First Search for Heavy Neutral Leptons with IceCube DeepCore

Dissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)

im Fach: Physik
Spezialisierung: Experimentalphysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Leander Fischer M. Sc.
geboren am 24. Oktober 1992
in Heidelberg

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr. Julia von Blumenthal

Dekanin der Mathematisch-Naturwissenschaftlichen Fakultät
Prof. Dr. Caren Tischendorf

Colophon
This document was typeset with the help of KOMA-Script and LaTeX using the open-source kaobook template class.

The source code of this thesis is available at:
https://github.com/LeanderFischer/phd_thesis

Zusammenfassung

Zusammenfassung . . .

Abstract

Abstract . . .

Todo list

Contents

List of Figures

List of Tables

Standard Model Background Simulation and Data Processing

1

Adapt chapter to reflect switched chapter order (RED)

SB: Have you considered instead to have a chapter on all Simulation (signal and background) and then another chapter on Data processing and reconstruction? The model independent simulation description can either go into the simluation chapter OR it can go into the double cascade reconstruction chapter at an appropriate location/subsection (ORANGE)

The analysis presented in this thesis is highly dependent on an efficient event selection to reduce the raw IceCube trigger data to a usable atmospheric neutrino sample. Based on this selection, a precise estimation of both expected SM background and expected BSM signal events can be made using MC simulations. This chapter describes the current simulation and event selection chain used for state-of-the-art IceCube neutrino oscillation measurements like [1]. The whole chain can be broadly split into 4 steps:

Step 1 Event Generation: The initial step for all particle (non-noise) simulation is the generation of events from selected initial distributions and fluxes. Events are the primary particle and the particles produced in the interaction with the ice.

Step 2 Detector Simulation: The particles from the first step are propagated through the ice, producing Cherenkov photons, which are then propagated further until they reach a DOM or are absorbed. If they hit a DOM the detector response (acceptance and PMT) is simulated.

Step 3 Processing: Starting from the PMT output, both real data and simulation are processed through the in-ice trigger, the online filter and processing, and the low energy event selection to produce a neutrino dominated sample.

Step 4 Reconstruction: Once the sample is small enough for more sophisticated reconstruction techniques to be feasible to run, the events are reconstructed using a CNN and some high level variables are computed. Based on these variables the final event selection is applied.

This chapter only describes the event generation for the SM background simulation (neutrinos and muons), while the signal simulation is described in Chapter **??**. The detector simulation is identical for both signal and background events while processing and reconstruction are applied to all simulation and data in the same way. Splitting the simulation steps has the advantage of reusing the outputs of for example the generation step to propagate the particles with different ice model, in order to estimate the systematic impacts of uncertainties of the ice properties. Similar approach can be taken for varying detector response and through this a more efficient (reduced) use of computing resources can be achieved. The following sections describe the different steps in more detail and the last section, Section 1.5, describes the related systematic uncertainties considered for this work.

1.1 Event Generation

The MC is used in the analysis by applying a method called *forward folding*, where a very large number of events (signal and background) is produced using sampling distribution that are tuned to have a large selection efficiency. Those distributions don't have to be physically correct distributions, but they need to cover the full parameter space of interest for the analysis. To produce a physical distribution, the events are weighted given a specific

[1]: Abbasi et al. (2023), "Measurement of atmospheric neutrino mixing with improved IceCube DeepCore calibration and data processing".

Table 1.1: Cylinder volumes used for GENIE neutrino simulation generation. Cylinder is always centered in DeepCore at $(x, y, z) = (46.29, -34.88, -330.00)$ m.

Flavor	**Energy** [GeV]	**Radius** [m]	**Length** [m]	**Events/File**	**Files**
$\nu_e + \bar{\nu_e}$	1-4	250	500	450000	650
	4-12			100000	
	12-100	350	600		
	100-10000	550	1000	57500	
$\nu_\mu + \bar{\nu_\mu}$	1-5	250	500	408000	1550
	5-80	400	900	440000	
	80-1000	450	1500	57500	
	1000-10000	550		6700	
$\nu_\tau + \bar{\nu_\tau}$	1-4	250	500	1500000	350
	4-10			300000	
	10-50	350	600	375000	
	50-1000	450	800	200000	
	1000-10000	550	1500	26000	

choice of physics and nuisance parameters. The large number of raw MC events ensures a good estimation of the expected numbers and weighted distributions.

The analysis itself is then performed by comparing the weighted MC distributions to the observed data. This is done by binning them as described in Chapter 3 and calculating a loss function comparing the bin expectations to the data. The physics and nuisance parameters that best correspond to the observed data are estimated by minimizing this loss function. In order to achieve a reliable result with this method the MC needs to be precise and as close to the data as possible (at least at the final event selection).

1.1.1 Neutrinos

Due to the very low interaction rate of neutrinos, the event generation is performed in a way that forces every event to interact in a chosen sampling volume. The weight of each event is then calculated as the inverse of the simulated neutrino fluence

$$w_{\text{gen}} = \frac{1}{F_{\text{sim}}} \frac{1}{N_{\text{sim}}} \, , \tag{1.1}$$

where F_{sim} is the number of neutrino events per energy, time, area, and solid angle and N_{sim} is the number of simulated events. If this weight is multiplied by the livetime and the theoretically expected neutrino flux for a given physical model, it results in the number of expected events in the detector for this particular MC event. The baseline neutrino flux used in this thesis, computed for the South Pole, is taken from Honda *et al.* [2].

[2]: Honda et al. (2015), "Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model"

The simulation volume is a cylinder centered in DeepCore with radius and height chosen such that all events possibly producing a signal are contained. The different sizes, chosen depending on energy and neutrino flavor, are shown in Table 1.1. The directions of the neutrinos are sampled isotropically and the energies are sampled from an E^{-2} power law. The number of simulated events is chosen such that the livetime is more than 70 years for each flavor. Neutrinos and antineutrinos are simulated with ratios of 70% and 30%, respectively.

To simulate the neutrino interaction with the ice, the GENIE event generator [3] (version 2.12.8) is used, resulting in the secondary particles and the kinematic and cross-section parameters. As input, the outdated GRV98LO [4] *parton distribution functions (PDFs)* was used, because it was the only option that could incorporate extrapolations to lower Q^2 [5]. Muons produced in these interactions are propagated using PROPOSAL [6], also simulating their Cherenkov light output. The shower development of gamma rays, electrons, and positrons below 100 MeV and hadronic showers below 30 GeV is simulated using GEANT4 [7] while for higher energies an analytical approximation from [8] is used.

[3]: Andreopoulos et al. (2015), "The GENIE Neutrino Monte Carlo Generator: Physics and User Manual"

[4]: Glück et al. (1998), "Dynamical parton distributions revisited"

[5]: Bodek et al. (2003), "Higher twist, xi(omega) scaling, and effective LO PDFs for lepton scattering in the few GeV region"

[6]: Koehne et al. (2013), "PROPOSAL: A tool for propagation of charged leptons"

[7]: Agostinelli et al. (2003), "Geant4—a simulation toolkit"

[8]: Rädel et al. (2012), "Calculation of the Cherenkov light yield from low energetic secondary particles accompanying high-energy muons in ice and water with Geant4 simulations"

[9]: Becherini et al. (2006), "A parameterisation of single and multiple muons in the deep water or ice"

[10]: Heck et al. (1998), "CORSIKA: A Monte Carlo code to simulate extensive air showers"

[11]: Gaisser (2012), "Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio"

[12]: Engel et al. (2017), "The hadronic interaction model Sibyll – past, present and future"

1.1.2 Muons

Atmospheric muons are generated on a cylinder surface enclosing the full IceCube detector array. The cylinder has a height of 1600 m and a radius of 800 m. The energy is sampled from an E^{-3} power law while the other sampling distributions (position, direction) are found from parameterizations based on [9]. This work uses full CORSIKA [10] simulations of muons to tailor the parameterizations, starting from *cosmic ray (CR)* interactions with atmospheric nuclei using the CR flux model from [11] and producing the muons applying the *hadronic interaction (HI)* model SIBYLL 2.1 [12]. After the generation, they are propagated through the ice with PROPOSAL producing photons, treating them exactly like the muons produced in neutrino interactions.

Since the offline processing and selection steps described in Section 1.3.2 and Section 1.4 reduce the muon contamination to an almost negligible level, the statistical uncertainty on the number of expected muon events at the final selection level is large and therefore two separate sets of muon simulation are produced. **A first set** including all events resulting from the above described generation to tune the lower level selection (up to L4) and **a second set** to estimate the muon contamination at higher levels (above L5), which only accepts muon events if they pass through a smaller cylinder centered in DeepCore (height of 400 m and radius of 180 m) and rejects events based on a KDE estimated muon density at L5 (in energy and zenith) increasing the simulation efficiency at L5 significantly .

put a number on this significant increase? (YELLOW)

1.2 Detector Simulation

The detector simulation is performed after the event generation, where the initial particles and the resulting photons and secondary particles from their propagation were produced. This part of the simulation chain is applied to all muon and neutrino simulation as well as the HNL signal simulation explained in detail in Chapter **??**. The detector simulation can be split into two parts, the propagation of the photons and the simulation of the detector response (including internal noise).

1.2.1 Photon Propagation

Any photon that was produced in the event generation is individually traced through the ice, simulating scattering and absorption processes. The

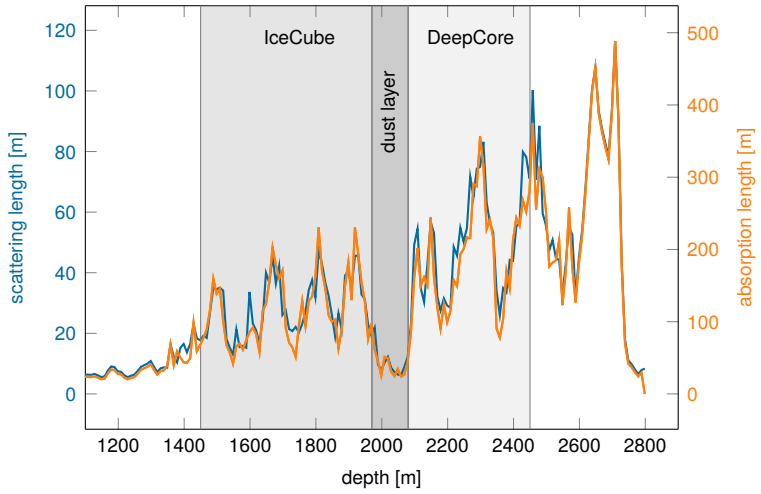


Figure 1.1: Scattering and absorption lengths in the SPICE model used for simulation production as a function of depth, modified from [16].

propagation is done using CLSIM [13] which is an implementation of the *Photon Propagation Code (PPC)* [14] in OpenCL. It is optimized to be run very efficiently on GPUs, which is what is done for IceCube simulation production. The ice is modeled as a set of 10 m thick, almost horizontal layers with specific absorption and scattering lengths. The *South Pole ice (SPICE)* model [15] accounts for the layers being tilted by a small amount (() and the absorption and scattering lengths having a non-uniformity with respect to the azimuth direction. Figure 1.1 shows the values of this model for the different depths, indicating the location of IceCube, DeepCore, and the dust layer.

[14]: Chirkin et al. (2019), "Photon Propagation using GPUs by the IceCube Neutrino Observatory"

[15]: Aartsen et al. (2013), "Measurement of South Pole ice transparency with the IceCube LED calibration system"

put a number on the tilt angle? (YELLOW)

In an initial step, each photon's absorption length is sampled from an exponential distribution with the expectation value at the current layer's absorption length. The following propagation steps are performed in parallel for all photons. In each of those steps, corresponding to a single scattering event, the photon travels a length that is sampled from an exponential distribution with the expectation value at the scattering length of the current layer and the scattering angle chosen based on a combination of a simplified Mie scattering distribution [17] and a Henyey-Greenstein distribution [18]. The parameters defining the shape of these distributions were calibrated using data from *in-situ* LED calibration runs. These steps are continuously repeated until each photon reached a DOM or was absorbed[1]. After all photons have been propagated in that manner, the final step is to output the photons that reached a DOM for further processing.

[17]: Mie (1908), "Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen"

[18]: Henyey et al. (1941), "Diffuse radiation in the Galaxy."

1: A photon is absorbed, when it traveled its full absorption length, sampled in the initial step of the photon propagation.

1.2.2 Detector Responses

The second part of simulating the IceCube detector is the DOM response. Whether a photon that reached a DOM produces a signal depends on the total efficiency and the angular acceptance curve of the specific DOM. The total efficiency includes effects of the DOM glass, PMT quantum and photoelectron collection efficiencies, and it is wavelength dependent. Additionally, there is another angle dependent effect called *hole ice* [19]. This effect is due to varied ice properties resulting from the re-freezing process of the water column inside the borehole after deployment of the string. Accepted photons are converted into a so-called *Monte Carlo photo-electron (MCPE)*. The amount of charge measured for each MCPE is determined by sampling

[19]: Fiedlschuster (2019), "The Effect of Hole Ice on the Propagation and Detection of Light in IceCube"

from a mixture of two exponential distributions and a normal distribution. This *single photo-electron (SPE)* distribution was tuned to match the observed distribution in each DOM in an *in-situ* calibration study [20]. Figure **??** shows the distribution compared to a lab measurement. Based on the sampled charges and times of MCPEs, the voltage waveforms for the (two) different readout channels are simulated and passed on to the trigger simulation starting with *WaveDeform*, which was already mentioned in Section **??**.

[20]: Aartsen et al. (2020), "In-situ calibration of the single-photoelectron charge response of the IceCube photomultiplier tubes"

Add SPE distribuiton plot (RED)

Besides the Cherenkov photons, IceCube also observes photons that are produced in radioactive decays inside the DOMs, both in the glass housing sphere and the PMT glass itself. To simulate this internal noise, the *Vuvuzela* module [21, 22] is used to create additional MCPEs that are fed into the same simulation chain described above. This module takes into account thermal and non-thermal components and their times are sampled using parameterizations of the measured distributions, where the thermal noise component is uncorrelated photons and the non-thermal component is from burst of photons. The noise hits are simulated by drawing the times from a constant rate Poisson process and the number of photons from a Poisson distribution. Then the time differences between the individual photons per hit is found, based on a Log-Normal distribution. The simulation is defined by 5 parameters that are calibrated for each DOM individually. Table 1.2 shows the average values for these parameters.

Parameter	Value
Therm. rate λ_{th}	180 Hz
Decay rate λ_{dec}	80 Hz
Decay hits η	8.5
Decay μ	4.3 $\log_{10}$(ns)
Decay σ	1.8 $\log_{10}$(ns)

Table 1.2: Typical parameter values used in the vuvuzela noise simulation. Averaged over all DOMs.

[21]: Larson (2013), "Simulation and Identification of Non-Poissonian Noise Triggers in the IceCube Neutrino Detector"
[22]: Larson (2018), "A Search for Tau Neutrino Appearance with IceCube-DeepCore"

1.3 Processing

After the detector simulation is performed, all MC and data are processed in exactly the same way. This section explains the trigger and event selection that is applied starting from the raw voltage measured by the PMTs. Most parts of this processing are identical to the procedure already described in [16, 23]. It is split in different steps run inside the ice, at the South Pole, and after the data was transferred to the North. The complexity and computational cost of the processing increases with each step, while the total number of events reduces, making it feasible and reducing the use of computational resources on events that are not of interest for the analysis.

[16]: Trettin (2023), "Search for eV-scale sterile neutrinos with IceCube DeepCore"
[23]: Lohfink (2023), "Testing non-standard neutrino interaction parameters with IceCube-DeepCore"

[24]: Aartsen et al. (2017), "The IceCube Neutrino Observatory: Instrumentation and Online Systems"

1.3.1 Trigger and Filter

Before the data can be sent to the North, the initial signal coming from the PMT is a voltage waveform that has to be digitized (for data) and then information of photon hits has to be extracted (also for the MC coming from the detector response simulation). The trigger and filter explained here are tailored to select events that passed through the DeepCore volume, while rejecting background events (either from atmospheric muons or from random noise). There are other filters used in IceCube which will not be explained here, since they are not relevant for this work. A full description of the instrumentation and the online systems can be found in [24].

Include some low level plots like the trigger efficiency for the HNL simulation (ORANGE)

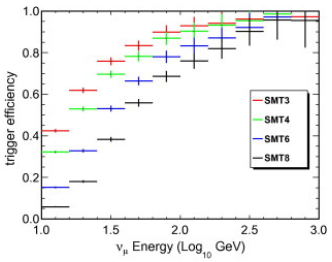


Figure 1.2: Efficiencies of different IceCube and DeepCore triggers, taken from [25].

In-ice Trigger

The trigger is applied inside the DOM in the ice before sending the information to the ICL on the surface. The time dependent voltage curves are

captured if a pre-defined threshold value is exceeded. Once the threshold set to the equivalent of 0.25 PE is crossed, 6.4 µs of the waveform are coarsely digitized by a *Fast Analog-to-Digital Converter (FADC)* with a sampling rate of 40 MHz. Additionally, the first 427 ns are digitized using an *Analog Transient Waveform Recorder (ATWD)* with a sampling rate of 300 MHz [26], but only if some trigger condition is met, because this readout frequency is too high to be sampled directly and requires some buffering. For DeepCore, the HLC condition already mentioned in Section **??** has to be met for three DOMs inside the fiducial volume within a time window of 5 µs. If this is the case, all waveforms that crossed the threshold within a 20 µs time window around the trigger are digitized and sent to the ICL for further processing. This trigger is called *Simple Multiplicity Trigger 3 (SMT-3)*. The DOM hits that are read out in this process, but do not meet the HLC condition, are called *soft local coincidence (SLC)* hits. The rate of the DeepCore SMT-3 trigger is ~250 Hz [27], accepting ~70 % of ν_μ-CC events at 10 GeV and ~90 % at 100 GeV [25]. The trigger efficiencies for different SMT triggers, including the DeepCore SMT-3, are shown in Figure 1.2.

[26]: Abbasi et al. (2009), "The IceCube data acquisition system: Signal capture, digitization, and timestamping"

[27]: Aartsen et al. (2017), "The IceCube Neutrino Observatory: instrumentation and online systems"

[25]: Abbasi et al. (2012), "The design and performance of IceCube DeepCore"

Online Filter

The digitized waveforms are sent to the ICL, where a further filter is applied *online*[2]. First, the WaveDeform algorithm is run to extract photon arrival times and charge from the waveforms, then the DeepCore filter is applied, which is an iterative hit cleaning starting from HLC hits and removing any hits outside a 125 m radius and a 500 ns time window (called *radius-time cleaning (RT-cleaning)*) of the initial hit. This mainly rejects unphysical SLC hits, which are potentially caused by random noise. The following selection steps are done using the resulting cleaned pulses.

2: Where *online* means running on hardware at the South Pole.

Next, an additional cut is applied to reject events that are likely to be caused by atmospheric muons. This is done by splitting the hits depending on whether they were inside the DeepCore fiducial volume or outside and then calculating the speed of each hit outside the fiducial volume towards the *center of gravity (COG)* of the hits inside. If one of them has a speed close to the speed of light, the whole event is rejected, because this is a strong indication for a muon event.

As input for the further selection levels, a few event properties, like vertex position and direction, are determined using fast and simple event reconstructions. After the DeepCore online filter, the rate is about 15 Hz, which can be sent to the North via satellite for further processing.

1.3.2 Event Selection

After the data was sent to the North, the *offline* filters and selection are applied to further recude reduce the background of atmospheric muons and noise. The selection is split into three levels referred to as *Level 3-5 (L3-L5)*, which bring down the neutrino and muon rate to ~1 mHz, while the remaining fraction of random noise is below 1 %.

Level 3

At the first offline filtering level, Level 3, 1D cuts are used to reduce atmospheric muons, pure noise, and coincident muons. These cuts are targeting regions where the data/MC agreement is poor, so that more sophisticated *machine learning (ML)* techniques can be applied at later levels. The cuts are made using 12 control variables, that are inexpensive to compute for the very large sample at this stage. The variables are related to position, time, and overall number of hits in the event.

Pure noise hits, that are temporally uncorrelated, are cleaned by applying a 300 ns sliding window, requiring the containment of more than 2 hits at its maximum. Additionally, an algorithm is run to check whether the hits show some directionality, accepting them only if they do.

To reduce the amount of muons a series of cuts is applied using spatial and temporal information. Events that have more than 9 hits observed above −200 m or the first HLC hit above −120 m are rejected as well as events where the fraction of hits in the first 600 ns of the event is above 0.37, ignoring the first two hit DOMs. Additionally, the ratio between hits in the veto region and the DeepCore fiducial volume is required to be below 1.5.

If a muon enters the detector after the data acquisition was already triggered, it causes events that span over a much larger time range. To reduce those coincident events, the time difference between first and last pulse cannot be above 5000 ns. This cut mainly affects a region of very poor data to MC agreement, because coincident events are not simulated at all.

The L3 cuts remove 95 % of the atmospheric muons and >99 % of pure noise hits, while keeping >60 % of the neutrino events. The sample now roughly contains muons/neutrinos/noise at a ratio of 100:10:1 with a total rate of ~0.5 Hz.

add example plots (2?) for L3 cut variables and applied cuts (YELLOW)

Level 4

After the total rate was reduced by the simple cuts of L3 and the overall agreement between data and MC is established, ML techniques can be applied to further reduce the background. For Level 4, two *Boosted Decision Trees (BDTs)* [28] classifier are trained to separate neutrino events from atmospheric muons and noise hits, separately. The output of each classifier, a probability score, can be seen in Figure 1.3. The noise filter is applied first and an event passes the score if it is larger than 0.7, reducing the noise hits by a factor of 100, while keeping 96 % of neutrinos. Then the second BDT classifier is applied to reject muons. It was trained partly on unfiltered data, which consists of >99 % atmospheric muons, to reject the data and keeping the neutrinos from the simulation. Rejecting events with a score smaller than 0.65 removes 94 % of atmospheric muons while keeping 87 % of neutrinos. This fraction varies depending on the flavor and interaction type, ν_μ-CC events for example, which have a muon in the final state, are therefore reduced to 82.5 %. After applying the L4 cuts based on the BDT classifier outputs, the sample is still dominated by atmospheric muons, while the noise rate dropped to below most neutrino types.

[28]: Friedman (2002), "Stochastic gradient boosting"

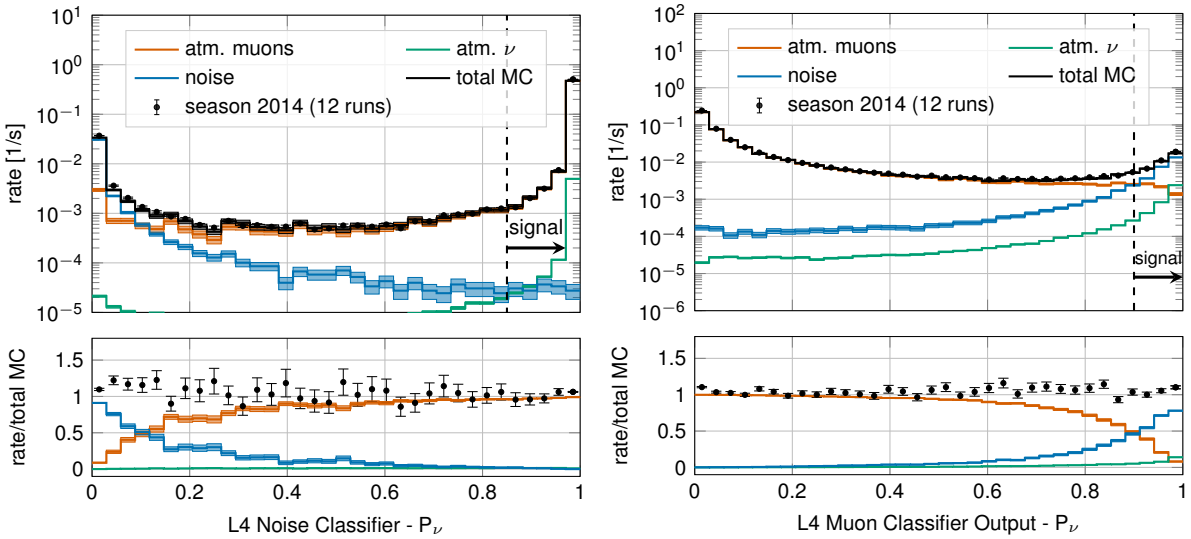


Figure 1.3: Distributions of Level 4 noise classifier output (left) and muon classifier output (right), where larger values indicate more neutrino-like and lower values more noise-like/muon-like. Taken from [1].

Level 5

Level 5 is the final selection level, before event reconstructions are applied. This level aims to reduce the remaining atmospheric muon rate below the rate of neutrinos. Muons not rejected by the earlier levels are those that produced little or no light in the veto regions. One possible reason is that they passed through one of the un-instrumented regions between the strings called *corridors*. To reject those, special corridor cuts, based on the number of hits they produced close to a potential corridor they passed through. The potential corridor in questions is identified based on a simple infinite track fit. In addition to the corridor cuts, starting containment cuts are applied to reject events that start at the edge of the fiducial volume. Events with more than seven hits in the outermost strings of the detector or those that have a down going direction in the uppermost region are rejected. This further reduces the fraction of muons by 96 % while keeping 48 % of neutrinos. The rates after this level are 1 mHz and 2 mHz for neutrinos and muons, respectively, making it a neutrino dominated sample.

add some figure showing the corridors? (YELLOW)

add table with rates per level (split in flavor) - maybe better in analysis chapter to also show signal? (RED)

1.4 Reconstruction

In the energy range most relevant for this work, between 10 GeV and 100 GeV, the light deposition is very low and only a few DOMs detect light, making the reconstructions difficult. In [29] two classical methods are described, which have partly been applied in one recent IceCube atmospheric neutrino oscillation measurement using a sub-sample of the DeepCore sample [1]. The algorithm used in this work on the other hand, is a newer method that applies a *convolutional neural network (CNN)* to reconstruct the events and determine some discriminating quantities. The latest muon neutrino disappearance result from IceCube [30] is based on this reconstruction.

[29]: Abbasi et al. (2022), "Low energy event reconstruction in IceCube DeepCore"

[1]: Abbasi et al. (2023), "Measurement of atmospheric neutrino mixing with improved IceCube DeepCore calibration and data processing"

[30]: Yu et al. (2023), "Recent neutrino oscillation result with the IceCube experiment"

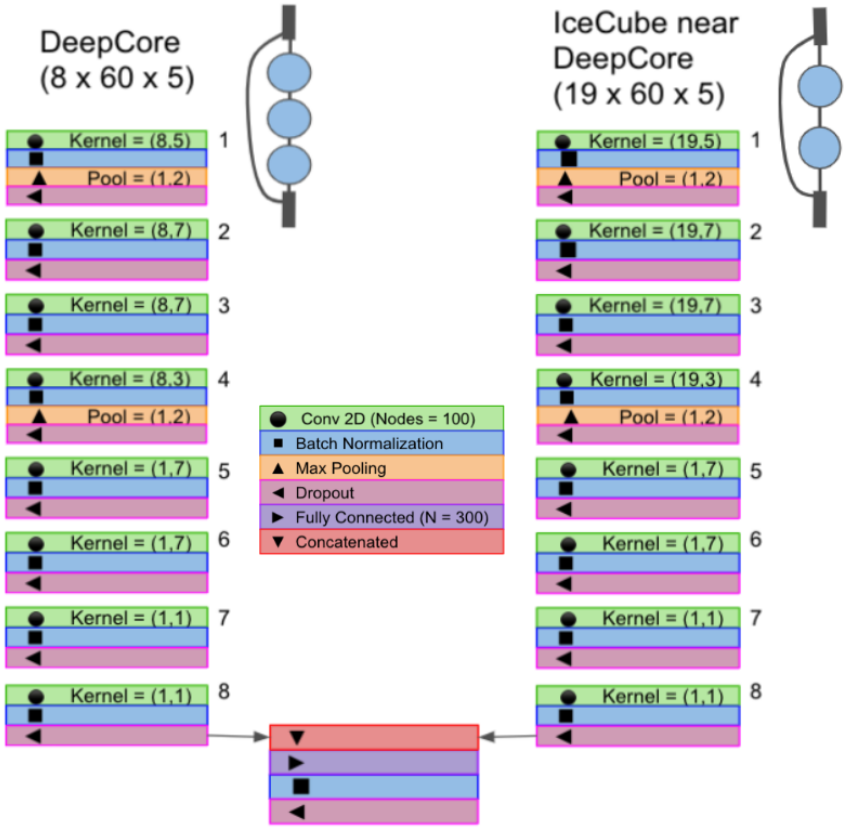


Figure 1.4: Architecture of the FLERCNN neural networks, taken from [31].

1.4.1 Fast Low Energy Reconstruction using Convolutional Neural Networks

As the name *Fast Low Energy Reconstruction using Convolutional Neural Networks (FLERCNN)* already indicates, the FLERCNN reconstruction [31, 32] is a CNN optimized to reconstruct IceCube events at low energies (<100 GeV) in a fast and efficient manner, by leveraging the approximate translational invariance of event patterns within the detector. The architecture of the network is very similar to the preexisting IceCube CNN event reconstruction [33], but optimized on low energy events and specifically tailored to include the DeepCore sub-array. Only the eight DeepCore strings and the central 19 IceCube strings are used for the reconstruction (compare to Figure **??**). Because of the different z-positions of the DeepCore and IceCube DOMs, they are divided into two networks that are combined in the final layer of the network. The full architecture is shown in Figure 1.4. The first dimension of the network is the string index, while the second dimension is the order of the DOMs along the vertical axis. The horizontal position of the DOMs is not used, since the strings are arranged in an irregular pattern. The information from the DOM hits is summarized into five charge and time variables, which make up the last dimension of the input layer. The variables are the total summed charge, the time of the first hit, the charge weighted mean time of the hits, the time of the last hit, and the charge weighted standard deviation of the hit times.

[31]: Yu et al. (2021), "Direction reconstruction using a CNN for GeV-scale neutrinos in IceCube"
[32]: Micallef (),

[33]: Huennefeld (2017), "Deep Learning in Physics exemplified by the Reconstruction of Muon-Neutrino Events in IceCube"

add image with selected strings used for flercnn IC and DC (YELLOW)

Five different networks are trained using this architecture. Three networks do the regression of the events' energy, zenith angle, and the starting vertex (x, y, z position), while two of them are used for classification. One is trained to predict the probability of the event being a track (used as PID) and the

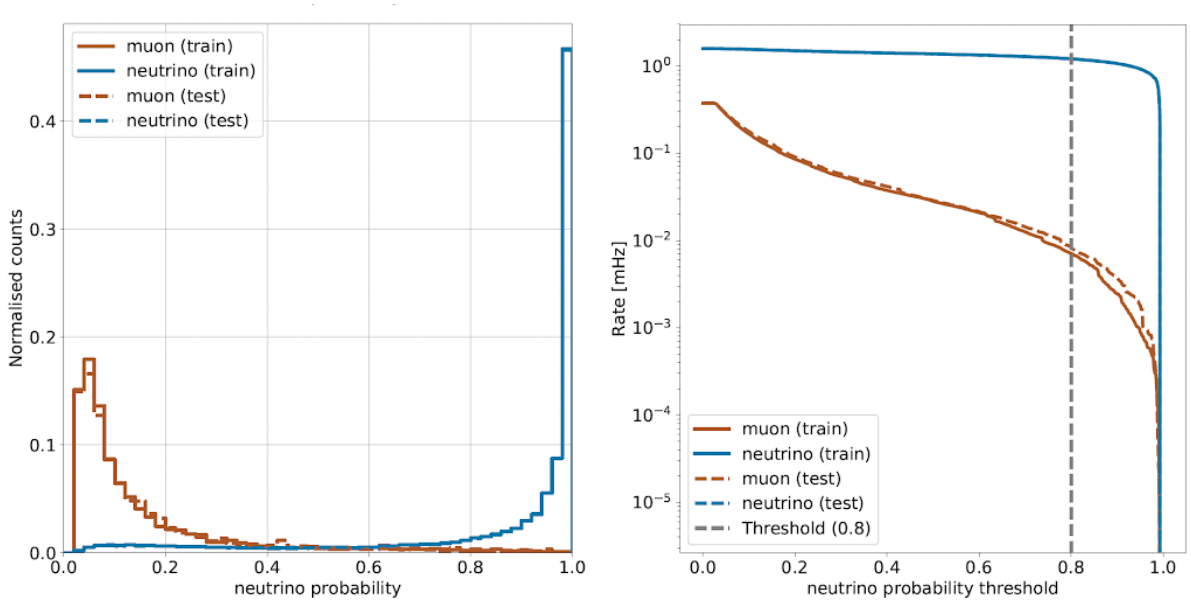


Figure 1.5: FLERCNN muon classifier output score (left) and rate of neutrinos and muons as function of muon classifier cut (right). Taken from **[flercnn_analysis_internal_note]**

other to predict the probability of the event being a muon. Each network is trained with an MC sample modified to have a flat distribution in the target variable, to be unbiased for that variable and ideally extending outside the target reconstruction region. For the classification tasks the loss function is the *binary cross entropy* and the activation function is a *sigmoid*. To perform the regression of zenith and vertex position, the loss function is the *mean squared error (MSE)*, while for the energy it is the *mean absolute percentage error*. The activation for all regression tasks is *linear*.

add some performance plots of the FLERCNN reconstruction (ORANGE)

There is more information on pre-processing the samples and preparing the input features, and training each cnn, but I'm not sure if that might be too much detail? (YELLOW)

1.4.2 Analysis Selection

Before the reconstruction is applied a few additional high level variables are computed, which are from fast and inexpensive algorithms. Then the reconstruction is performed by applying the trained FLERCNN networks to get the output quantities. After that, another BDT classifier is trained to further reduce the muon background for the final sample. The BDT is trained on five high level variables, where three are FLERCNN reconstruction variables (vertex z, ρ_{36}[3], and muon probability) and two are lower level variables (L4 muon classifier output and L5 corridor cut variable). To train the BDT, the FLERCNN nominal simulation set is used, only using events with $\cos(\theta_{zenith}) \leq 0.3$. The output of the BDT is the neutrino probability and a cut at 0.8 is applied to reject events with a high probability of being a muon. Figure 1.5 shows the output of the BDT classifier, where the neutrinos in both training and testing sets are gathered at 1 and muons are around 0, which shows great classification power.

3: A radial variable that is often used in IceCube, is the horizontal distance to string 36 called ρ_{36}, which is basically the distance to the center of IceCube.

add reference for flercnn analysis internal note (ORANGE)

To get the final, pure sample of well reconstructed neutrinos another set of cuts is applied. The first cuts are meant to reject events with poor reconstruction quality, by requiring the events to fall into the DeepCore volume, where the denser, better instrumented detector leads to enhanced resolution. The cuts are applied on the vertex z and ρ_{36} and are listed in Table 1.3. The FLERCNN reconstruction was optimized for atmospheric

Variable	Threshold	Removed
Number of hit DOMs	≥ 7	1.05 %
Radial distance	< 200 m	0.09 %
Vertical position	−495 m < z < −225 m	5.48 %
Energy	5 GeV < E < 100 GeV	20.70 %
Cosine of zenith angle	< 0.04	19.66 %
Number of direct hits	> 2.5	10.50 %
Number of hits in top layers	< 0.5	0.03 %
Number of hits in outer layer	< 7.5	0.001 %
Muon classifier score	≥ 0.8	23.90 %

Table 1.3: Cuts performed to select the final analysis sample. Parts of the cuts are meant to increase the data/MC agreement, while others are meant to reject events with poor reconstruction quality.

neutrino analyses which are mainly in the region below 100 GeV and there are very few events with energies below 5 GeV, so the reconstructed energy is required to be in that range. Additionally, rejecting events with fewer than seven hits in the selected DOMs used for FLERCNN showed to increase the resolution.

Another set of cuts is applied to make sure the agreement between data and MC is good. To remove coincident muon and neutrino events, cuts are applied to the number of hits in the top 15 layers of IceCube DOMs and the number of hits in the outermost IceCube strings. Coincident random noise events are removed by requiring more than three hit DOMs from direct photons[4]. Neither of the two coincident event types are simulated, which can be seen as bad agreement between data and MC. The last cut is on the reconstructed cosine zenith, which is required to be smaller than 0.04 to reject down-going muons.

4: *Direct photons* are photons that were not scattered on their way from the interaction vertex to the DOM.

1.5 Systematic Uncertainties

SB: I think its better to discuss the sources of uncertainty when you introduce each component, e.g. when you talk about cross-sections you can explain the existing measurements and models, and how the uncertainty is derived. Same for the detector calibration uncertainties - you can discuss the magnitude of the errors and how they are derived when you discuss the detector. (ORANGE)

There are multiple sources of systematic uncertainties related to the event generation and processing explained in this chapter. All uncertainties considered in this work need to be implemented with parameters that can be varied continuously so that a simultaneous fit of the physics and systematic parameters can be performed. Where possible, a correct model of the effect is used, but in many cases the variations are captured by effective parameters. Uncertainties that solely scale the total event rate are not included individually, since the analysis only uses the relative distribution of events and a single scaling parameter N_ν is used to scale the total neutrino rate instead.

1.5.1 Atmospheric Flux Uncertainties

The flux of atmospheric neutrinos is influenced by multiple factors, the spectrum and composition of CRs, the assumed atmospheric conditions, and the HI model used to describe the air showers development. Uncertainties of the neutrino flux are therefore dictated by the uncertainties on these components, where the variations in atmospheric conditions were found to have negligible effect [1].

[1]: Abbasi et al. (2023), "Measurement of atmospheric neutrino mixing with improved IceCube DeepCore calibration and data processing"

Cosmic ray flux: The selected sample of atmospheric neutrinos lies around energies of up to 100 GeV. The initial primary particles in the CR flux can have 100 times larger energies and therefore the CR flux between 10 GeV and 10 TeV is important, which dominantly consists of hydrogen and helium

nuclei [34]. The uncertainty in this CR flux component can be described as a power law correction [35, 36]

$$\Phi'_\nu = \Phi_\nu \left(\frac{E}{E^\star} \right)^{\Delta\gamma} , \tag{1.2}$$

where $E^\star$ is the pivot energy and $\Delta\gamma$ is the correction to the power law exponent. This modification propagates into the neutrino flux, which is therefore corrected in the same way. $E^\star$ was chosen to be 24 GeV as to minimize the dependence of the overall flux scale on $\Delta\gamma$ [1].

Hadronic interaction model: Neutrinos are produced in the decaying hadrons in CR air showers, spanning a large parameter space that is sparsely evaluated by experimental data. To include uncertainties based on energy, direction, and neutrino flavor, the MCEq package [37] is used to compute the distribution of atmospheric leptons and to estimate the impact of varying their contributions. The calculations result in the change in flux $d\Phi_l/dB$ for a variation dB of some parameter B. Scaling this variation by some value b, the modified total flux, s is then given by

$$\Phi'_l = \Phi_l + \left(b \cdot \frac{d\Phi_l}{dB} \right) . \tag{1.3}$$

Matching the work in [38], the parameter space is divided in regions of the primary energy, E_i, and the energy fraction of the secondary meson, x_{lab}, with varying uncertainties, derived from fixed target experiment data. The Sibyll2.3c [39] HI model and the GSF CR flux [34] were used to calculate the related flux changes[5] for the different regions in E_i and x_{lab}, resulting in 17 variables, encoding the possible changes. Figure 1.6 shows the selected regions of the parameter space and the names given to the uncertainties. The variational term in Equation 1.3 is applied for each of these parameters and the total variation is the sum of all individual variations.

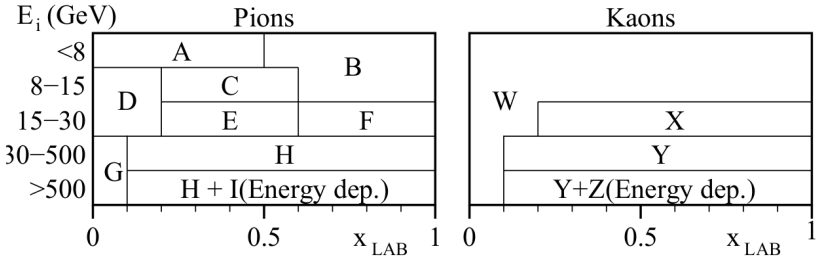


Figure 1.6: Flux uncertainty regions of the hadronic interaction model in the phase space of the primary energy E_i and the energy fraction of the secondary meson x_{lab}. Taken from [38].

1.5.2 Cross-Section Uncertainties

The uncertainties related to the cross-sections are split into low and high energy components, since there is no coherent model to explain both DIS interactions, which are the dominant processes above 20 GeV, and *charged current resonance production (CCRES)* and *charged current quasi elastic scattering (CCQE)*, which are relevant below 20 GeV where interactions with the nucleons as a whole are important. Three parameters are included to account for all relevant cross-sections uncertainties.

[34]: Dembinski et al. (2017), "Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV"
[35]: Barr et al. (2006), "Uncertainties in atmospheric neutrino fluxes"
[36]: Evans et al. (2017), "Uncertainties in atmospheric muon-neutrino fluxes arising from cosmic-ray primaries"
[1]: Abbasi et al. (2023), "Measurement of atmospheric neutrino mixing with improved IceCube DeepCore calibration and data processing"
[38]: Barr et al. (2006), "Uncertainties in Atmospheric Neutrino Fluxes"
[39]: Riehn et al. (2020), "Hadronic interaction model sibyll 2.3d and extensive air showers"
[34]: Dembinski et al. (2017), "Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV"
5: The choice of flux and HI model have minor impact on the variations.

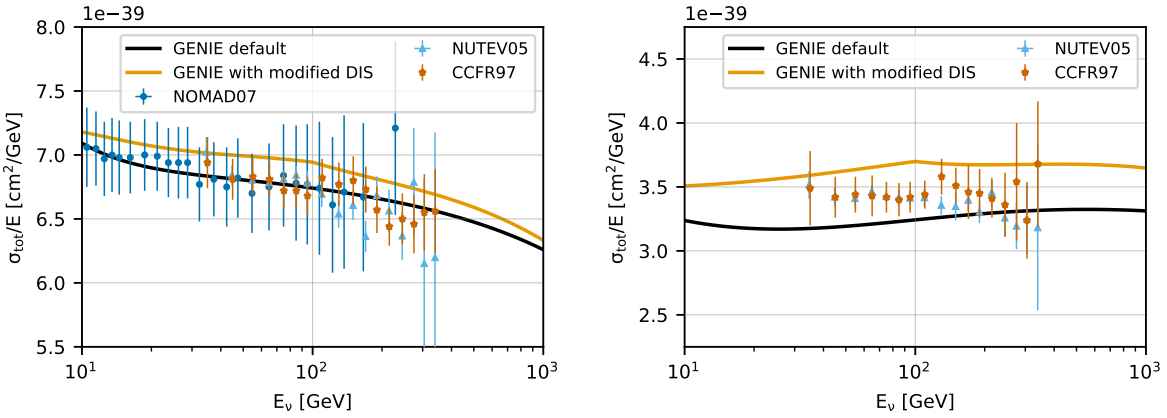


Figure 1.7: Inclusive total neutrino-nucleon cross-sections on an isoscalar target (black) for neutrinos (left) and antineutrinos (right) calculated with GENIE, comparing to measurements from NOMAD [41], NUTEV [42], and CCFR [43]. The scaled GENIE cross-section (orange) is also shown. Taken from [1].

At low energies two parameters are included accounting for uncertainties in form factors of CCQE and CCRES events. These uncertainties are due to uncertainties in the *axial mass* M_A, which enters the form factor as in

$$F(Q^2) \sim \frac{1}{(1-(\frac{Q}{M_A})^2)^2}\,, \tag{1.4}$$

where Q^2 is the momentum transfer squared. The axial mass can be determined experimentally and to include uncertainties on the values of M_A^{CCQE} and M_A^{CCRES}, the cross-sections are computed with GENIE, where the form factors are calculated varying the axial mass by $\pm 20\%(1\sigma)/\pm 40\%(1\sigma)$ around the nominal value. This is an approximation of the recommended uncertainties by the GENIE collaboration, which are -15%, $+25\%$ for M_A^{CCQE} and $\pm 20\%$ for M_A^{CCRES} [3]. To apply a continuous uncertainty variation of the axial mass in a fit, the total cross-section is fit with a quadratic function to interpolate between the cross-sections computed with the different axial masses.

which experiments measure the axial mass? (ORANGE)

Even though the DIS interactions can be calculated very precisely, there are still uncertainties in the input PDF, describing the probability of finding a specific parton (quark) with a specific momentum fraction x inside a nucleon. To account for differences between the used method and more sophisticated methods using newer PDFs seen at high energies, an uncertainty parameter is introduced. The parameter is based on the discrepancy between the cross-sections computed with GENIE and the ones computed with CSMS [40] above 100 GeV. The included parameter scales the cross-section from the GENIE values to the CSMS values, which are considered more accurate above 100 GeV. The scaling is done as a function of energy and inelasticity and to guarantee continuity, the scaling is extrapolated linearly below 100 GeV. The parameter is designed such that a value of 0.0 corresponds to the GENIE cross-sections and a value of 1.0 gives an approximation of the CSMS cross-sections. A comparison of the total cross-sections GENIE (scaled/unscaled) with the data is shown in Figure 1.7.

[40]: Cooper-Sarkar et al. (2011), "The high energy neutrino cross-section in the Standard Model and its uncertainty"

1.5.3 Detector Calibration Uncertainties

The detection process of neutrinos in IceCube has several sources of uncertainties, where the effects of the properties of the ice itself and the optical efficiency of the DOMs are dominant for this analysis. None of these uncertainties can be described by an analytic expression, so they have to be estimated using MC simulation. This is done by producing additional systematic simulation samples at discrete values of those parameters. The five relevant uncertainty parameters are the absolute efficiency of the DOMs, a global scaling of bulk ice scattering and absorption lengths, and variations of the relative angular acceptance due to hole ice variations in two parameters. To perform the fit, continuous variations with respect to these parameters, will be derived with a method explained in Section 3.2.3.

DOM efficiency: As was already mentioned in Section **??**, the absolute efficiency of the DOMs, ϵ_{DOM} is calibrated using minimum ionizing muons from air showers, due to the lack of a calibrated light source in the detector. Using the muons as a steady, controlled source of light, the efficiency can be estimated by comparing simulated muon data sets with varied DOM response to the measured data. Since the uncertainties found in multiple iterations of this study [44, 45] are at the order of 10 %, this systematic is highly relevant and has to be included in the analysis.

[44]: Feintzeig (2014), "Searches for Point-like Sources of Astrophysical Neutrinos with the IceCube Neutrino Observatory"
[45]: Kulacz (2019), "In Situ Measurement of the IceCube DOM Efficiency Factor Using Atmospheric Minimum Ionizing Muons"

Bulk ice scattering and absorption: Absorption and scattering length are the most important properties that govern the propagation of photons through the ice. The simulation principle and how the depth dependent absorption and scattering coefficients are used was already explained in Section 1.2.1. To account for uncertainties on this model of the bulk ice coefficients, a global scaling for each of the two parameters (global absoprtion, global scatteriong) is applied.

Hole ice angular acceptance: Due to bubble formation in the re-freezing process of the boreholes, the hole ice seems to be less transparent in the center of the columns [46]. This effectively decreases the chance of photons hitting the DOMs directly from below, which can be described as an additional angular modification of the DOM acceptance. The modification is parameterized by a two dimensional, normalized[6] function, where the two dominant of the parameters (p_0, p_1), dictating its form, are enough to describe all past and the current hole ice models from both *in-situ* and laboratory measurements. Figure 1.8 shows the acceptance modification as a function of the incident photon angle $\cos(\eta)$. The current baseline model, the variations achieved through modifying p_0 and p_1, and a laboratory measurement can be seen.

[46]: Rongen, Martin (2016), "Measuring the optical properties of IceCube drill holes"

6: The hole ice angular acceptance modification is normalized so that it does not affect the total charge.

explain the BFR ice thingie (RED)

Ice Model:

1.5.4 Muon Uncertainties

The muon fraction in the final level selection (see Section 1.4.2) is below 1 %, therefore additional muon systematic uncertainties apart from the spectral

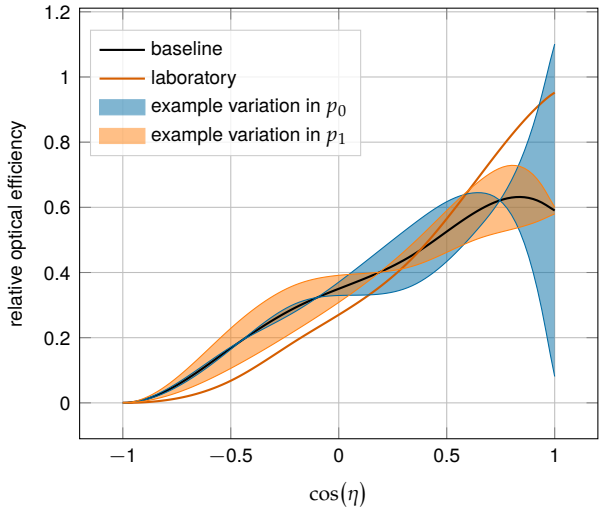


Figure 1.8: Relative angular acceptance modification due to hole ice. Shown is the current baseline model, the variations from changing p_0 and p_1, and a laboratory measurement. Modified from [16].

index are not implemented, but rather a total muon scaling parameter is added. This total scale is somewhat degenerate with the DOM efficiency, since an increased DOM efficiency leads to better muon rejection. Both the total muon scaling and the muon spectral index have a very small impact on the analysis as will be shown in Section 3.2.3.

cite this? (YELLOW)

Detecting Low Energy Double Cascades

2

2.1 Reconstruction

All existing reconstruction algorithms applied for low energy atmospheric neutrino events mentioned in Section 1.4 are either assuming a single cascade hypothesis or a track and cascade hypothesis, which are the two SM morphologies observable at these energies, as was described in Section **??**. A HNL being produced and decaying inside the IceCube detector however, will produce two cascade like light depositions. The morphology, spatial separation between the cascades, and their individual properties depend on the model parameters discussed in Section **??**. To investigate the performance of the detector to observe and identify these events, a low energy double cascade reconstruction algorithm was developed. It is based on a pre-existing algorithm used to search for double cascades produced from high energy astrophysical tau neutrinos [47] that was established in [48], but first mentioned in [49].

[47]: Abbasi et al. (2020), "Measurement of Astrophysical Tau Neutrinos in IceCube's High-Energy Starting Events"

[48]: Usner (2018), "Search for Astrophysical Tau-Neutrinos in Six Years of High-Energy Starting Events in the IceCube Detector"

[49]: Hallen (2013), "On the Measurement of High-Energy Tau Neutrinos with IceCube"

2.1.1 Table-Based Minimum Likelihood Algorithms

The aforementioned reconstruction is relying on a minimum likelihood algorithm, which is the *classical* approach to IceCube event reconstructions, as opposed to ML based methods. A Poissonian likelihood is constructed, which compares the observed photon numbers, n, with their arrival times to the expected light depositions, μ, for a given even hypothesis as

$$\ln(L) = \sum_j \sum_t n_{j,t} \cdot \ln\left(\mu_{j,t}(\Theta) + \rho_{j,t}\right) - \left(\mu_{j,t}(\Theta) + \rho_{j,t}\right) - \ln\left(n_{j,t}!\right) \,, \quad (2.1)$$

where ρ are the number of expected photons from noise, Θ are the parameters governing the source hypothesis, and the likelihood is calculated summing over all DOMs j splitting observed photons into time bins t. The light expectations are calculated using look-up tables [50] that contain the results from MC simulations of cascade events or track segments. By varying the parameters defining the event hypothesis, the likelihood of describing the observed light pattern by the expected light depositions is minimized to find the reconstructed event. Algorithms of this kind used in IceCube are described in great detail in [51]. For the table production a specific choice of ice model has to be made, while the calibrated DOM information is taken from the measurement itself.

[50]: Whitehorn et al. (2013), "Penalized splines for smooth representation of high-dimensional Monte Carlo datasets"

[51]: Aartsen et al. (2014), "Energy Reconstruction Methods in the IceCube Neutrino Telescope"

Based on the tabulated light expectations for cascades and track segments, various event hypothesis can be constructed, like the common cascade only or the track and cascade hypotheses. The hypothesis describing the double cascade signature of the HNL is using two cascades that are separated by a certain distance. The whole hypothesis is defined by 9 parameters and assumes that the two cascades are aligned with each other, which is a safe assumption for strongly forward boosted interactions. The parameters are the position of the first cascade ,x, y, z, the direction of both cascades, ϕ, θ,

Elaborate whether this is the case (show it in a plot?). Discuss directionality of cascades in general. (ORANGE)

and its time, t, as well as the decay length, L, between the two cascades. Assuming the speed of the HNL to be the speed of light, c, this already defines the full hypothesis. Note here, that the HNL particle does not produce any light while traveling, as it is electrically neutral. Since the likelihood only sums over DOMs that have observed photons, the non-observation of light is used as information and will exclude hypotheses with light expectation in those DOMs. The full 9 parameters describing the event are $\Theta = (x, y, z, t, \theta, \phi, E_0, E_1, L)$. To compute the full likelihood the term in Equation 2.1 is summed over both cascade parts, i, as $\sum_i \ln(L_i)$.

2.1.2 Optimization for Low Energies

Optimizing the double cascade reconstruction for low energy events was done in parallel to the development of the model dependent simulation generator introduced in Section **??**. A preliminary sample of HNL events was used, containing a continuum of masses between 0.1 GeV and 1.0 GeV and lab frame decay lengths sampled uniformly in the range from 5 m to 500 m. Even though this sample is not representative of a physically correct model and therefore not useful to predict the event expectation, it can still be used to optimize the reconstruction. The double cascade nature of the individual events and the evenly spaced decay length distribution are especially useful for this purpose.

The simulation is processed up to Level 5 of the selection chain described in Section 1.3.2 and one of the reconstructions from [29] is applied to the events, fitting a cascade and a track and cascade hypothesis. The results from this reconstruction are used as an input for the double cascade reconstruction, where the position of the vertex, the direction of the event, and its interaction time are used as the input quantities for the first cascade, and the length of the track reconstruction is used as a seed for the distance between the two cascades.

[29]: Abbasi et al. (2022), "Low energy event reconstruction in IceCube DeepCore"

Decay Length Seeds

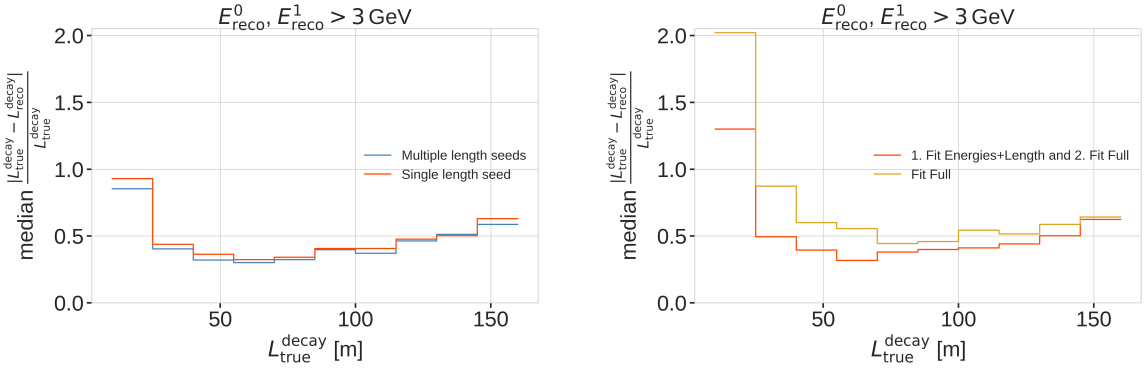


Figure 2.1 fix caption of this figure (RED)

The full 9 dimensional likelihood space is very complex and can have many local minima, depending on the specific event and its location in the detector. Especially the seed value of the length between the two cascades was found to have a very strong impact on whether the global minimum was found

during the minimization. To mitigate this effect, multiple fits are performed, seeding with variations of the input length at different orders of magnitude. The best result is used, selected based on the total likelihood value of the best fit parameter set. A small improvement in the decay length resolution can be found by using this approach as compared to a single length seed. The effect can be seen in the left part of Figure 2.1, which shows the median, absolute, fractional decay length resolution as a function of the true decay length for a single length seed and multiple length seeds. Only events that have reconstructed cascade energies above 3 GeV are used for this comparison, where this threshold was roughly chosen to select well reconstructed double cascade events.

Fit Routine

Because the length seed showed to have such a large impact on the reconstruction performance, a more sophisticated fit routine than fitting all 9 parameters at once was tested. In a first fit iteration, some parameters are fixed and the resulting best fit point is used to fit all 9 parameters in a second iteration. In the right part of Figure 2.1 it can be seen how a fit split into two consecutive steps, where the first step fits only both cascade energies and the decay length and the second step fits the full 9 parameters, performs better as compared to a single, full 9 parameter fit. The initial seed for both routines is the same.

Minimizer Settings

To investigate the effect of the minimizer used to find the best fit parameters, the reconstruction was performed using three different minimizers, which were easily accessible within the reconstruction framework. The minimizers used were Minuit1 Simplex, Minuit2 Simplex, and Minuit2 Migrad. The initial idea was to test a global minimizer, or a routine that can find the rough position of the global minimum first and then a local minimizer to find the exact minimum, but unfortunately this was not possible with the minimizers available in the framework. As can be seen in Figure 2.2, Minuit1 Simplex performed best and was chosen as the default for the reconstruction.

cite these buggers again, and or point back to where they were mentioned? (RED)

I think I could show these 3 plots side by side, if I remake them with better dimensions (ORANGE)

fix caption of this figure (RED)

2.1.3 Performance

The chosen reconstruction chain used to test the performance of the detector to observe low energy double cascades is the following; Minuit1 Simplex is used as the minimizer, the decay length is seeded with 3 different values, 0.5x, 1.0x, and 1.5x the length of the preceding track reconstruction, and the fit routine is split into two steps, where the first step fits the energies and the decay length and the second step fits the full 9 parameters. In the first step, the number of time bins in Equation 2.1 is set to 1, so just the number of photons and their spatial information is used. The second step is seeded with the best results from the first fit and here the number of time bins is chosen such that each photon falls into a separate time bin, which means all time information is used. The average runtime per event is ~16 s on a single CPU core, but is very dependent on the number of photons observed

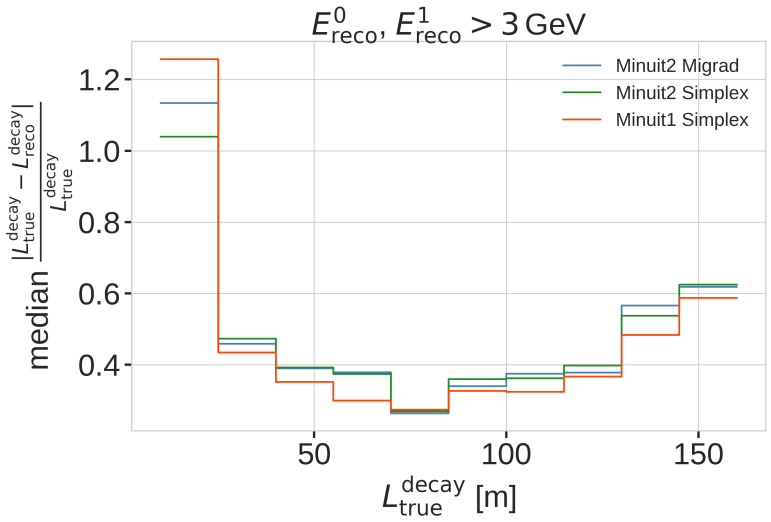


Figure 2.2: title

in the event, since the likelihood calculation in the second step scales with this number and a table lookup has to be performed for each photon.

To get a more realistic estimate of the reconstruction performance, it is run on a second preliminary sample of HNL events, containing masses between 0.1 GeV and 3.0 GeV and the lab frame decay length is sampled from an inverse distribution in the range from 1 m to 1000 m, which is a better approximation of the expected exponential decay distribution of the HNL. The performance is shown for events where the reconstruction chain was successfully run, the event selection criteria up to the final selection level of low energy analyses are fulfilled, and the reconstructed energy of both cascades is above 3.0 GeV.

Energy Resolutions

The energy resolution is inspected by looking at the 2-dimensional distribution of reconstructed energy versus the true energy as shown in Figure 2.3. The bin entries are shown as well as the median and ±25 % calculated per vertical column, to get an idea of the distribution for a given energy slice. The color scale is showing the PDF along each true energy slice, which is the full information highlighted by the median ±25 % quantile lines. The reconstructed energy is only the energy that is observable from photons, while the true energy is the total cascade energy, including the parts that go into EM neutral particles that do not produce light. It is therefore expected that the reconstructed energy is lower than the true and the median therefore does not line up with the axis diagonal.

The histogram for the first cascade energy is shown on the left and above an energy of ~10 GeV the reconstruction performs well, with the median being parallel to the diagonal and the spread being small. Below this energy the reconstruction is over-estimating the true energy, which is a known effect in IceCube, where the reconstruction is biased towards higher energies around the energy detection threshold, because events that enter the sample are events with an over fluctuation in their light deposition, which makes them pass into the selection and being reconstructible in the first place.

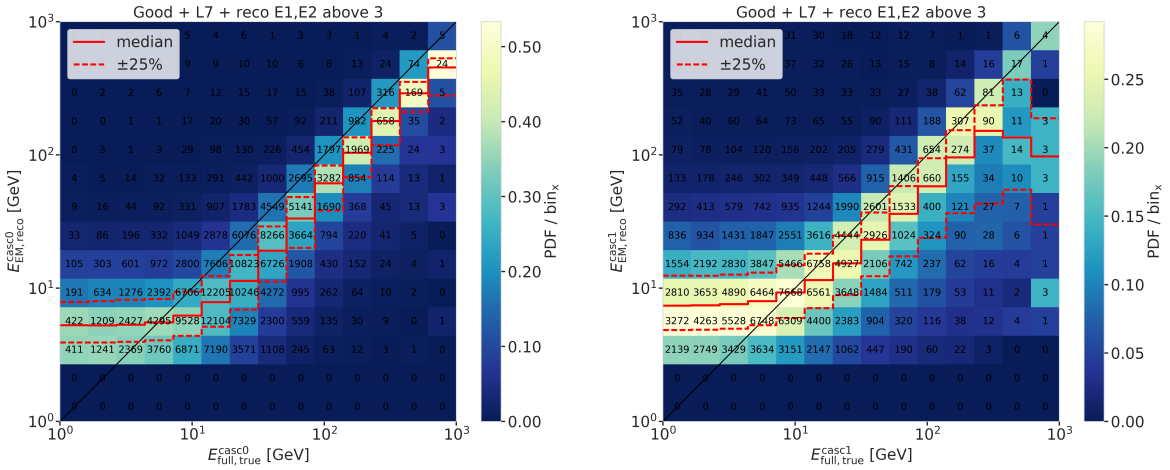


Figure 2.3: Reconstructed (EM) energy versus true energy (full) energy for the first cascade (left) and second cascade (right). The color scale is according to the PDF in each vertical true energy slice, with the solid and dashed lines showing the median ±25 % quantiles. The bin entries are shown as numbers.

For the second cascade the overall behavior is similar, only that the energy where the reconstruction starts to perform well is higher around ~20 GeV. The spread around the median is also larger and starts to expand a lot above 200 GeV, where the statistics are lower as can be seen from the bin counts. It is also very apparent that the majority of energies of the second cascade are at lower true energy values between 1 GeV and 20 GeV, which can be seen by the large bin counts.

For both cascade resolutions the effect of the reconstruction being biased towards lower values can be seen. This is due to the comparison of the full true energy to the reconstructed EM energy as mentioned before.

Length Resolutions

The decay length resolution is also investigated by looking at the 2-d histogram, where the reconstructed decay length is plotted versus the true decay length. The left part of Figure 2.4 shows the distributions after the same selection criteria from Section 2.1.3 are applied. It can be observed that for short true lengths the reconstruction is over-estimating the length, while for long true lengths the reconstruction is strongly under-estimating the length. There is a region between true lengths of 20 m and 80 m where the median reconstruction is almost unbiased, but the 50 % interquartile range is large and increasing from ~50 m to ~70 m with true decay lengths.

The over-estimation at small true lengths can be explained by multiple factors, one being that the shortest DOM spacing is ~7 m, vertically for DeepCore strings, but mostly larger than that, so resolving lengths below this is very complicated, and the reconstruction tends to be biased towards estimating the length around where the light was observed. Another reason is a similar argument to why the energies are over-estimated at small true values, namely that events that passed the selection and were reconstructed in those cases, probably have an over fluctuation in light deposition, extending further out from the vertices, so the reconstructed length is larger. Additionally, approaching a length of 0.0, the reconstructed length will of course always be a one-sided distribution, because the lengths have to be positive.

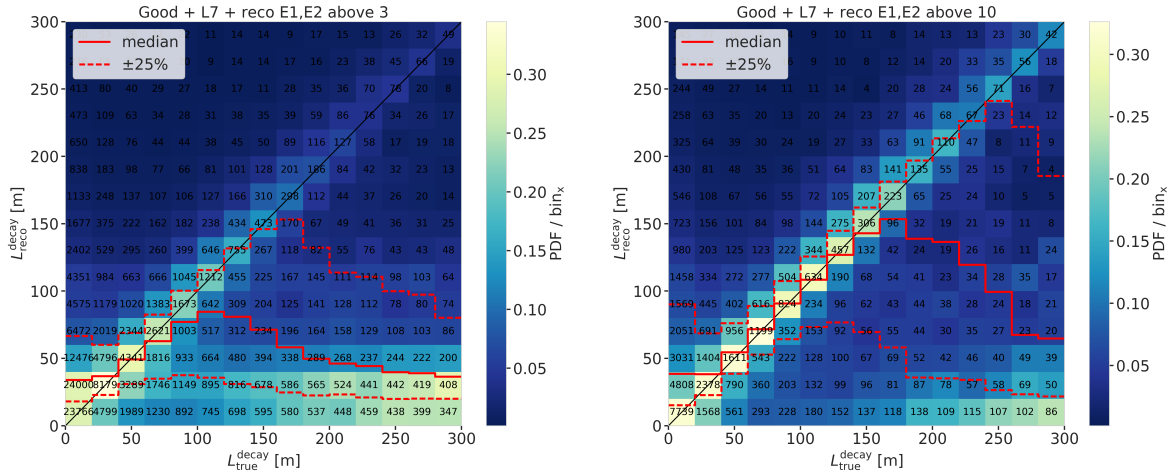


Figure 2.4: Reconstructed decay length versus true decay length for ~3 GeV (left) and ~10 GeV (right) minimum reconstructed cascade energies. The color scale is according to the PDF in each vertical true length slice, with the solid and dashed lines showing the median ±25 % quantiles. The bin entries are shown as numbers.

The under-estimation at large true lengths is more puzzling, and it seems like the distribution becomes bimodal in the reconstructed lengths, with one population around the diagonal, meaning that they are properly reconstructed, and one population at very short reconstructed lengths, which are badly reconstructed. Above 150 m the badly reconstructed population starts to dominate, and the median resolution drops off strongly. The assumption is that for these events, only one cascade was observed with enough light to be reconstructed, and the reconstruction describes the one observed cascade in two parts, separated by a short distance, driven by similar factors as mentioned before. A quick check to confirm whether this is the case, was to increase the selection criteria to minimum reconstructed cascade energies of 10 GeV, which is shown in the right part of Figure 2.4. It can be seen that the median resolution is already much better, aligning with the expectation between 40 m and 160 m. Judging from the median resolution and the spread in this range, there is very few events with an over-expectation in the energy, since both of them are aligning with the diagonal. Towards lower reconstructed lengths on the other hand, the spread is still very large, and above 200 m the badly reconstructed population starts to dominate again.

make on figure, where the two selected regions are highlighted? (RED)

fix caption (RED)

Badly Reconstructed Cascade Population

To investigate the badly reconstructed population further, a rough separation was made to find out what the cause of the difference is. It was already established that a larger reconstructed energy in both cascades, which is related to a larger true energy in form of more deposited light, leads to a better reconstruction in more events. To select the two populations, only events with true decay length larger than 80 m are used as shown in Figure 2.5, and the populations are split by the reconstructed decay length being larger or smaller than 80 m. Based on assumptions of a few potential reasons for the bad reconstruction, several variables were checked to see if there is a difference between the two populations.

The left part of Figure 2.6 shows the true horizontal distance of the second cascade from string 36. The distance is denoted as ρ_{36} and is a very good

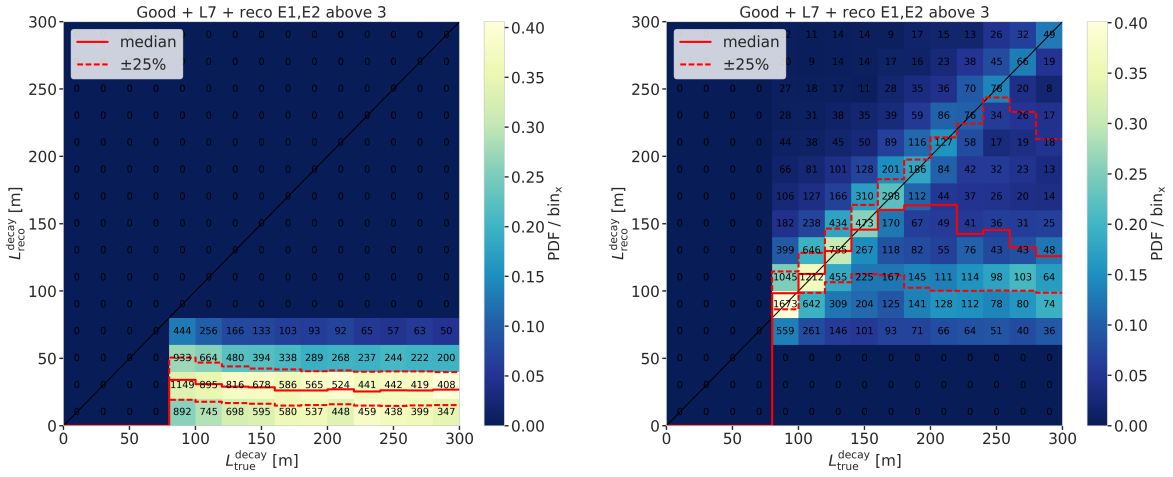


Figure 2.5

proxy for the distance to the center of the detector, because string 36 is almost at the center. While the distributions looks very similar for the first cascade (not shown), for the second cascade the badly reconstructed population extends to larger values. Considering that the DeepCore strings are roughly inside a 70 m radius from the center, and the next layer of IceCube strings is at a radius of 125 m, this is a plausible explanation for a worse reconstruction, because for the badly reconstructed population the second cascades are more often in regions without DOMs, so less or no light is observed from them.

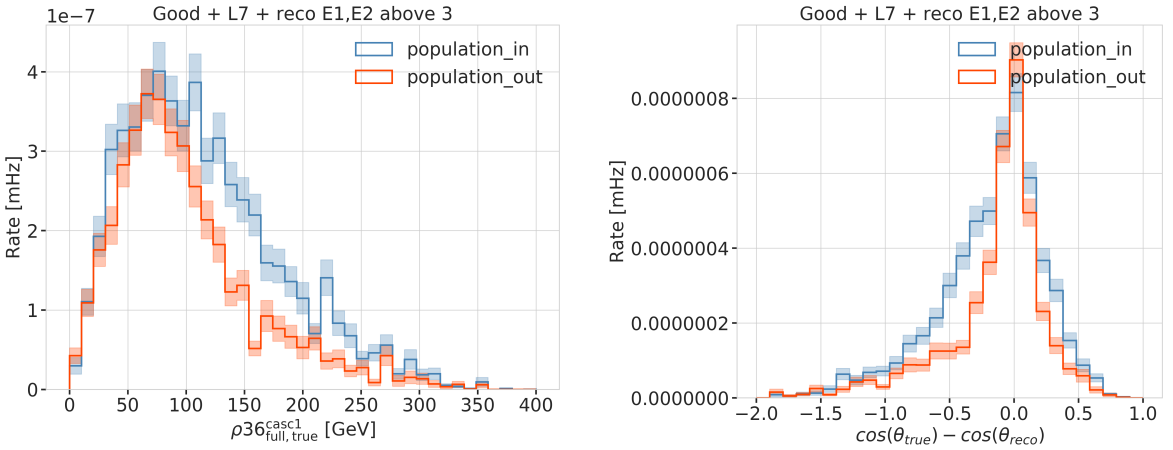


Figure 2.6

re-make normalized? (ORANGE)

fix caption (RED)

Another possible reason why the reconstruction underperforms could be that the initial seed direction was off and therefore one of the cascades cannot be found properly. Looking at the error of the cosine of the reconstructed zenith angle shown in the right of Figure 2.6, we see that the badly reconstructed population has a larger error, and is less peaked around 0.0. This could be a hint that the direction is worse for the badly reconstructed population, which could be due to a bad seed direction, or just the result of one cascade not being observed properly.

re-make normalized? (ORANGE)

fix caption (RED)

The true energies of both cascades are shown in Figure 2.5, where it can

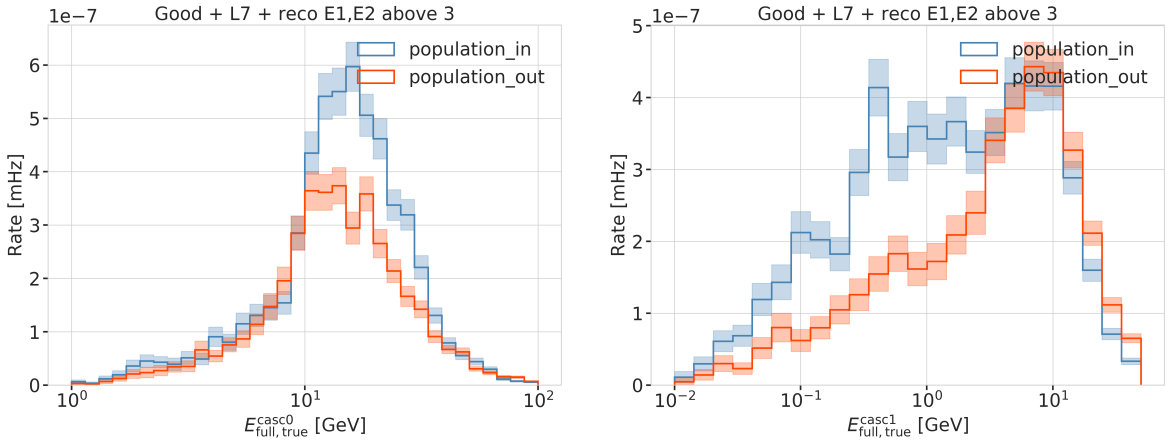


Figure 2.7

be observed that the first cascade energy is generally much larger than the second, peaking between 10 GeV and 20 GeV, while the second cascade peaks below 10 GeV. For the first cascade there is no significant difference beteween the two populations, but for the second cascade the badly reconstructed population has a larger fraction of events with lower energies and the distribution is almost uniform in the range of 2 GeV to 10 GeV, while the well reconstructed population has a peak around 10 GeV and falls off faster towards lower energies. This is a strong indication that the main reason for the bad reconstruction is the low energy of the second cascade.

Despite the fact that the split into the two populations was very rudimentary, it is clear that the main reason for the bad reconstruction is the low energy of the second cascade, while other factors, like the position of the second cascade, or the potentially bad input seed direction are also contributing. For a thorough investigation, a more sophisticated separation would be needed, but this is sufficient to conclude that the main reason for the bad reconstruction is the low energy of the second cascade.

circle back to the energy distributions I showed somewhere else, and state how the cascade energies are distributed in general (ORANGE)

2.2 Double Cascade Classification

Even though the performance results show that it is very complicated to reconstruct these low energy double cascade events, the attempt to identify them in the background of SM neutrino events was made. For this purpose a classifier was trained to distinguish between HNL *signal* events and SM neutrino *background* events, using the same preliminary sample of HNL events as was used to assess the reconstruction performance. To mitigate the effect of the bad reconstruction, a set of cuts was applied to make sure the classifier is trained on well reconstructed events, which were chosen as events with a minimum reconstructed energy of both cascades of 5 GeV and a minimum reconstructed decay length of 40 m. These cuts were applied to both signal and background events.

fix caption (RED)

Additionally, some cuts on the true energies and decay length were applied for the signal, which are a minimum true energy of both cascades of 5 GeV, and a true decay length between 40 m and 150 m. These were chosen to make sure the HNL events were theoretically double cascade like and at a

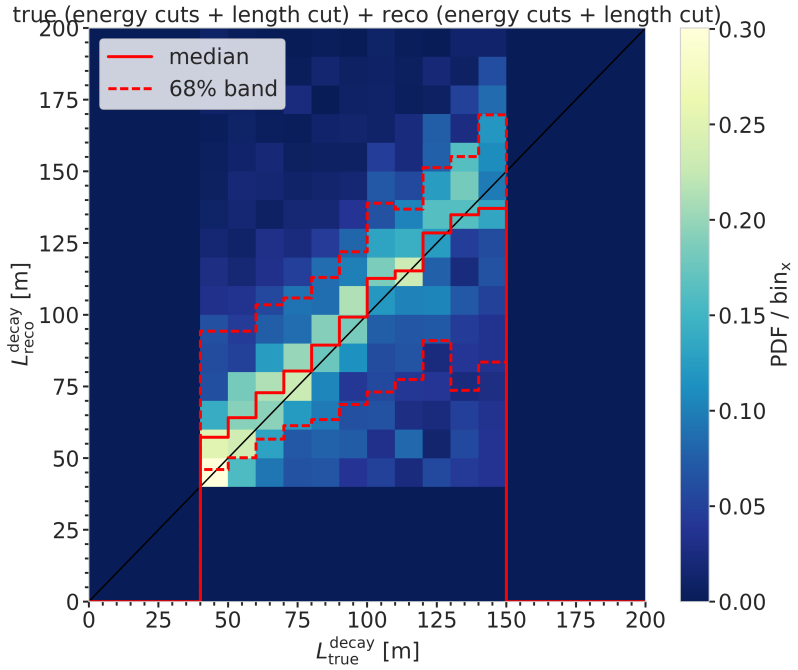


Figure 2.8

sensible length scale inside DeepCore. Figure 2.8 shows the decay length 2-d histogram after the cuts were applied.

add table with cuts? (YELLOW)

The classifier used was a *Boosted Decision Tree (BDT)* from the *scikit-learn (sklearn)* package [52] and the input features are taken from the double cascade reconstruction explained in Section 2.1 as well as some additional variables from earlier levels of the processing explained in Section 1.3. Figure 2.9 shows the distributions of two example input features, where the left plot shows the output probability of the classifier trained to distinguish track from cascade like events, which is used in the oscillation analysis, and the right plot shows the reconstructed decay length from the double cascade reconstruction. Shown are the distributions for the the HNL signal, the individual background components, and the total background.

[52]: Pedregosa et al. (2011), "Scikit-learn: Machine Learning in Python"

fix caption of this figure (RED)

Both a single classifier trained to distinguish between HNL signal events and all SM background events was tested as well as two classifiers, one trained to distinguish signal from tracks, and the other to distinguish signal from cascades, separately. Since the SM neutrino events at these energies are either track like or cascade like, the latter approach was expected to perform better. Despite the fact that several combinations of features and classifier hyperparameters were tested, it was not possible to identify a pure double cascade region with a single classifier.

Applying the two classifiers trained to distinguish signal from track and signal from cascade, it is possible to select a region with only signal events. This is visualized in Figure 2.10, where the probabilities of 1 mean very signal like, and only the regions close to 1 are shown for both outputs, to highlight the interesting region, where a pure HNL sub-sample can be selected. When applying a weight to the signal events though, the expected rate of the selected events is very low. Assuming a highly optimistic mixing of 1, it would take more than 20 years of data taking to observe a single event and with this low simulation statistics the prediction is not very reliable, either. Making a weaker cut to select a signal like region will contain a

Make a single plot, showing S/B or so? (RED)

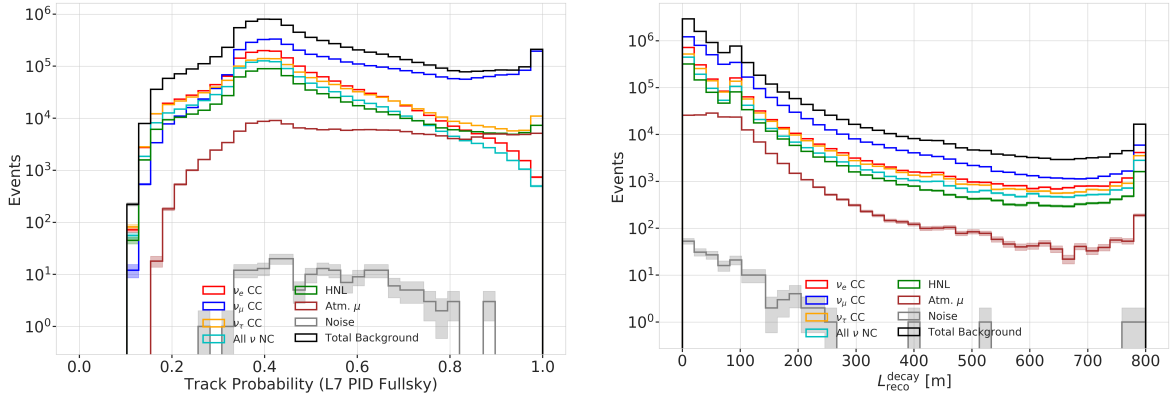


Figure 2.9

large amount of background events, which dominate over the signal at ~ 2 orders of magnitude for a mixing of 0.1. The conclusion from this is, that with the current selection and reconstruction chain and a classical BDT, it is not possible to distinguish signal events at a level feasible to perform an analysis.

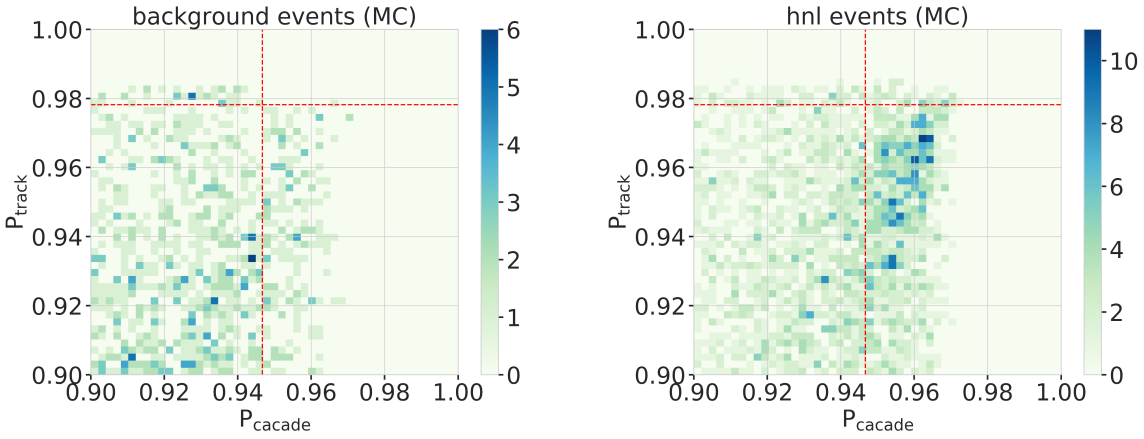


Figure 2.10

fix caption of this figure (RED)

describe the "good fit" selection when I first show plots that use it (ORANGE)

2.3 Generalized Double Cascade Performance

All the above results were obtained using preliminary development versions of the model dependent HNL simulation. To investigate the effect of the low energy event selection and the double cascade reconstruction performance in a more generic way, the model independent simulation introduced in Section **??** was used to perform the same tests and a series of additional checks. The important advantage of the model independent samples is the controllable parameter space, especially in cascade energies and decay length, because the event kinematics are not coupled to the underlying HNL model, but can be chosen freely. This means that some benchmark edge cases can be investigated, and the performance can also be assessed for a realistic scenario in addition to mapping out the effects of the event selection and where the reconstruction breaks down.

2.3.1 Idealistic Events

The *best case* scenario to observe an event is to be directly on top of a string with a straight up-going direction. Using the simulation sample introduced in Section **??** and running the double cascade reconstruction from Section 2.1 on these events, it is possible to estimate the performance limit of the reconstruction. Figure 2.11 shows one example event view from that sample, where the cascade energies are 2.4 GeV and 4.9 GeV, and the decay length is 20 m. It can be seen that despite the low energies, both cascades deposit light in the DOMs and the reconstruction is expected to work.

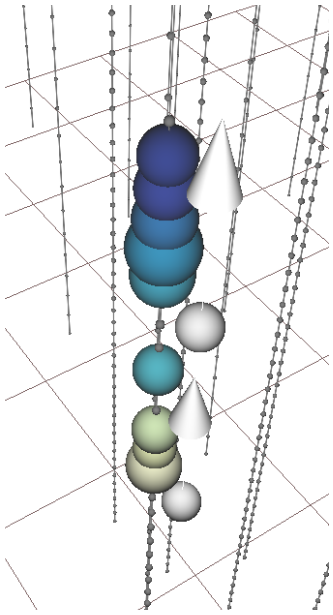


Figure 2.11

fix caption (RED)

Check this number (RED)

The performance of the length reconstruction is shown in Figure 2.12, where the median of the absolute, fractional decay length resolution is shown on the left and the 2-d histogram of the reconstructed versus the true decay length is shown on the right. For these results and the following, all events that were reconstructed with non-zero cascade energies and non-zero decay length are used, and the events are unweighted. The length is very well reconstructed, with the median resolution being below 30 % above a true decay length of ~10 m, and falling off with increasing true length, down to ~10 % at 100 m.

The 2-d histogram shows that there is no under-estimation of the length up to a true decay length of ~210 m, which shows that if there are DOMs in the region between the two cascades that have not observed any light, the reconstruction is very stable. Considering the underlying Poisson likelihood in Equation 2.1 used for the reconstruction, this makes sense, since DOMs being present, but not observing any light is affecting the light expectation that goes into the likelihood and therefore makes these hypotheses unlikely and therefore incompatible with the data.

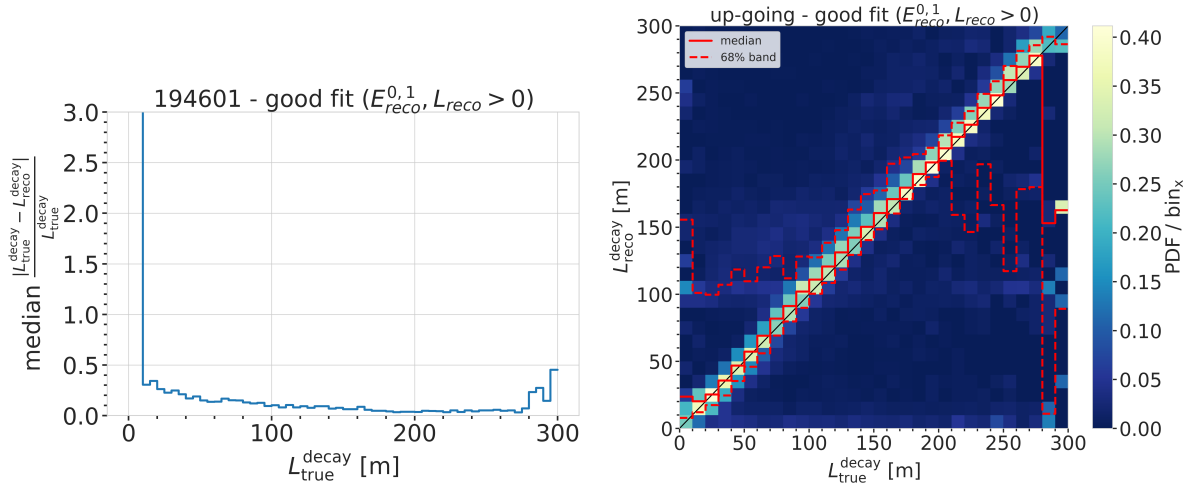


Figure 2.12

fix caption of this figure (RED)

2.3.2 Realistic Events

The sample of HNL events introduced in Section **??**, which is a more realistic representation of the expected HNL events, but still offers more controlled energy and length distributions, is used to investigate the selection efficiency, to cross check the reconstruction performance, and to benchmark the limits

where the reconstruction breaks down. An example event view is shown in Figure 2.13, for cascade energies of 30.8 GeV and 25.3 GeV, and a decay length of 100 m. Here it can be seen that even for these higher energies, if the cascades are further away from the strings, only individual photons are observed in the DOMs, and detecting and reconstructing them gets significantly more challenging.

Check this number (RED)

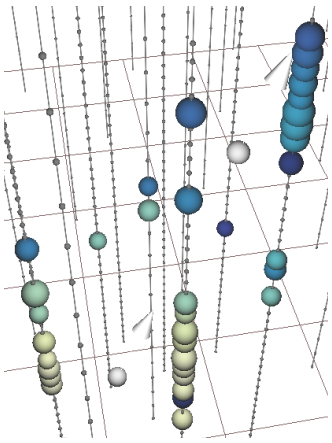


Figure 2.13

To assess the efficiency of the low energy event selection introduced in Section 1.3, the energy and length distributions are shown across the different selection levels in Figure **??**. Table **??** shows the total efficiency of the selection, where it can be seen that at level xx it is reduced the most and only xx% of the events pass the selection to level 5.

Looking at the energy distributions in Figure 2.14 shows a similar behavior to the results discussed in Section 2.1.3, but here there is no bias in the reconstructed energy, because the events are simulated as EM cascades, which means all energy is deposited in light and can be reconstructed. Above around 5 GeV to 6 GeV the median is very stable, and the 1-sigma resolution band is 50 % narrow and decreasing with energy down to 20 % at 100 GeV. Interestingly, the second cascade energy reconstruction performs slightly worse, although they have the same energy ranges for this sample. This could hint at an asymmetry in the reconstruction process, which might relate to how the two cascades are parameterized, or be due to the different positions and the dominantly up-going direction used in the sampling combined with the DOMs looking down. The total energy resolution shown in the left part of Figure 2.15 is very good, above 10 GeV it is unbiased and the 1-sigma resolution band is below 20 %.

fix caption (RED)

Make/add relevant plots here (RED)

Plot energy (true total)

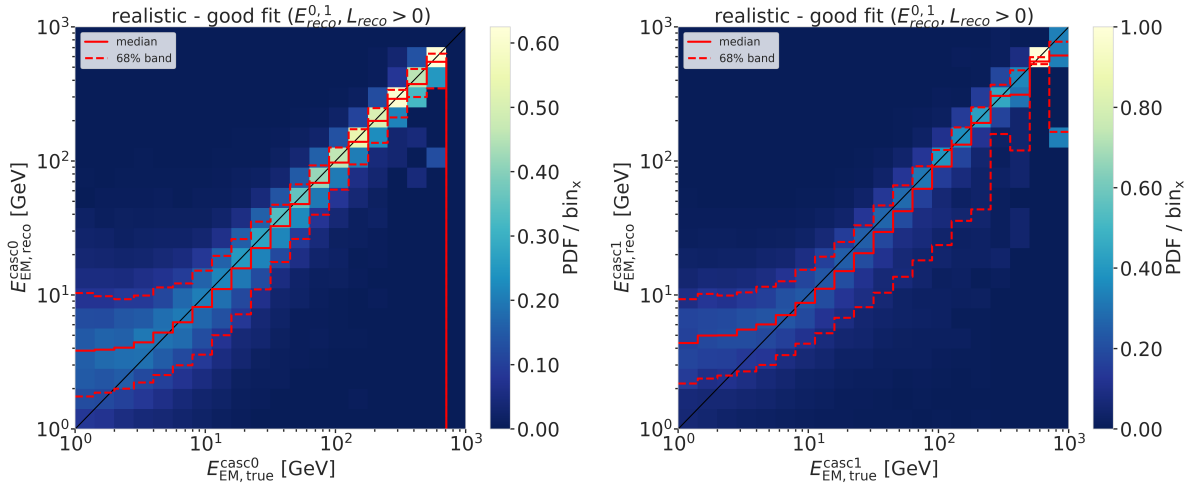


Figure 2.14

fix caption of this figure (RED)

The decay length resolution shown in the right part of Figure 2.15 looks similarly bad to the results discussed in Section 2.1.3 and shows the same features with a region between 20 m and 80 m where it is roughly unbiased, but the 1-sigma resolution band is wide with a lot of outliers towards short reconstructed lengths. Below 20 m the reconstructed lengths are always over-estimating the true and above 80 m a population of events start to dominate where the decay lengths is not getting reconstructed at all, as investigated before.

fix caption of this figure (RED)

To get an estimate of what minimum energies are necessary for the recon-

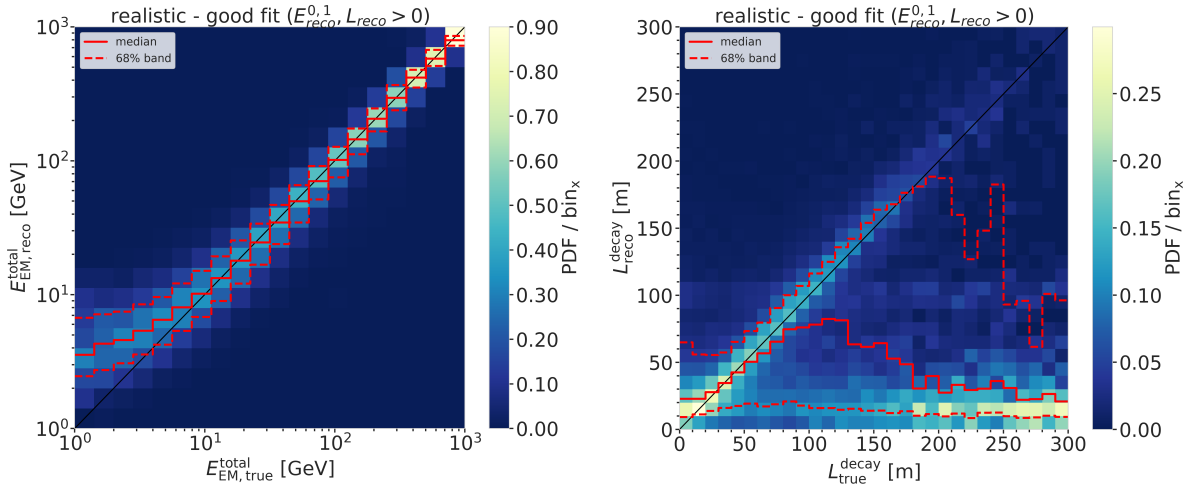


Figure 2.15

struction to perform reasonably well, the fractional decay length resolution is shown as a function of the total true energy and the minimum energy of both individual cascades in Figure 2.16. In the left part it can be seen that the median of the decay length resolution stabilizes around 0 for a total energy above 20 GeV, but the spread of the distribution is still quite large with a 1-sigma band of 80 % to 100 %, decreasing down to ~60 % at 100 GeV. Based on the right part of the figure, the decay length resolution starts to be unbiased for a minimum energy of any cascade of 7 GeV, with an equivalently large spread. A rough takeaway from this is that the decay length reconstruction is not very reliable for events with one cascade energy below 7 GeV and with a total energy below 20 GeV.

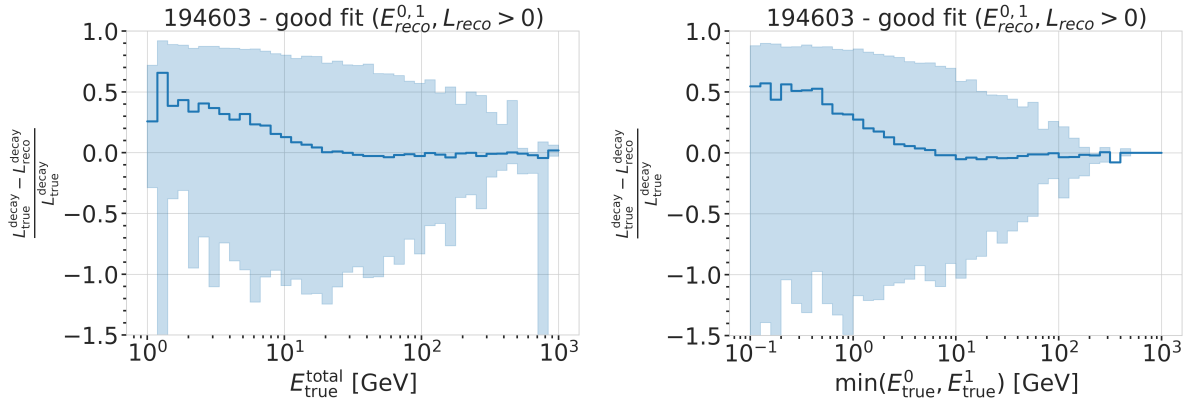


Figure 2.16

fix caption of this figure (RED)

Search for Tau Neutrino Induced Heavy Neutral Lepton Events

3

This chapter describes the search for HNL events using 10 years of IceCube DeepCore data. The expected number of HNL events in the data sample depends on the mass of the additional heavy state, m_4, and the mixing element $|U^2_{\alpha 4}|$, with $\alpha = e, \mu, \tau$, between the SM flavors and the new mass state. As discussed in Section **??**, this work focuses on the mixing to the tau sector, $|U^2_{\tau 4}|$, which has the weakest constraints to date. Since the mass itself influences the production and decay kinematics of the event and the accessible decay modes, individual mass samples were produced as described in Section **??**. The mass influences the decay length and energy distributions, while the mixing both changes the overall expected rate of the HNL events and the shape in energy and length. We perform three independent searches for each mass sample, where the mixing is measured in each of the fits.

3.1 Final Level Sample

The final level simulation sample of this analysis consists of the neutrino and muon MC introduced in Section 1.1 and one of the three HNL samples explained in Section **??**, while the data are the events measured in 10 years of IceCube DeepCore data taking. All simulation and the data are processed through the full event selection chain described in Section 1.3 and Section 1.4 leading to the final level sample. As described in Section 1.4.2, event triggers consisting purely of random coincidences induced by noise in the DOMs have been reduced to a negligible rate, and will not be discussed further.

To get the neutrino expectation, the MC events are weighted according to their generation weight introduced in Section 1.1.1, multiplied by the total lifetime, and the expected neutrino flux. For the correct expectation at the detector, the events have to be weighted by the oscillation probability, depending on their energy and their distance traveled from the atmosphere to the detector. The oscillation probabilities are calculated using a PYTHON implementation of the calculations from [53], which use the matter profile of the Earth following the *Preliminary Reference Earth Model (PREM)* [54] as input. Apart from the energy and the distance, the two relevant parameters defining the oscillation probabilities are the atmospheric neutrino oscillation parameters θ_{23} and Δm^2_{31}. Since the HNL events originate from the tau neutrinos that were produced as muon neutrinos in the atmosphere and then oscillated into ν_τ, this weighting is also applied in addition to the specific weighting scheme for the HNL events described in Section **??**, which itself is defined by the mixing $|U^2_{\tau 4}|$ and the mass m_4.

[53]: Barger et al. (1980), "Matter effects on three-neutrino oscillations"
[54]: Dziewonski et al. (1981), "Preliminary reference Earth model"

3.1.1 Expected Rates/Events

The rates and the expected number of events for the SM background are shown in Table 3.1 with around 175000 total events expected in the 10 years. Only data marked as good is used for the analysis, where *good* refers to

measurement time with the correct physics run configuration and without other known issues. The resulting good detector livetime in this data taking period was 9.28 years. The rates are calculated by summing the weights of all events in the final level sample, while the uncertainties are calculated by taking the square root of the sum of the weights squared. The expected number of events is calculated by multiplying the rate with the livetime. The individual fractions show that this sample is neutrino dominated where the majority of events are ν_μ-CC events.

Table 3.1: Final level rates and event expectation of the SM background particle types.

Type	**Rate** [mHz]	**Events** (9.28 years)	**Fraction** [%]
ν_μ^{CC}	0.3531	103321 ± 113	58.9
ν_e^{CC}	0.1418	41490 ± 69	23.7
ν^{NC}	0.0666	19491 ± 47	11.1
ν_τ^{CC}	0.0345	10094 ± 22	5.8
μ_{atm}	0.0032	936 ± 15	0.5
total	0.5992	175332 ± 143	100.0

Table 3.2 shows the rates and expected number of events for the HNL signal simulation. The expectation depends on the mass and the mixing and shown here are two example mixings for all the three masses that are being tested in this work. A mixing of 0.0 would result in no HNL events at all. It can already be seen that for the smaller mixing of $|U_{\tau 4}|^2 = 10^{-3}$ the expected number of events is very low, while at the larger mixing of $|U_{\tau 4}|^2 = 10^{-1}$ the number is comparable to the amount of atmospheric muons in the background sample.

Table 3.2: Final level rates and event expectations of the HNL signal for all three masses and two example mixing values.

HNL mass	**Rate** [µHz]	**Events (in** 9.28 years**)**
$\lvert U_{\tau 4}\rvert^2 = 10^{-1}$		
0.3 GeV	3.3	975 ± 2
0.6 GeV	3.1	895 ± 2
1.0 GeV	2.5	731 ± 2
$\lvert U_{\tau 4}\rvert^2 = 10^{-3}$		
0.3 GeV	0.006	1.67 ± 0.01
0.6 GeV	0.022	6.44 ± 0.01
1.0 GeV	0.025	7.27 ± 0.01

3.1.2 Analysis Binning

[30]: Yu et al. (2023), "Recent neutrino oscillation result with the IceCube experiment"

An identical binning to the analysis performed in [30] is used. In total, there are three bins in PID (cascade like, mixed, and track like), 12 bins in reconstructed energy, and 8 bins in cosine of the reconstructed zenith angle as specified in Table 3.3. Extending the binning towards lower energies or

Add fractions of the different particle types in the bins for benchmark mass/mixing (another table?) (ORANGE)

Table 3.3: Three-dimensional binning used in the analysis. All variables are from the PLERCNN reconstruction) explained in Section 1.4.

Variable	N_{bins}	**Edges**	**Spacing**
P_ν	3	[0.00, 0.25, 0.55, 1.00]	linear
E	12	[5.00, 100.00]	logarithmic
$\cos(\theta)$	8	[-1.00, 0.04]	linear

increasing the number of bins in energy or cosine of the zenith angle did not improve the HNL sensitivities significantly, because the dominant signal region is already covered with a sufficiently fine binning to observe the shape and magnitude of the HNL events on top of the SM background. This

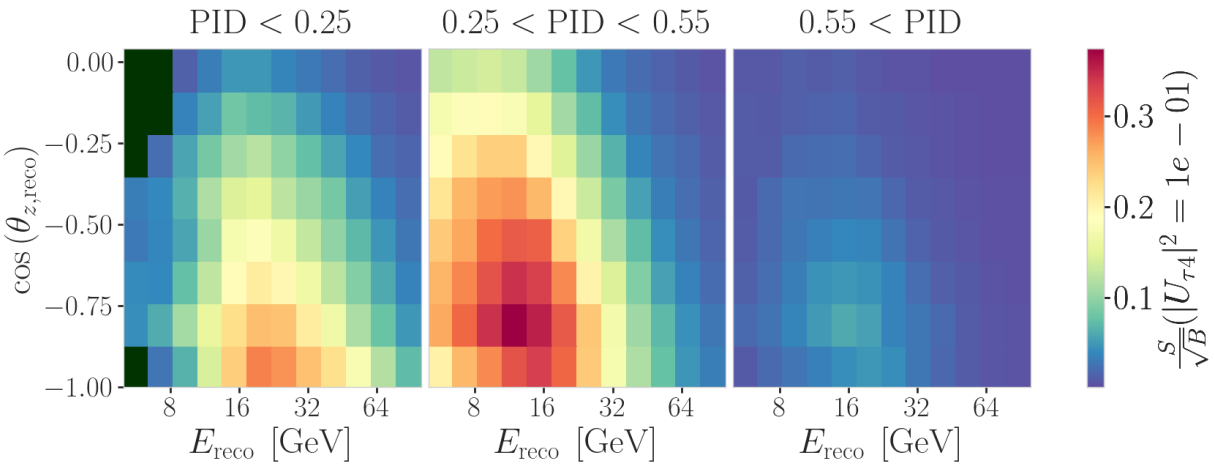


Figure 3.1: Signal over square root of background expectation in 9.28 years for the 1.0 GeV mass sample at a mixing of 0.1, while all other parameters are at their nominal values.

can be seen in the middle panel of Figure 3.1, which shows the expected signal events divided by the square root of the expected background events for every bin used in the analysis. The signal expectation is using the 1.0 GeV mass sample at a reference mixing of 0.1, with the corresponding three dimensional histogram shown in Figure B.1. Both the nominal background expectation used to calculate the signal to square root of background ratio and the detector data can be seen in Figure 3.2.

Some low energy bins in the cascade like region have very low MC expectations (<1 event) and are therefore not taken into account in the analysis, to prevent unwanted behavior in the fit. Those are shown in dark green in the three dimensional histograms, and both background and data histograms show a strong decrease of events towards low energies in the cascade like bin. This background expectation is not necessarily supposed to agree with the data, because this is the distributions assuming nominal parameter values, before performing the fit to find the parameters that describe the data best. All parameters used in the analysis are discussed in Section 3.2.2, and post-fit data to MC comparisons are shown in Section 3.3.3.

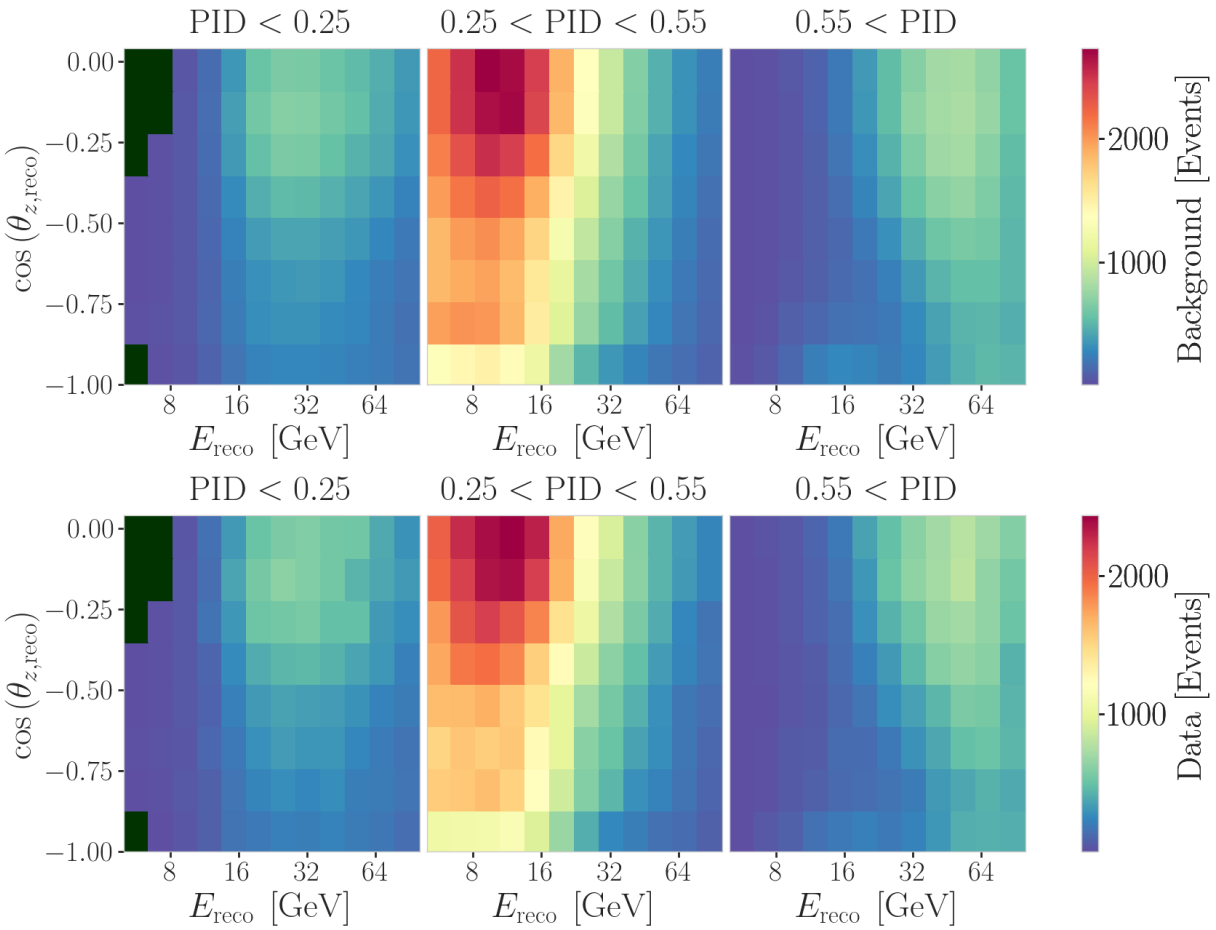


Figure 3.2: Background expectation in 9.28 years for all other parameters are at their nominal values (top) and observed data (bottom).

3.2 Statistical Analysis

3.2.1 Test Statistic

The measurements are performed by comparing the weighted MC to the data. Through variation of the nuisance and physics parameters that govern the weights, the best matching set of parameters can be found, by optimizing a fit metric. The comparison is done using a modified χ^2, defined as

$$\chi^2_{\text{mod}} = \sum_{i \in \text{bins}} \frac{(N_i^{\text{exp}} - N_i^{\text{obs}})^2}{N_i^{\text{exp}} + (\sigma_i^{\nu})^2 + (\sigma_i^{\mu})^2 + (\sigma_i^{\text{HNL}})^2} + \sum_{j \in \text{syst}} \frac{(s_j - \hat{s}_j)^2}{\sigma_{s_j}^2}, \tag{3.1}$$

as the fit metric. It is designed such that taking the difference between a free fit and a fit with fixed parameters based on a chosen hypothesis, $\Delta\chi^2_{\text{mod}}$, can directly be used as a *test statistic (TS)* for hypothesis testing, due to its asymptotic behavior. The total even expectation is $N_i^{\text{exp}} = N_i^{\nu} + N_i^{\mu} + N_i^{\text{HNL}}$, where N_i^{ν}, N_i^{μ}, and N_i^{HNL} are the expected number of events in bin i from neutrinos, atmospheric muons, and HNLs, while N_i^{obs} is the observed number of events in the bin. The expected number of events from each particle type is calculated by summing the weights of all events in the bin $N_i^{\text{type}} = \sum_i^{\text{type}} \omega_i$, with the statistical uncertainty being $(\sigma_i^{\text{type}})^2 = \sum_i^{\text{type}} \omega_i^2$. The additional term in Equation 3.1 is included to apply a penalty term for

prior knowledge of the systematic uncertainties of the parameters where they are known. s_j are the systematic parameters that are varied in the fit, while $\hat{s_j}$ are their nominal values and σ_{s_j} are the known uncertainties.

3.2.2 Physics Parameters

The variable physics parameter in this analysis is the mixing between the HNL and the SM τ sector, $|U_{\tau 4}|^2$. It is varied continuously in the range of [0.0 , 1.0] by applying the weighting scheme described in Section **??**. The fit is initialized at an off-nominal value of 0.1. The other physics parameter, the mass m_4 of the HNL, is implicitly fixed to one of the three discrete masses to be tested, by using the corresponding sample of the HNL simulation described in Section **??**.

3.2.3 Nuisance Parameters

All systematic parameters introduced in Section 1.5 apart from the detector calibration uncertainties, are already parameterized in a continuous way and can be varied in the fit. To be able to do the same with the detector uncertainties, a novel method is applied that will briefly be introduced here before going into the selection of the free parameters.

Treatment of Detector Systematic Uncertainties

Since the variations related to the detector calibration uncertainties introduced in Section 1.5.3 are estimated by simulating MC at discrete values of the systematic parameters, a method to derive continuous variations is needed to perform the fit. The method applied here was initially introduced in [55] and first used in the low energy sterile neutrino search in [16] (section 7.4.3). Using a *likelihood-free inference* technique, re-weighting factors are found for every event in the nominal MC sample, given a specific choice of detector systematic parameters. These factors quantify how much more or less likely the event would be for the corresponding change in detector response from the nominal parameters. Without going into the details of the method, which were already exhaustively discussed in [55] and [16], the performance is assessed here for the HNL signal simulation. In order to do so, the weights are applied to the nominal MC samples, choosing the detector systematic values used to produce the discrete samples and the resulting event expectations are compared to the expectations from the individual, discrete MC samples. The bin counts are compared by calculating the pull defined as

[55]: Fischer et al. (2023), "Treating detector systematics via a likelihood free inference method"
[16]: Trettin (2023), "Search for eV-scale sterile neutrinos with IceCube DeepCore"

$$p = \frac{N_{\text{reweighted}} - N_{sys}}{\sqrt{\sigma^2_{\text{reweighted}} + \sigma^2_{sys}}} , \tag{3.2}$$

where N are the bin-wise event expectations and σ are their MC uncertainty. For the SM BG simulation, the performance was already investigated in [23] (section 7.4.4, appendix B5) and the re-weighted nominal MC was shown to be in agreement with the discrete systematic sets at a sufficient level. Figure **??** shows the bin-wise pulls for the 1.0 GeV HNL mass sample at a mixing of 0.1 for a selection of the discrete systematic samples, where the DOM efficiency and the bulk ice absorption was varied by ±10%. The distributions

[23]: Lohfink (2023), "Testing non-standard neutrino interaction parameters with IceCube-DeepCore"

should follow a standard normal distribution, and no strong clustering or systematic deviations. It can be seen that spread of the distribution is slightly larger than 1.0 and the center is close to 0.0 as expected. A similar performance is found for the additional systematic sets and can be found in Section **??**.

add bin-wise pulls and pull distribution for selection of sets and rest to backup (RED)

Free Parameters

To decide which systematic uncertainties should be included in the fit, we test the potential impact they have on the TS if they are neglected. The test is performed by creating Asimov data using the BG simulation and the HNL simulation of the 1.0 GeV mass sample at a mixing value of 0.1, which is chosen as a benchmark physics parameter, but the explicit choice does not have a significant impact on the test. The systematic parameter of interest is set to a value above its nominal expectation, either pulled up by $+1\sigma$ or by an educated estimate for parameters without a well-defined uncertainty. A fit is performed fixing the systematic parameter of interest and leaving all additional parameters free. The resulting TS is the fit metric difference between between this fit and a fit with all parameters free, which would result in a fit metric of 0.0 for this Asimov test. This difference is called mis-modeling significance and parameters below a significance of 0.1 σ are fixed. The test is performed in an iterative manner until the final set of free parameters is found.

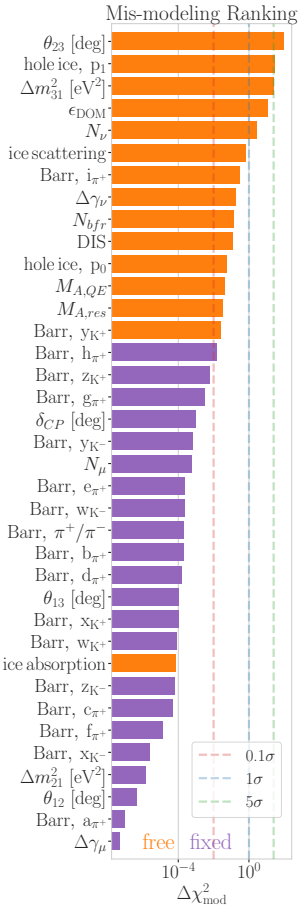


Figure 3.3: Mis-modeling impact ranking of the systematic parameters. The mis-modeling is calculated as the fit metric difference between a fit with the parameter fixed at its nominal value and a fit with the parameter pulled up by $+1\sigma$. The test was performed using Asimov data of the 1.0 GeV mass sample at a reference mixing of 0.1.

Figure 3.3 shows the resulting significances of one of these tests. The parameters tested are the systematic parameters introduced in Section 1.5 and the atmospheric oscillation parameters mentioned in Section 3.1. In the final selection of free parameters the Barr h_{π^+} parameter was also left free, to sufficiently cover the relevant energy production range of the Pions, as can be seen in Figure 1.6, where both for Kaons and Pions the uncertainties are included for primary energies above 30 GeV and $x_{\mathrm{lab}} > 0.1$. Additionally, the ice absorption is still kept free, despite showing a small significance, which is done because the bulk ice parameters are not well constrained and are known to have a large impact, which might be concealed in this idealized test, due to correlations with the other parameters. In this test, the effect of correlations is challenging to consider, because only the impact of one parameter is tested at a time, using the overall mis-modeling significance as a measure. The mis-modeling could be reduced by a correlated parameter capturing the effect of the parameter of interest. For this reason a very conservative threshold of 0.1 σ is chosen and some parameters below the threshold are still left free in the fit.

All nuisance parameters that are left free in the fit are summarized in Table 3.4, showing their nominal values, the allowed fit ranges, and their Gaussian prior, if applicable. The scaling parameter N_ν is included to account for the overall normalization of the neutrino rate, and it has the identical effect on the SM neutrino events and the BSM HNL events, because they both originate from the same neutrino flux. Despite being known to ~5 % in this energy range, there is no prior applied to this parameter, because the fit itself is able to constrain it well, which can be seen by the large impact it shows in Figure 3.3. Concerning the atmospheric neutrino flux, the CR power law flux correction factor $\Delta\gamma_\nu$ introduced in Section 1.5.1 is included with nominal value of 0.0 which corresponds to the baseline flux model by Honda *et al* [2]. A slightly conservative prior of 0.1 is applied to the parameter, while latest

Need cite here! (RED)

[2]: Honda et al. (2015), "Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model"

Parameter	Nominal	Range	Prior
$\theta_{23}[^\circ]$	47.5047	[0.0, 90.0]	-
$\Delta m^2_{31}[\text{eV}^2]$	0.002475	[0.001, 0.004]	-
N_ν	1.0	[0.1, 2.0]	-
$\Delta\gamma_\nu$	0.0	[-0.5, 0.5]	0.1
Barr h_{π^+}	0.0	[-0.75, 0.75]	0.15
Barr i_{π^+}	0.0	[-3.05, 3.05]	0.61
Barr y_{K^+}	0.0	[-1.5, 1.5]	0.3
DIS	0.0	[-0.5, 1.5]	1.0
$M_{\text{A,QE}}$	0.0	[-2.0, 2.0]	1.0
$M_{\text{A,res}}$	0.0	[-2.0, 2.0]	1.0
ϵ_{DOM}	1.0	[0.8, 1.2]	0.1
hole ice p_0	0.101569	[-0.6, 0.5]	-
hole ice p_1	-0.049344	[-0.2, 0.2]	-
bulk ice absorption	1.0	[0.85, 1.15]	-
bulk ice scattering	1.05	[0.9, 1.2]	-
N_{bfr}	0.0	[-0.2, 1.2]	-

Table 3.4: Systematic uncertainty parameters that are left free to float in the fit. Their allowed fit ranges are shown with the nominal value and the Gaussian prior width if applicable.

measurements show an uncertainty of 0.05 [36]. The Barr parameters are contrained by a Gaussian prior, taken from [38]. All the detector systematic uncertainties discussed in Section 1.5.3 are included in the fit. The DOM efficiency ϵ_{DOM} is constrained by a Gaussian prior with a width of 0.1, which is a conservative estimate based on the studies of the optical efficiency using minimum ionizing muons from [44, 45]. The two atmospheric neutrino oscillation parameters θ_{23} and Δm^2_{31} are also included in the fit with nominal values of 47.5° and 2.48×10^{-3} eV^2 [30], respectively. Since they govern the shape and the strength of the tau neutrino flux, by defining the oscillation from ν_μ to ν_τ, they are also relevant for the HNL signal shape.

[36]: Evans et al. (2017), "Uncertainties in atmospheric muon-neutrino fluxes arising from cosmic-ray primaries"

[38]: Barr et al. (2006), "Uncertainties in Atmospheric Neutrino Fluxes"

[44]: Feintzeig (2014), "Searches for Point-like Sources of Astrophysical Neutrinos with the IceCube Neutrino Observatory"

[45]: Kulacz (2019), "In Situ Measurement of the IceCube DOM Efficiency Factor Using Atmospheric Minimum Ionizing Muons"

[30]: Yu et al. (2023), "Recent neutrino oscillation result with the IceCube experiment"

I could add some final level effects of some systematics on the 3D binning and maybe discuss how they are different from the signal shape, or so? (ORANGE)

3.2.4 Low Energy Analysis Framework

The analysis is performed using the PISA [56] [57] software framework, which was developed to perform analyses of small signals in high-statistics neutrino oscillation experiments. It is used to generate the expected event distributions from several MC samples, which can then be compared to the observed data. The expectation for each MC sample is calculated by applying physics and nuisance parameter effects in a stage-wise manner, before combining them to the final expectation.

[56]: Aartsen et al. (2020), "Computational techniques for the analysis of small signals in high-statistics neutrino oscillation experiments"

3.3 Analysis Checks

Fitting to data is performed in a *blind* manner, where the analyzer does not immediately see the fitted physics and nuisance parameter values, but first checks that a set of pre-defined *goodness of fit (GOF)* criteria are fulfilled. This is done to circumvent the so-called *confirmation bias* [58], where the analyzer might be tempted to construct the analysis in a way that confirms their expectation. After the GOF criteria are met to satisfaction, the fit results are unblinded and the full result can be revealed. Before these blind fits to data are performed, the robustness of the analysis method is tested using pseudo-data that is generated from the MC.

[58]: Nickerson (1998), "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises"

3.3.1 Minimization Robustness

To find the set of parameters that best describes the data, a staged minimization routine is used. In the first stage, a fit with coarse minimizer settings is performed to find a rough estimate of the *best fit point (BFP)*. In the second stage, the fit is performed again in both octants[1] of θ_{23}, starting from the BFP of the coarse fit. For each individual fit the *MIGRAD* routine of IMINUIT [59] is used to minimize the χ^2_{mod} fit metric defined in Equation 3.1. Iminuit is a fast, python compatible minimizer based on the MINUIT2 C++ library [60]. The individual minimizer settings for both stages are shown in Table 3.5.

1: There is a degeneracy between the lower octant ($\theta_{23} < 45°$) and the upper octant ($\theta_{23} > 45°$), which can lead to fit metric minima (local and global) at two positions that are mirrored around 45° in θ_{23}.

[59]: Dembinski et al. (2022), *scikit-hep/iminuit: v2.17.0*

[60]: James et al. (1975), "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations"

Fit	Err.	Prec.	Tol.
Coarse	1e-1	1e-8	1e-1
Fine	1e-5	1e-14	1e-5

Table 3.5: Migrad settings for the two stages in the minimization routine. *Err.* are the step size for the numerical gradient estimation, *Prec.* is the precision with which the LLH is calculated, and *Tol.* is the tolerance for the minimization.

To test the minimization routine and to make sure it consistently recovers any physics parameters, pseudo-data sets are produced from the MC by choosing the nominal nuisance parameters and specific physics parameters, without adding any statistical or systematic fluctuations to it. These so-called *Asimov*[2] data sets are then fit back with the full analysis chain. This type of test is called *Asimov inject/recover test*. A set of mixing values between 10^{-3} and 10^{0} is injected and fit back. Without fluctuations the fit is expected to always recover the injected parameters (both physics and nuisance parameters). The fitted mixing values from the Asimov inject/recover tests are compared to the true injected values in Figure 3.4 for all three mass samples. As desired, the fit is always able to recover the injected physics parameter and the nuisance parameters within the statistical uncertainty or at an insignificant fit metric difference.

FInd first occurance of "Asimov" and add reference and explain it there (RED)

2: A pseudo-data set without statistical fluctuations is called Asimov data set.

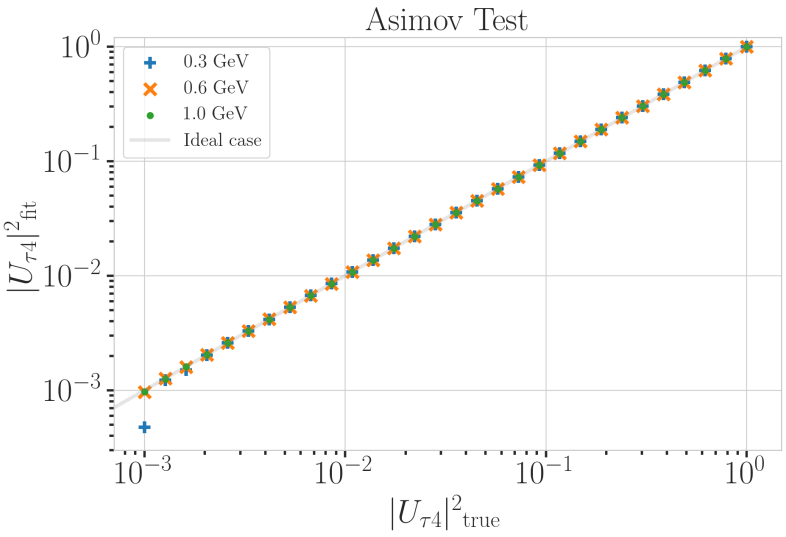


Figure 3.4: Asimov inject/recover test results for all three mass samples. Mixing values between 10^{-3} and 10^{0} are injected and fit back with the full analysis chain. The injected parameter is always recovered within the statistical uncertainty or at an insignificant fit metric difference.

3.3.2 Goodness of Fit

To estimate the GOF, pseudo-data is generated from the MC by injecting the BFP parameters as true parameters and then fluctuating the expected bin counts to account for MC uncertainty and Poisson fluctuations in data. First, the expectation value of each bin is drawn from a Gaussian distribution centered at the nominal expectation value with a standard deviation corresponding to the MC uncertainty of the bin. Based on this sampled expectation value, each bin count is drawn from a Poisson distribution, independently, to get the final pseudo-data set. These pseudo-data sets are

analyzed with the same analysis chain as the real data, resulting in a final fit metric value for each pseudo-data set. By comparing the distribution of fit metric values from this *ensemble* of pseudo-data trials to the fit metric of the fit to real data, a p-value can be calculated. The p-value is the probability of finding a value of the fit metric at least as large as the one from the data fit. Figure 3.5 shows the distribution from the ensemble tests for the 0.6 GeV mass sample and the observed value from the fit, resulting in a p-value of 28.5 %. The p-values for the 0.3 GeV and 1.0 GeV are 28.3 % and 26.0 %, respectively, and the corresponding plots are shown in Section **??**. Based on this test, it is concluded that the fit result is compatible with the expectation from the ensemble of pseudo-data trials.

Add 3D BFP-data pull distribution for one mass (they look the same, no?) (RED)

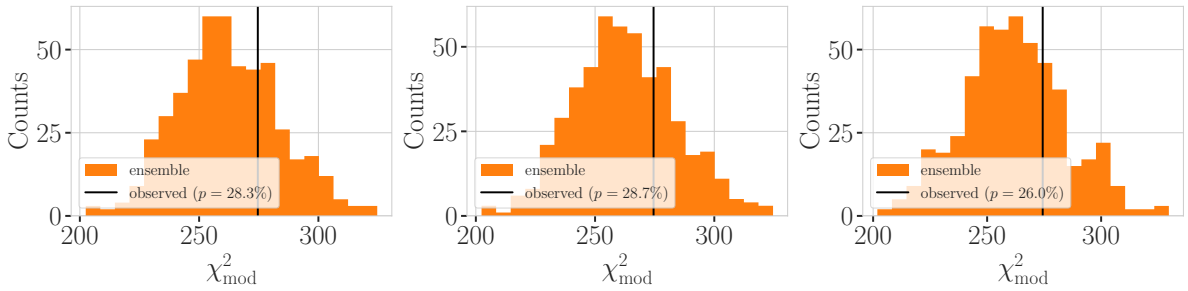


Figure 3.5: Observed fit metric (data fit) and fit metric distribution from pseudo-data ensemble generated around the best fit point. Shown are the results for all three mass samples, with the ensemble distribution on orange, the observed value in black, and the p-value in the legend.

3.3.3 Data/MC Agreement

At the BFP, the agreement between the data and simulation is probed by comparing the one dimensional analysis distributions for PID, energy, and cosine of the zenith angle. As an example, two distributions for the 0.6 GeV mass sample are shown in Figure **??**. The data is compared to the total MC expectation, which is also split up into the individual signal and background components for illustration. Good agreement can be observed in the pull distributions, and is quantified by a reduced χ^2, which is close to 1.0 for all distributions. The reduced χ^2 for all investigated distributions is listed in Table **??**, while the distributions themselves can be found in Section **??**.

specify which they are, once I have them (RED)

add 1-d data/mc agreement for example mass sample (0.6?) and all 3 analysis variables (RED)

add table with reduced chi2 for all 1-d distributions (RED)

3.4 Results

3.4.1 Best Fit Nuisance Parameters

The resulting nuisance parameter values from the fits are illustrated in Figure 3.6, where the differences to the nominal values are shown, normalized by the distance to the closest boundary. The results from all three fits are shown in the same plot and the fits prefer values of the same size for all three mass samples. For parameters that have a Gaussian prior, the 1 σ range is also displayed. As was already confirmed during the blind fit procedure, all fitted parameters are within this range. The effective ice model parameter, N_{bfr}, prefers a value of ~0.74 , indicating that the data fits better to an ice model that includes real birefringence effects. For completeness, the explicit results are listed in Table B.1. There, the nominal values and the absolute differences to the best fit value are also presented.

Cite (again)! (RED)

Show best fit hole ice angular acceptance compared to nominal and flasher/in-situ fits, maybe? (YELLOW)

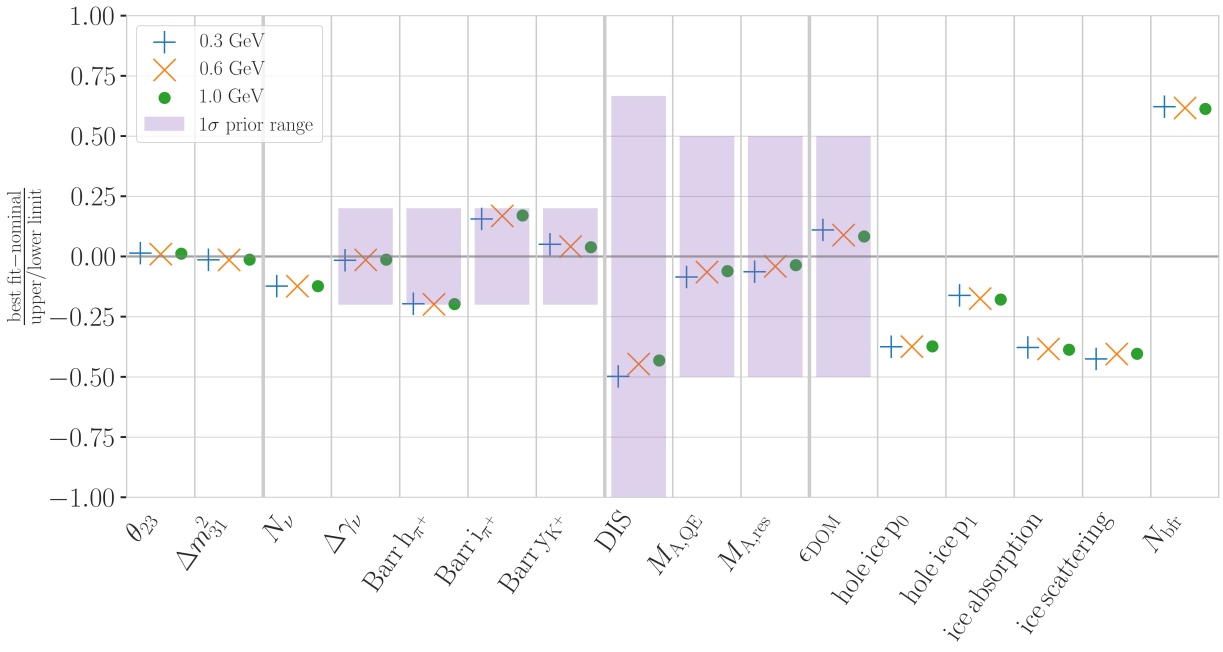


Figure 3.6: Best fit nuisance parameter distances to the nominal values, normalized by the distance to the closest boundary. For parameters with a Gaussian prior, the +1 σ range is also shown.

3.4.2 Best Fit Parameters and Limits

The fitted mixing values are

$$|U_{\tau 4}|^2(0.3\,\text{GeV}) = 0.003^{+0.084}\,,$$
$$|U_{\tau 4}|^2(0.6\,\text{GeV}) = 0.080^{+0.134}\,,\text{and}$$
$$|U_{\tau 4}|^2(1.0\,\text{GeV}) = 0.106^{+0.132}\,,$$

with their +1 σ uncertainty. All of them are compatible with the null hypothesis of 0.0 mixing, although the 0.6 GeV and 1.0 GeV fits indicate a mixing value of 0.08 and 0.106 , respectively. The best fit mixing values and the corresponding upper limits at 68 % and 90 % *confidence level (CL)* are listed in Table 3.6, also showing the p-value to reject the null hypothesis. The CLs and p-value are estimated by assuming that *Wilks' theorem* [61] holds, meaning that the TS follows a χ^2 distribution with one degree of freedom.

[61]: Wilks (1938), "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses"

Table 3.6: Best fit mixing values and the corresponding upper limits at 68 % and 90 % confidence level, as well as the p-value to reject the null hypothesis, estimated by assuming that Wilks' theorem holds.

HNL mass	$\lvert U_{\tau 4}\rvert^2$	68 % CL	90 % CL	NH p-value
0.3 GeV	0.003	0.09	0.19	0.97
0.6 GeV	0.080	0.21	0.36	0.79
1.0 GeV	0.106	0.24	0.40	0.63

Figure 3.7 shows the observed TS profiles as a function of $|U_{\tau 4}|^2$ for all three fits. The TS profile is the difference in χ^2_{mod} between the free fit and a fit where the mixing is fixed to a specific value. Also shown is the expected TS profile, based on 100 pseudo-data trials, produced at the BFP and then fluctuated using both Poisson and Gaussian fluctuations, to include the data and the MC uncertainty as was explained in Section 3.3.2. The Asimov expectation and the 68 % and 90 % bands are shown and the observed TS profiles lie within the 68 % band for all three, confirming that they are compatible with statistical fluctuations of the observed data. For the 0.3 GeV fit, the observed contour is slightly tighter than the Asimov expectation, meaning that the

observed upper limits in $|U_{\tau 4}|^2$ are slightly stronger than expected. For the 0.6 GeV the opposite is the case and the observed upper limit is therefore slightly weaker than expected. For the 1.0 GeV fit, the observed upper limit is very close to the Asimov expectation in the region where the 68 % and 90 % CLs thresholds are crossed. The observed upper limits are also shown in Table 3.6.

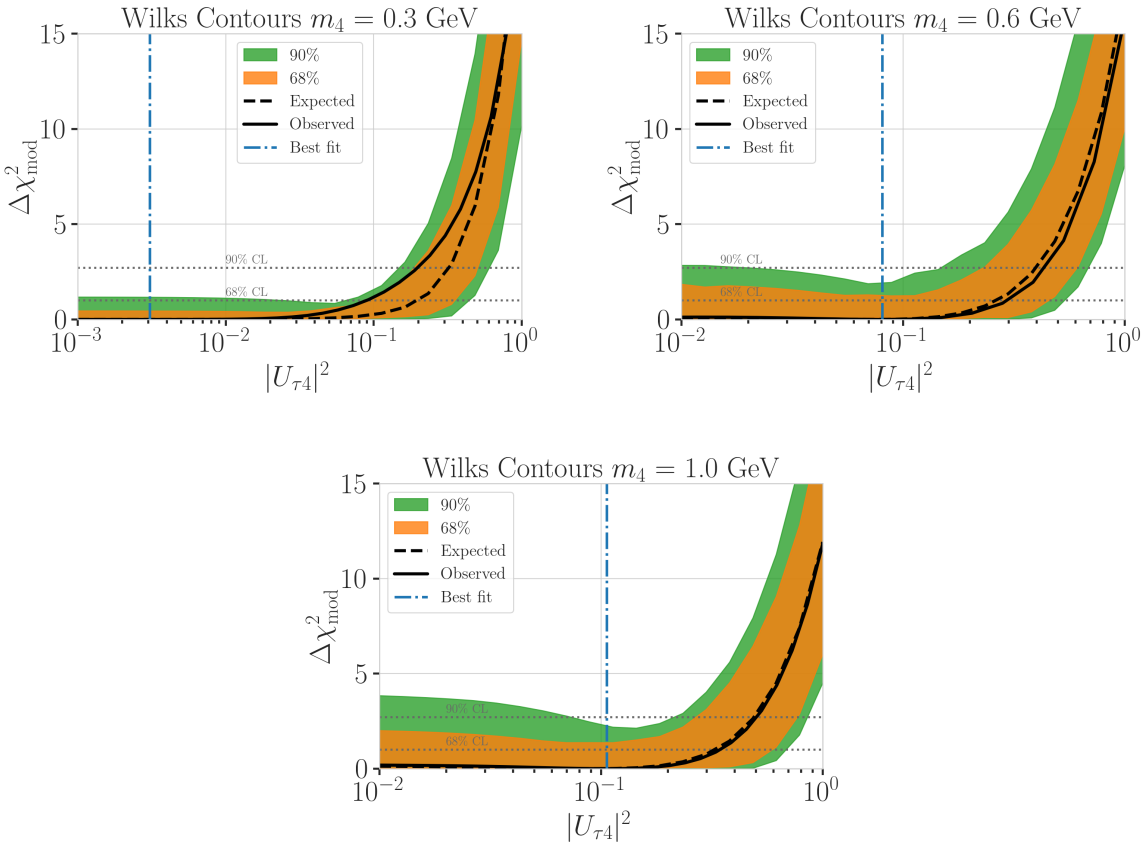


Figure 3.7: Best fit point TS profiles as a function of $|U_{\tau 4}|^2$ for the 0.3 GeV, 0.6 GeV, and 1.0 GeV mass samples. Shown are the observed profiles, the Asimov expectation at the best fit point, and the 68 % and 90 % bands, based on 100 pseudo-data trials. Also indicated are the 68 % and 90 % CL levels assuming Wilks' theorem.

Appendix

A

Heavy Neutral Lepton Signal Simulation

A.1 Model Independent Simulation Distributions

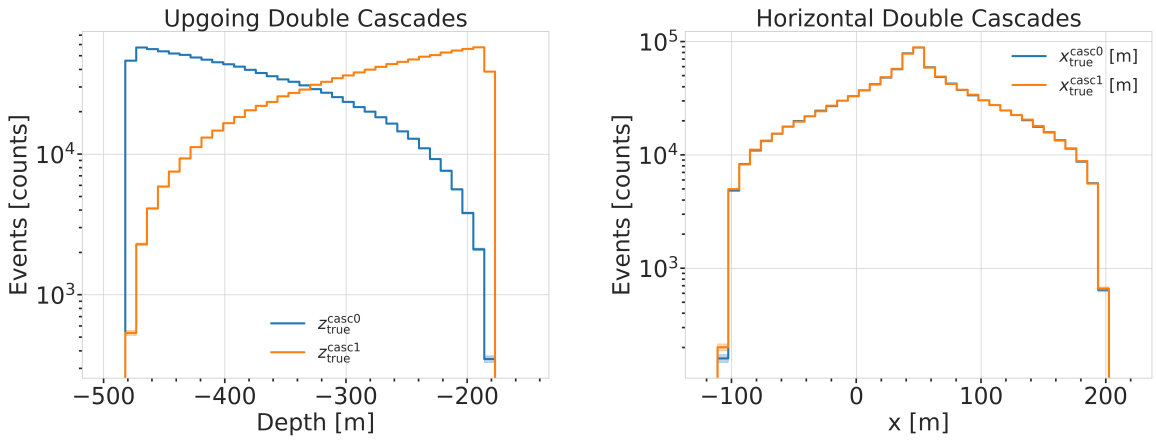


Figure A.1: Generation level distributions of the simplistic simulation sets. Vertical positions (left) and horizontal positions (right) of both sets are shown.

- Re-make plot with x,y for horizontal set one plot!
- Re-make plot with x, y, z for both cascades in one.
- Re-arrange plots in a more sensible way.

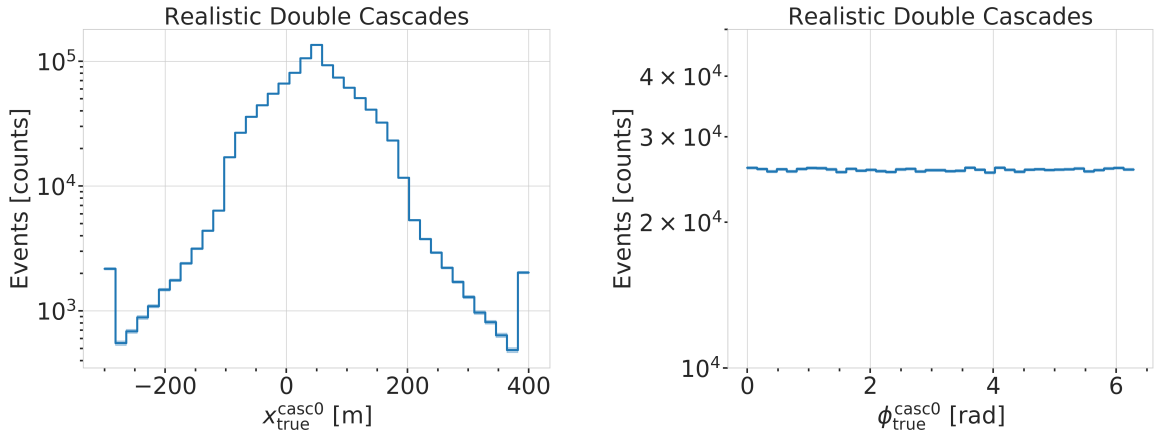
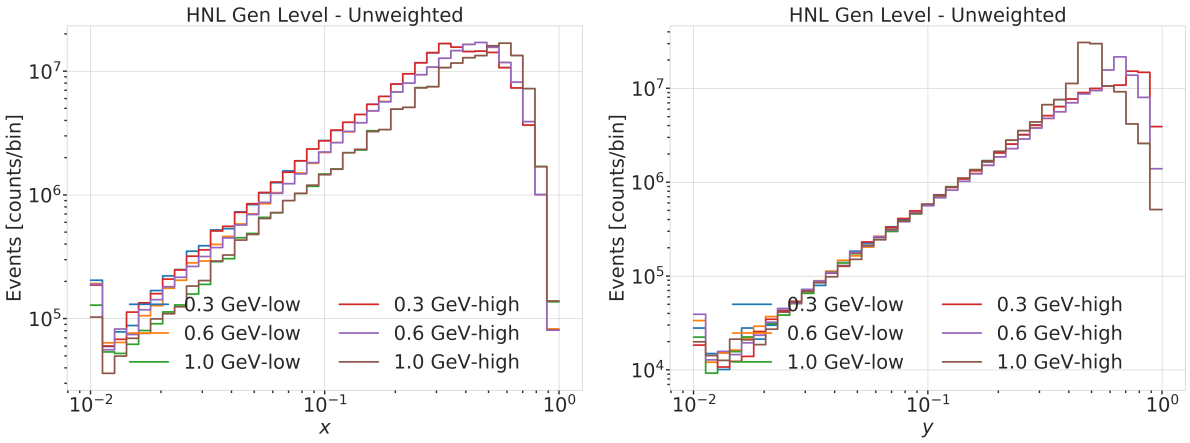


Figure A.2: Generation level distributions of the realistic simulation set. Shown are the cascade x, y, z positions (left) and direction angles (right).

A.2 Model Dependent Simulation Distributions



(a) Bjorken x

(b) Bjorken y

Figure A.3: Generation level distributions of the model dependent simulation.

B

Analysis Results

B.1 Final Level Simulation Distributions

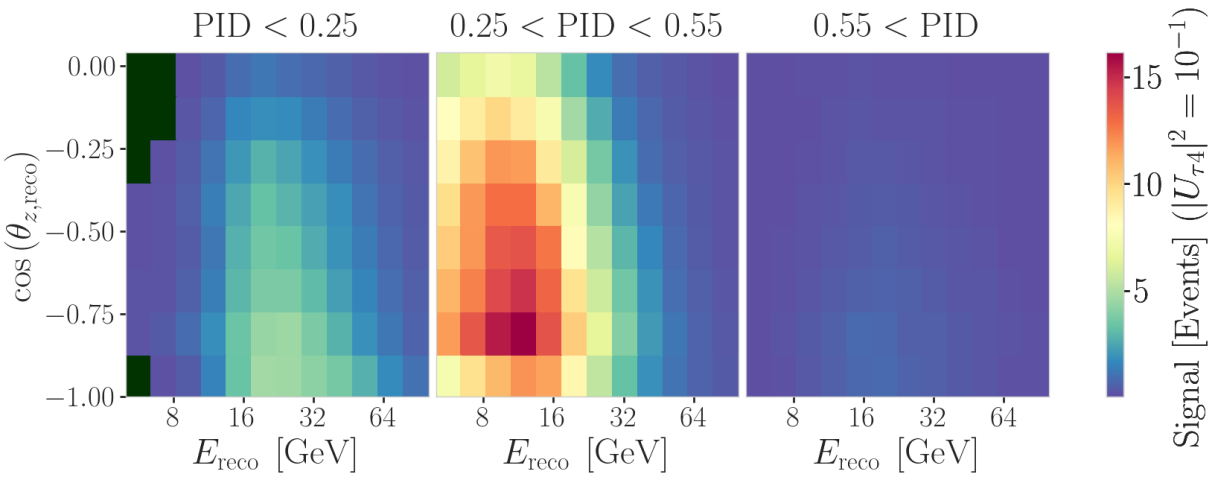


Figure B.1: Signal expectation in 9.28 years for the 1.0 GeV mass sample at a mixing of 0.1, while all other parameters are at their nominal values (top) and observed data (bottom).

B.2 Best Fit Nuisance Parameters

fix design + significant digits to show (ORANGE)

maybe show range/prior and then deviation in sigma, or absolute for the ones without prior

Table B.1: Best fit nuisance parameters for the three mass samples. Also shown is the nominal value and the difference between the nominal and the best fit.

Parameter	Nominal	Best Fit			Nominal - Best Fit		
		0.3 GeV	0.6 GeV	1.0 GeV	0.3 GeV	0.6 GeV	1.0 GeV
$\|U_{\tau 4}\|^2$	-	0.003019	0.080494	0.106141	-	-	-
$\theta_{23}[^\circ]$	47.5047	48.117185	47.918758	48.010986	−0.612485	−0.414058	−0.506286
$\Delta m^2_{31}[\mathrm{eV}^2]$	0.002475	0.002454	0.002454	0.002455	0.000020	0.000021	0.000019
N_ν	1.0	0.889149	0.889055	0.889559	0.110851	0.110945	0.110441
$\Delta\gamma_\nu$	0.0	−0.007926	−0.006692	−0.006596	0.007926	0.006692	0.006596
Barr h_{π^+}	0.0	−0.147475	−0.148481	−0.148059	0.147475	0.148481	0.148059
Barr i_{π^+}	0.0	0.475448	0.513393	0.521626	−0.475448	−0.513393	−0.521626
Barr y_{K^+}	0.0	0.076176	0.062893	0.057548	−0.076176	−0.062893	−0.057548
DIS	0.0	−0.248709	−0.223302	−0.215666	0.248709	0.223302	0.215666
$M_{\mathrm{A,QE}}$	0.0	−0.170528	−0.128150	−0.120345	0.170528	0.128150	0.120345
$M_{\mathrm{A,res}}$	0.0	−0.125855	−0.080875	−0.070716	0.125855	0.080875	0.070716
ϵ_{DOM}	1.0	1.021984	1.017789	1.016689	−0.021984	−0.017789	−0.016689
hole ice p_0	0.101569	−0.161341	−0.161051	−0.160129	0.262910	0.262620	0.261698
hole ice p_1	−0.049344	−0.073701	−0.075596	−0.076261	0.024357	0.026252	0.026917
ice absorption	1.00	0.943261	0.942463	0.942000	0.056739	0.057537	0.058000
ice scattering	1.05	0.986152	0.989289	0.989438	0.063848	0.060711	0.060562
N_{bfr}	0.0	0.746684	0.740255	0.736215	−0.746684	−0.740255	−0.736215

Bibliography

Here are the references in citation order.

[1] R. Abbasi et al. "Measurement of atmospheric neutrino mixing with improved IceCube DeepCore calibration and data processing". In: *Phys. Rev. D* 108 (1 July 2023), p. 012014. DOI: 10.1103/PhysRevD.108.012014 (cited on pages 1, 8, 11–13).

[2] M. Honda et al. "Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model". In: *Phys. Rev. D* 92 (2 July 2015), p. 023004. DOI: 10.1103/PhysRevD.92.023004 (cited on pages 2, 36).

[3] C. Andreopoulos et al. "The GENIE Neutrino Monte Carlo Generator: Physics and User Manual". In: (2015) (cited on pages 3, 13).

[4] M. Glück, E. Reya, and A. Vogt. "Dynamical parton distributions revisited". In: *The European Physical Journal C* 5 (Sept. 1998), pp. 461–470. DOI: 10.1007/s100529800978 (cited on page 3).

[5] A. Bodek and U. K. Yang. "Higher twist, xi(omega) scaling, and effective LO PDFs for lepton scattering in the few GeV region". In: *Journal of Physics G: Nuclear and Particle Physics* 29.8 (2003), p. 1899. DOI: 10.1088/0954-3899/29/8/369 (cited on page 3).

[6] J.-H. Koehne et al. "PROPOSAL: A tool for propagation of charged leptons". In: *Computer Physics Communications* 184.9 (2013), pp. 2070–2090. DOI: https://doi.org/10.1016/j.cpc.2013.04.001 (cited on page 3).

[7] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nucl. Instr. Meth. Phys. Res.* 506.3 (July 2003), pp. 250–303. DOI: 10.1016/s0168-9002(03)01368-8 (cited on page 3).

[8] L. Rädel and C. Wiebusch. "Calculation of the Cherenkov light yield from low energetic secondary particles accompanying high-energy muons in ice and water with Geant4 simulations". In: *Astroparticle Physics* 38 (Oct. 2012), pp. 53–67. DOI: 10.1016/j.astropartphys.2012.09.008 (cited on page 3).

[9] Y. Becherini et al. "A parameterisation of single and multiple muons in the deep water or ice". In: *Astroparticle Physics* 25.1 (2006), pp. 1–13. DOI: https://doi.org/10.1016/j.astropartphys.2005.10.005 (cited on page 3).

[10] D. Heck et al. "CORSIKA: A Monte Carlo code to simulate extensive air showers". In: (Feb. 1998) (cited on page 3).

[11] T. K. Gaisser. "Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio". In: *Astropart. Phys.* 35 (2012), pp. 801–806. DOI: 10.1016/j.astropartphys.2012.02.010 (cited on page 3).

[12] R. Engel et al. "The hadronic interaction model Sibyll – past, present and future". In: *EPJ Web Conf.* 145 (2017). Ed. by B. Pattison, p. 08001. DOI: 10.1051/epjconf/201614508001 (cited on page 3).

[13] C. Kopper et al. https://github.com/claudiok/clsim (cited on page 4).

[14] D. Chirkin et al. "Photon Propagation using GPUs by the IceCube Neutrino Observatory". In: *2019 15th International Conference on eScience (eScience)*. 2019, pp. 388–393. DOI: 10.1109/eScience.2019.00050 (cited on page 4).

[15] M. G. Aartsen et al. "Measurement of South Pole ice transparency with the IceCube LED calibration system". In: *Nucl. Instrum. Meth. A* 711 (2013), pp. 73–89. DOI: 10.1016/j.nima.2013.01.054 (cited on page 4).

[16] A. Trettin. "Search for eV-scale sterile neutrinos with IceCube DeepCore". PhD thesis. Berlin, Germany: Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2023. DOI: https://github.com/atrettin/PhD-Thesis (cited on pages 4, 5, 15, 35).

[17] G. Mie. "Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen". In: *Annalen der Physik* 330.3 (1908), pp. 377–445. DOI: https://doi.org/10.1002/andp.19083300302 (cited on page 4).

[18] L. G. Henyey and J. L. Greenstein. "Diffuse radiation in the Galaxy." In: *apj* 93 (Jan. 1941), pp. 70–83. DOI: 10.1086/144246 (cited on page 4).

[19] S. Fiedlschuster. "The Effect of Hole Ice on the Propagation and Detection of Light in IceCube". In: (Apr. 2019) (cited on page 4).

[20] M. G. Aartsen et al. "In-situ calibration of the single-photoelectron charge response of the IceCube photomultiplier tubes". In: *Journal of Instrumentation* 15.6 (June 2020), P06032. DOI: 10.1088/1748-0221/15/06/P06032 (cited on page 5).

[21] M. Larson. "Simulation and Identification of Non-Poissonian Noise Triggers in the IceCube Neutrino Detector". Available at https://ir.ua.edu/handle/123456789/1927. MA thesis. University of Alabama, Tuscaloosa, AL, USA, 2013 (cited on page 5).

[22] M. Larson. "A Search for Tau Neutrino Appearance with IceCube-DeepCore". available at https://discoverycenter.nbi.ku.dk/teaching/thesis_page/mjlarson_thesis.pdf. PhD thesis. University of Copenhagen, Denmark, 2018 (cited on page 5).

[23] E. Lohfink. "Testing nonstandard neutrino interaction parameters with IceCube-DeepCore". PhD thesis. Mainz, Germany: Johannes Gutenberg-Universität Mainz, Fachbereich für Physik, Mathematik und Informatik, 2023. DOI: http://doi.org/10.25358/openscience-9288 (cited on pages 5, 35).

[24] M. G. Aartsen et al. "The IceCube Neutrino Observatory: Instrumentation and Online Systems". In: *JINST* 12.03 (2017), P03012. DOI: 10.1088/1748-0221/12/03/P03012 (cited on page 5).

[25] R. Abbasi et al. "The design and performance of IceCube DeepCore". In: *Astropart. Phys.* 35.10 (2012), pp. 615–624. DOI: 10.1016/j.astropartphys.2012.01.004 (cited on pages 5, 6).

[26] R. Abbasi et al. "The IceCube data acquisition system: Signal capture, digitization, and timestamping". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 601.3 (2009), pp. 294–316. DOI: https://doi.org/10.1016/j.nima.2009.01.001 (cited on page 6).

[27] M. G. Aartsen et al. "The IceCube Neutrino Observatory: instrumentation and online systems". In: *Journal of Instrumentation* 12.3 (Mar. 2017), P03012. DOI: 10.1088/1748-0221/12/03/P03012 (cited on page 6).

[28] J. H. Friedman. "Stochastic gradient boosting". In: *Computational Statistics & Data Analysis* 38 (2002), pp. 367–378 (cited on page 7).

[29] R. Abbasi et al. "Low energy event reconstruction in IceCube DeepCore". In: *Eur. Phys. J. C* 82.9 (2022), p. 807. DOI: 10.1140/epjc/s10052-022-10721-2 (cited on pages 8, 18).

[30] S. Yu and J. Micallef. "Recent neutrino oscillation result with the IceCube experiment". In: *38th International Cosmic Ray Conference*. July 2023 (cited on pages 8, 32, 37).

[31] S. Yu and on behalf of the IceCube collaboration. "Direction reconstruction using a CNN for GeV-scale neutrinos in IceCube". In: *Journal of Instrumentation* 16.11 (Nov. 2021), p. C11001. DOI: 10.1088/1748-0221/16/11/C11001 (cited on page 9).

[32] J. Micallef. https://github.com/jessimic/LowEnergyNeuralNetwork (cited on page 9).

[33] M. Huennefeld. "Deep Learning in Physics exemplified by the Reconstruction of Muon-Neutrino Events in IceCube". In: *PoS* ICRC2017 (2017), p. 1057. DOI: 10.22323/1.301.1057 (cited on page 9).

[34] H. Dembinski et al. "Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV". In: *PoS* ICRC2017 (2017), p. 533. DOI: 10.22323/1.301.0533 (cited on page 12).

[35] G. D. Barr et al. "Uncertainties in atmospheric neutrino fluxes". In: *Phys. Rev. D* 74 (9 Nov. 2006), p. 094009. DOI: 10.1103/PhysRevD.74.094009 (cited on page 12).

[36] J. Evans et al. "Uncertainties in atmospheric muon-neutrino fluxes arising from cosmic-ray primaries". In: *Phys. Rev. D* 95 (2 Jan. 2017), p. 023012. DOI: 10.1103/PhysRevD.95.023012 (cited on pages 12, 37).

[37] A. Fedynitch et al. https://github.com/afedynitch/MCEq (cited on page 12).

[38] G. D. Barr et al. "Uncertainties in Atmospheric Neutrino Fluxes". In: *Phys. Rev. D* 74 (2006), p. 094009. DOI: 10.1103/PhysRevD.74.094009 (cited on pages 12, 37).

[39] F. Riehn et al. “Hadronic interaction model sibyll 2.3d and extensive air showers”. In: *Phys. Rev. D* 102 (6 Sept. 2020), p. 063002. DOI: 10.1103/PhysRevD.102.063002 (cited on page 12).

[40] A. Cooper-Sarkar, P. Mertsch, and S. Sarkar. “The high energy neutrino cross-section in the Standard Model and its uncertainty”. In: *JHEP* 08 (2011), p. 042. DOI: 10.1007/JHEP08(2011)042 (cited on page 13).

[41] Q. Wu et al. “A precise measurement of the muon neutrino–nucleon inclusive charged current cross section off an isoscalar target in the energy range 2.5<E<40 GeV by NOMAD”. In: *Physics Letters B* 660.1 (2008), pp. 19–25. DOI: https://doi.org/10.1016/j.physletb.2007.12.027 (cited on page 13).

[42] M. M. Tzanov. “Precise measurement of neutrino and anti-neutrino differential cross sections on iron”. PhD thesis. University of Pittsburgh, Pennsylvania, Jan. 2005 (cited on page 13).

[43] W. G. Seligman. “A next-to-leading-order QCD analysis of neutrino-iron structure functions at the Tevatron”. PhD thesis. Columbia University, New York, Aug. 1997 (cited on page 13).

[44] J. Feintzeig. “Searches for Point-like Sources of Astrophysical Neutrinos with the IceCube Neutrino Observatory”. PhD thesis. University of Wisconsin, Madison, Jan. 2014 (cited on pages 14, 37).

[45] N. Kulacz. “In Situ Measurement of the IceCube DOM Efficiency Factor Using Atmospheric Minimum Ionizing Muons”. MA thesis. University of Alberta, 2019 (cited on pages 14, 37).

[46] Rongen, Martin. “Measuring the optical properties of IceCube drill holes”. In: *EPJ Web of Conferences* 116 (2016), p. 06011. DOI: 10.1051/epjconf/201611606011 (cited on page 14).

[47] R. Abbasi et al. “Measurement of Astrophysical Tau Neutrinos in IceCube’s High-Energy Starting Events”. In: *arXiv e-prints* (Nov. 2020) (cited on page 17).

[48] M. Usner. “Search for Astrophysical Tau-Neutrinos in Six Years of High-Energy Starting Events in the IceCube Detector”. PhD thesis. Berlin, Germany: Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2018. DOI: 10.18452/19458 (cited on page 17).

[49] P. Hallen. “On the Measurement of High-Energy Tau Neutrinos with IceCube”. MA thesis. Aachen, Germany: RWTH Aachen University, 2013 (cited on page 17).

[50] N. Whitehorn, J. van Santen, and S. Lafebre. “Penalized splines for smooth representation of high-dimensional Monte Carlo datasets”. In: *Computer Physics Communications* 184.9 (2013), pp. 2214–2220. DOI: https://doi.org/10.1016/j.cpc.2013.04.008 (cited on page 17).

[51] M. G. Aartsen et al. “Energy Reconstruction Methods in the IceCube Neutrino Telescope”. In: *JINST* 9 (2014), P03009. DOI: 10.1088/1748-0221/9/03/P03009 (cited on page 17).

[52] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cited on page 25).

[53] V. Barger et al. “Matter effects on three-neutrino oscillations”. In: *Phys. Rev. D* 22 (11 Dec. 1980), pp. 2718–2726. DOI: 10.1103/PhysRevD.22.2718 (cited on page 31).

[54] A. M. Dziewonski and D. L. Anderson. “Preliminary reference Earth model”. In: *Physics of the Earth and Planetary Interiors* 25.4 (1981), pp. 297–356. DOI: https://doi.org/10.1016/0031-9201(81)90046-7 (cited on page 31).

[55] L. Fischer, R. Naab, and A. Trettin. “Treating detector systematics via a likelihood free inference method”. In: *Journal of Instrumentation* 18.10 (2023), P10019. DOI: 10.1088/1748-0221/18/10/P10019 (cited on page 35).

[56] M. G. Aartsen et al. “Computational techniques for the analysis of small signals in high-statistics neutrino oscillation experiments”. In: *Nucl. Instrum. Meth. A* 977 (2020), p. 164332. DOI: 10.1016/j.nima.2020.164332 (cited on page 37).

[57] https://github.com/icecube/pisa (cited on page 37).

[58] R. S. Nickerson. “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises”. In: *Review of General Psychology* 2 (1998), pp. 175–220 (cited on page 37).

[59] H. Dembinski et al. *scikit-hep/iminuit: v2.17.0*. Version v2.17.0. Sept. 2022. DOI: 10.5281/zenodo.7115916 (cited on page 38).

[60] F. James and M. Roos. "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations". In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9 (cited on page 38).

[61] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360 (cited on page 40).