reconstructed physical quantities themselves. The first interesting variable that was used as a PID variable in previous analyses [10, 31] is the ratio between the goodness of fit for a track hypothesis and a cascade hypothesis in the directional reconstruction explained

in 3.4.2. The variable referred to as the χ^2 -ratio (Figure 4.2a) is calculated as

$$\chi^2\text{-ratio} = \frac{\chi^2_{\text{track}}}{\chi^2_{\text{cascade}}} \cdot \frac{\text{DoF}_{\text{cascade}}}{\text{DoF}_{\text{track}}}, \quad (4.1)$$

where $\text{DoF}_{\text{cascade}}$ and $\text{DoF}_{\text{track}}$ are the degrees of freedom in the respective fit. They are $\text{DoF}_{\text{cascade}} = N_{\text{DOMs}} - 3$ and $\text{DoF}_{\text{track}} = N_{\text{DOMs}} - 4$. N_{DOMs} is the number of DOMs with direct photons used for the reconstruction. Small values of the χ^2 -ratio indicate track-like events and large values cascade-like events. This behavior can clearly be seen in Figure 4.2a.

Another variable, which is related to the energy reconstruction described in Section 3.4.3, is the difference of negative log-likelihoods (compare Equation 3.11) for the track- and cascade-only hypothesis fit, defined as

$$\text{LLH-difference} = \text{LLH}_{\text{track}} - \text{LLH}_{\text{cascade}}. \quad (4.2)$$

Since small LLH values imply a good fit, small values of the LLH-difference (Figure 4.2c) suggest track-like events and large values cascade-like events. Figure 4.2c shows the indicated behavior of the LLH-difference.

The other features are related to the reconstructed position of the interaction vertex and the reconstructed track length. The track length (Figure 4.2b) itself also holds potential in discriminating tracks from cascades, as can be seen in Figure 4.2b. Events from ν_μ -CC interactions (tracks) have longer reconstructed track lengths than events coming from ν_e -CC + ν_e -NC + ν_μ -NC interactions (cascades). This is especially true for higher energy ν_μ -CC interactions where the muon carries a large amount of energy. Figure 4.2 also shows the horizontal, radial distance to the center of the detector for the interaction vertex (Figure 4.2d) and the track endpoint (Figure 4.2e) as well as the depth of the two positions (Figure 4.2f, Figure 4.2g). The position of the endpoint of the track combines the information of the direction of the event and the reconstructed track length from the energy reconstruction. It seems like these variables are not useful in discriminating tracks from cascades, because the track and cascade distributions are very similar. This is certainly true when using them as a single discriminator, but in combination with other variables by applying a multivariate technique they are useful. This will be further discussed in Section 4.2.1. Another reason why the spatial variables are used is that the instrumentation of IceCube is much denser around the center as shown in Figure 3.2. As a result, the reconstruction performs better in the center leading to more accurate reconstructed quantities. This could improve the classification for events occurring in the that region.

4.2 Gradient Boosting Classifier

As already mentioned, the goal is to develop a new event classification algorithm for PID. As discussed in Section 4.1, there are multiple variables that hold some separation power between

tracks and cascades. But, since there is no direct functional relation between these variables and the neutrino flavor, we need to build a non-parametric classification model. For this purpose, we utilize the Monte-Carlo (MC) simulation of IceCube events that is available. There are a variety of ways to build a multivariate model. Here we choose to apply a Gradient Boosting Machine (GBM) [32].

GBMs are machine learning techniques that combine weak prediction models in a stage-like fashion to produce a strong prediction model. GBMs can be used to optimize any differentiable loss function and offer the possibility to tune the parameters of the training process for custom use. In this work, we use the Gradient Tree Boosting algorithm provided by the scikit-learn package [33]. In this case, the weak prediction models are formed by simple decision trees, which are models that predict a discrete output variable dependent on a set of input variables. Figure 4.3 shows a visualization of an example decision tree. Following the paths from the root at the top to the leaves at the bottom, at every interior node the data is split by applying a straight cut on one of the input variables. The cuts are chosen to achieve the best possible separation of the data into the output classes, measured by some metric. The cuts and the variables are specified at the top of each node. Also shown is the percentage of the sample chosen by the applied cuts. Each leaf provides a value for the output variable dependent on the input variables. The different paths from the root to the leaves form the classification rules. Figure 4.3 shows what fractions are tracks and cascades and a classification into the output classes, based on the class of the majority of events. The boosting algorithm fits multiple such decision tree models in an iterative way, where at each step a loss function is calculated from the ensemble of previously fit weak models. The next decision tree will be constructed so that the ensemble loss is reduced along its negative gradient, leading to a reduction of the residual loss. This process is repeated until the loss is reduced below a certain threshold or the maximum number of trees is reached.

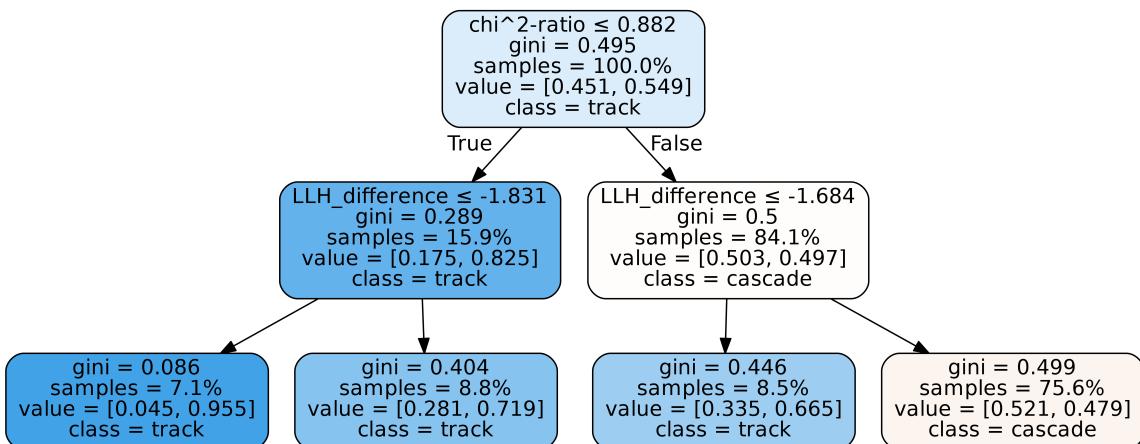


Figure 4.3: Example decision tree showing the variable and the chosen cut at the top of each node, the fraction of the full sample that is selected by each cut and the fractions that are track and cascade. Based on the bigger fraction a classification into track or cascade is displayed at the bottom of each node.

To use a GBM classifier on real data, we first need to train it on a dataset for which we know the true interaction type. For this work, the training is performed on a set of simulated neutrino events in the energy range of $1 - 1000$ GeV. To get a realistic distribution of events and to take into account the atmospheric neutrino production explained in Section 2.3 and the detector volume, the events are weighted depending on their energy and direction. The weights also include the oscillation probability, as will be discussed in Section 5.1.1. The units of the weights are chosen as events/s, such that summing over the weights of all events in a dataset gives the overall event rate. The events are reconstructed using the methods explained in Section 3.4, but also carry the true information from the simulation production. The following sections illustrate the process of selecting the input variables and performing some data pre-processing.

4.2.1 Feature Selection

The classifier is trained on a set of selected input variables. These so-called *features* are selected by investigating if they have some potential in separating tracks from cascades.

A first step is to check if the feature distributions for track and cascade vary significantly as already done in Section 4.1. Features that show a separation are useful to feed into a GBM, but features that do not can also be useful.

As a second step, we look at the Pearson correlations [34] of the features. The correlations for the selected features presented in Section 4.1 are shown in Figure 4.4. Again, they are split in ν_μ -CC (track) and ν_e -CC+ ν_e -NC+ ν_μ -NC (cascade). Features can help the classification if

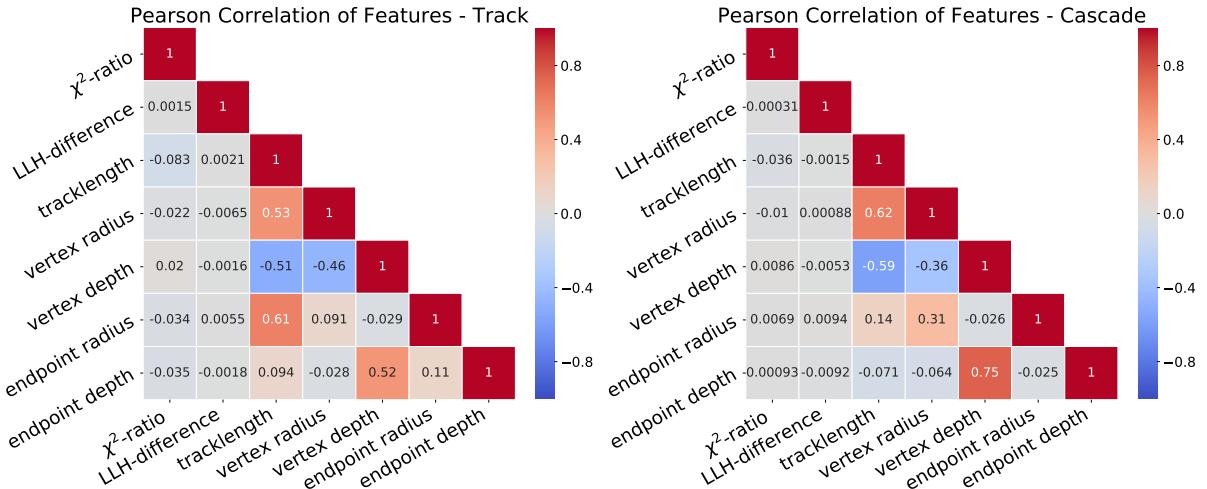


Figure 4.4: Pearson correlation of the selected features for ν_μ -CC (track-like) events (left) and ν_e -CC+ ν_e -NC+ ν_μ -NC (cascade-like) events (right). The color code indicates the strength of the correlation.

their correlations with other features are different for tracks and cascades. For example, we see that for tracks the correlation between track length and χ^2 -ratio is -0.083, while for cascades it is only -0.036. Also interesting are correlations between features that show a strong separation in track and cascade and features that do not. For example, with a value of 0.00093, there is

effectively no correlation between the χ^2 -ratio and the track endpoint depth for cascades but for tracks the correlation is -0.035.

As a third step, we need to guarantee that the classification based on the features will perform in the same way when applied to real data. Although the MC simulation set reflects the current best knowledge of the neutrino interactions and fluxes as well as the detector properties, there could still be inaccuracies. This could lead to a mismatch between variable distributions in data and simulation as well as wrong correlations between the variables. We therefore check if the feature distributions agree between data and simulation. For the purpose of checking the agreement, 10% of the data of the years 2011-2018 are used and a larger MC dataset. The weights are calculated assuming standard oscillations with $\Delta m_{32}^2 = 2.42 \cdot 10^{-3}$ eV² and $\sin^2 \theta_{23} = 0.538$. The detailed comparison plots for every feature variable are listed in appendix A. The agreement is reasonable for all of the features. It should be noted that this comparison is done without having performed a fit on the oscillation parameters and the overall normalization.

Finally, the selection of input features is determined after checking the performance of the classifier depending on some chosen metric. The importances of the individual features can be extracted for a trained classifier. This shows which of the features has the strongest effect on the classification. Based on the importances, features can be removed and new feature combinations tested. The set of features that results in the best performance is chosen. The features discussed here are the best combination found and, therefore, the selection chosen in this work.

4.2.2 Data Preprocessing

To prepare the data for the training, some pre-processing steps are necessary. Only events that passed the directional reconstruction cleaning procedure will have the features chosen for the training and in some cases the energy reconstruction fails even though the directional reconstruction succeeded. For this reason, all events are removed where some of the features are missing. After removing these incomplete events, a boolean variable is created indicating whether the event is a true track (ν_μ -CC) or a true cascade (ν_e -CC+ ν_e -NC+ ν_μ -NC). This boolean variable will be used as a target variable for the training step. Events coming from any ν_τ interaction are not taken into account here, because to first order they are true cascades. For the training step the ν_τ weights are, therefore, set to zero. The training of the classifier is always performed on a subsample of the data called the *training set*, while the remaining part of the data - called the *test set* - is used to estimate its performance and test its robustness. In this thesis, the data is split in ~50% training set and ~50% test set. To compensate for the fact that there are significantly more track-like events in the sample than there are cascade-like events, the weights of the events are scaled in a way so that the total rate of tracks and cascades is the same. Additionally, the weights of all events are scaled to lie between 0 and 1. Otherwise, the classifier can struggle to handle very small numerical values.

4.3 Classifier Results

The results of the best-performing classifier are reported here. For every event that is classified by the trained algorithm, a probability score is calculated that represents the probability to be a true track event. Figure 4.5 shows the probability score distribution for training and test sets split in ν_μ -CC (tracks) and ν_e -CC+ ν_e -NC+ ν_μ -NC (cascades). There is a clear peak at

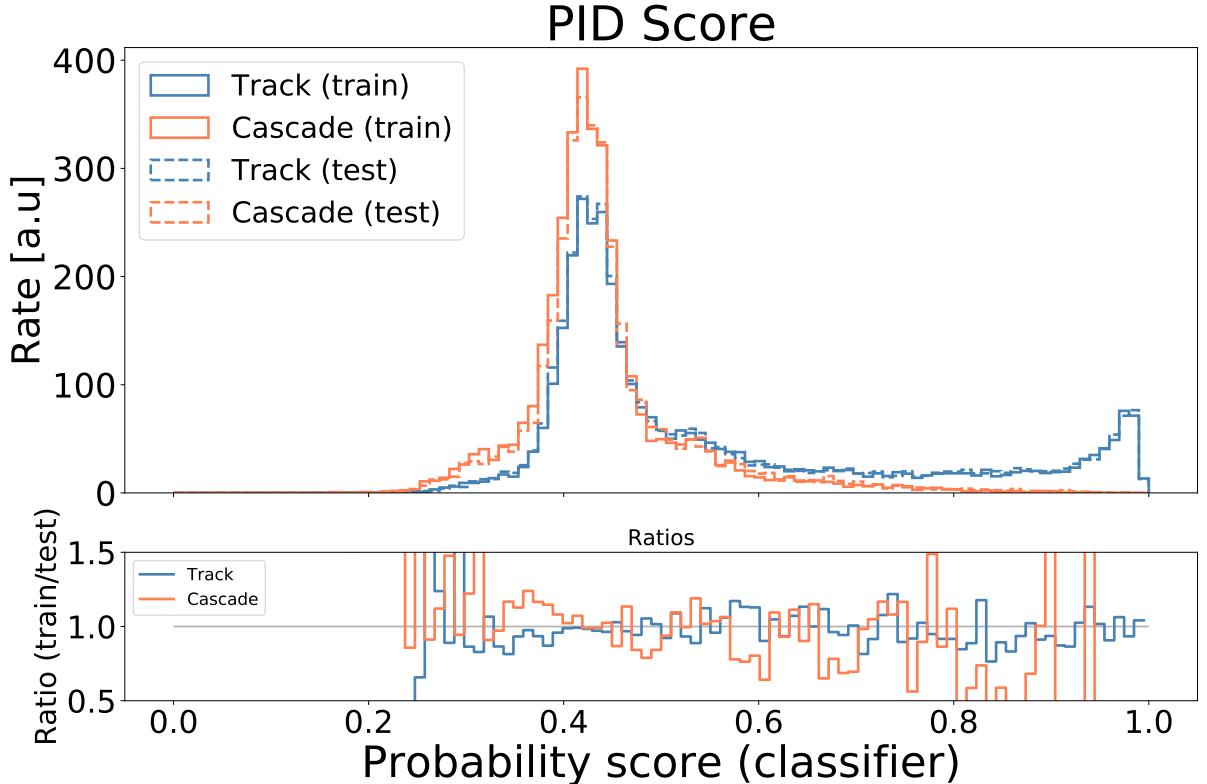
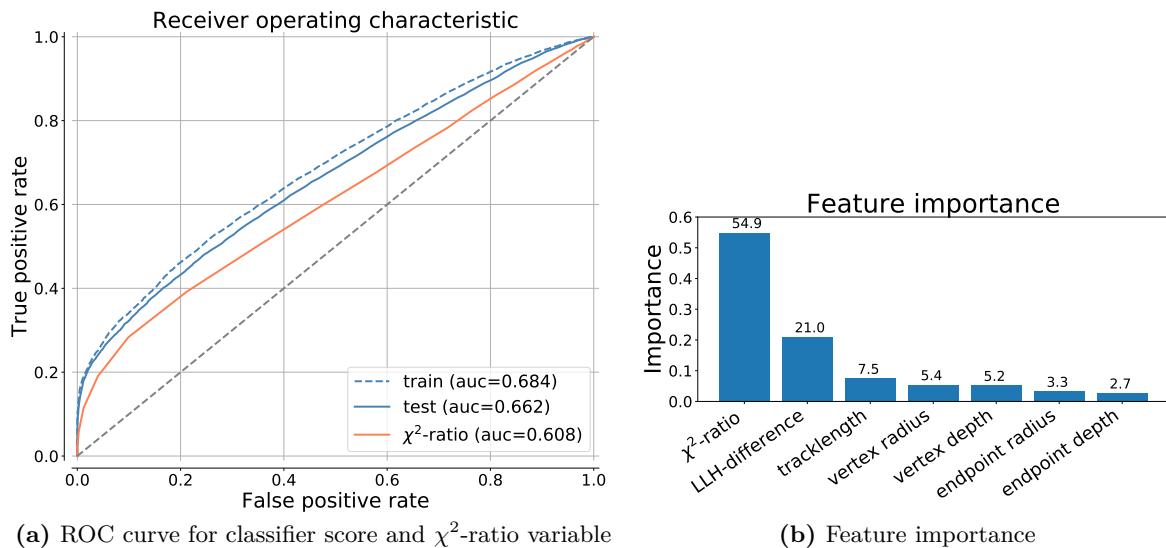


Figure 4.5: Classifier output probability of the final, trained classifier for training/test set split in ν_μ -CC (track) and ν_e -CC+ ν_e -NC+ ν_μ -NC (cascade).

values close to 1, which indicates track-like events. For lower probability scores the cascades dominate the distribution, but there is still a significant track population. These events are not separated very efficiently by the classifier. Based on the probability distribution, we can divide events into subsamples that are track- and cascade-dominated. Historically, there have only been two bins, but in principle any number of PID bins could be chosen. The bins are selected by applying cuts to the probability distribution. Ultimately, the optimal cut values will depend on the explicit analysis being performed. As a test case, we consider the impact this classifier has on the sensitivity to atmospheric neutrino oscillations in Chapter 5. Several cut values and number of PID bins are tested for this type of analysis.

The receiver operating characteristic curve (ROC curve) of the trained classifier is shown in Figure 4.6a compared to the ROC curve using only the χ^2 -ratio variable. This graphical

evaluation method is a visualization of how the probability score distribution for track and cascade is separated by straight cuts. The curve shows the performance of the classifier for various thresholds, where the threshold is the cut on the probability score placed to split true and false (e.g. track and cascade). As shown in Figure 4.6a the true positive rate (TPR) is plotted on the y-axis against the false positive rate (FPR) on the x-axis. TPR is the fractional number of correctly classified true events, while FPR is the fractional number of misclassified false events. TPR, therefore, means the fraction of tracks that were correctly identified as tracks and FPR the fraction of cascades that were falsely identified as tracks. Every point on the curve belongs to a different threshold value. For a perfect classifier, the curve goes as close as possible to the top left corner of the plot, which is the point where 100 % tracks are correctly classified while having 0 % misclassification of cascades.



(a) ROC curve for classifier score and χ^2 -ratio variable

(b) Feature importance

Figure 4.6: Receiver operating characteristic and feature importances for final, trained classifier.

In Figure 4.6b the relative importances of the various input features are shown. As already observed from the distributions in Figure 4.2, the χ^2 -ratio and the LLH-difference distribution have the strongest effect in separating tracks and cascades with an importance of 55 % and 21 %. The track length and the spatial coordinates play a lesser role with 3 – 8 %. The effect of removing them is tested, but the result is a worse performance of the algorithm, due to the reasons discussed in Section 4.2.1.

To illustrate how the trained GBM classifies events coming from the different flavor interactions, the probability score is plotted for the different interaction types individually in Figure 4.7. On the left, the normalized distribution is shown with a logarithmic y-axis and on the right the cumulative distribution. As already discussed before, the ν_μ -CC event distribution has a clear peak at values of 1 but also spreads towards lower probability values. It does not have the stronger visible rising edge at probability score values around 0.2-0.3 that the all ν -NC event distributions show. This is also weakly observed for the ν_e -CC event distribution. The

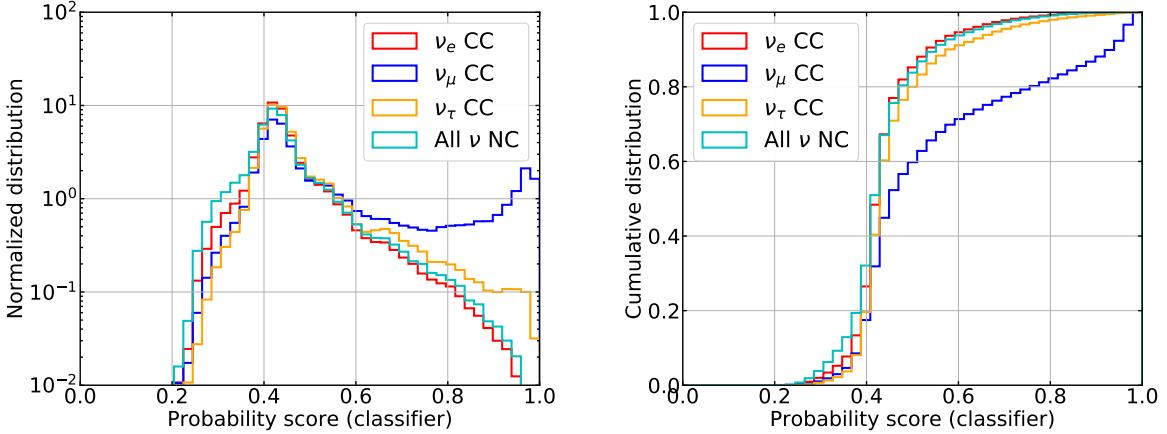


Figure 4.7: The expected PID score distribution (left) and the cumulative distribution (right) for the different interaction types.

shapes of ν_e -CC event distribution and the all ν -NC event distribution are very similar. They have their peak at ~ 0.42 and afterward fall off towards higher probability values. This behavior makes sense since both ν_e -CC events and the all ν -NC events produce cascade signatures only. The shape of the ν_τ -CC event distribution at lower probability values follows the shape of the ν_μ -CC event distribution. At higher probability values the ν_τ -CC event distribution also shows some excess as compared to the ν_e -CC and the all ν -NC event distributions. This can be explained by the fact that the tau leptons produced in ν_τ -CC interactions decay to muons with a 17% BR [9]. These events can potentially produce real, visible tracks and the classifier can identify them.

An additional study was performed in order to test the effect of training the classifier on a sub-set of ν_μ -CC events. We trained the classifier on ν_μ -CC events above a certain threshold track length as true tracks, while adding the remaining fraction of ν_μ -CC events below the threshold to the cascades. This approach is motivated by the fact that low energy ν_μ -CC events have very short tracks and there is a limit below which the detector cannot identify them due to the minimum spatial separation of DOMs. Although the classifier results achieved with this approach were good, there was no significant improvement in the oscillation parameters sensitivities. For this reason, the results are not presented here.

4.3.1 Classifier Evaluation

It is necessary to perform certain checks on the trained classifier to guarantee that it performs well. For example, an algorithm is said to *overfit* when it starts to classify events from known data better than events from unknown data. Overfitting occurs when the algorithm starts to learn features of the training set that are not general features of the whole dataset, but rather statistical fluctuations. As a result, the classification of the test set becomes worse.

We can prevent overfitting by comparing the outputs for the training and the test sets. In Figure 4.5 this was done for the distribution of probability scores for true tracks and true cascades separately. To check whether the distributions match, we calculate the agreement using a two-sided Kolmogorov-Smirnov-test (KS-test) [35]. The two-sided KS-test checks if the null hypothesis, that two samples are drawn from the same continuous distribution, cannot be rejected. A high p-value, therefore, means that the two distributions are similar. For the tracks a p-value of 67 % was found while for the cascades it was 99 %, from which we confirm that there is no overtraining. In addition to the KS-test, the distributions are compared by their ratio, as shown in the lower part of Figure 4.5. Apart from the regions where the statistics are low, the ratio is close to 1 and does not show any significant bias. As it was already done with the input feature distributions, we also check the agreement between data and simulation for the output probability score distribution. Figure 4.8 shows the comparison plot for the classifier output which shows good agreement between data and simulation.

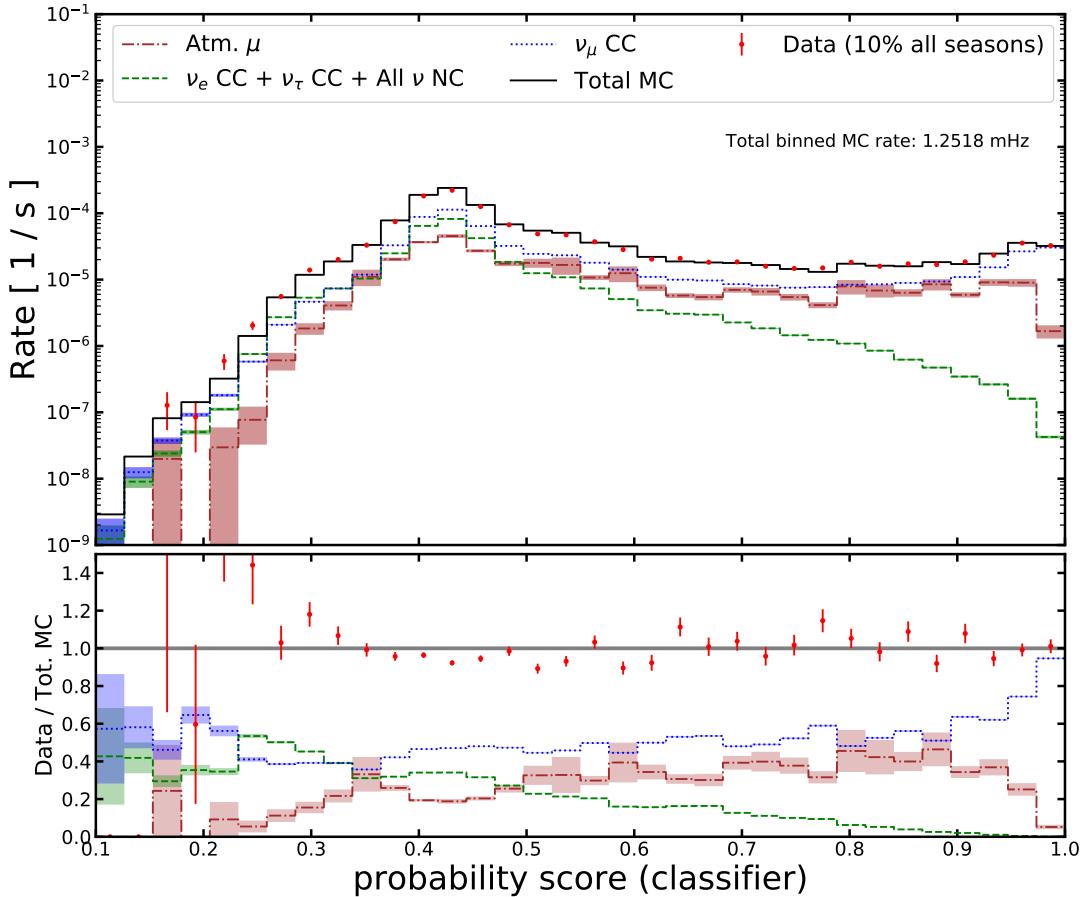


Figure 4.8: Data/MC agreement for classifier output probability with $\Delta m_{32}^2 = 2.42 \cdot 10^{-3} \text{ eV}^2$ and $\sin^2 \theta_{23} = 0.538$.

Another way to monitor overtraining is related to the ROC curve shown in Figure 4.6a. For this work, the abilities of the classifier are measured by calculating the area under the curve

(AUC) for the ROC curve. In the ideal case, AUC=1. A classifier that always classifies the same number of tracks and cascades has a ROC curve following the diagonal between x- and y-axis, which results in an AUC of 0.5. Such a random classifier has no ability to separate tracks from cascades. In this work, the AUC is the main metric used to assess the performance of the trained GBM. Additionally, it provides a method to check for overfitting by comparing the values of the training and test sets. If we achieve much higher AUC values for the training set than for the test set, it is a clear sign of overfitting. The AUC values achieved with the trained classifier are 0.684 and 0.662 for training and test sets, respectively, and 0.608 for the χ^2 -ratio. It can be seen that the classifier performs better than the χ^2 -ratio and that there exists a region where the TPR is increasing, although the FPR stays close to zero. This means that for some threshold values, a set of events can be clearly identified as tracks without having any cascade contamination in that chosen track bin. The AUC score of the training set is not significantly larger than the AUC score of the test set and the classifier, therefore, does not show signs of overfitting.

4.3.2 Hyper-Parameter Tuning

The classifier training process is governed by a set of training variables called hyper-parameters. Some of them apply to the individual trees while others are related to the boosting algorithm as a whole. For most datasets, it is useful to modify the parameters of the classifier to achieve the best performance. In this work we perform a grid-search for every parameter to find the best training output, which is evaluated using AUC as a metric. The best performing parameters are then used to train the classifier and the usual checks for overtraining and behavior of the classifier are made. At the start, a set of baseline parameters is chosen to have a comparison when trying to improve the performance. The tuning of the parameters is essentially done in four steps. In the first step, the learning rate is kept fixed and the best number of estimators is found. The maximum depth of the trees and the minimum samples required in a node to perform a split are tuned in a second step. The minimum samples required on a leaf are optimized in a third step to achieve the best training AUC score. Altering the maximum number of features considered for every split did not improve the score after that, neither did decreasing the learning rate and increasing the number of estimators.

The default parameters, as well as the tuned parameters are shown in Table 4.1. Also shown is the overall score on training and test sets after the classifier training with the new parameters. Evidently, tuning the hyper-parameters did increase the AUC score on the training set, but for the test set, it led to degradation. This is a clear sign of overtraining since the tuning solely affects the performance on the training set. We also see large discrepancies when checking the probability score distributions for training and test sets shown in Figure 4.9. The probability distributions for the training and test sets using the tuned parameter set do not agree with each other. This is particularly the case for the cascade distributions.

This exercise demonstrates the trade-off between hyper-parameter tuning and overtraining.

	Default	Tuned
Learning rate	0.275	0.275
Number of estimators	50	80
Maximum depth	3	5
Minimum samples to split	2	1100
Minimum samples in leaf	1	5
Maximum features	7	7
AUC(train)	0.684	0.746
AUC(test)	0.662	0.654

Table 4.1: Classifier hyper-parameters before and after the tuning procedure. Also shown is the resulting AUC score for training and test sets.

It also shows that to perform a proper hyper-parameter tuning the simulation statistics are not sufficient. Therefore, we use the default parameters listed in Table 4.1 since they do not result in overtraining.

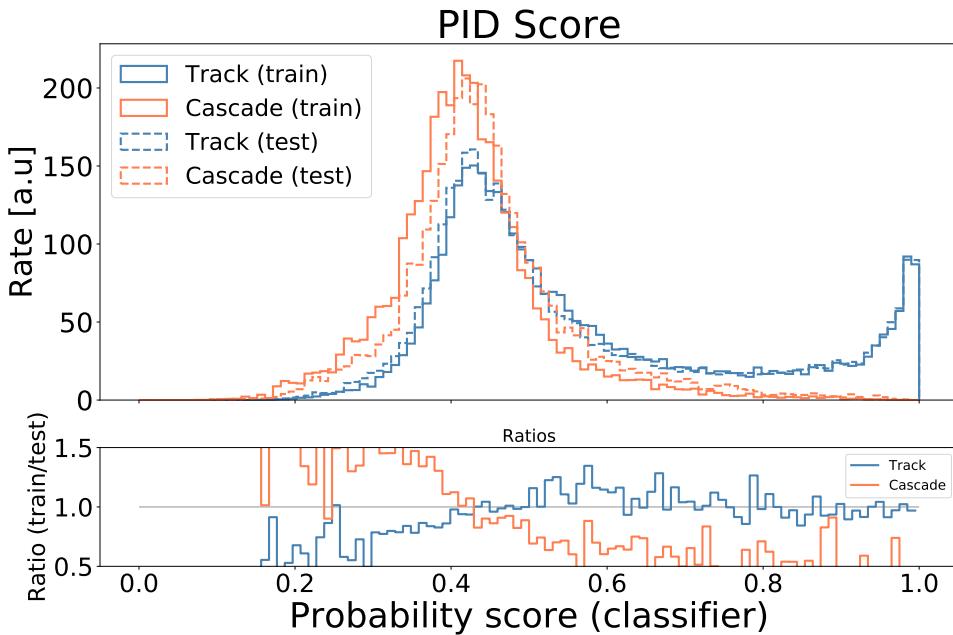


Figure 4.9: Classifier output probability with tuned parameters for training/test set split in ν_μ -CC (track) and ν_e -CC+ ν_e -NC+ ν_μ -NC (cascade).

Chapter 5

Sensitivity to Atmospheric Neutrinos Oscillations

This chapter shows the impact that the improved PID has on a measurement of atmospheric neutrino oscillations. The newly developed PID variable *GBM probability* based on the multivariate classification method explained in Chapter 4 is compared to the univariate PID discriminator χ^2 -ratio that was used in the past to show the improvement. Section 5.1 describes the binned likelihood analysis strategy and how the parameter uncertainties are estimated. Section 5.2 illustrates the binning optimization procedure. In Section 5.3, the sensitivities to the atmospheric mixing parameters Δm_{32}^2 and θ_{23} are presented.

5.1 Analysis Strategy

The extraction of neutrino oscillation parameters is based on the dependence of the oscillation probability on the neutrino energy and the traveled distance of the neutrino, as explained in Section 2.4. The traveled distance of a neutrino that is detected in IceCube is related to its incoming zenith angle and therefore we can illustrate the oscillation probability as a function of energy E and cosine of the zenith angle θ_z , $\cos \theta_z$. Figure 5.1 shows the muon neutrino survival probability for a chosen set of oscillation parameters. A zenith angle of $\theta_z = \pi$ ($\cos \theta_z = -1$) corresponds to a neutrino coming from the negative z direction. The green bands are the oscillation pattern that we want to observe. Due to the minimum neutrino detection energy threshold of approximately 5 GeV in DeepCore, the signal we can detect is in the energy region of 20-40 GeV for up-going neutrinos.

5.1.1 Binned Poisson Likelihood

A binned likelihood method is used to determine the neutrino oscillation parameters. For this method, the observed events are binned in three-dimensional histograms by E , $\cos \theta_z$ and PID. Assuming a Poisson distribution of event counts in each bin we can fit the oscillation parameters

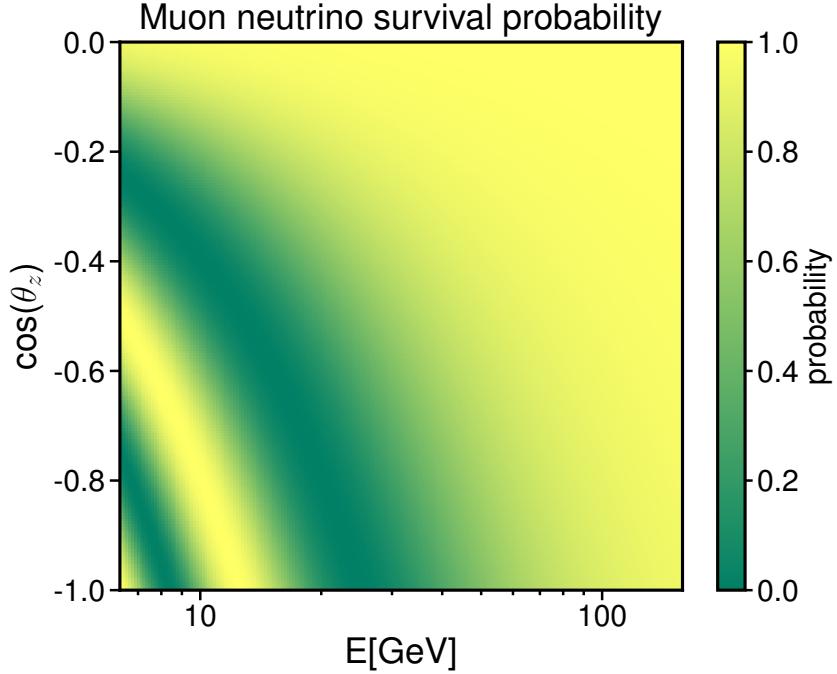


Figure 5.1: Muon neutrino survival probability in two-flavor approximation as in Equation 2.14 with $\Delta m_{32}^2 = 2.42 \cdot 10^{-3} \text{ eV}^2$ and $\sin^2 \theta_{23} = 0.538$.

by minimizing the logarithmic likelihood

$$-\log(\mathcal{L}) = \sum_{i=1}^N \mu_i(\theta, \phi) - n_i \log(\mu_i(\theta, \phi)) + \sum_k \frac{(\phi_k - \phi_k^0)^2}{2\sigma_{\phi_k}^2}, \quad (5.1)$$

where n_i is the number of observed data events and $\mu_i(\theta, \phi)$ is the expected number of events in the i^{th} bin depending on the physics parameters of interest θ and the nuisance parameters ϕ . The second term in Equation 5.1 accounts for prior knowledge on the nuisance parameters. It is the sum of Gaussian penalties, where ϕ_k^0 is the central value of the k^{th} nuisance parameter and σ_{ϕ_k} its Gaussian standard deviation. The priors are taken from model predictions or from independent measurements. As minimizer we use the Sequential Least SQuares Programming (SLSQP) method provided by the SciPy library [36].

The expected number of events of the bin content is calculated from simulated events using forward folding [10]. Forward folding means that the weight of each simulated event is modified to match the expected distribution under assumption of a set of parameters. The expected oscillated flux of a given neutrino flavor ν_α is calculated from the k^{th} simulated event as the composition of the atmospheric fluxes multiplied by the probabilities to oscillate into the neutrino flavor of interest. The flux is given by

$$\Phi_{\nu_\alpha}^{k,\text{osc}} = \Phi_{\nu_e}^k \cdot P_{\nu_e \rightarrow \nu_\alpha}^{\text{osc}} + \Phi_{\nu_\mu}^k \cdot P_{\nu_\mu \rightarrow \nu_\alpha}^{\text{osc}}, \quad (5.2)$$

where $\Phi_{\nu_e}^k$ and $\Phi_{\nu_\mu}^k$ are the atmospheric fluxes corresponding to the k^{th} event and $P_{\nu_e \rightarrow \nu_\alpha}^{\text{osc}}$ and $P_{\nu_\mu \rightarrow \nu_\alpha}^{\text{osc}}$ are the oscillation probabilities to the flavor ν_α . The probabilities are modified by the neutrino mixing parameters, while the atmospheric fluxes depend on the flux model and its governing parameters [12]. Multiplying the flux for a given neutrino flavor by the weight factor that incorporates interaction effects (detector response, cross-section etc.) yields the total weight for each event k and flavor α

$$\omega^{k,\alpha} = \omega_0^{k,\alpha} \cdot \Phi_{\nu_\alpha}^{k,\text{osc}}, \quad (5.3)$$

where $\omega_0^{k,\alpha}$ is the weight factor. This is the weight factor that was already mentioned in Section 4.2 used to re-weight the events to get a realistic distribution. The total weight $\omega^{k,\alpha}$ has units of events per second. Binning all simulated events, taking into account their total weight, and multiplying the summed weight by the detector livetime, yields the expected number of events $\mu_i(\theta)$.

5.1.2 Sensitivity Estimation

To estimate the uncertainty on the physics parameters, we use an Asimov dataset [37]. This means that the expected number of events and the number of observed data events in Equation 5.1 are both produced from the same simulation dataset. We choose a set of physics parameters $\hat{\theta}$ for the production of the data events. As a result, these injected parameters $\hat{\theta}$ should be perfectly recovered when performing the maximum likelihood fit. The shape of the uncertainty region is estimated by profiling the likelihood with respect to the physics parameters of interest θ . For any set of θ the likelihood value is calculated using Equation 5.1. With this likelihood we can construct the *test statistic* (TS) used in this work, which is defined as

$$-2\Delta \log \mathcal{L}(\theta) = -2 \left(\log \mathcal{L}(\theta) - \log \mathcal{L}(\hat{\theta}) \right), \quad (5.4)$$

where $\hat{\theta}$ are the injected parameters and θ are the set of parameters to be probed. Furthermore, we apply Wilk's theorem [38] by assuming that the TS is approximated by a χ^2 -distribution. For a χ^2 -distribution with two degrees of freedom the 68 % and 90 % confidence levels (C.L.) are found at values 2.28 and 4.61, respectively. Parameters with a TS exceeding these values can be rejected with the given C.L..

5.1.3 Systematic Uncertainties

The impact of systematic uncertainties is incorporated in the analysis by treating them as nuisance parameters implemented in Equation 5.1. They have an effect on the shape and normalization of the expected event distributions. The systematic uncertainties can be roughly categorized in the following groups depending on their origin: The initial, un-oscillated flux of atmospheric neutrinos, the neutrino-nucleon interaction cross-section, the neutrino flavor oscillation parameters and the detector response. The discussion of the impact of the systematic

uncertainties follows closely the description in [39].

Atmospheric neutrino flux: A detailed study of the uncertainties resulting from the calculation of the neutrino flux has been performed in [40]. Depending on their origin, they can depend on E , $\cos \theta_z$ and particle species. The uncertainties on the neutrino energy spectra resulting from uncertainties in modelling the cosmic ray energy spectra are purely E dependent. Uncertainties in the π - and k -meson production lead to uncertainties that are E , $\cos \theta_z$ and particle species dependent. The explicit parametrization of the atmospheric neutrino flux uncertainties is not discussed here but can be found in [39]. Additionally, there are uncertainties that are independent of E and $\cos \theta_z$ that effect the total scaling of the neutrino flux. Since these uncertainties are very large, they are not constrained but instead the total normalization of the neutrino flux is left as a free fit parameter.

Neutrino-nucleon interactions: From all the cross-section systematics that were tested, only two were not already degenerate with other systematic uncertainties. The two are the axial mass form factors for CC quasi-elastic and resonant scattering. They parameterize the quasi-elastic and resonant scattering cross-sections as described in [11]. Both of them are included as systematic uncertainties based on corrections calculated using GENIE [41], which is the tool used to produce the neutrino interaction simulation. Possible uncertainties in the hadronization processes of NC interactions are taken into account by adding a normalization factor for NC interactions.

Oscillation parameters: The oscillation model in this work uses the full, three-flavor formalism including three mixing angles, two mass-squared splittings and a CP violating phase (compare Equation 2.6). Since we are not sensitive to Δm_{21}^2 and θ_{12} nor the CP violating phase δ_{CP} , they are fixed. The angle θ_{13} is treated as a systematic uncertainty. The atmospheric mixing parameters Δm_{32}^2 and θ_{23} are free to fit. The neutrino mass ordering is not known and for this work only normal mass ordering is considered.

Additionally, there are uncertainties related to the detector response and the atmospheric muon flux. However, these systematics are not taken into account in this work. Two sensitivity studies for atmospheric neutrino mixing are performed in this work: one without taking into account any systematic uncertainties and one taking into account the systematic uncertainties previously discussed.

5.2 Analysis Binning

In this work we use the same E and $\cos \theta_z$ binning as in previous analyses [10]. The binning is the following: 11 logarithmically spaced bins in the E range of $[10^{0.8}, 10^{2.2}]$ GeV ($[6.31, 158.49]$ GeV)

and 10 linearly spaced bins in the $\cos \theta_z$ range of $[-1., 0.]$. The binning in the GBM probability must be optimized. This is done by performing a scan over varying GBM probability cut values and choosing the value which yields the best result in terms of sensitivities. The sensitivity studies are executed in the full three flavor model assuming vacuum oscillation. The oscillation parameters injected are taken from NuFit 2.0 [42] with maximum mixing $\theta_{23} = 45^\circ$ and 8.0 years of livetime assuming normal mass ordering (i.e $m_3 > m_2 > m_1$). The optimization is performed for the GBM probability as well as for the χ^2 -ratio. Figure 5.2 shows the results of the optimization processes.

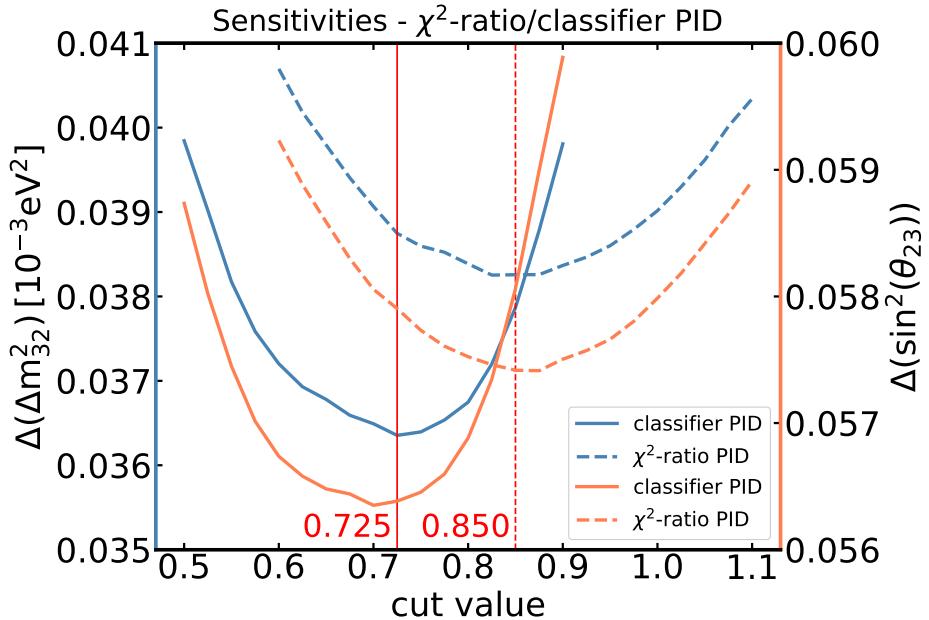


Figure 5.2: Sensitivity results of the two-bin PID cut optimization for the GBM probability (solid line) and the χ^2 -ratio (dashed line) with optimal cut value indicated in red. $\Delta(\Delta m_{32}^2)$ (blue) and $\Delta(\sin^2(\theta_{23}))$ (orange) are the sizes of the 68% C.L. for mixing angle and mass splitting, respectively.

For a binary binning, the optimal GBM probability cut is 0.725 and the optimal χ^2 -ratio cut is 0.85. As can be observed, the new PID achieves better sensitivities at the chosen cut value. In the following, these values will be used to split the sample in track bin and cascade bin to perform an oscillation measurement.

To understand why the chosen binning results in the best sensitivities, the track/cascade purity of the bins is shown in Figure 5.3. The true track fraction in the track bin and the true cascade fraction in the cascade bin are shown. It would be ideal if both were maximized, but a high purity of the track bin is more important since this is where we observe the primary disappearance signal. For the GBM probability, the track purity increases with the PID cut, while the cascade purity decreases slightly. Judging from this behavior it seems best to choose a large cut value which would lead to a very pure track sample. However, the left plot in Figure 5.4 shows that with the chosen GBM probability cut at 0.725 we maintain a reasonable event rate

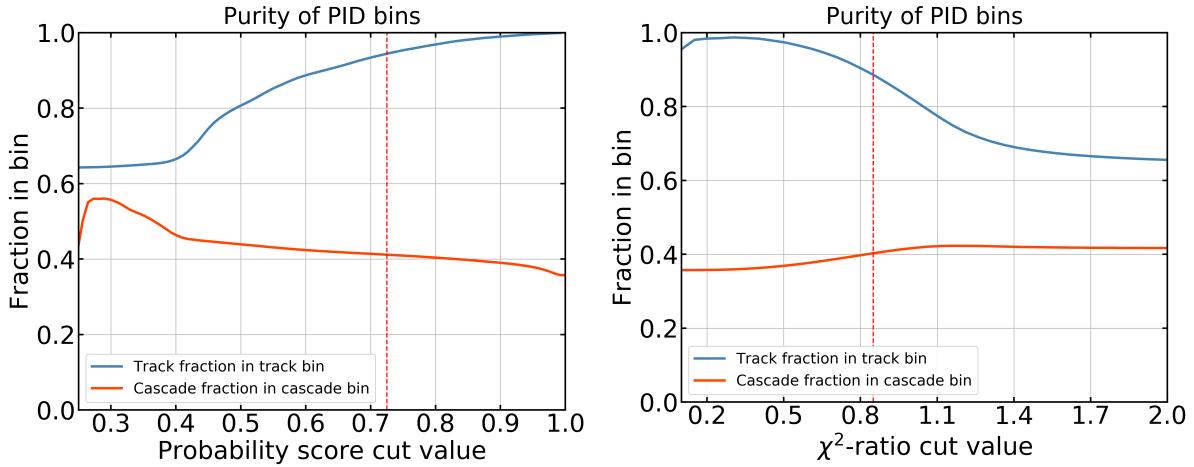


Figure 5.3: Purity of the PID binning depending on PID cut value for GBM probability (left) and χ^2 -ratio (right). The chosen binning is indicated as a red, dashed line.

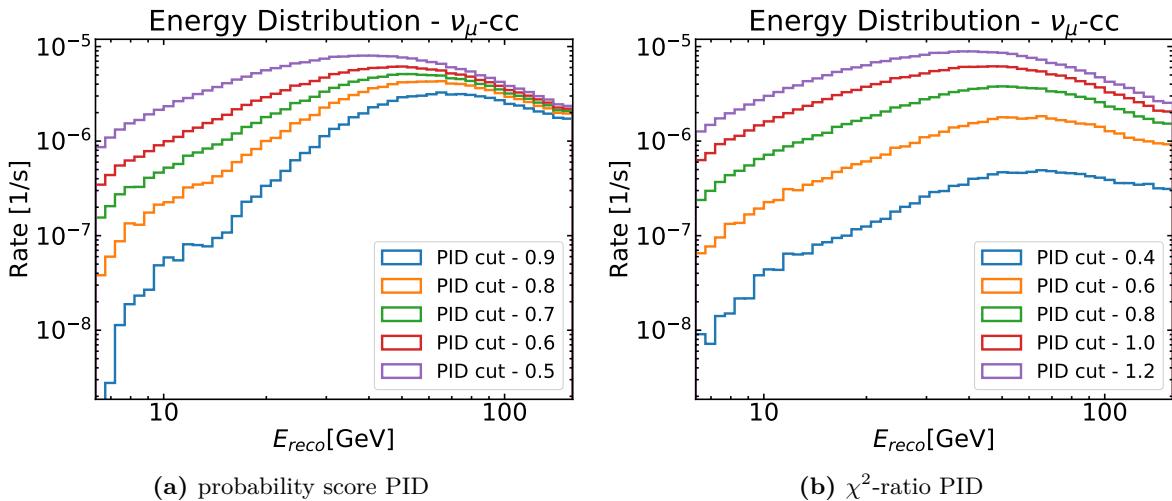


Figure 5.4: Energy distribution of ν_μ -CC events in the track bin cut on GBM probability (left) and on χ^2 -ratio (right).

in the energy range of 20-40 GeV, where the most dramatic oscillation effects are observed. For larger cut values the rate in this energy region decreases strongly. For the chosen cut value on the probability score, the resulting purities are 94.4% track purity (track bin) and 41.2% cascade purity (cascade bin).

The purity behavior for the cut on the χ^2 -ratio is reversed, due to the fact that low χ^2 -ratio values indicate tracks and high values indicate cascades. The cascade purity is almost constant while the track purity decreases with increasing values of χ^2 -ratio. This seems to indicate that a small PID cut value would be best, but comparing with the energy distributions for different cut values shown in the right plot of Figure 5.4, we see that below a cut value of ~ 0.8 the rate decreases strongly in the region of the oscillation signal. For the chosen cut of 0.85 on the

χ^2 -ratio PID we get 88.6 % track purity and 40.3 % cascade purity.

With these optimize cuts, we divide events into track-like or cascade-like samples. The expected E and $\cos \theta_z$ distributions for each sample are shown in Figure 5.5 and Figure 5.6 for the χ^2 -ratio PID and the GBM probability PID, respectively. The color scale indicates the ratio

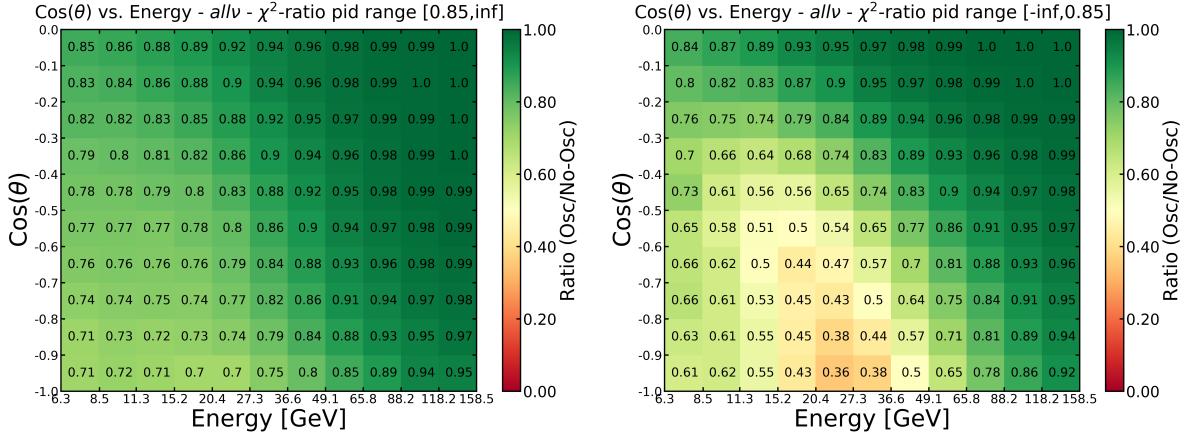


Figure 5.5: Expected neutrino oscillation effects for two-bin case with χ^2 -ratio PID. Shown is the ratio between un-oscillated and oscillated event rates for the cascade bin (left) and the track bin (right).

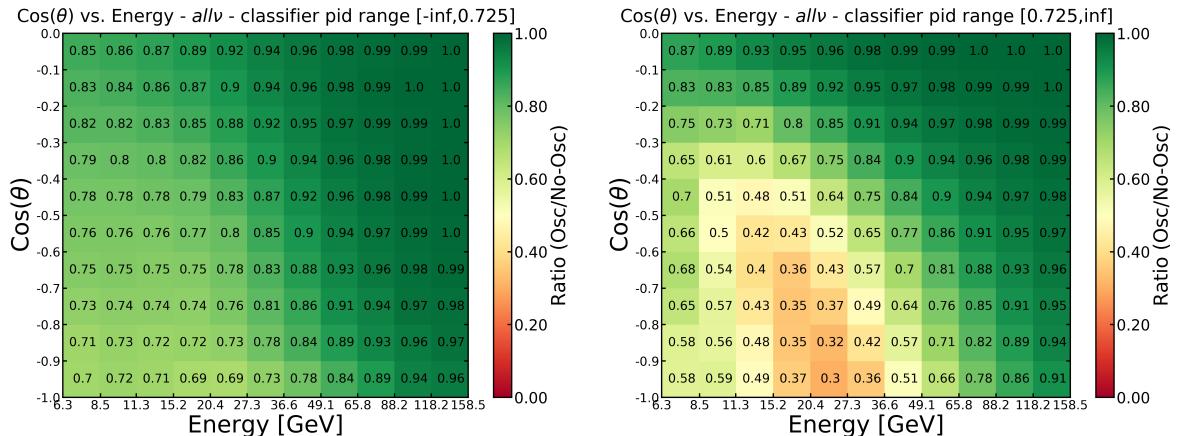


Figure 5.6: Expected neutrino oscillation effects for two-bin case with GBM probability. Shown is the ratio between un-oscillated and oscillated event rates for the cascade bin (left) and the track bin (right).

of the oscillated rate to the un-oscillated rate which shows the oscillation effect. The oscillation signal is observed in the region where we expect it comparing with Figure 5.1. We also observe that the oscillation signal is stronger and more localized using the GBM probability PID and extends further into the region of horizontal events. In Appendix B all the binned event rates are shown for the un-oscillated flux and the oscillated flux.

We also investigate the sensitivity that can be achieved using three PID bins. The performance of the individual three-bin studies is visualized in Figure 5.7, where small uncertainties are dark blue. There is a region where several cut combinations perform similarly good. The

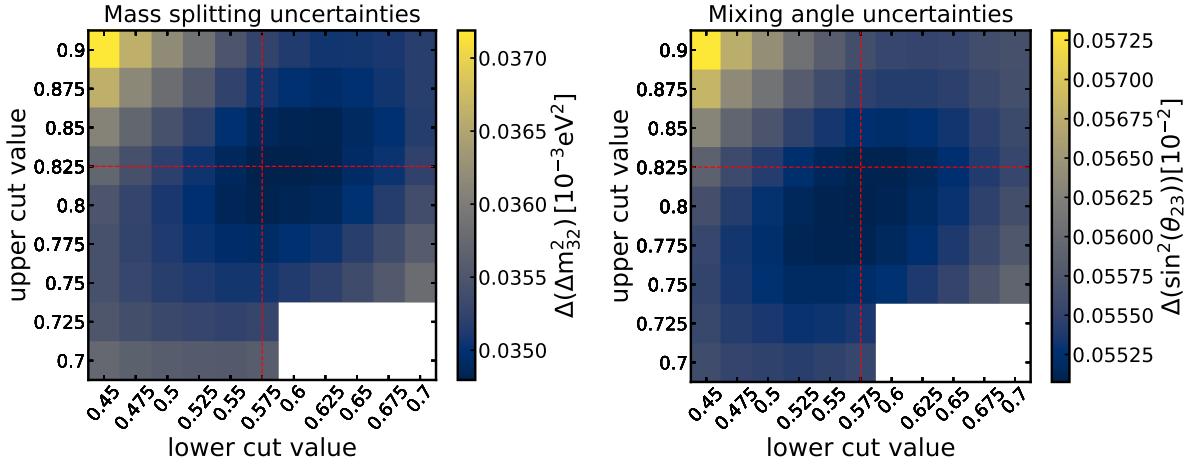


Figure 5.7: Sensitivity results of the three-bin PID cut optimization for the GBM probability. The optimal cut values are indicated as red, dashed lines. $\Delta(\Delta m_{32}^2)$ (left) and $\Delta(\sin^2(\theta_{23}))$ (right) are the sizes of the 68 % C.L. for mixing angle and mass splitting, respectively.

best result with three bins is found for cut values 0.575 and 0.825, resulting in 97.6 % track purity (track bin), 77.0 % track purity (intermediate bin) and 43.7 % cascade purity (cascade bin).

Similar to how it was done for the two-bin cases, we apply these optimized cuts to divide the events into a track-like, a cascade-like, and a mixed sample. We investigate the observed oscillation effects by taking the ratio of the oscillated rate to the un-oscillated rate as shown in Figure 5.8. In the intermediate bin we see a weak oscillation effect, while in the track bin the effect is much stronger than in Figure 5.5. Additionally, the strength of the signal is larger than in Figure 5.5 and Figure 5.6. The binned event rates for the un-oscillated flux and the oscillated flux are shown in Appendix B.

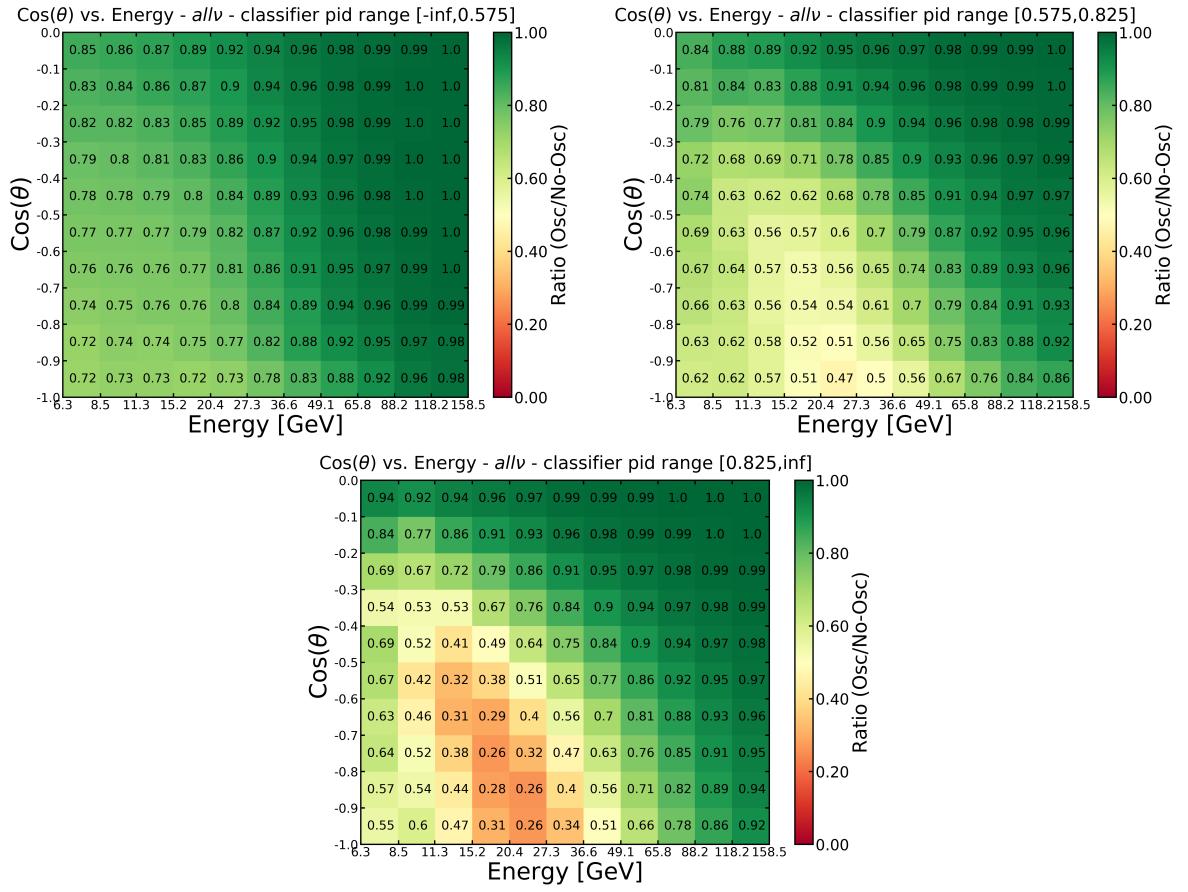


Figure 5.8: Expected neutrino oscillation effects for three-bin case with GBM probability. Shown is the ration between un-oscillated and oscillated event rates for the cascade bin (top left), the intermediate bin (top right) and the track bin (bottom center).

5.3 Sensitivity Improvement

The performance of the GBM probability used with two bins and with three bins will be compared to the χ^2 -ratio PID with two bins by calculating the Asimov sensitivities for each case. The same binning, livetime and mass ordering as in Section 5.2 is chosen, but this time the injected oscillation parameters are taken from a more recent fit [43] and the oscillation probabilities are calculated in the full three flavor model with matter effects using the algorithm from [44, 45] and PREM [46] as Earth matter profile. Two sensitivity studies to atmospheric neutrino mixing are performed; one without systematic uncertainties and one including systematic uncertainties. The values of the widths of the 68 % C.L. and the relative improvement compared to the study using the χ^2 -ratio PID are listed in Table 5.1. The result using two bins with the GBM probability shows marginal improvement in the sensitivity in the mass splitting. The version using three bins shows better performance for both the mass splitting and the mixing angle, reducing the uncertainty by 8.7 % and 3.1 %, respectively.

The same procedure is performed including systematic uncertainties and the results are

	$\Delta(\Delta m_{32}^2) [10^{-5} \text{eV}^2]$	improv. [%]	$\Delta(\sin^2(\theta_{23})) [10^{-2}]$	improv. [%]
χ^2 -ratio PID - 2 bins	3.67	NA	6.46	NA
GBM probability - 2 bins	3.49	4.9	6.43	0.6
GBM probability - 3 bins	3.35	8.7	6.26	3.1

Table 5.1: Sensitivity improvement results for the study without systematics. Improvement is the relative fraction compared to the sensitivity using χ^2 -ratio as PID. $\Delta(\Delta m_{32}^2)$ and $\Delta(\sin^2(\theta_{23}))$ are the sizes of the 68 % C.L. for mixing angle and mass splitting, respectively.

summarized in Table 5.2. The results show a similar trend. For the two-bin version there is significant improvement in both quantities, but it is much stronger in the mass splitting. Splitting the sample in three bins based on the GBM probability leads to a larger improvement in the sensitivities of the mass splitting and the mixing angle, but the effect is stronger on the mixing angle.

	$\Delta(\Delta m_{32}^2) [10^{-5} \text{eV}^2]$	improv. [%]	$\Delta(\sin^2(\theta_{23})) [10^{-2}]$	improv. [%]
χ^2 -ratio PID - 2 bins	4.97	NA	7.65	NA
GBM probability - 2 bins	4.50	9.5	7.41	3.1
GBM probability - 3 bins	4.32	13.0	7.10	7.2

Table 5.2: Sensitivity improvement results for the study including systematics. Improvement is the relative fraction compared to the sensitivity using χ^2 -ratio as PID. $\Delta(\Delta m_{32}^2)$ and $\Delta(\sin^2(\theta_{23}))$ are the sizes of the 68 % C.L. for mixing angle and mass splitting, respectively.

The best results are achieved using the GBM probability PID with three bins which yields $\Delta(\Delta m_{32}^2) = 4.32 \cdot 10^{-5} \text{eV}^2$ and $\Delta(\sin^2(\theta_{23})) = 7.10 \cdot 10^{-2}$. This is a 13.0 % improvement in $\Delta(\Delta m_{32}^2)$ and a 7.2 % improvement in $\Delta(\sin^2(\theta_{23})) = 7.10 \cdot 10^{-2}$ relative to the uncertainties estimated using the χ^2 -ratio PID. The improvement that can be achieved by applying the new PID variable demonstrates the strength of the multivariate classification algorithm. As compared to the formerly used univariate discriminator, the new discriminator maximizes the use of information gained from the set of input variables.

Chapter 6

Summary and Conclusions

Discriminating tracks caused by ν_μ charged-current interactions from cascades caused by both neutral-current interactions of all flavors and charged-current interactions of electron and tau neutrinos is essential to perform atmospheric neutrino oscillation measurements in IceCube as well as other experiments. Within this thesis, we develop a new method to distinguish tracks from cascades in IceCube DeepCore. The method utilizes a multivariate machine learning algorithm to maximize the information gained from the detector response.

Interesting reconstructed variables are investigated to examine how useful they are in distinguishing tracks from cascades. The Gradient Tree Boosting algorithm is chosen to enhance the separation in tracks and cascades in comparison to a previously used univariate method. The algorithm is trained on selected reconstructed variables. The process of tuning the hyperparameters of the algorithm is tested but is found to be not feasible due to insufficient amount of simulation statistics. The performance of the trained classifier is estimated resulting in an AUC of 0.662, whereas the formerly used PID variable χ^2 -ratio only reaches an AUC of 0.608.

As a test case, the impact of the new PID method on the sensitivity to the atmospheric neutrino oscillation parameters Δm_{32}^2 and θ_{23} is considered. For this purpose, the output probability score of the classifier is utilized as a new PID variable called GBM probability. The PID binning in the GBM probability is optimized for the two bin case and the three bin case by choosing the PID cut that leads to the largest reduction in the uncertainty of the atmospheric neutrino mixing parameters calculated relative to the uncertainties estimated using the χ^2 -ratio. Making use of the GBM probability and the optimized binning leads to significant improvement in the sensitivity to atmospheric neutrino oscillation parameters. The optimal configuration includes three PID bins and for an analysis including all systematic uncertainties apart from the detector systematics, the size of the 68 % confidence region is reduced by 13 % for Δm_{32}^2 and by 7.2 % for $\Delta(\sin^2(\theta_{23}))$. The resulting uncertainties are $\Delta(\Delta m_{32}^2) = 4.32 \cdot 10^{-5} \text{ eV}^2$ and $\Delta(\sin^2(\theta_{23})) = 7.10 \cdot 10^{-2}$.

The developed method can be extended to incorporate additional features from other reconstruction algorithms as well as low-level observables from the detector response. The GBM

CHAPTER 6. SUMMARY AND CONCLUSIONS

probability will be used as PID discriminator in the upcoming 8-year IceCube DeepCore oscillation analysis. The relative improvement achieved with this multivariate method over a univariate classifier is expected to increase further with the inclusion of detector systematic uncertainties.

Bibliography

- [1] W. Pauli, Phys. Today **31N9**, 27 (1978).
- [2] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, Science **124**, 103 (1956).
- [3] G. Danby, J.-M. Gaillard, K. Goulian, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger, Phys. Rev. Lett. **9**, 36 (1962).
- [4] K. Kodama *et al.*, Physics Letters B **504**, 218 (2001).
- [5] R. Davis *et al.*, *Proceedings of the Neutrino '72 Europhysics Conference*, Vol. 1 (Balatonfured, Hungary, 1972.) p. 29.
- [6] M. G. Aartsen *et al.* (IceCube Collaboration), Journal of Instrumentation **12**, P03012 (2017), arXiv:1612.05093 [astro-ph.IM].
- [7] M. Thomson, *Modern particle physics* (Cambridge University Press, Cambridge [u.a.], 2013) pp. XVI, 554 S.
- [8] S. L. Glashow, Nuclear Physics **22**, 579 (1961).
- [9] M. Tanabashi *et al.* (Particle Data Group), Phys. Rev. D **98**, 030001 (2018).
- [10] A. Terliuk, *Measurement of atmospheric neutrino oscillations and search for sterile neutrino mixing with IceCube DeepCore*, Ph.D. thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, Berlin, Germany (2018).
- [11] J. A. Formaggio and G. P. Zeller, Rev. Mod. Phys. **84**, 1307 (2012).
- [12] M. Honda, M. S. Athar, T. Kajita, K. Kasahara, and S. Midorikawa, Phys. Rev. D **92**, 023004 (2015).
- [13] A. Fedynitch, R. Engel, T. K. Gaisser, F. Riehn, and T. Stanev, in European Physical Journal Web of Conferences, European Physical Journal Web of Conferences, Vol. 99 (2015) p. 08001, arXiv:1503.00544 [hep-ph].
- [14] M. Honda, T. Kajita, K. Kasahara, S. Midorikawa, and T. Sanuki, Phys. Rev. D **75**, 043006 (2007).
- [15] S. Bilenky and B. Pontecorvo, Physics Reports **41**, 225 (1978).
- [16] P. A. M. Dirac, Proceedings of the Royal Society of London Series A **114**, 243 (1927).

BIBLIOGRAPHY

- [17] R. Abbasi *et al.* (IceCube Collaboration), *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **601**, 294 (2009).
- [18] R. Abbasi *et al.* (IceCube Collaboration), *Astropart. Phys.* **35**, 615 (2012), arXiv:1109.6096 [astro-ph.IM] .
- [19] P. A. Cherenkov, *Phys. Rev.* **52**, 378 (1937).
- [20] S. Euler, *Observation of oscillations of atmospheric neutrinos with the IceCube Neutrino Observatory*, Ph.D. thesis, Rheinisch-Westfälischen Technischen Hochschule, Aachen, Germany (2014).
- [21] I. Frank and I. Tamm, *C. R. Acad. Sci. USSR* **14**, 109 (1937).
- [22] L. Rädel and C. Wiebusch, *Astroparticle Physics* **38**, 53 (2012), arXiv:1206.5530 [astro-ph.IM] .
- [23] D. Chirkin and W. Rhode, (2004), arXiv:hep-ph/0407075 [hep-ph] .
- [24] L. Raedel, *Simulation Studies of the Cherenkov Light Yield from Relativistic Particles in High-Energy Neutrino Telescopes with Geant4*, Master's thesis, Rheinisch-Westfälischen Technischen Hochschule, Aachen, Germany (2012).
- [25] J. Lundberg, P. Miocinovic, T. Burgess, J. Adams, S. Hundertmark, P. Desiati, K. Woschnagg, and P. Niessen, *Nucl. Instrum. Meth.* **A581**, 619 (2007), arXiv:astro-ph/0702108 [ASTRO-PH] .
- [26] J. P. Y. Garza, *Measurement of neutrino oscillations in atmospheric neutrinos with the IceCube DeepCore detector*, Ph.D. thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I, Berlin, Germany (2014).
- [27] J. A. Aguilar *et al.*, *Astroparticle Physics* **34**, 652 (2011), arXiv:1105.4116 [astro-ph.IM] .
- [28] A. Terliuk, *Energy reconstruction of ν_μ events in DeepCore*, Master's thesis, Taras Shevchenko National University of Kyiv, Ukraine (2014), defended in Ukrainian, english version available from andrey.terlyuk@gmail.com.
- [29] R. Nahnhauer, A. Terliuk, and J. P. Yanez, IceCube Internal Report **201304001-v2** (2013), available upon request from members of IceCube Collaboration.
- [30] D. E. Groom, N. V. Mokhov, and S. I. Striganov, *Atomic Data and Nuclear Data Tables* **78**, 183 (2001).
- [31] IceCube Collaboration, M. G. Aartsen, *et al.*, (2017), arXiv:1702.05160 [hep-ex] .
- [32] J. H. Friedman, *Ann. Statist.* **29**, 1189 (2001).
- [33] F. Pedregosa *et al.*, *Journal of Machine Learning Research* **12**, 2825 (2011).
- [34] S. M. Stigler, *Statist. Sci.* **4**, 73 (1989).
- [35] A. Kolmogorov, *Inst. Ital. Attuari, Giorn.* **4**, 83 (1933).
- [36] P. Virtanen *et al.*, (2019), arXiv:1907.10121 [cs.MS] .

- [37] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *European Physical Journal C* **71**, 1554 (2011), [arXiv:1007.1727 \[physics.data-an\]](https://arxiv.org/abs/1007.1727).
- [38] S. S. Wilks, *Ann. Math. Statist.* **9**, 60 (1938).
- [39] M. G. Aartsen *et al.* (IceCube Collaboration 1), *Phys. Rev. D* **99**, 032007 (2019).
- [40] G. D. Barr, S. Robbins, T. K. Gaisser, and T. Stanev, *Phys. Rev. D* **74**, 094009 (2006).
- [41] C. Andreopoulos, C. Barry, S. Dytman, H. Gallagher, T. Golan, R. Hatcher, G. Perdue, and J. Yarba, (2015), [arXiv:1510.05494 \[hep-ph\]](https://arxiv.org/abs/1510.05494).
- [42] M. C. Gonzalez-Garcia, M. Maltoni, and T. Schwetz, *Journal of High Energy Physics* **2014**, 52 (2014).
- [43] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, I. Martinez-Soler, and T. Schwetz, *Journal of High Energy Physics* **2017**, 87 (2017).
- [44] V. Barger, K. Whisnant, S. Pakvasa, and R. J. N. Phillips, *Phys. Rev. D* **22**, 2718 (1980).
- [45] R. Wendell, (2009), <http://www.phy.duke.edu/~raw22/public/Prob3++/>.
- [46] A. M. Dziewonski and D. L. Anderson, *Physics of the Earth and Planetary Interiors* **25**, 297 (1981).

Appendices

Appendix A

Data/MC Checks

This section shows the plots made to check whether the distributions of the input variables agree between data and simulations.

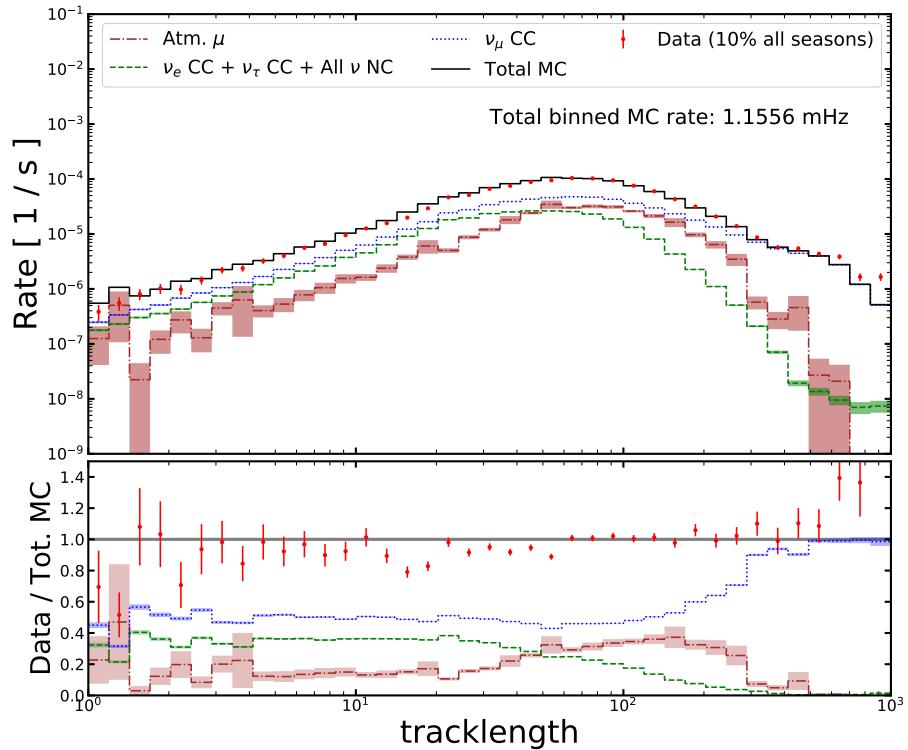


Figure A.1: The simulated distributions of track length compared to 10 % of the OscNext data sample (red points). Shown are the contributions from atmospheric muons (dark red), track-like events (blue), cascade-like events (green) and the combined simulation in black.

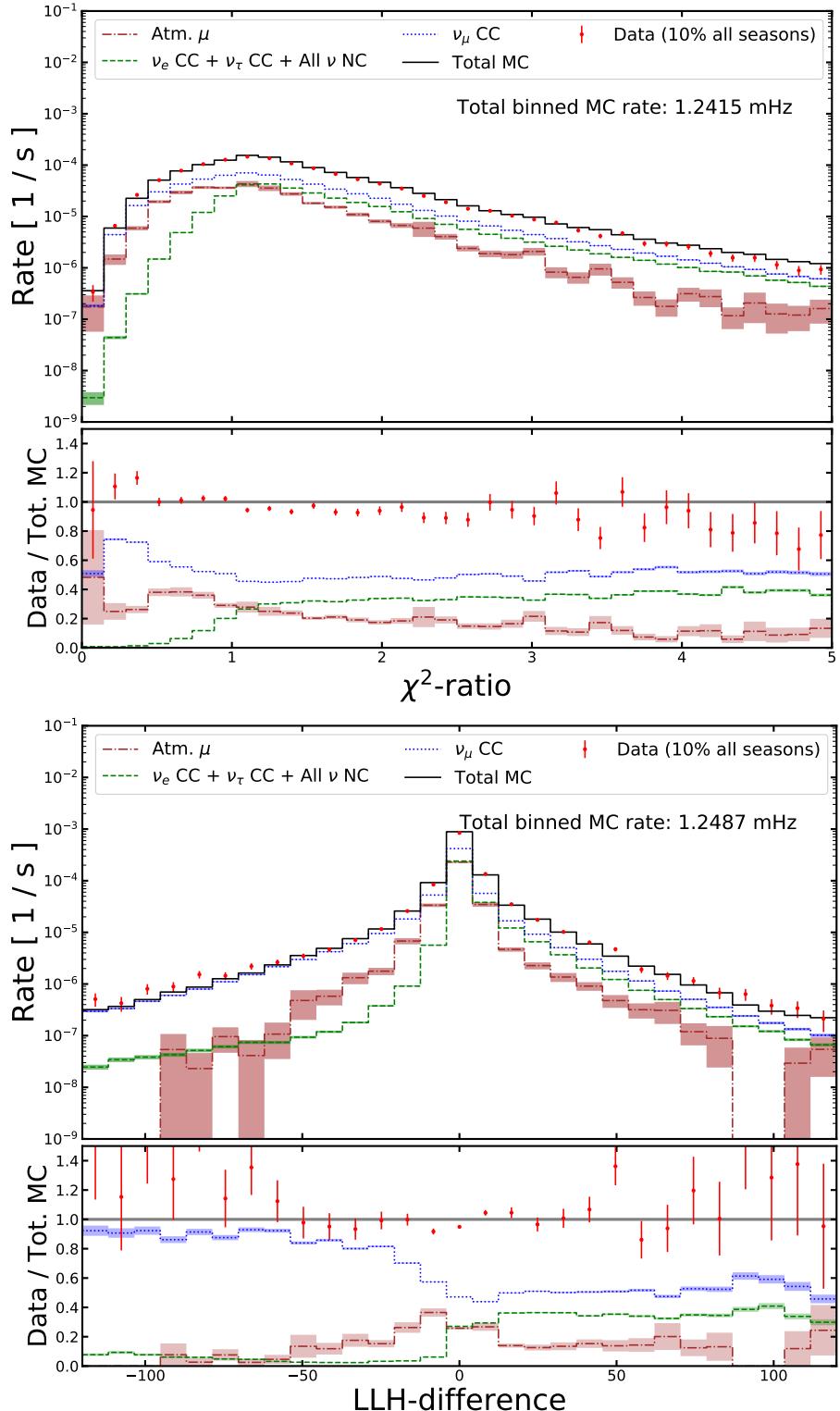


Figure A.2: The simulated distributions of χ^2 -ratio (left) and LLH-difference (right) compared to 10 % of the OscNext data sample (red points). Shown are the contributions from atmospheric muons (dark red), track-like events (blue), cascade-like events (green) and the combined simulation in black.

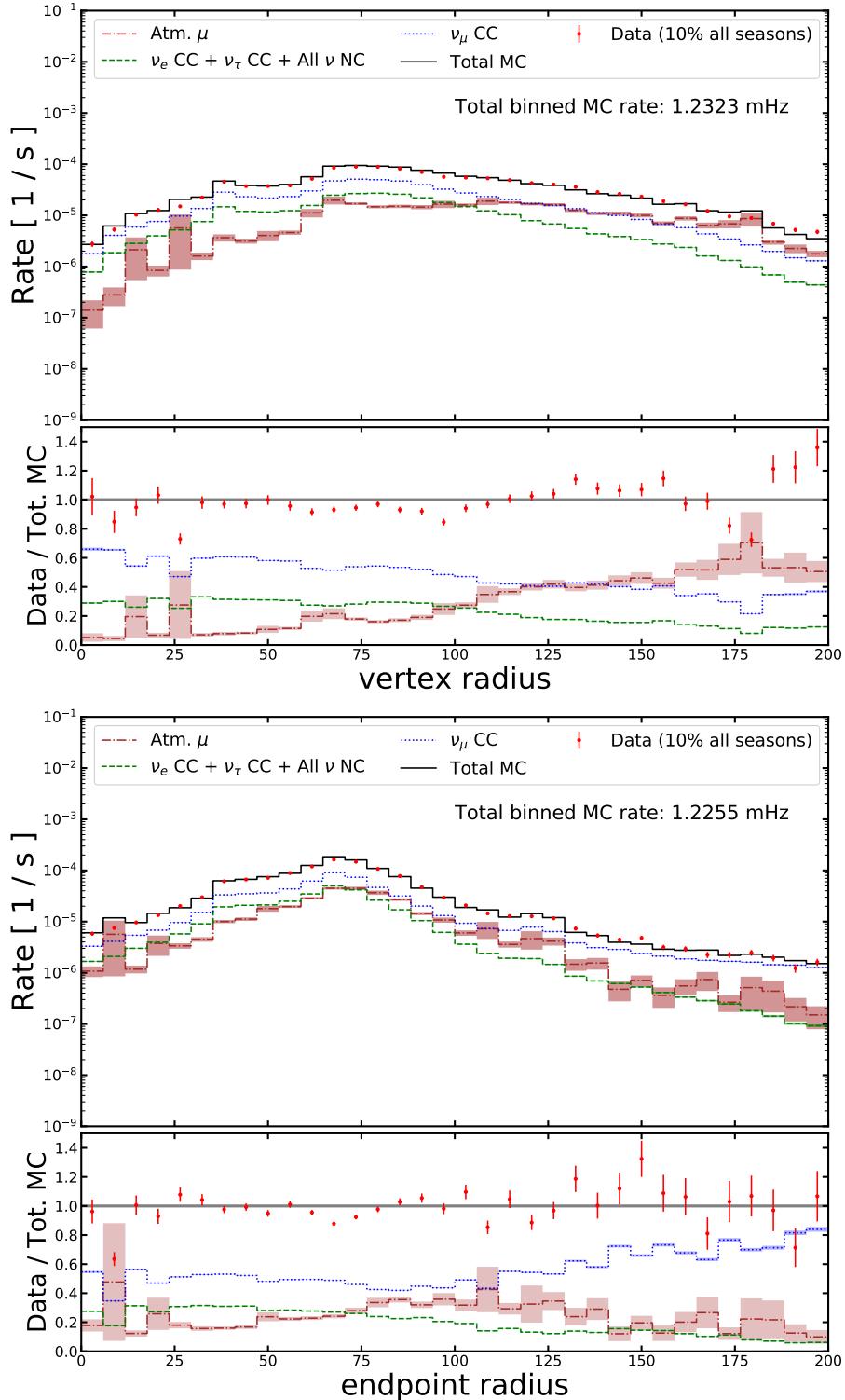


Figure A.3: The simulated distributions of vertex radius (left) and endpoint radius (right) compared to 10 % of the OscNext data sample (red points). Shown are the contributions from atmospheric muons (dark red), track-like events (blue), cascade-like events (green) and the combined simulation in black.

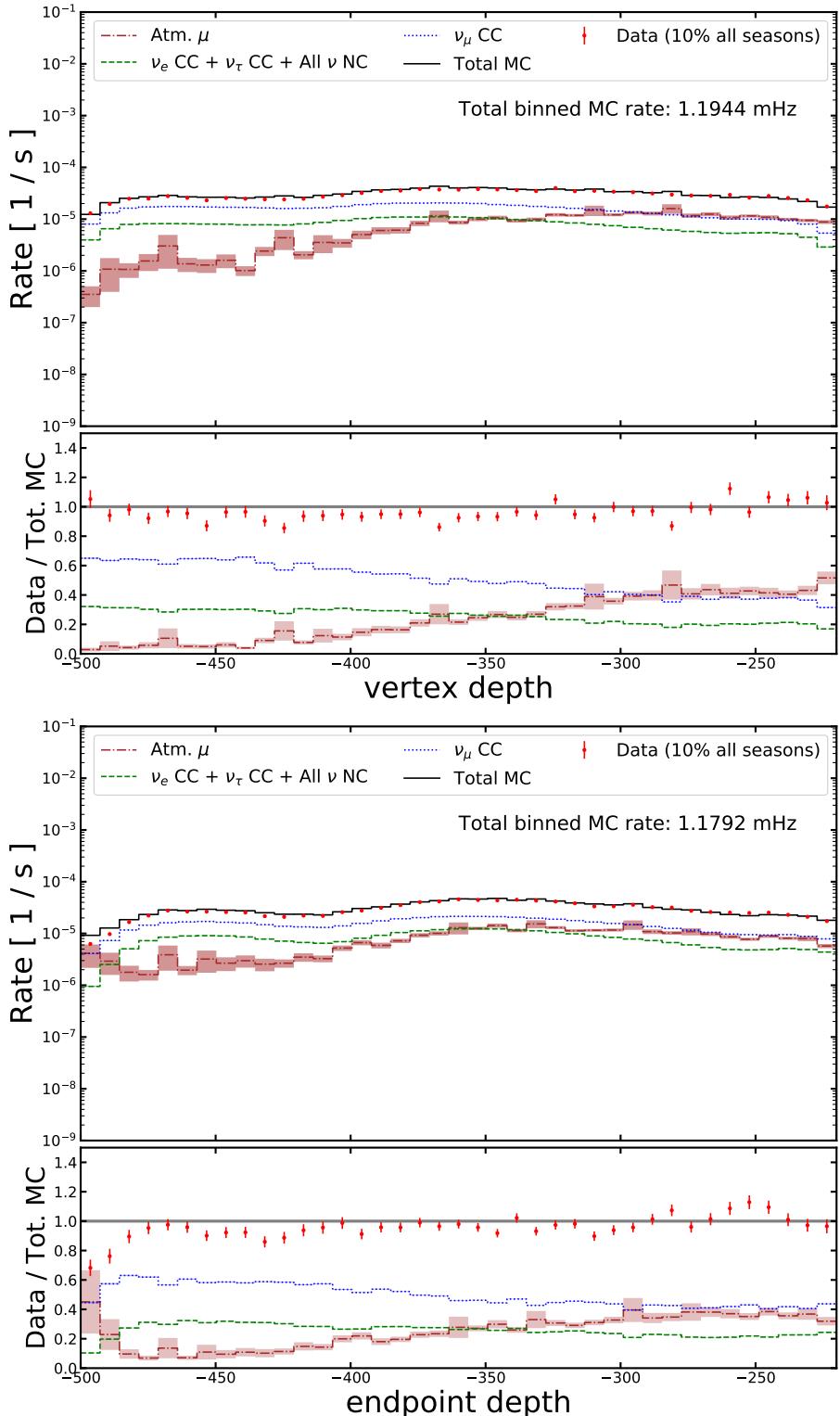


Figure A.4: The simulated distributions of vertex depth (left) and endpoint depth (right) compared to 10 % of the OscNext data sample (red points). Shown are the contributions from atmospheric muons (dark red), track-like events (blue), cascade-like events (green) and the combined simulation in black.

Appendix B

Expected Event Distributions

This section shows the expected histograms with the event rates with and without oscillations for the different PID bins.

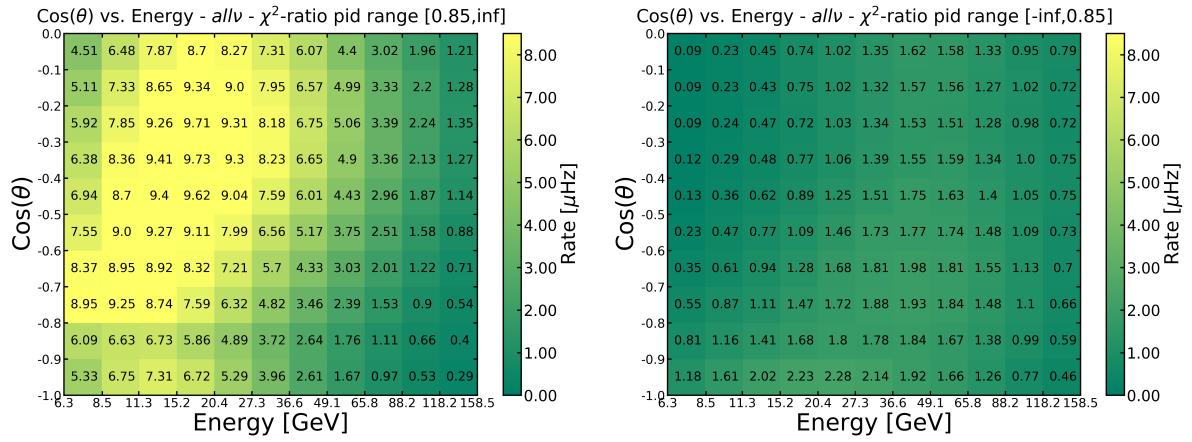


Figure B.1: Expected unsoscillated event rates for two-bin case with χ^2 -ratio PID.

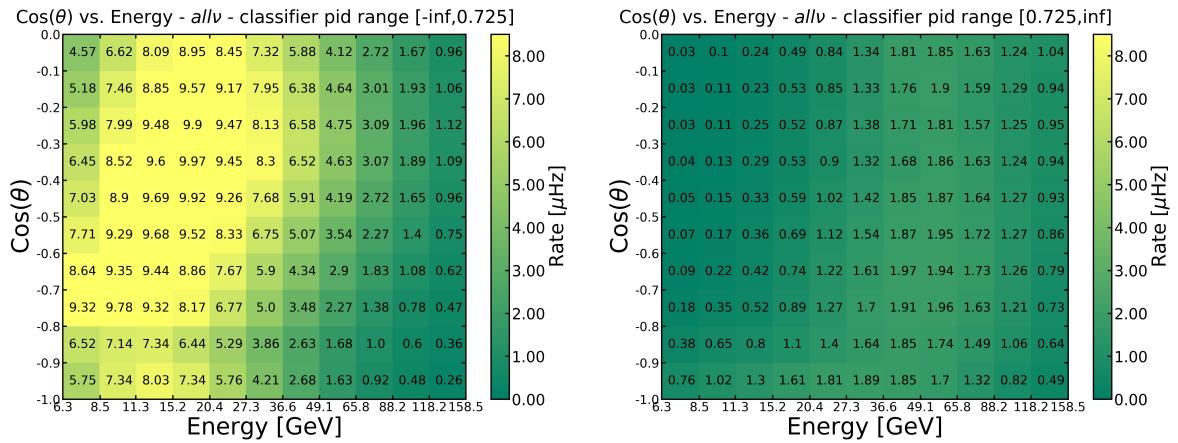


Figure B.2: Expected unsoscillated event rates for two-bin case with classifier PID.

APPENDIX B. EXPECTED EVENT DISTRIBUTIONS

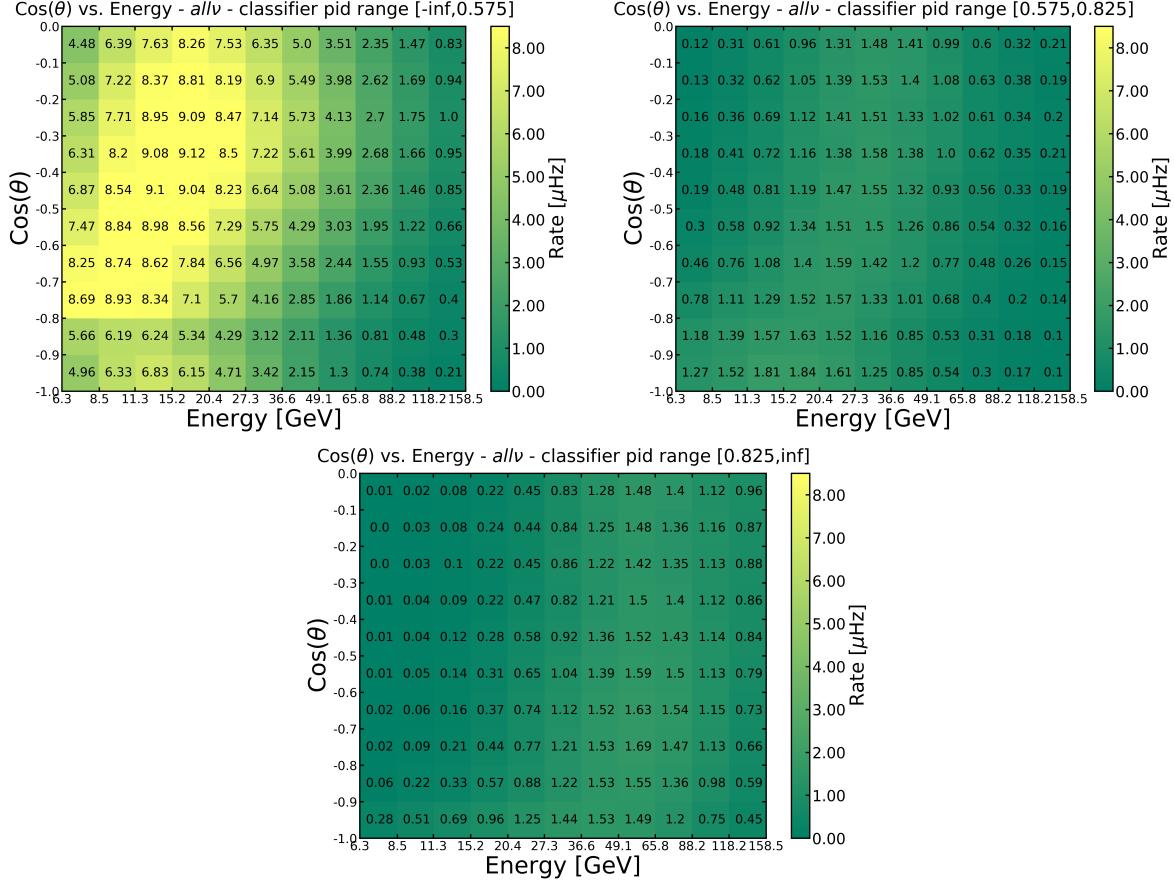


Figure B.3: Expected unsoscillated event rates for three-bin case with classifier PID.

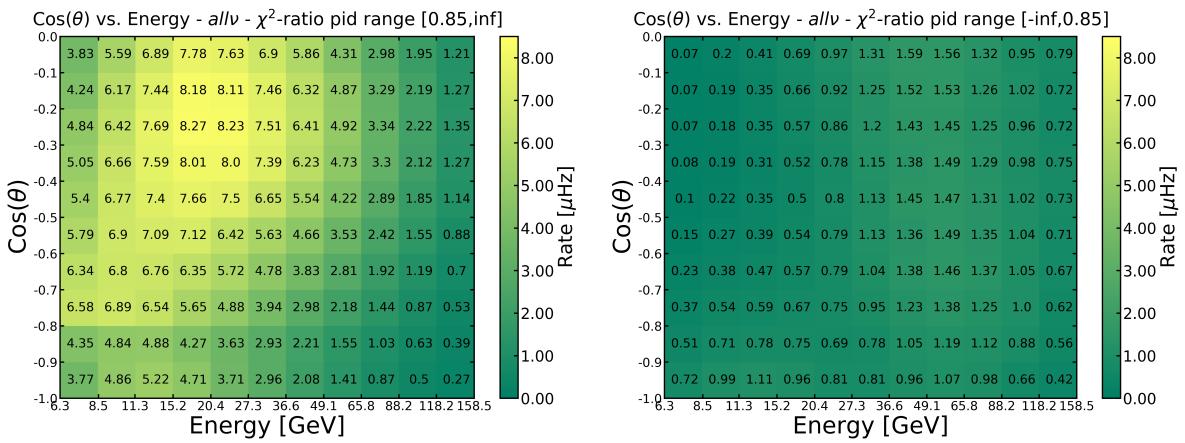


Figure B.4: Expected oscillated event rates for two-bin case with χ^2 -ratio PID.

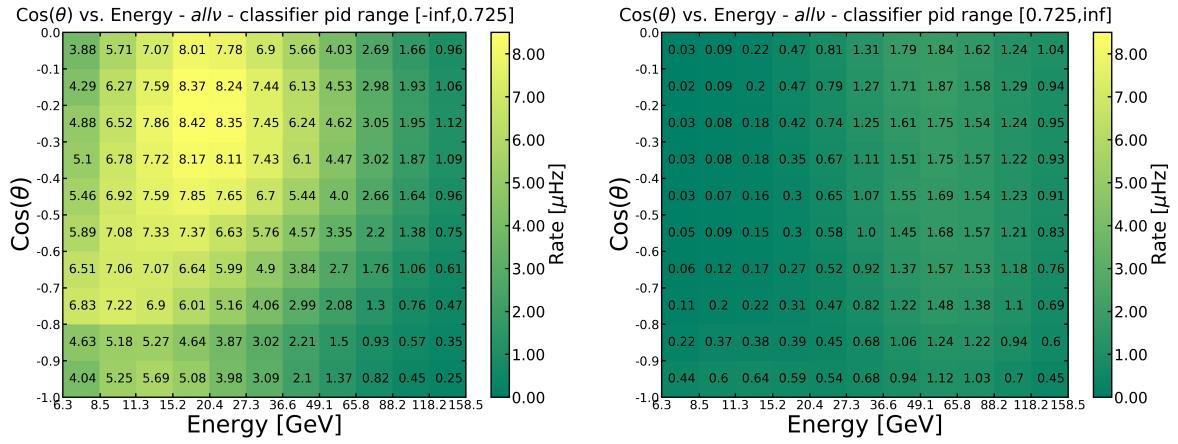


Figure B.5: Expected oscillated event rates for two-bin case with classifier PID.

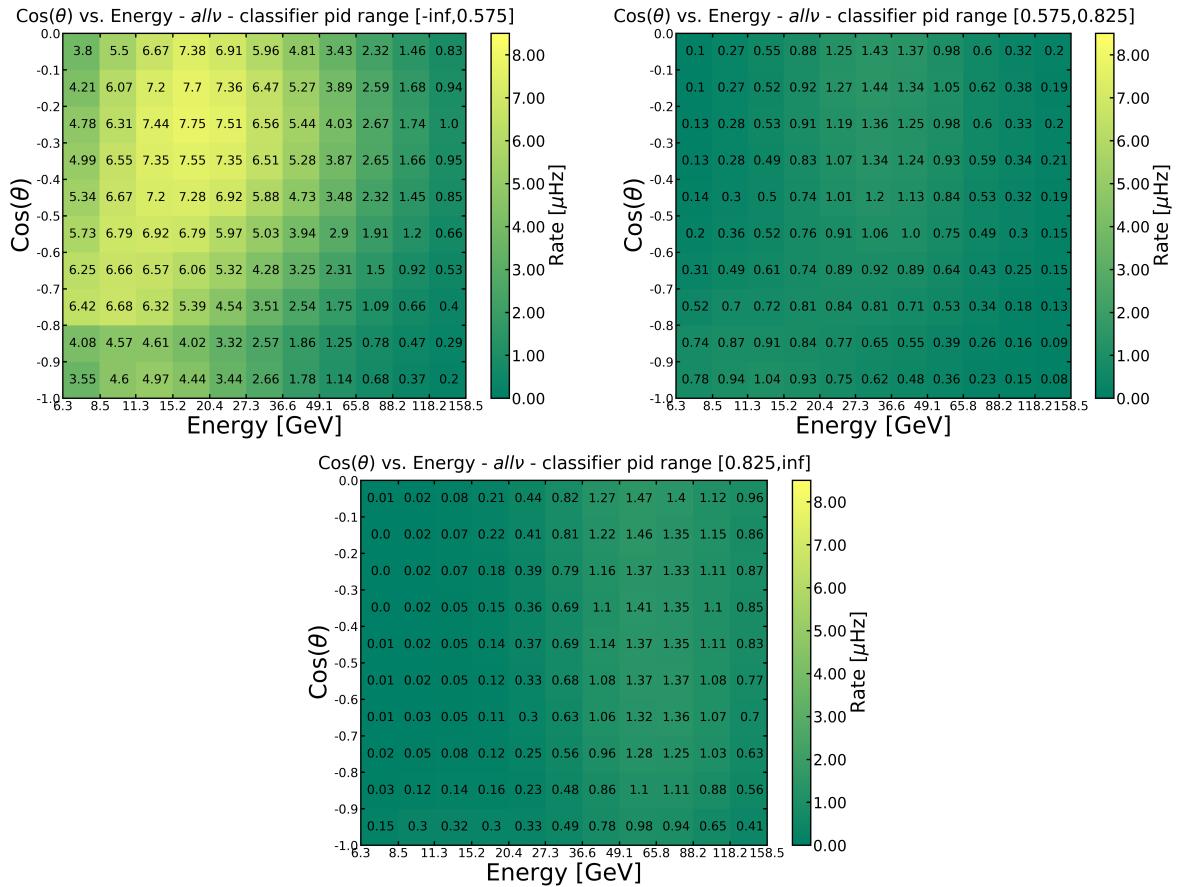


Figure B.6: Expected oscillated event rates for three-bin case with classifier PID.

List of Figures

2.1	Feynman diagrams of neutrino weak interactions, taken from [10]	6
2.2	Total inclusive neutrino-nucleon cross-sections, taken from [11].	6
2.3	Neutrino-nucleon deep inelastic scattering, taken from [10]	8
2.4	Atmospheric neutrino fluxes, taken from [12]	9
2.5	Vacuum oscillation probabilities in the two-neutrino approximation	11
3.1	IceCube Neutrino Observatory and Digital Optical Module	13
3.2	IceCube top view	14
3.3	Cherenkov light front	16
3.4	Hyperbolic time-depth pattern of muon passing a string, taken from [26]	20
3.5	Track endpoint reconstruction, taken from [29]	21
4.1	Examples of low energy track and cascade events	25
4.2	Distributions of the selected input features	26
4.3	Example decision tree	28
4.4	Pearson correlation of the selected input features	29
4.5	Classifier output probability of final, trained classifier	31
4.6	Receiver operating characteristic and feature importances for the final, trained classifier	32
4.7	Expected PID score distribution and the cumulative distribution for the different interaction types	33
4.8	Data/MC agreement for classifier output probability	34
4.9	Classifier output probability with tuned parameters	36
5.1	Muon neutrino survival probability in two-flavor approximation	38
5.2	Sensitivity results of the two-bin PID cut optimization	41
5.3	Purity of the PID binning depending on PID cut value	42
5.4	Energy distribution of ν_μ -CC events in the track bin	42
5.5	Expected neutrino oscillation effects for two-bin case with χ^2 -ratio PID	43
5.6	Expected neutrino oscillation effects for two-bin case with GBM probability	43
5.7	Sensitivity results of the three-bin PID cut optimization	44

LIST OF FIGURES

5.8	Expected neutrino oscillation effects for three-bin case with GBM probability	45
A.1	The simulated distribution of track length compared to 10 % of the OscNext data sample	VII
A.2	The simulated distributions of χ^2 -ratio (left) and LLH-difference (right) compared to 10 % of the OscNext data sample	VIII
A.3	The simulated distributions of vertex radius (left) and endpoint radius (right) compared to 10 % of the OscNext data sample	IX
A.4	The simulated distributions of vertex depth (left) and endpoint depth (right) compared to 10 % of the OscNext data sample	X
B.1	Expected unoscillated event rates for two-bin case with χ^2 -ratio PID	XI
B.2	Expected unoscillated event rates for two-bin case with classifier PID	XI
B.3	Expected unoscillated event rates for three-bin case with classifier PID	XII
B.4	Expected oscillated event rates for two-bin case with χ^2 -ratio PID	XII
B.5	Expected oscillated event rates for two-bin case with classifier PID	XIII
B.6	Expected oscillated event rates for three-bin case with classifier PID	XIII

List of Tables

3.1	Event signatures in IceCube and their underlying interactions, taken from [10]	23
4.1	Classifier hyper-parameters	36
5.1	Sensitivity improvement results for the study without systematics	46
5.2	Sensitivity improvement results for the study including systematics	46

Acknowledgements

I would also like to express my gratitude to Prof. Dr. Ulrich Uwer for offering to take responsibility for my work as my primary supervisor at Heidelberg University and for offering his time to examine it. I would like to thank Marek Kowalski for his support as a secondary supervisor and for making it possible for me to work on such an inspiring research topic as part of the IceCube group at DESY.

I am especially grateful to Summer Blot for the guidance in finding the topic of my thesis and for the strong support throughout the whole project. Without her experience and her constant feedback the work contained in this thesis would not have been possible.

I would like to thank Jakob van Santen and Alexander Trettin for proof-reading parts of my thesis, and Summer Blot for reading the entire thesis and providing valuable feedback.

I would like to thank my family: my parents Christine and Wolfgang and my brother Simon for supporting me throughout writing this thesis and special thanks to my dad for proof-reading the final version of my thesis. Last but not least i would like to thank Lotti for always being there for me.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den (Datum)