UNIVERSITY OF COPENHAGEN

In-situ relative efficiency calibration of digital optical modules in the IceCube detector

Sofus Kjærsgaard Stray

Supervisor: D. Jason Koskinen

Niels Bohr Institute

University of Copenhagen

May 20, 2021

Abstract

The IceCube Neutrino Observatory is a kilometer-scale detector located deep in the ice at the South Pole. It is a world-class experiment in multi-messenger astronomy with discoveries of high-energy neutrino sources and competitive precision measurements in neutrino physics. IceCube detects neutrinos using Cherenkov light produced by charged high-energy particles from interactions in the ice. The light is captured by photomultiplier tubes (PMTs) built into Digital Optical Modules (DOMs) embedded into the ice, forming the detector. In the centre of the detector is a densely instrumented region known as DeepCore, whose PMTs have higher quantum efficiency than those in the rest of IceCube. Laboratory tests have been conducted on the PMTs but once they are placed in the ice, localised ice properties and efficiency variations will impact the DOM efficiencies.

Through emitted Cherenkov light, minimum ionising atmospheric muons can be used as calibration sources to characterise the relative individual DOM Efficiency (RIDE) in the detector. This study uses two Monte Carlo simulation datasets matching IceCube configuration years from 2012 and 2016, with input relative quantum efficiencies of 1.0 and 1.35 for the IceCube normal quantum efficiency (NQE) and DeepCore high quantum efficiency (HQE) DOMs respectively. Through optimising cuts on the zenith angle of the muons and the distance from the muon tracks to the DOMs, we calculate the RIDE for NQEs and HQEs separately. For the first dataset, we find an NQE RIDE centred at 1.01 ± 0.10 and an HQE RIDE centred at 1.31 ± 0.10 with a slope of 0.06 per 100 meters in DeepCore. For the second dataset, we find an NQE RIDE centred at 1.01 ± 0.08 and an HQE RIDE centred at 1.24 ± 0.06 which is inconsistent with input DOM efficiencies.

A new machine learning-based reconstruction method was developed to classify stopped muons within the detector in order to characterise minimum ionising muons. The 2012 dataset classifier gave a precision of precision of 97.9% with a sensitivity of 39.1%, and the 2016 dataset classifier gave a precision of 72.6% with a sensitivity of 20.7%. Machine learning regression models for predicting the endpoint of a stopped muon were also developed. The 2012 dataset regressor had a mean absolute prediction error of 61.7 meters and a coefficient of determination of 0.79. The 2016 dataset regressor had a mean absolute prediction error of 94.9 meters and a coefficient of determination of 0.67. The lower performance on the 2016 dataset is thought to originate in poor predictive ability of the muon track depth.

Acknowledgements

Author's Contributions

The main result of this paper is the development of the RIDE analysis on dCORSIKA simulation datasets, see Section 2.5. Using the IceTray framework, I created a method of scanning over a list of dCORSIKA files and output HDF5 tables with relevant DOM and particle data, including improvements to computation time over code developed by Étienne Bourbeau.

Additionally, I developed a robust method of scanning over relevant parameters and calculate the RIDE for each parameter combination. This helped precisely quantify the direct effect of varying the zenith angle and distance from the DOMs to the muon track on the RIDE value, which has not been done previously. This is detailed in Section 2.4.

I also developed a neural network classifier to tag stopped muons inside the detector. This is the first attempt of stopped muon classification using neural networks and I have shown the potential of machine learning as a possible reconstruction. In addition to the classifier, I also developed a regression neural network to predict the endpoint of stopped muons. Discussion on the neural network implementation is in Section 3.2.1 and the performance of the classifier and regressor in Section 3.3.

Contents

1 Introduction

The Standard Model of particle physics describes all known elementary particles and three of the four fundamental forces. Physics beyond the Standard Model is currently one of the the frontiers of science and includes a fundamental mystery in modern particle physics: The existence of neutrino mass and the nature of neutrino oscillation[1].

1.1 Neutrino Physics

Neutrinos are incredibly light electrically neutral leptons that come in three flavours corresponding to the three charged leptons: electron, muon, and tau. Possessing no electric charge or colour charge, the neutrino does not interact via the electromagnetic or strong force. Having very little mass[2], there is practically zero gravitational forces on the neutrino, leaving only the weak interaction. The weak interaction involves the exchange of the neutral Z^0 boson in neutral current (NC) interactions or the charged $W^{\pm}$ boson in charged current (CC) interactions. In NC interactions, the neutrino simply transfers energy and momentum to a charged particle in the medium. Any combination of neutrino and lepton flavour can participate in this interaction, but no information about the neutrino flavour can be known from it. CC interactions requires the flavour to be preserved. See Figure 1 for three example Feynman diagrams. Due to the short interaction range[3] and strength of the weak coupling constant[4], neutrino interactions are in general heavily suppressed, allowing them to travel huge distances unimpeded. Being able to travel through the cosmos without changing directions allow neutrinos to point directly towards their source, making them excellent astrophysical messengers[5].

1.2 Neutrino Oscillations

In the Standard Model, neutrinos are described as massless particles and occurring in three flavour states, ν_e, ν_μ, and ν_τ, corresponding to the three charged leptons. However, observations in the 1970s suggested that neutrinos oscillate between these three flavour states. This cannot occur for massless particles, suggesting a gap in the Standard Nodel. The theory of neutrino oscillation stems from assuming that neutrinos exist in a superposition of three mass states. The mathematical description of this superposition is given by the Pontecorvo-Maki-Nakagawa-Sakata[6, 7] (PMNS)
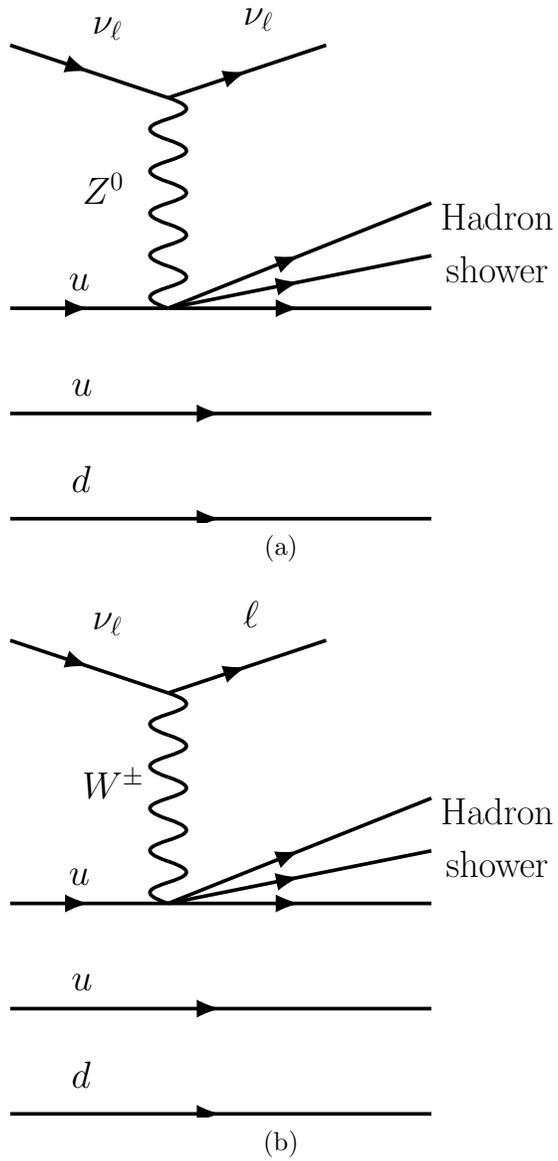
Figure 1: Two example Feynman diagrams of (a) a neutral current interaction, and (b) a charged current interactions. ℓ represents an arbitrary charged lepton.

matrix,

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

where ν_1, ν_2, and ν_3 are the three mass eigenstates. If the neutrino was massless, then the off-diagonal elements in the PMNS matrix would be zero. Each neutrino flavour state ν_α is described by the mass state ν_j and the PMNS matrix element $U_{\alpha j}^*$ as

$$
|\nu_\alpha\rangle = \sum_j U_{\alpha j}^* |\nu_j\rangle.
\tag{2}
$$

Equation 2 describes a stationary neutrino. As neutrinos propagate, the time evolution operator dictates how the flavour states change. As neutrinos are extremely light they will travel at relativistic speeds and we can assume that their energy is approximately equal to their momentum, $E \approx p$, and the time it takes to travel a distance is approximately equal to the distance travelled, $t \approx L$, since the mass is much smaller than the momentum, $m \ll p$. Using this approximation, the neutrino flavour state after travelling a length L becomes[8]

$$
|\nu_\alpha\rangle = \sum_j U_{\alpha j}^* |\nu_j\rangle = \sum_j U_{\alpha j}^* e^{-i\dfrac{m_j^2 L}{2E}} |\nu_j, 0\rangle.
\tag{3}
$$

We can then calculate the probability of a flavour α oscillating into a flavour β,

$$
P\left(\nu_\alpha \rightarrow \nu_\beta\right)_{\alpha \neq \beta} = 4 \sum_{i<j} \mathrm{Re}\left(U_{\alpha i}^* U_{\beta i} U_{\alpha j} U_{\beta j}^*\right) \sin^2\left(\frac{\Delta m_{ji}^2 L}{4E}\right),
\tag{4}
$$

where $\Delta m_{ji}^2 = m_j^2 - m_i^2$, the squared mass differences between two mass eigenstates m_j and m_i. Equation 4 tells us that if the neutrino was massless and $\Delta m_{ji}^2 = 0$ then no neutrino oscillations would occur. Experimental confirmation of neutrino oscillation[9] is evidence of at least two of the neutrino mass eigenstates being non-zero. Neutrino oscillation is thus described by the 9 matrix elements and 2 neutrino squared mass differences. However, since the PMNS matrix has to be unitary it can

3

be written as

$$U_{PMNS} = \begin{bmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{bmatrix}, \tag{5}$$

$$= \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{CP}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{CP}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{CP}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{CP}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{CP}} & c_{23}c_{13} \end{bmatrix}, \tag{6}$$

where $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$. The PMNS matrix is then fully parameterised by the three mixing angles θ_{12}, θ_{23}, θ_{13}, and the charge-parity violating phase angle δ_{CP}. Latest fits to the mixing angles and squared mass differences are[10]

$$\theta_{12} = 33.44^{+0.78}_{-0.75} \,^{\circ}, \tag{7}$$

$$\theta_{23} = 49.0^{+1.1}_{-1.4} \,^{\circ}, \tag{8}$$

$$\theta_{13} = 8.57^{+0.13}_{-0.12} \,^{\circ}, \tag{9}$$

$$\Delta m^2_{21} = 7.42^{+0.21}_{-0.20} \cdot 10^{-5} \text{ eV}^2, \tag{10}$$

$$\Delta m^2_{31/32} = 2.514^{+0.028}_{-0.027} \cdot 10^{-3} \text{ eV}^2. \tag{11}$$

Determining the neutrino oscillation parameters is one of the key fields of neutrino physics and observatories dedicated to neutrino detection will help further this goal. Due to their heavy interaction suppression, neutrinos are incredibly difficult to detect. Weak interactions occur with numerous types of particles other than neutrinos and any detector will have to be heavily shielded from background cosmic rays and be massive in scale in order to detect a sizeable number of neutrinos over time[11].

1.3 Detecting Neutrinos

There are various ways of detecting neutrinos. Some methods employ a relatively high-density transparent interaction medium which neutrinos interact with. This produces secondary particles that in turn produce detectable signals. Certain materials produce light when struck by the secondary particles called scintillation light, while superluminal particles in the medium produce the light-equivalent of a sonic boom known as Cherenkov light. The latter method will be the focus of this thesis. Cherenkov-based detectors are numerous, and include Super Kamiokande[12], ANTARES[13], KamLAND[14], Sudbury[15], and IceCube[16].

1.4   Cherenkov Light

Cherenkov light is produced when a charged particle travels through the medium at speeds faster than the speed of light in said medium. Cherenkov light can be directly observed when the medium is transparent, such as water or ice. In general, the neutrino interactions that produce Cherenkov light come in two different event types. The first type, known as "tracks", occur when a muon from a charged current muon neutrino interaction passes through the detection medium. Tracks have small angular resolutions and allow detection of the particle direction through the medium. Electron and tau neutrinos typically do not produce tracks. Electrons have a high charge-to-mass ratio, causing strong scattering and radiation effects. Tauons have a relatively high mass and typically decay too quickly to form a visible track. Tauons can occasionally decay into a muon and two neutrinos to produce a track. The second type are spherical "cascades" that are produced by the aforementioned electron neutrino interactions, as well as tau neutrino and hadronic interactions. Due to spherical nature of these interactions, the angular resolution of the lepton and neutrino is poor. There is another type known as a "double-bang" where a high-energy interaction creates a tau track with sufficient energy to travel a long enough distance that its decay produces a cascade[17]. All three event types can be seen in Figure 2.

Photon detectors directly capture the Cherenkov light from high-energy charged particles in order to reconstruct the neutrino interactions. Due to the requirement of a large volume of transparent material, some detectors such as the Sudbury Neutrino Observatory or Super Kamiokande use large tanks of water or heavy water, whereas other detectors such as ANTARES or IceCube use the natural mediums of the Mediterranean sea and the ice cap in Antarctica respectively, the latter of which will be the focus of this thesis.

1.5   IceCube Neutrino Observatory

The IceCube Neutrino Observatory is a neutrino detector located in the Antarctic. A cubic kilometer of ice is fitted with 5160 digital optical modules (DOMs). This transforms the ice into one huge Cherenkov detector that allows the observatory to detect neutrinos through Cherenkov light from charged particles. The IceCube Neutrino Observatory is sensitive to neutrino energies down to ∼10 GeV[19] and up to the PeV range[20].

1.5.1   Detector Geometry

The IceCube detector extends to a depth of around 2500 meters and is composed of 5160 DOMs separated into 86 vertical strings with 60 DOMs per string. The strings

Charged-current ν$_μ$ **Neutral-current / ν$_e$**   **Charged-current ν$_τ$**

(data)   (data)   (simulation)

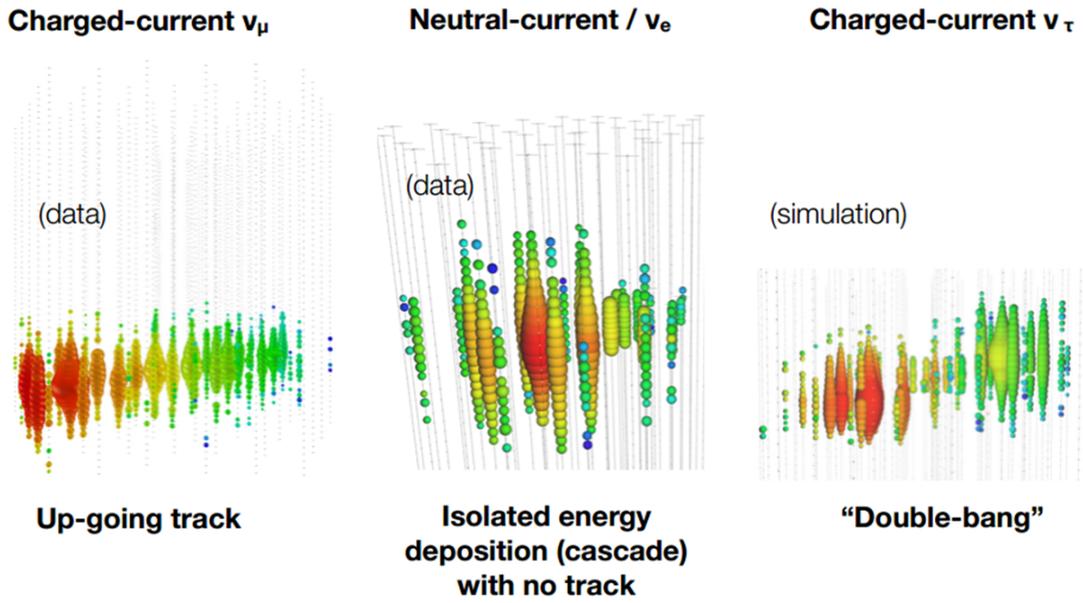**Up-going track**   **Isolated energy deposition (cascade) with no track**   **"Double-bang"**

Figure 2: The three event types in IceCube. The diameter of the spheres is proportional to the number of detected photons and the colour shows the time passed, red first through blue last. The double bang is only a simulation whereas the tracks and cascades are from data. Taken from Ref. [18].
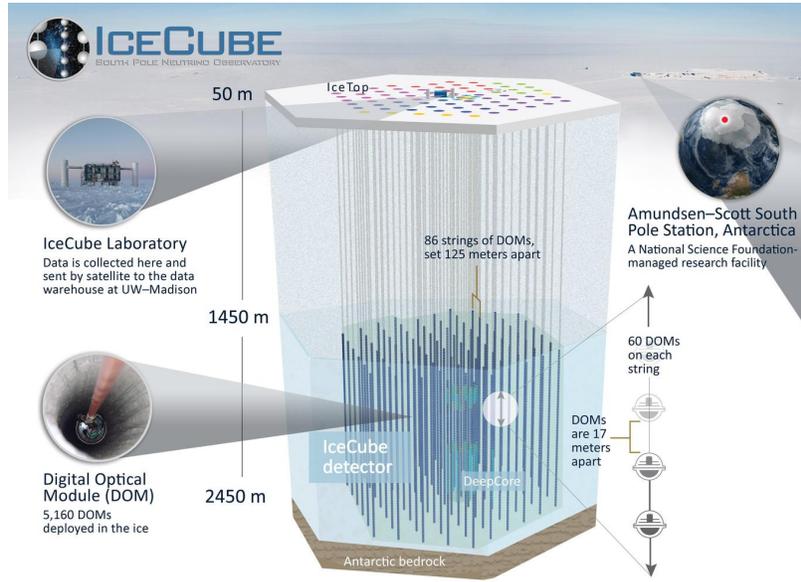
Figure 3: An illustration of the IceCube Neutrino Observatory. Shown is the main IceCube detector and the two subdetectors IceTop and DeepCore, and the IceCube Laboratory sitting on top of the detector. An illustration of the DOMs inside their borehole is also shown, with the number of strings, DOM spacing, and depth beneath the surface. Taken from Ref. [22].

are deployed at 125 meter spacings in a hexagonal grid, while the DOMs along each string are separated by a distance of 17 meters, with the exception of eight strings, that form the DeepCore detector, which have a separation of 7 meters. This geometry configuration creates a Cherenkov detector over a square kilometer in breadth and in a depth of 1450 m to 2450 m, with DeepCore covering 2100 m to 2450 m. In addition, there are 81 stations, separate from the strings, on top of the detector known as IceTop. A diagram of the detector can be seen in Figure 3. As of 2016, 98.4% of the DOMs in IceCube are still functioning after their initial deployment in the ice[21]. In our analysis, we primarily consider the main IceCube detector as well as the DeepCore subdetector. IceCube employs two types of depth coordinate systems. The first fixes the origin at the ice surface and increases with depth. The second fixes the origin in the centre of the instrumented volume of IceCube, about 2000 m below the surface, and decreases with depth[21]. We will refer to the former as surface coordinates and the latter as IceCube coordinates.

Penetrator

HV Divider

LED
Flasher
Board

DOM
Mainboard

Mu-metal
grid
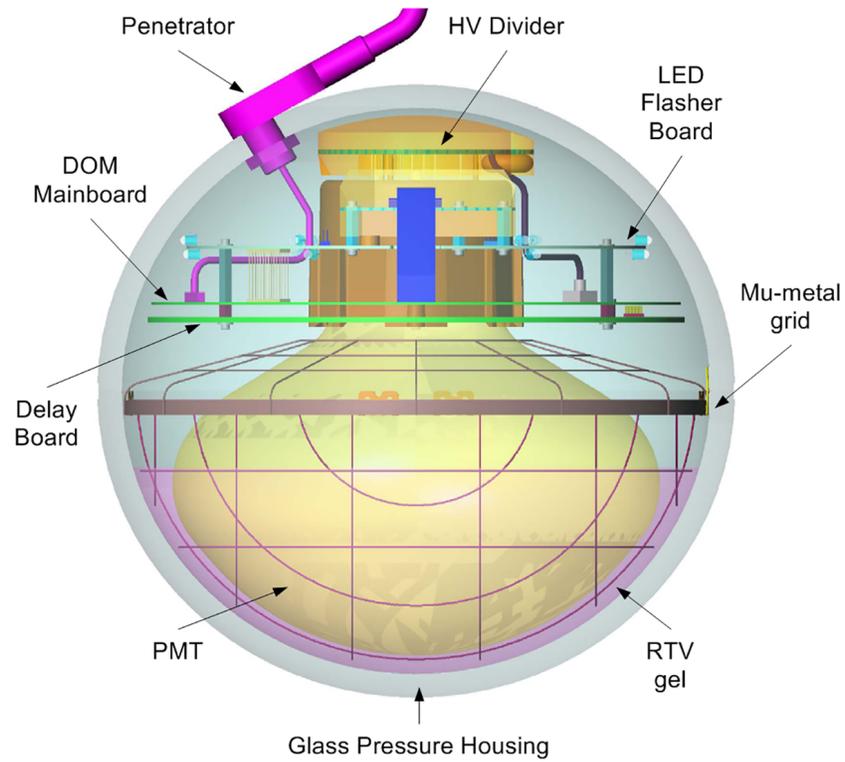
Delay
Board

PMT

RTV
gel

Glass Pressure Housing

Figure 4: Diagram of the Digital Optical Module (DOM) used in the IceCube detector. Shown are all the key components and the housing apparatus. Taken from Ref. [16].

1.5.2 Digital Optical Module

The digital optical modules (DOMs) are the instruments that measure and collect charge and timing data in the IceCube detector. The DOMs are spherical modules containing a Hamamatsu R7081-02 (R7081-02MOD in DeepCore)[23] photomultiplier tube (PMT) that rests in an RTV gel surrounded by a mu-metal grid for magnetic shielding[24]. The DOMs contain the PMT voltage generators and divider circuits, and the DOM mainboard which contains the core functionality of the DOMs. They also contain a flasher board, 6 horizontal and 6 tilted LEDs which can be programmed to flash with specific combinations, light levels, and pulse lengths[25]. This allows simulation of physical events, investigation of the optical ice properties, and DOM efficiency calibrations[16]. All instruments are encased in a glass pressure housing capable of withstanding the immense pressures of the ice. A schematic of the DOM can be seen in Figure 4.

1.5.3 Quantum Efficiency

Quantum efficiency refers to the ratio of charge carriers produced by the circuits to photons hitting the PMT. It provides a good description of how much light gets translated into a charge response by the DOMs. For IceCube, the quantum efficiency is related to the PMT circuitry. As mentioned in Section 1.5.2, DeepCore contains DOMs with modified PMTs that are more efficient than the ones in the standard DOMs. The standard Normal Quantum Efficiency (NQE) DOMs are specified to have a peak quantum efficiency of ∼25% at 390 nm whereas these High Quantum Efficiency (HQE) DOMs have a specified peak quantum efficiency of ∼34%, later confirmed by laboratory measurements[21]. This gives the HQEs a relative quantum efficiency of ∼1.35. Combined with the denser geometry, the deployment in the clearest part of the ice, and the usage of the surrounding modules as background veto, this reduces the DeepCore energy threshold to about 10 GeV compared to the 100 GeV in the rest of the detector[19]. IceCube boasts achievements including the first identified high-energy astrophysical neutrino observation[26], competitive neutrino oscillation parameter estimations[27], and measurement of atmospheric tau neutrino appearance[28]. These discoveries require extremely sensitive and well calibrated instruments. This work aims to improve the systematic uncertainties in IceCube to further increased sensitives in present astrophysical discoveries. In particular, systematic uncertainty tests on the tau neutrino appearance analysis show that fixing the DOM efficiency parameter has the third largest detector-relevant effect, see Figure 5.

It is evident that improvements to the systematic uncertainty on the DOM efficiency will have remarkable effects on the sensitivity of IceCube experiments.
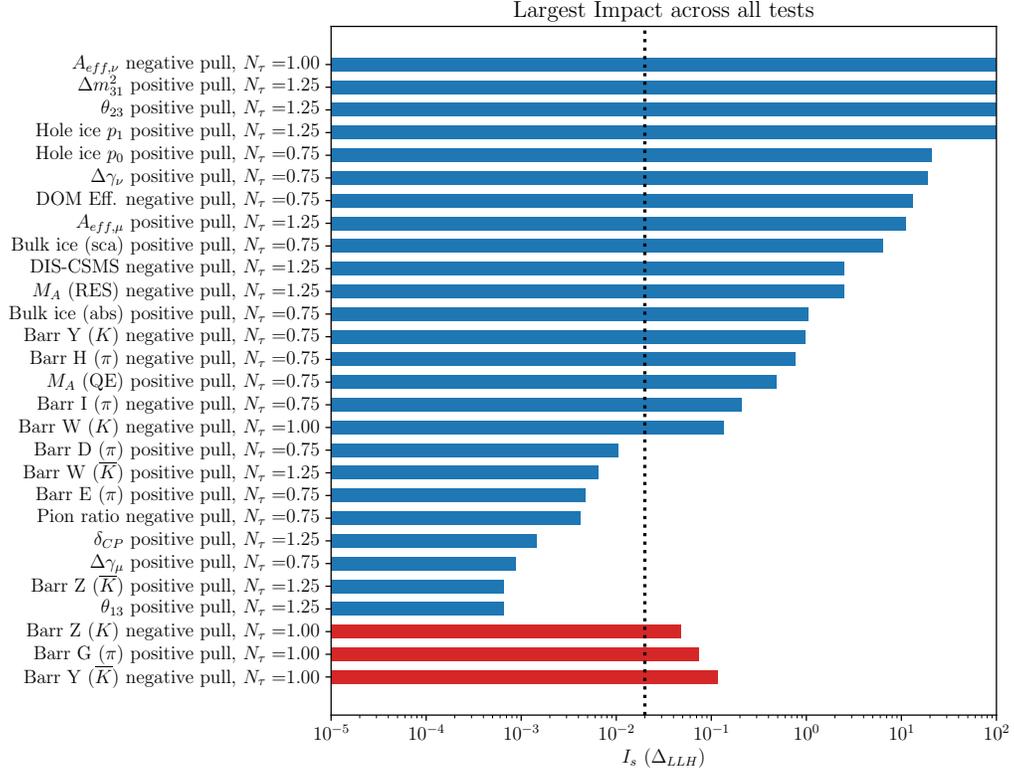
Figure 5: Impacts of systematic uncertainties on the tau neutrino appearance analysis, carried out by fixing a single parameter instead of fitting it as a floating parameter. The DOM efficiency is the seventh largest systematic and the third largest systematic directly related to the detector. Taken from Ref. [29].

1.6 DOM Efficiency Factor

The laboratory-measured DOM efficiency can not be expected to be identical to the in situ DOM efficiency after the DOMs are placed in the ice at IceCube. We expect effects on each individual DOM based on localised ice properties and individual PMT efficiency variations in the DOMs. The characteristics of the ice environment is not yet fully understood, and models for the DOM efficiency factor must be based on Monte Carlo events and simulated responses. Flasher data[30] is not a sufficient calibration tool as they are not consistent light sources. Repeated light flashes using the same flasher settings have been observed to give different charge on the DOMs. A good calibration source candidate is atmospheric muons, see Section 1.7. As this is an external light source, the amount of photons hitting the PMTs is heavily dependent on the photons reaching the PMTs from the source, which can be inconsistent. Current modelling and analysis suggest that the primary cause for inconsistent photon propagation is the absorption lengths of the ice which changes significantly with depth[31]. In particular, there exists a "dust layer" between 2000 m and 2100 m which heavily suppresses the travel length of photons[31]. Figure 6 shows the absorption length as a function of depth. Given that quantum efficiency is calculated from photon hits on the PMTs which depend on photon propagation, there is a direct relationship between the DOM efficiency and photon absorption in the ice. It is of interest to develop a robust method to check the relatively efficiency between DOMs of expected similar amount of photon hits.

Studying this relative DOM efficiency in greater detail requires a comparison between increasingly more complex detector simulations and data from in situ measurements. Monte Carlo methods allow us to simulate photon propagation and detector response from millions of events and calculate the individual DOM efficiency from the mean charge of the total amount of events. In order to take the depth dependent absorption length into account, we divide the DOMs into groups based on depth with the expectation that DOMs at similar depths will have similar amount of photon hits, see Figure 7. After binning each DOM into groups, we now define the Relative Individual DOM efficiency, RIDE, as

$$RIDE_i = \frac{\left(\dfrac{\sum_{events} q}{\sum_{events} hit}\right)_i}{\left(\dfrac{\sum_{events} q}{\sum_{events} hit}\right)_{monitor}}, \tag{12}$$

where the monitor refers to the median NQE DOM charge of the group the ith DOM is in. A DOM is only part of an event if it is expected to have a photon hit based on its distance from the light source, see Section 2.3. $\sum_{events} hit$ is the total number of
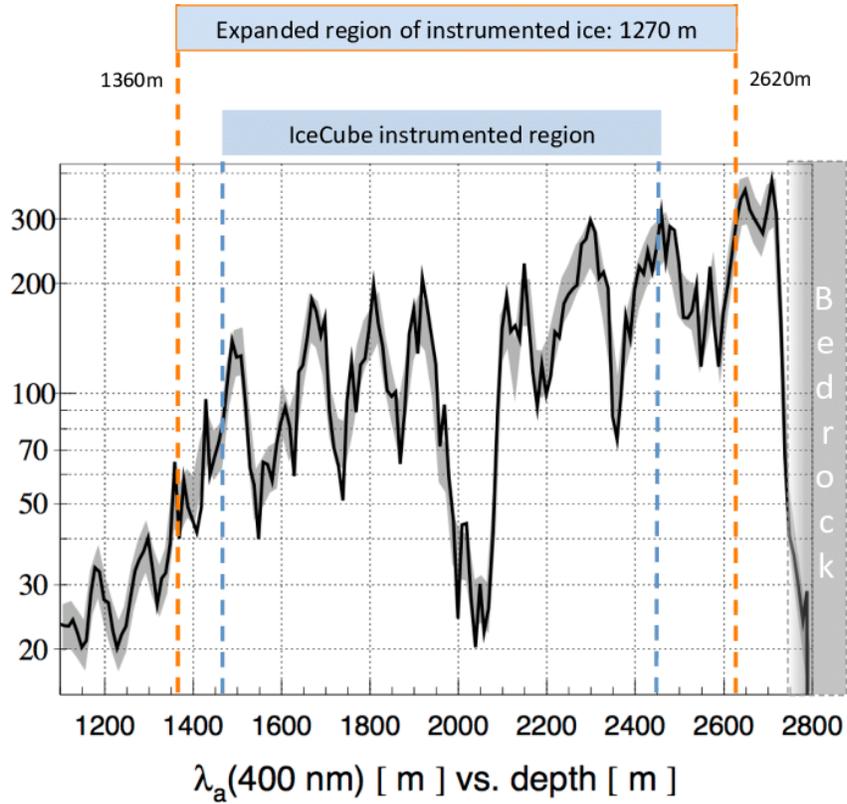
11

Figure 6: Absorption length in the ice as a function of depth in surface coordinates. At around 2000 m depth the absorption length drastically decreases. Indicated is also the current instrumented region of IceCube. Taken from Ref. [32].
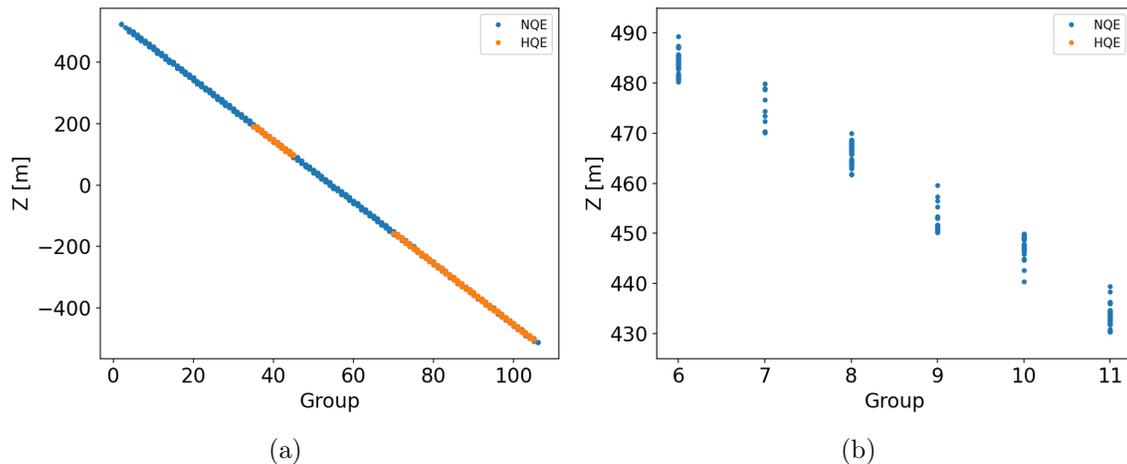
12

Figure 7: The binned DOM groups that the monitor is calculated from with NQEs and HQEs individually marked. The bins are defined in fixed 10 m depth spacings. (a) the DOM depth as a function of group in the entire detector. (b) a zoom-in of the top of the detector, showing the individual DOMs in 5 groups. Note the difference in DOM density per group.

events the DOM is part of and $\sum_{events} q$ is the total charge deposited by those events. RIDE is a normalised value that will return 1 for every DOM if the mean charge of the DOMs in each group is exactly the same. Based on the measured quantum efficiency, we expect the NQE DOMs to be centred around the monitor of 1 and the HQE DOMs to be centred around 1.35. Previous papers on this topic[33] were able to reproduce the expected quantum efficiency using uniformly produced muons and a uniform bulk ice model. This confirmed the fundamental usefulness of the calculation method and allows continued analysis with simulated atmospheric muons and depth-dependent ice models that include tilted anisotropic ice layers.

The simulated mean charge response of the DOMs is illustrated in Figure 8. We see immediate (if mirrored) correspondence between the mean charge deposited in each DOM and the absorption length from Figure 6. The principal analysis goal is to reproduce the expected quantum efficiency in simulation files. By directly comparing simulated input DOM efficiencies (1.0 for NQEs and 1.35 for HQEs) and output RIDE values, this analysis will allow us to pinpoint the properties that causes deviation from the current models.
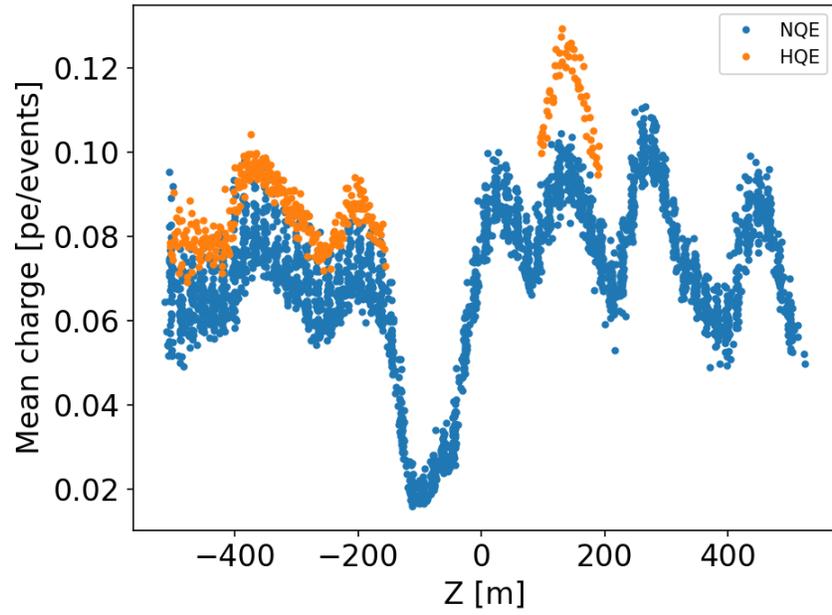
Figure 8: The mean charge $\sum_{events} q / \sum_{events} hit$ of each DOM as a function of depth in IceCube coordinates. Note the decrease in mean charge around $z = -100$ and the general increased mean charge of high quantum efficiency DOMs.

14

1.7 Minimum Ionising Muons

As mentioned previously, the in-built flasher boards of the PMTs are not sufficient calibration sources. What is needed is another light source which should be in large abundance and present in the entire detector. The light sources should travel long distances in straight lines through the ice medium. They should also radiate a near constant amount of light as they travel through the detector, which is known as a standard candle in astrophysics. The atmospheric muon tracks described in Section 1.4 fulfil these criteria and will be the focus of this study.

As muons travel through the ice, they lose energy via several dominant mechanisms, including ionisation, bremsstrahlung, pair production, and photonuclear interactions. Ionisation refers to the excitation of nearby electrons as a particle travels through the medium. Bremsstrahlung is the radiation that occurs when a particle is decelerated by the electric field of nearby charged particles. This type of radiation is continuous. Pair production refers to the creation of an electron-positron pair from a high energy photon. Photonuclear interactions describe a high-energy photon absorbed by an atomic nucleus which then emits a subatomic particle. The last three are examples of radiative energy loss. The cumulative effect of these losses is known as the mass stopping power and is well described by the Bethe equation,

$$\left\langle -\frac{dE}{dx} \right\rangle = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A} \frac{z^2}{\beta^2} \left[\frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \quad (13)$$

where N_A is Avogadro's constant, r_e and m_e is the electron radius and mass respectively, c is the speed of light, Z and A is the atomic number and mass of the absorbing material respectively, z is the charge of the incident particle, β is the velocity of the incident particle over the speed of light, γ is the Lorentz factor, W_{max} is the maximum possible energy transfer to an electron in a single collision, I is the mean excitation energy, and $\delta(\beta\gamma)$ is the density effect correction. In the region $0.1 \lesssim \beta\gamma \lesssim 1000$, this equation is accurate to within a few percent[34]. A plot of the Bethe equation for muons on copper can be seen in Figure 9. At low energies, the ionisation losses dominate. The effect of ionisation loss decreases as the muons become more energetic. Meanwhile, the radiative losses at high energy begin to dominate as the effects of bremsstrahlung, e^+e^--pair production, and photonuclear contributions increase.

In-between these two extremes, at energies between ~ 100 MeV and ~ 100 GeV, we reach a point where both the ionisation and radiative losses are minimal and the mass stopping power becomes comparatively near-constant as a function of muon energy. The radiative effects are less than 1% and the small loss we see is almost entirely due to ionisation of the surrounding medium. This region of nearly constant minimum ionisation will radiate a near constant amount of light which makes for an
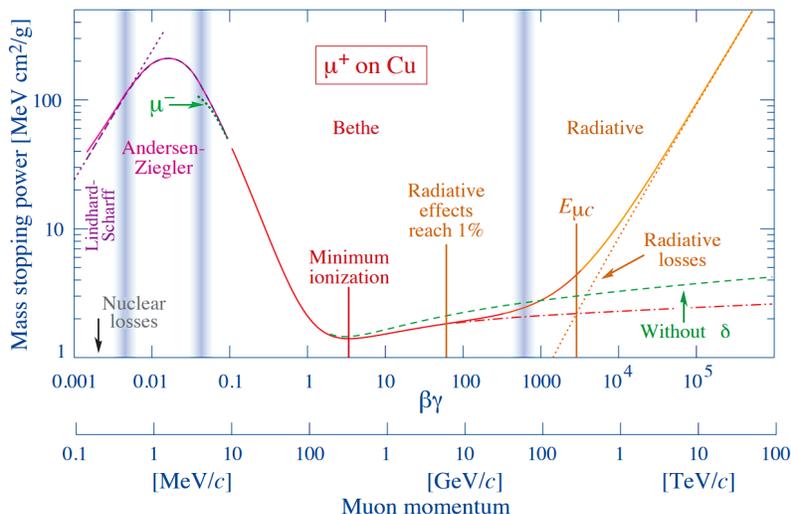
Figure 9: The stopping power of muons in copper as a function of momentum. Shown are 12 orders of magnitude in kinetic energy. Solid curves indicate total stopping power with dotted lines indicating individual effects. The colour indicates the main source of energy loss. Taken from Ref. [34].

excellent calibration source and will function as our standard candle for this study.

In order to mark muons as minimum ionising in IceCube, we choose to approximate this characteristic with stopping muons. Stopping muons refer to any muons that stop inside the detector. At a stopping power in the regime of 1, a muon will lose roughly 1 MeV per centimeter travelled in the medium of ice where $\rho \approx 1$ g cm^{-3}. The last few thousand meters will then correspond to a loss of a few hundred GeV. Any muon that stops within the detector will then at most have a few hundred GeV of energy while in the detector and, since IceCube is not sensitive to events with energies below 100 MeV, any detected muon in this regime will be a minimum ionising muon.

Even though atmospheric muons have high energy and would ordinarily not come to a stop easily, travelling through the ice reduces the energy of the muons to the point of being able to stop within the detector. The atmospheric muon flux covers a range of six orders of magnitude and is primarily limited by the minimum amount of energy required to penetrate the ice, see Figure 10. The largest amount of muons seen in IceCube have just enough energy to reach the detector and are thus likely to stop within it. We therefore expect a large amount of minimum ionising muons to be available for our analysis.
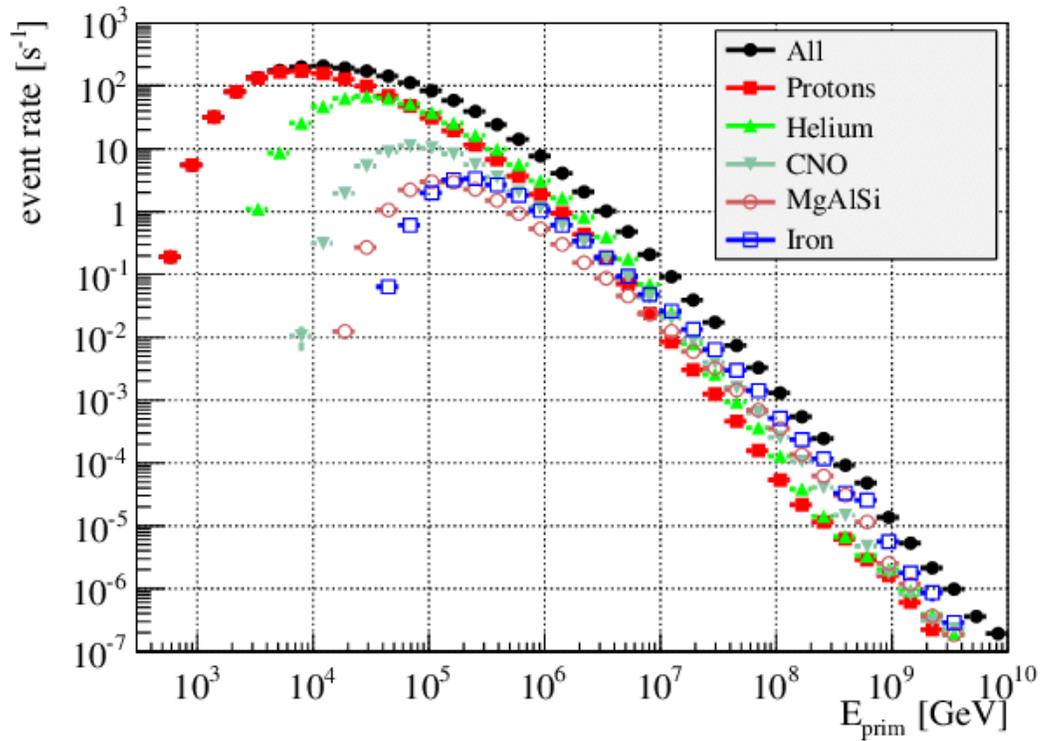
Figure 10: The atmospheric muon flux as a function of energy based on the simulated primary types. The downward turn around 10^4 GeV indicates that at these energies the muons cannot penetrate the ice to reach the detector. Taken from Ref. [35].

17

2 Relative Individual DOM Efficiency

Analysis will be done on simulated Monte Carlo files where the DOM efficiency is known prior in order to show that our methods will give back the expected RIDE results. As the goal here is to get agreement between simulated input and output DOM efficiency, reconstruction techniques will not be used. This means that the phenomenology of the primary particles, muon direction and position, and other reconstructable variables will be taken directly from the simulation as true values as opposed to reconstructed values. This decreases the amount of unknowns in our analysis and allows us to ascertain whether we can get the expected RIDE values assuming perfect knowledge of the particles involved.

2.1 Simulations

IceCube employs a large variety of different simulations based on analysis needs. These models exist within the IceTray framework, a highly modular and flexible framework that allows a large variety of services and modules to be processed in a single environment[36]. The simulation of muon and neutrino interactions are done in four steps: Generation of primary particles, propagation of light emitting tracks and cascades, generation of photoelectrons on the PMTs, and finally simulation of the detector response.

As mentioned in Section 1.6, previous methods[33] made use of two highly simplified models: Directly producing uniformly distributed muons across the entire detector, and a basic, uniform bulk ice model. Said method confirmed the usefulness of the RIDE model in principle. In this thesis, we aim to implement a full muon simulation as outlined above with modern ice models that closer match observed ice properties.

2.1.1 CORSIKA

CORSIKA stands for COsmic Ray SImulations for KAscade and is a Monte Carlo simulation tool[37]. It simulates particle air showers generated by high energy cosmic rays. dCORSIKA, a modified release based on CORSIKA optimised for IceCUBE and AMANDA, is the primary simulation tool used by IceCube. dCORSIKA is used for down-going muon simulations and thus contains improvements to the original CORSIKA code such as a curved Earth surface and more accurate muon simulations at zenith angles close to 90°[38].

The default dCORSIKA uses the poly-gonato model, which simulates all elements up to iron[39]. We use a simplified 5-component model that only simulates protons, helium, nitrogen, aluminium, and iron. The simulation initiates and propagates air

18

showers of cosmic rays into the IceCube detector. After generating the primary particles through cosmic ray air showers, the photons are propagated using CLSim.

2.1.2 CLSim

CLSim is a photon propagation software package specifically designed to work with the IceTray framework[40]. It works by tracking every single photon generated by a source in the detector and propagating those. This is a very time-consuming process that is made 100 times faster by using a GPU with the OpenCL framework that CLSim uses.

CLSim simulates the photons from a particle at steps under the assumption of a constant speed each step. At each step, the Cherenkov angle and number of photons is calculated, propagated, and checked if they hit a DOM.

2.1.3 Datasets

IceCube has currently generated two large-scale atmospheric muon datasets using dCORSIKA. These two datasets match the IceCube configurations in 2012 and 2016 respectively and will from now on be referred to by those configuration years. Both datasets generate air showers with a weighed $E^{-2.6}$ spectrum at an energy range of 600 GeV to 10^5 GeV and an angular range of $0°$ (straight-down the detector) to $90°$ (from the side). The 2012 dataset generates 10 million air showers, while the 2016 dataset generates 1 million air showers.

The primary differences between the datasets lie in the hadronic interaction models, ice models, wavelength dependent DOM efficiencies, and noise models. The 2012 dataset uses Sibyll 2.1, a widely used hadronic interaction model[41], while the 2016 dataset uses the newer version Sybyll 2.3c, which contains numerous improvements to Sibyll 2.1, including new cross section fits, increased rate of baryon-antibaryon pair production, and a phenomenological model for describing charm particle production[42]. For extensive air showers, the depth of the average maximum position of showers is decreased and the number of muons is larger by a factor ∼1.35, especially at low energies. This could lead to an overall larger ratio of stopped muons for the 2016 dataset, but the DOM efficiency itself is not expected to be greatly impacted by the hadronic interaction model.

The 2016 dataset also contains a large number of coincident events. In practical terms, this dataset will contain individual events with multiply primary cosmic rays, whereas the 2012 dataset separates all coincident air showers into individual events. Further discussion regarding coincident events can be seen in Sections 2.2.1 and 3.

The 2012 and 2016 datasets use the ice models SpiceLea and Spice3.2 respectively. SpiceLea includes the parameterisation of anisotropy in the bulk ice[43]. Spice3.2 uses
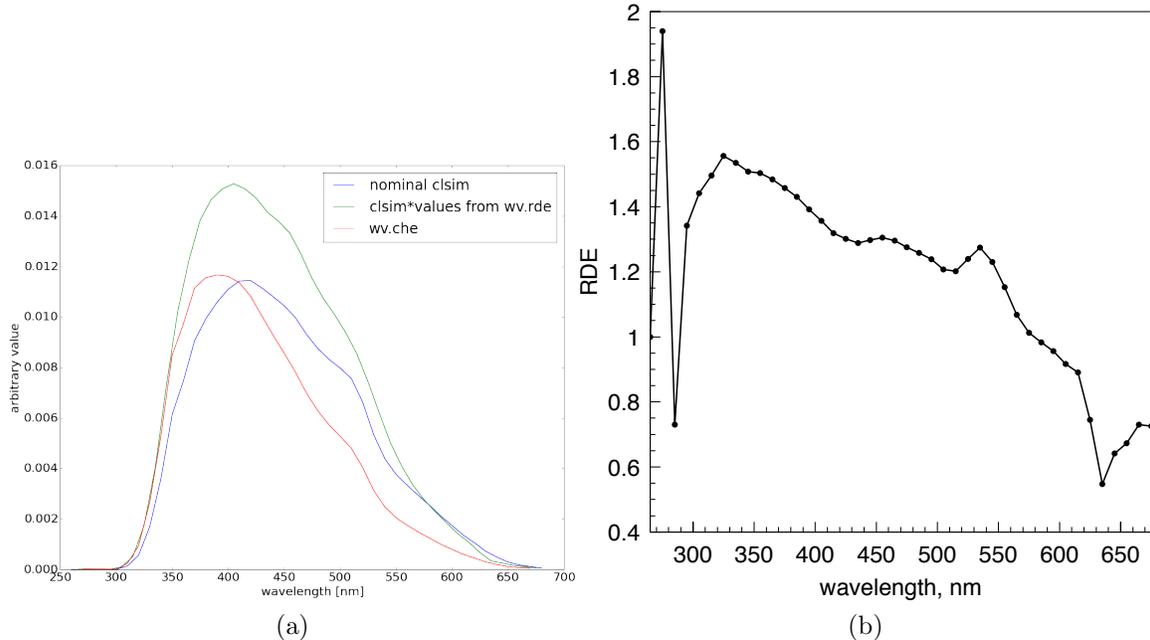
Figure 11: The effect of wavelength on the DOM efficiency value. (a) the *wv.rde* values adding a wavelength dependent factor on the DOM efficiency compared to nominal CLSim. (b) The relative DOM efficiency as a function of wavelength. Taken from Ref. [45].

flasher data from all 86 strings of the IceCube detector. It contains improvements to scattering and absorption, model error, and includes updated tilt maps. The overall effect compared to SpiceLea is 4% less absorption on average[44].

The 2016 dataset includes the introduction of a wavelength dependent DOM efficiency, which could modify the nominal HQE RIDE values away from the expected flat line of 1.35, see Figure 11. The RIDE calculation is relative and deviations from the nominal HQE value of 1.35 is expected to occur only if the wavelength dependency has a magnified or reduced effect on the HQEs compared to the NQEs. This analysis allows us to study the possible effects of wavelength dependency on the relative DOM efficiency.

In an attempt to better simulate real detector data, noise is added to the simulations. The 2012 dataset most likely uses a purely Poissonian noise model. The noise in the 2016 dataset is generated using the noise simulation module Vuvuzela[46]. Vuvuzela splits the noise profile into Poissonian and non-Poissonian noise, with the former describing thermal noise and radiative decays. The radiative decays produce a "burst" of noise that is modelled by the average number of photoelectrons emitted

20

during the burst, the mean time between them, and the standard deviation of the timing as non-Poissonian noise. The noise models help prepare the analysis for eventual attempts at calculating the RIDE on data. This analysis does not run any specific noise-cleaning algorithms, and we expect a large systematic error on the RIDE to come from the different noise models.

It should be noted that the simulation files used in this thesis is not generated by the author. Any analysis done is purely on pre-generated files.

2.2   Event Selection

The event selection's primary goal is ensuring that the chosen events are all atmospheric minimum ionising muons. These events will characterise the RIDE most accurately and allow us to compare meaningfully to data, while also giving us enough events to cover the vast majority of DOMs within the detector. The event selection will thus try to maximise the amount of events available to us within simulation files, while ensuring that all these events satisfy the minimum ionising criteria.

Event selection in simulation and data differ crucially in that we know the exact characteristics of every simulated muon, whereas reconstruction algorithms are required to predict the position of muons in data based on the DOM responses, see Section 3.

2.2.1   Initial Filtering

IceCube applies filters and triggers to each event which attempt to keep all physics related events while discarding noise events. As the current analysis uses simulated events and simulated DOM responses, a large amount of basic filtering can be done in the muon input values themselves. The first trigger, InIce, simply checks if the event is located within the detector. This separates out IceTop events. The second trigger confirms that the event is a muon, as muons are the only relevant particles for this analysis. The third trigger checks if the event has more than eight DOM responses, which every event passes. Specifically for the 2016 dataset, we implement a filter that removes all coincident events, events with more than one primary particle. These initial filters cut down the number of events early before the more complex filters begin. Note that these are all simulation filters that are not related to the filters and triggers run on actual data in IceCube.

2.2.2   Endpoint

As explained in Section 1.7, minimum ionised muons are approximated as stopped muons in this analysis. As such, the primary event selection is simply the selection of

stopped muons, characterised as having stopped within the detector boundary. This is calculated by considering the boundary formed by the outermost strings in IceCube (see Figure 12), and excluding all muons that are not within that boundary. In order to make sure that the stopping muons travel inside the detector for long enough to generate visible light, we additionally impose the restriction that the muon has to stop within 100 m of the detector boundary and above -400 m depth, measured in IceCube coordinates.

2.2.3   Angle of Incident

The angle of incident, or the zenith angle, of the muons refers to the angle at which the muons enter the detector, with 0° representing a muon travelling straight down into the detector and 180° represents a muon travelling from the earth straight up through the detector. Muons between 0° and 90° are known as down-going whereas muons between 90° and 180° are known as up-going. Due to the density of the earth heavily suppressing up-going muons, the simulation restricts the muon zenith angle to between 0° and 90°. In addition, previous studies[33] imposed a further cut between 40° and 70° in order to only keep muons which would produce tracks that directly hit the underside of the DOM, see Section 1.5.2. Other studies claimed that this inclusion is superfluous and favour removing it to increase statistics[47]. In Section 2.4, we analyse the exact effects of a zenith angle cut on the RIDE distribution.

2.2.4   Muon Bundles

The IceCube detector is triggered by approximately 3000 atmospheric muons produced in cosmic ray air showers per second. Cosmic rays often contain non-zero muon multiplicity, where a single primary particle produces multiple muons sharing similar opening angles. Simulations carried out by dCORSIKA and Sibyll characterise these multiplicities in single events known as muon bundles. The energy deposited by these bundles, and thus the photon propagation and charge deposited in the DOMs, is approximately proportional to the amount of muons contained in them[35]. This causes an inconsistency in the DOM efficiency, as deposited charge can no longer be traced back to the Cherenkov light from any single muon, rendering other event filtering useless. This analysis will thus only be looking at single-multiplicity events and we deploy a filter where all events that contain multiple muons are discarded.

2.3   DOM Exclusion

In addition to individual event cuts, the DOMs involved in the analysis of each event have to be restricted as well. All DOMs that are not considered functioning are
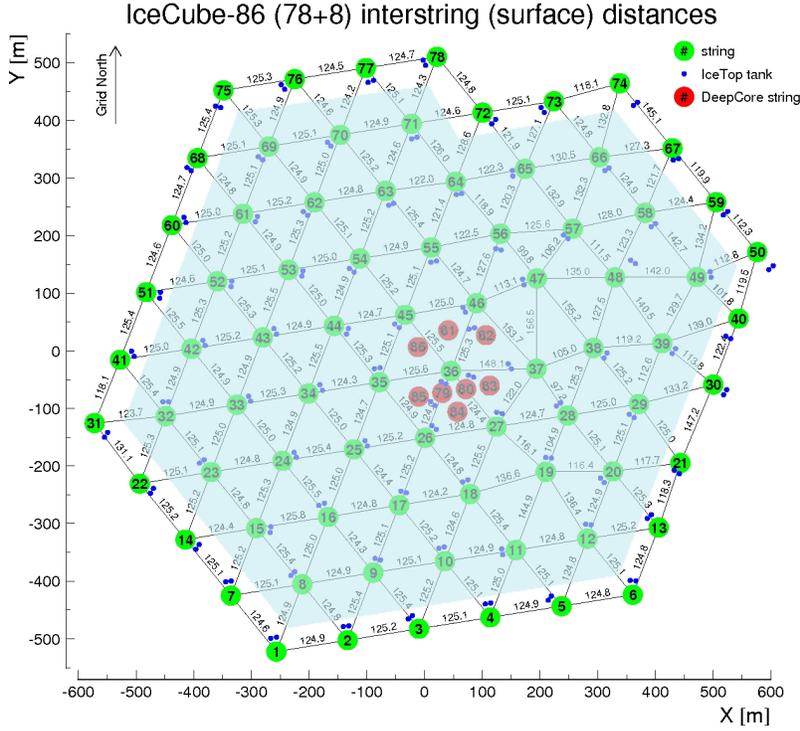
Figure 12: A top-down view diagram of the IceCube detector. Shown are the string numbers at their position in x and y with DeepCore strings labelled. The blue shaded area represents the RIDE analysis region. The unshaded outer strings are removed from the analysis and the boundary formed by them creates the stopped muon criteria. Taken from Ref. [48].

naturally excluded from the analysis entirely. The DOMs located in the outermost strings of the detector will generally see very little light, partly due to the endpoint restrictions of the muons, and are a source of error and bias on the RIDE analysis. As such, the DOMs in the strings outside the shaded area in Figure 12 are removed from the analysis.

As the distance from the track increases, Cherenkov light becomes increasingly more scattered. DOM responses far away from a muon track are likely to be uncorrelated noise or come from these highly scattered photons. In order to reduce this effect, we impose a maximum distance cut from the track. This cut will need to be balanced with the relative decrease in statistics.

Recent studies[47] predict that DOMs located very close to the track do not replicate the true charge information. A study on a minimum distance cut from the track is therefore also necessary. We will refer to these two quantities as the maximum and

minimum DOM-distance.

It should be noted that all DOMs that pass the outer-strings and DOM-distance cuts will be considered in the event even if they do not detect any charge.

2.4   Individual Group Analysis

Based on initial cuts on DOM characteristics, four parameters remain that need to be studied in detail: The minimum and maximum distance from any given DOM to the track in each event (DOM-distance), and the minimum and maximum opening zenith angle of said track (zenith angle). We are working with simulation files and know that each NQE and HQE DOM has a set relative input efficiency of 1.00 and 1.35 respectively. This efficiency should be mirrored in the RIDE values. Changing the zenith angle and DOM-distance cuts have no direct impact on the ability of each DOM to convert PMT photon hits into a charge response. The former will only impact what muons are allowed into the analysis, while the latter will only impact the probability of a photon from a muon track hitting a DOM. Since a photon from a track farther away from a DOM is less likely to hit the DOMs PMT, increasing the maximum DOM-distance should decrease the mean charge equally for all DOMs. So we expect neither the zenith angle nor the DOM-distance to impact the relative DOM efficiency. As we increase the maximum DOM-distance, the probability of any DOM within the track having charge decreases, but so does the probability of falsely discarding a DOM that did get a PMT hit from the event. A low DOM-distance has a larger ratio of charge-carrying DOMs but less statistics than a high DOM-distance. On the other hand, too large a DOM-distance increases the probability of noise hits. If the minimum and maximum DOM-distance is within only a few dozen meters of each other, we risk discarding the vast majority of DOM hits and losing a lot of statistics. As such, it is possible that too large maximum or minimum DOM-distances will introduce variations from the expected RIDE value through either an increase in noise or a decrease in statistics.

Based on these expectations, we want to study the exact effect on the RIDE when we vary the zenith angle and DOM-distance. To study these parameters, we scan over a range of possible values and calculate the RIDE for each value combination. Since the RIDE is a relative value depending only on the depth group any given DOM is in, we can separate the calculations into individual groups and analyse them in detail separately. Due to the need for high statistics to get meaningful RIDE values, this analysis requires high amounts of computation time and memory usage, up to hours and gigabytes of memory per group. Therefore the parameter scan will only be done over three groups. Since the study focuses on the high DOM efficiency PMTs located in DeepCore, we will focus on three groups located in the centre of DeepCore. Note that this analysis is carried out on events that have passed all initial, endpoint, and

[Minimum, Maximum] DOM depth		
Group	Z-bin group depth [m]	TIL group depth [m]
1	$[-200, -190]$	$[-205, -193]$
2	$[-250, -240]$	$[-255, -233]$
3	$[-290, -280]$	$[-300, -267]$

Table 1: A table of the minimum and maximum depths of the DOMs in the three selected groups for the z-bin and TIL groups. The two group definitions are not exactly alike but the groups are chosen such that the minimum and maximum depths of the TIL groups most closely matches the z-bin groups in each respective group.

muon bundle filters. Additionally, in order to only look at DOMs relevant to the overall study, a strict maximum distance cut of 200 meters between the DOMs and the track has been imposed. The study is thus focused on a DOM exclusion between 0 and 200 meters only. This value was chosen in order to have at least 50 meters of analysis room based on the chosen cuts of previous studies[33, 47].

The current group definition is based on uniform bulk ice layers from previous studies[33] and not the tilted anisotropic ice layer modelling of SpiceLea and Spice3.2. It is therefore of interest to analyse the RIDE with a group definition based on the ice layers instead of the uniform 10 m depth spacing previously used, see Figure 7. The new definition attempts to group DOMs into layers of similar ice properties instead of purely by depth in order to better follow the assumption of similar amount of photon hits. The DOMs in this definition are not grouped into similar z-bins. We will refer to the groups of uniform 10 m depth space binning as the z-bin groups, and the groups following the tilted ice layer models as the tilted ice layer (TIL) groups.

Based on the two group definitions, we pick three groups in the centre of DeepCore to analyse. In order to compare the two definitions meaningfully, we have chosen to pick groups that share a similar depth binning. The bin range of the two definitions can be seen in Table 1 and additionally illustrated in Figure 13 with the three chosen groups included.

With the groups selected, analysis can begin. Initially, we scan over the four parameters, the minimum and maximum DOM-distance and zenith angles, in pairs of two and plot the RIDE-value for each parameter-pair. We do this twice, once for the z-bin groups and once for the TIL groups, and repeat again for the NQEs and HQEs and also for the 2012 and 2016 datasets. We scan over the zenith angles from $0°$ to $90°$ in bins of $2°$ and use a DOM-distance of 0 m to 200 m. We scan over the DOM-distance from 0 m to 200 m in bins of 5 m and use the full angular spectrum ($0°$ to $90°$). A scan over all four parameters individually would be in four dimensional
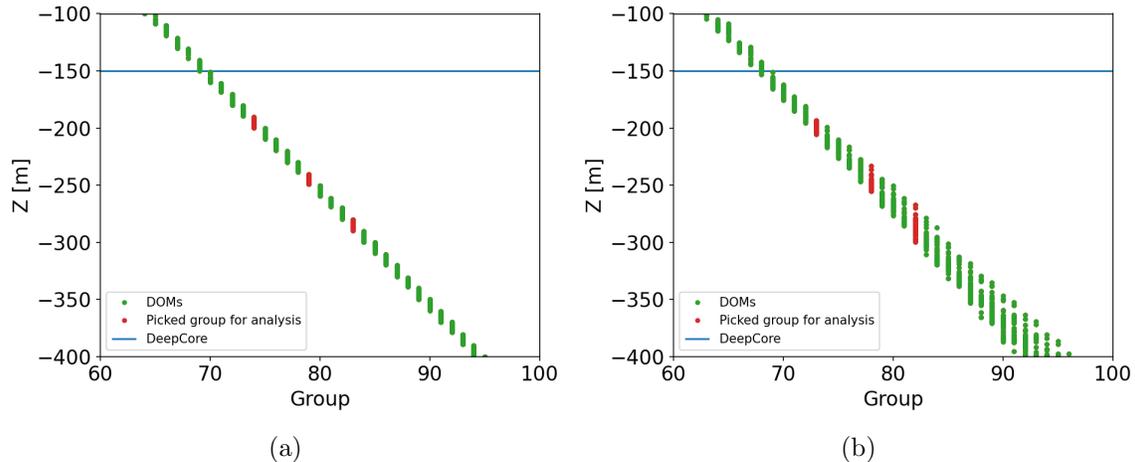
Figure 13: Plot of the depth as a function of the two group definitions. Shown are also the DOMs in the three analysis groups, and where DeepCore begins. (a) the z-bin groups, (b) the TIL groups. Note that neither plot shows the full detector, but only the DeepCore region relevant to the three groups.

space and require over 400 times as much computational power, which is out of scope of this thesis.

The aim is to find a combination of DOM-distances that allow a large amount of statistics while making sure that each charge-carrying DOM is from the photons produced by the muon track in the event. We first calculate the mean RIDE value across an entire group to get a single number that signifies the general RIDE value of the DOMs.

Starting with the 2012 dataset, we scan over the DOM-distance. We see a clear change in RIDE as a function of both minimum and maximum DOM-distance for all three HQE groups (Figure 14 top). In general, there is a region that returns the expected RIDE of 1.35 and as we decrease the minimum or maximum DOM-distance the RIDE also decreases. Increasing the maximum DOM-distance also increases the RIDE, whereas increasing the minimum DOM-distance increases the RIDE up to a certain point for the two upper groups. We expected some variation based on noise and lack of statistics, but a change from 1.2 to 1.6 as we vary the maximum DOM-distance from 75 meters to 175 meters is highly unexpected and points towards some as of yet unknown effect. By simply varying the distance from the muon track at which a DOM's PMT is expected to get hit by a photon from it, the RIDE value changes by up to ± 0.25.

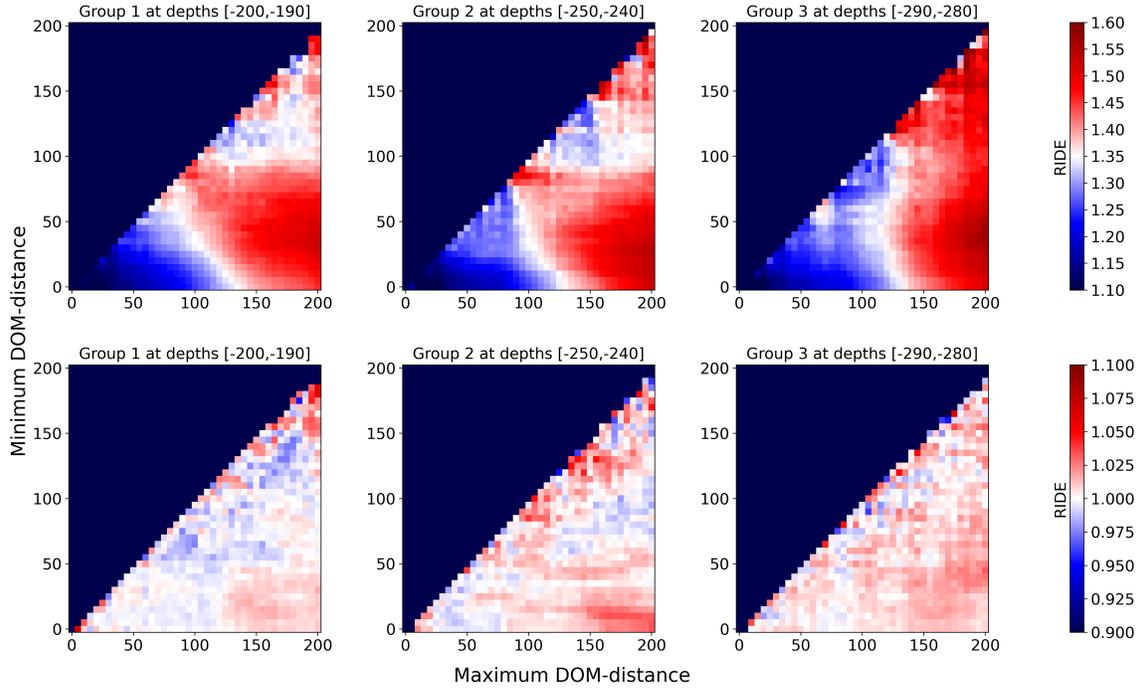As this study aims to reproduce the input simulation RIDE values of 1.35, the

26

Figure 14: The 2012 dataset mean RIDE value of all DOMs in each respective group as a function of the minimum and maximum DOM distance to the muon track with the z-bin groups. Top: HQE RIDE value centred around 1.35. Bottom: NQE RIDE values centred around 1.00. Regions with zero events are shown with a RIDE value of 0.

current course of action is picking the DOM-distance that reproduces this expectation. The HQE RIDE values (top) for all three groups converge at a 1.35 value in a line ending at approximately $[\text{min}, \text{max}] = [0, 130]$. There is also a region at $[\text{min}, \text{max}] = [110, 180]$ with convergence for the two upper groups, but is not shared by the lower group. It should be noted that as we scan over the minimum and maximum parameters, some bins will contain zero DOMs. These are represented by the smallest value on the colour scale.

The NQE RIDE values (bottom) behave more towards expectations. Since the RIDE value is normalised by the median of the NQEs only, we expect any effect that impacts the mean charge of the DOMs to be subdued for the NQEs. In the regions where the minimum and maximum DOM-distances are more than 10 meters apart, the NQE RIDE values never deviate more than 5% from 1.00. The optimal DOM-distance should then primarily be based on the HQE RIDE values instead.

To quantify the distribution of RIDE values around the mean, we calculate the

standard deviation of the RIDE in each group as well. We generally expect that in areas of low statistics, where the minimum and maximum DOM-distances are similar, the standard deviation increases. We also expect that the introduction of more noisy charge responses at larger maximum DOM-distances would increase the standard deviation as well. The HQEs are situated in the centre of IceCube even within each group, and there will generally be more muon tracks that move through the centre, as those near the edges are more likely to be cut. So the HQEs are more likely to have a mean charge closer to a central value, and we generally expect the HQE DOMs to have lower standard deviations as a result.

All these expectations are confirmed in Figure 15. At areas where the minimum and maximum DOM-distances are within 10 meters the standard deviation is above 0.2 for both HQEs and NQEs. There is also a slight increase as the maximum DOM-distance increases for the NQEs (bottom). In the areas where the HQE and NQE mean RIDE values were centred at 1.35 and 1.00 respectively, the standard deviation remains below 0.05 ($< 4\%$) for the HQEs (top) and below 0.1 ($< 10\%$) for the NQEs (bottom).

While the mean HQE RIDE value varies highly as the DOM-distance changes compared, the standard deviation does not. In addition, the HQEs retain a smaller standard deviation than the NQEs in general. This suggests that mean HQE RIDE value changes are systematic in nature and not due to a larger spread around the mean.

For the zenith angle scans, we expect the decrease in available muons to be the primary effect on the charge responses and so we expect a noisier but consistent mean RIDE value with decreasing standard deviation as the maximum zenith angle increases and the minimum zenith angle decreases. Because the zenith angle scans use a DOM-distance of 0 m to 200 m, the mean RIDE value should be similar to the RIDE values of the $[min, max] = [0, 200]$ bins in figure 14.

We see the zenith angle scans in Figure 16. At very low maximum or very high minimum zenith angles, no muons get past the zenith angle cut. The NQEs have a generally higher mean RIDE value owing to the DOM-distance, but still do not deviate more than 10%. The HQEs also behave as expected, with some increase in RIDE value near regions with small difference in minimum and maximum zenith angles. A zenith angle cut of 40° to 70° (see Section 2.2.3) retain similar RIDE values to one of 0° to 90°.

The standard deviation, see Figure 16, generally decreases for all groups and for both HQEs (top) and NQEs (bottom) as the zenith angle includes more of the sky. The standard deviations are again less than 4% and 10% for the HQEs and NQEs respectively. To maintain as much data as possible, we will not implement a zenith angle cut on the 2012 dataset going forward.

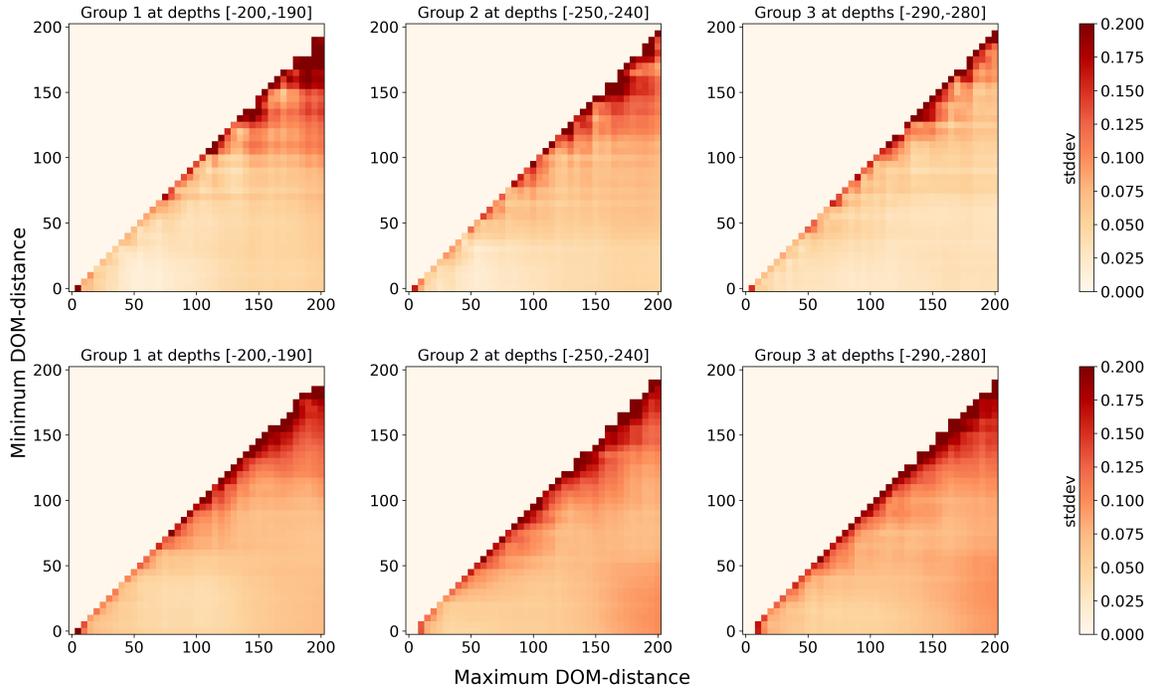The above scans were also carried out on the TIL groups, but the optimal RIDE

Figure 15: The 2012 dataset standard deviation of the RIDE values of all DOMs in each respective group as a function of the minimum and maximum DOM-distance with the z-bin groups. Top: HQE stddev values from 0 to 0.2. Bottom: NQE stddev values from 0 to 0.2. Regions with zero events are shown with an stddev of 0.
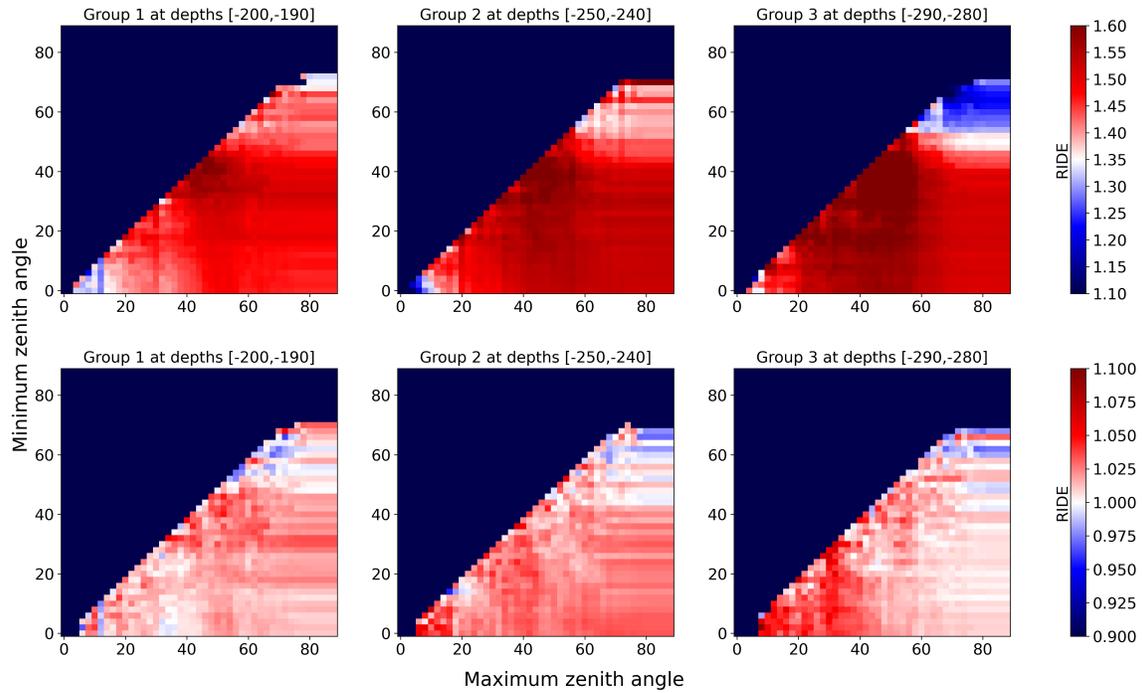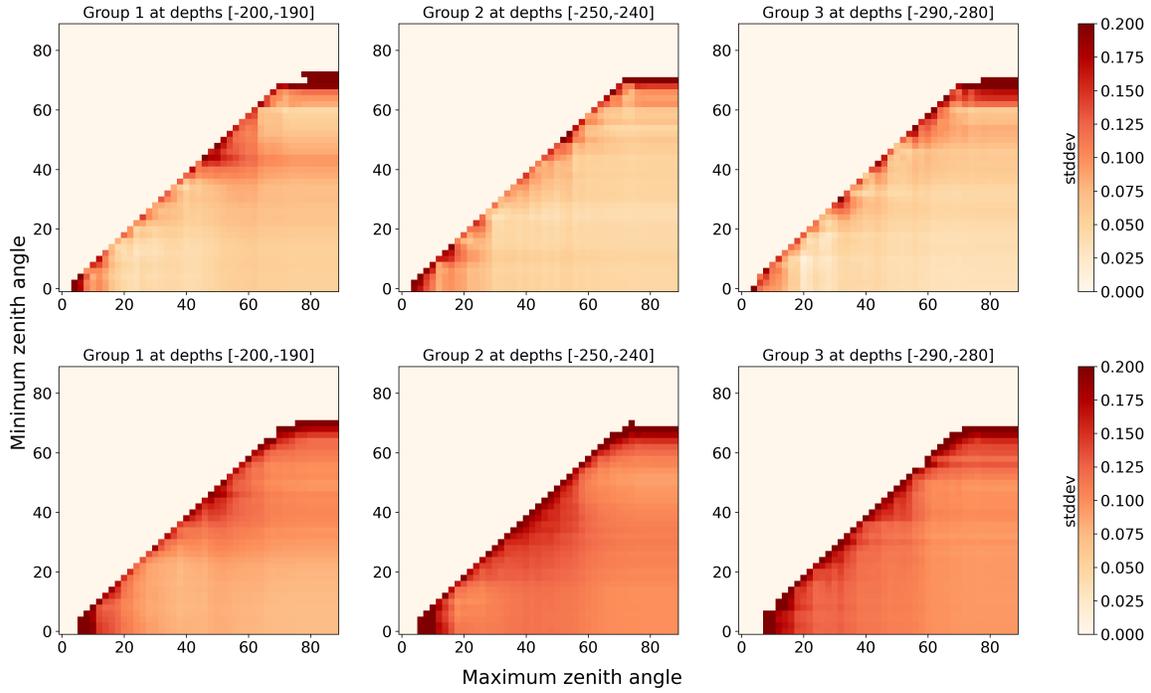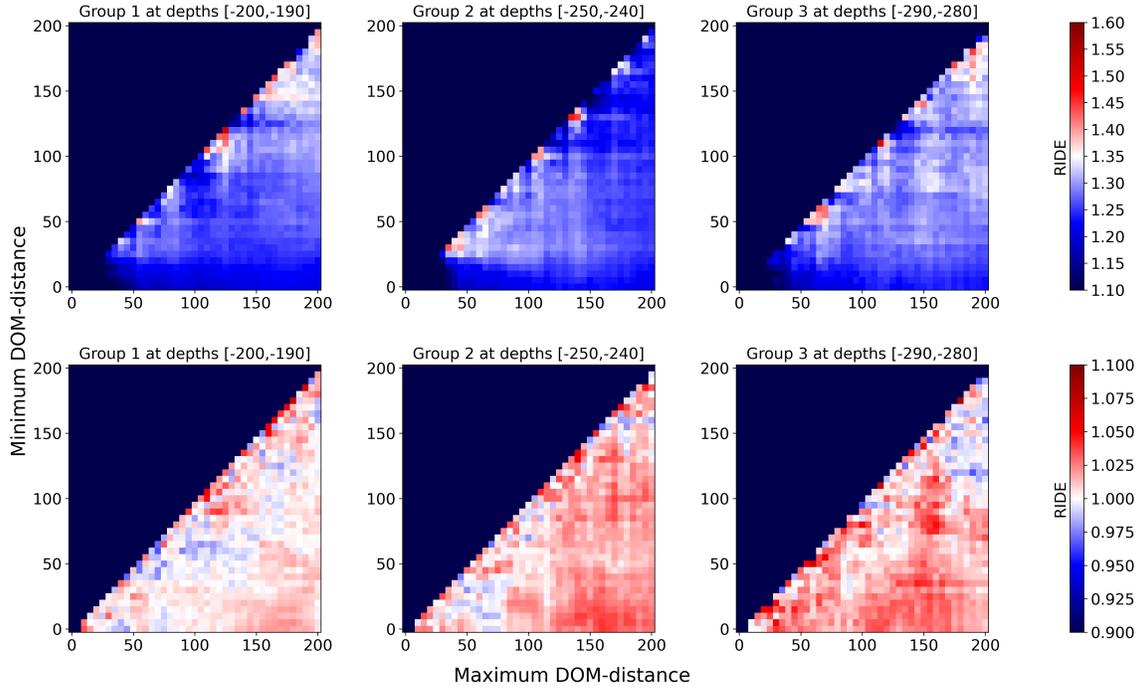
Figure 16: The 2012 dataset mean RIDE value of all DOMs in each respective group as a function of the minimum and maximum zenith angle with the z-bin groups. Top: HQE RIDE value centred around 1.35. Bottom: NQE RIDE values centred around 1.00. Regions with zero events are shown with a RIDE value of 0.

Figure 17: The 2012 dataset standard deviation of the RIDE values of all DOMs in each respective group as a function of the minimum and maximum zenith angle with the z-bin groups. Top: HQE stddev values from 0 to 0.2. Bottom: NQE stddev values from 0 to 0.2. Regions with zero events are shown with an stddev of 0.

Figure 18: The 2016 dataset mean RIDE value of all DOMs in each respective group as a function of the minimum and maximum DOM-distance with the z-bin groups. Top: HQE RIDE value centred around 1.35. Bottom: NQE RIDE values centred around 1.00. Regions with zero events are shown with a RIDE value of 0.

means and standard deviations did not change considerably, see Appendix A.

The 2016 dataset behaves more towards our initial expectations of a mean RIDE value that does not vary significantly with the DOM-distance, see Figure 18. However, the mean RIDE value that the DOMs are centred around is ∼1.25 rather than 1.35. Since no DOM-distance returns the input RIDE of 1.35, it is not possible to use this metric to gauge the best DOM-distance like we did with the 2012 dataset. The NQEs generally have higher RIDE values, towards 1.05. The standard deviations of the 2016 dataset is similar to the 2012 dataset, see Appendix A, which does not explain the decreased overall mean RIDE value across all DOM-distances.

The zenith angle scans (Figure 19) show similar effect on the mean RIDE values as the 2012 dataset and retains a mean RIDE of ∼1.25 for the HQEs and 1.00 to 1.05 for the NQEs.

Based on these figures, an agreed-upon DOM-distance and zenith angle combination that returns a RIDE value of 1.35 for all three groups is not attainable with the 2016 dataset. Due to the prior expectations discussed in Section 2.1.3, we will not
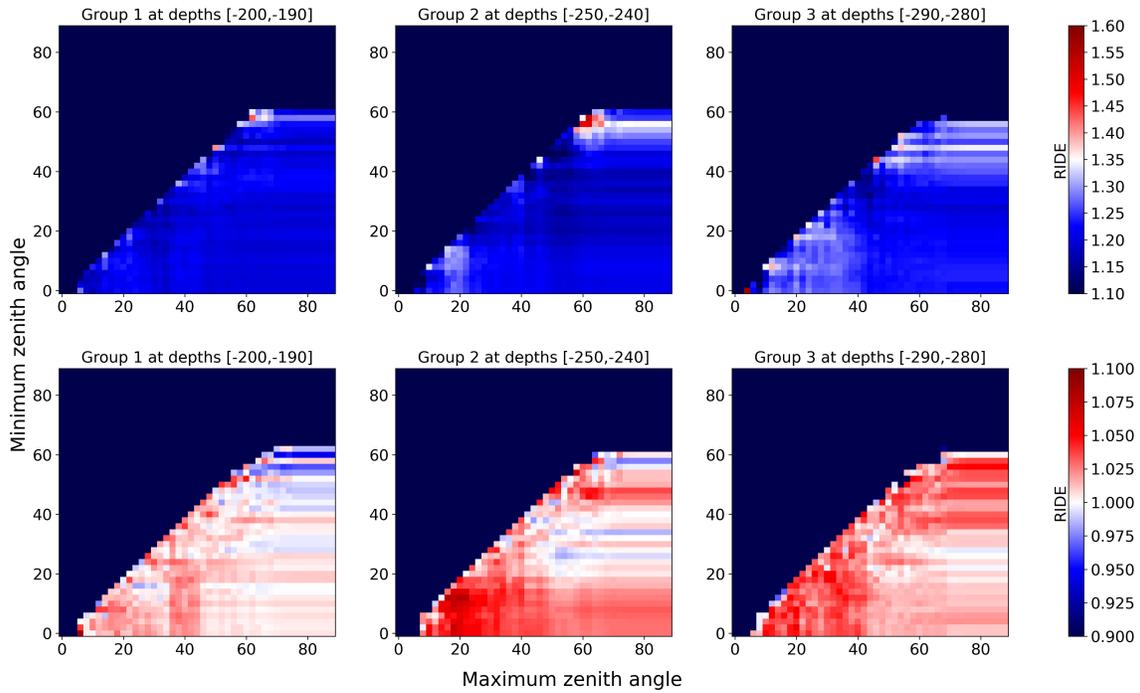
Figure 19: The 2016 dataset mean RIDE value of all DOMs in each respective group as a function of the minimum and maximum zenith angle with the z-bin groups. Top: HQE RIDE value centred around 1.35. Bottom: NQE RIDE values centred around 1.00. Regions with zero events are shown with a RIDE value of 0.
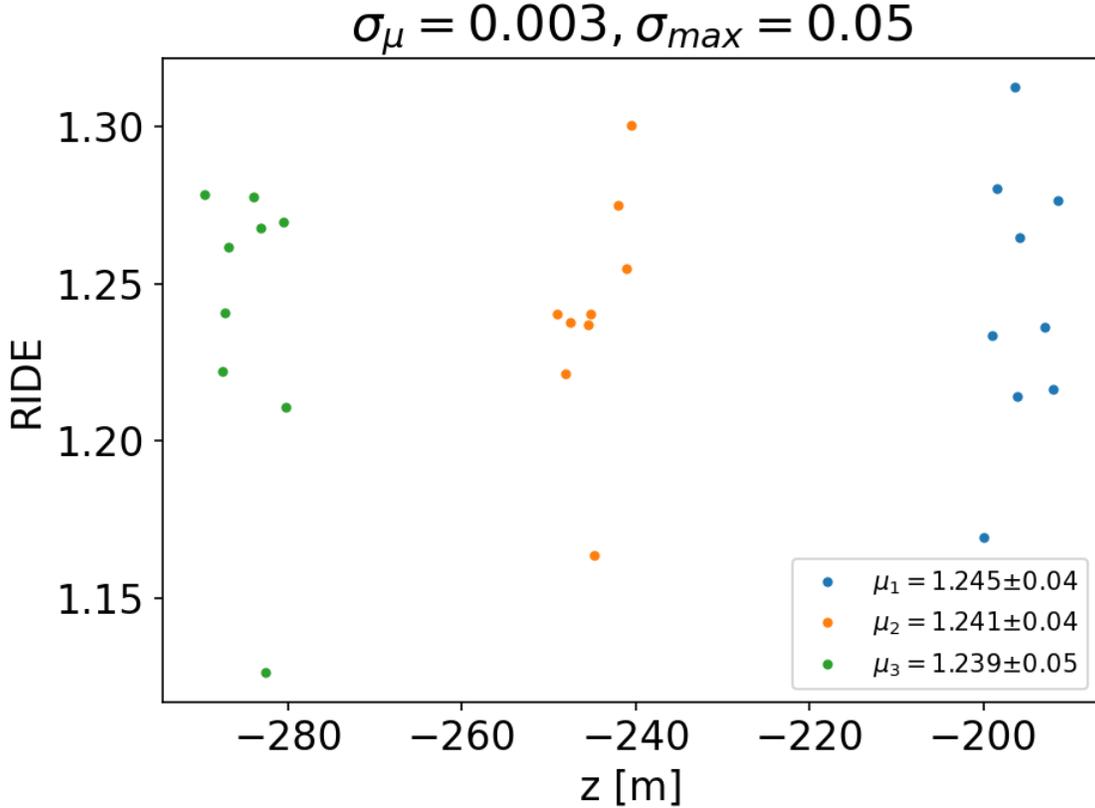
Figure 20: HQE DOMs of the three z-bin groups from the 2016 dataset with a full-sky zenith angle and DOM-distance of $[0, 200]$ m. The mean RIDE values are shown in the legend with the two metrics σ_μ and σ_{max} in the title.

use a mean RIDE value of 1.35 as the metric for the scans. Instead, we will focus on agreement across all three groups for both NQEs and HQEs while retaining the smallest standard deviation between the DOMs in each group. This will give a flat and consistent RIDE value for the largest amount of DOMs in DeepCore, but will not be centred around 1.35.

In order to find the most agreement between the three groups in our analysis, we introduce two new metrics: σ_μ, the standard deviation between the mean RIDE value of each group, and σ_{max}, the highest standard deviation between the DOMs in each group. This is calculated separately for HQEs and NQEs. An illustration of the two metrics can be seen in Figure 20.

We expect that σ_μ will be smaller than σ_{max} as the spread of the DOMs around the mean RIDE value from the standard deviation is larger than the deviation between

34

the mean RIDE of the three groups at any given DOM-distance or zenith angle combination. Because the standard deviation is larger for the NQEs than HQEs, we expect σ_{max} to be larger for the NQEs. However, since the mean RIDE between the NQE groups are in more agreement with each other, σ_μ is expected for be smaller for the NQEs.

Analysing the 2016 dataset using the two new metrics, we see that the standard deviation of the mean σ_μ remains below 0.05 for the HQEs (Figure 21 top left) and 0.025 for the NQEs (Figure 21 bottom left). The three groups generally agree on the mean RIDE for a given DOM-distance. The spread of the DOMs σ_{max} is below 0.1 for the HQEs and between 0.1 and 0.15 for the NQEs. At any given DOM-distance combination, the NQE group means are closer together than the HQE group means, but the NQE DOMs within each group are more scattered than the HQE DOMs. The former conforms to our initial findings with the 2012 dataset that the HQE mean RIDE is more sensitive to the change in parameters than the NQE mean RIDE. The latter agrees with initial expectations as well, as the HQE DOMs are more likely to be closer to their central value.

In both cases σ_μ is very noisy and picking a DOM-distance that minimises the value is not clear. The HQE σ_{max} has a more defined shape, with the minimisation occurring around a minimum DOM-distance of 0 m and a maximum DOM-distance of 125 m. This shape is not present in the NQE σ_{max} value.

The zenith angle scans are shown in Figure 22. We see a more distinct shape for the HQEs (top) where clear lines separate high and low σ_μ values at around 40° for both the minimum and maximum zenith angles. This line appears at lower maximum (20°) and higher minimum (50°) angles for σ_{max}. This separation mainly point us towards higher maximum zenith angles and lower minimum zenith angles being optimal for our analysis. This conclusion is shared by the NQE distributions (bottom), but the σ_μ shape is less defined here.

For consistency checks, the two metrics were also calculated for the 2012 dataset. Both the DOM-distance and zenith angle scans show no abnormalities in the chosen regions, excepting a smaller HQE σ_μ at higher minimum zenith angles. These checks were additionally carried out on both datasets using the TIL groups. These presented smaller overall values of σ_{max} but did not change where the minimum values occurred, see Appendix A.

Based on the above findings, we present the following conclusions: For the 2012 dataset, optimal parameters are a DOM-distance of $[0, 130]$ m and a zenith angle of $[0°, 90°]$. These values are found based on optimising the RIDE-value to be near 1.35. For the 2016 dataset, the optimal DOM-distance and zenith angles were found to be $[0, 125]$ m and $[0°, 90°]$ respectively. These values were found by optimising for a minimum deviation between groups and DOMs in each group. It is apparent that two different optimisation techniques yielded similar parameters. This might point
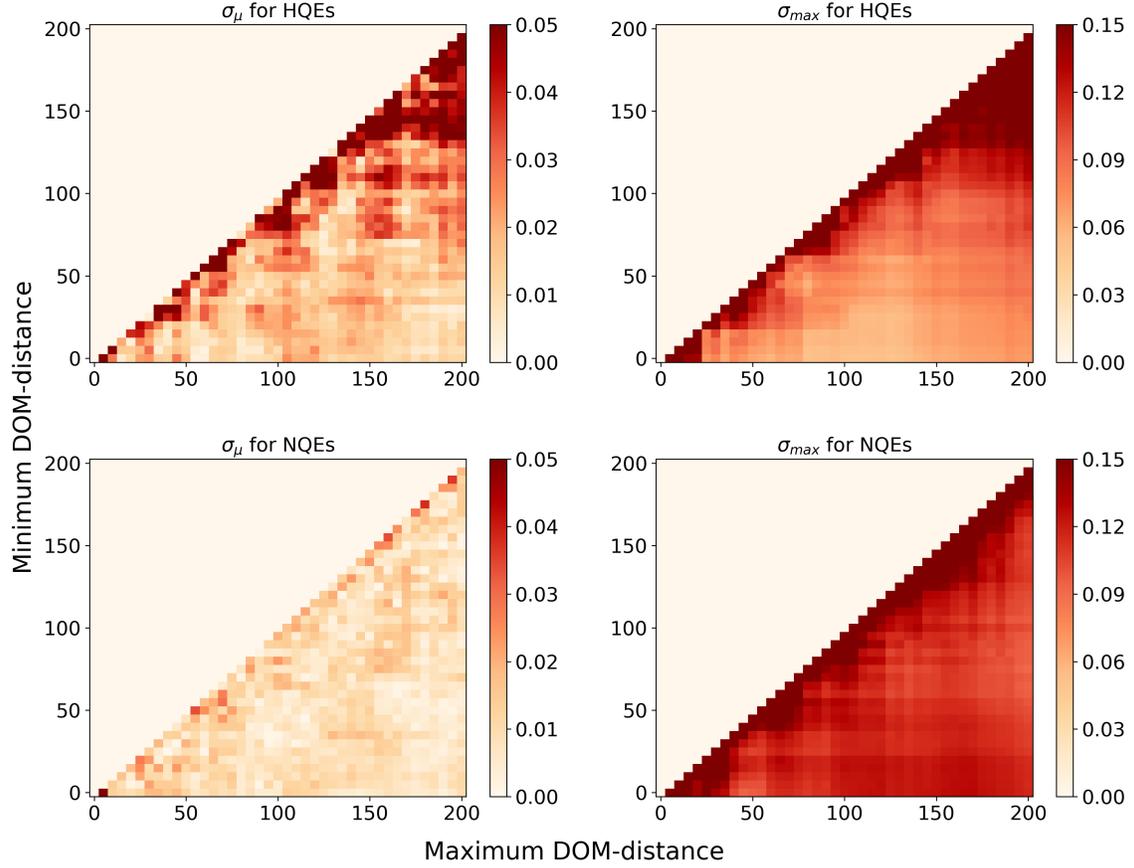
35

Figure 21: The 2016 dataset σ_μ (left) and σ_{max} (right) for HQEs (top) and NQEs (bottom) of all DOMs in each respective group as a function of the minimum and maximum DOM-distance with the z-bin groups. Regions with zero events are shown with a metric of 0.
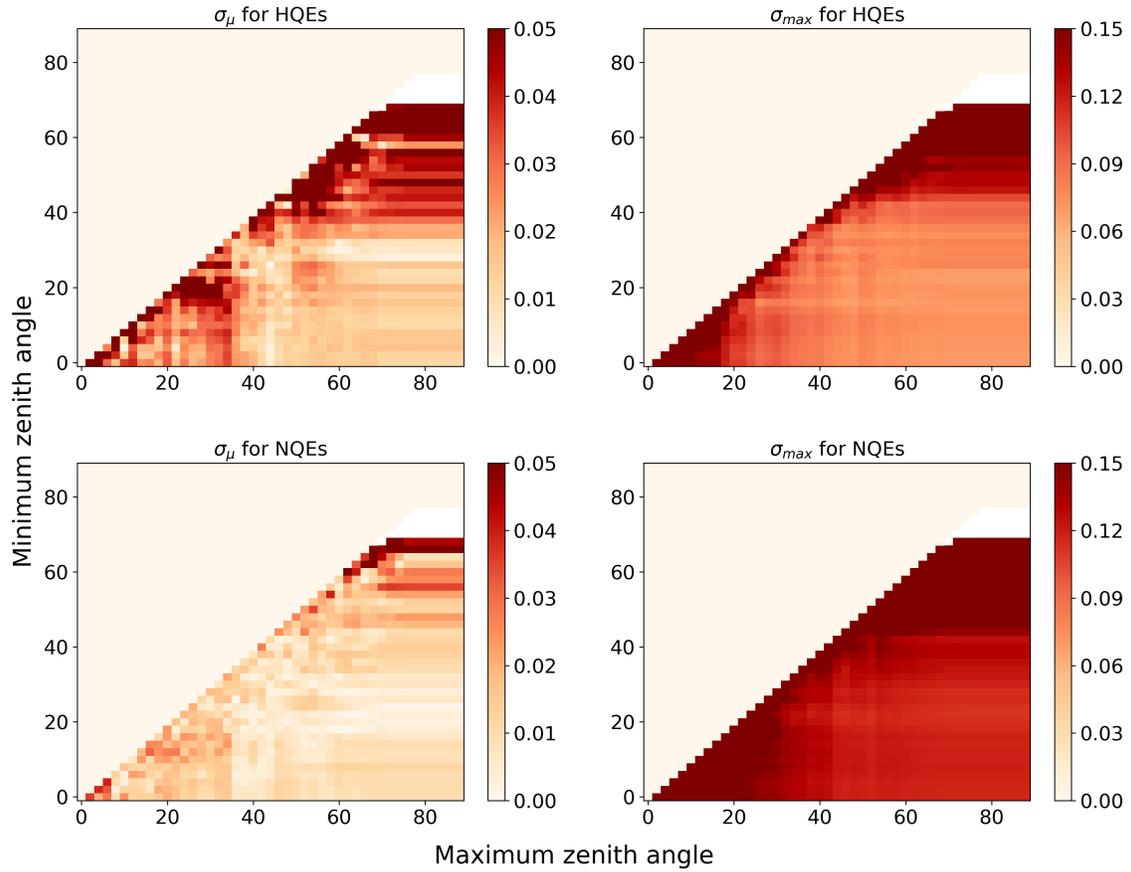
Figure 22: The 2016 dataset σ_μ (left) and σ_{max} (right) for HQEs (top) and NQEs (bottom) of all DOMs in each respective group as a function of the minimum and maximum zenith angle with the z-bin groups. Regions with zero events are shown with a metric of 0.

Event cuts
Only single multiplicity muons
No coincident events (2016 dataset)
Endpoint.xy > 100 m from detector boundary
Endpoint.z > -400 m (IceCube coordinates)
0° < Muon zenith angle < 90°

DOM cuts
No DOMs in outermost strings
0 m < DOM-distance < 130 m (2012 dataset)
0 m < DOM-distance < 125 m (2016 dataset)

Table 2: All present event and DOM cuts for the analysis. Some cuts only apply to one of the two datasets.

towards a possible physical reason for the optimised DOM-distance values.

A summary of the chosen event and DOM cuts can be seen in Table 2.

2.5 RIDE Results & Analysis

Section 2.4 gave us the optimal DOM-distance and zenith angle parameters to calculate the overall RIDE for the entire detector. With these set in place, we produce the RIDE results. We plot the RIDE for each DOM in the entire detector for a single combination of DOM-distances and zenith angles. This allows us to directly see the distribution of the RIDE across the detector instead of only in three selected groups. The goal is to return the input RIDE values of 1.00 and 1.35 for the NQEs and HQEs respectively. This should manifest as a normal distribution around these two centre values, with a strong separation between the central HQE RIDE and central NQE RIDE. Around the dust layer and near the bottom of the detector ($z < -400$ m) the spread around the central value should increase as the amount of photon hits on the DOMs near the muon tracks in these regions should be lower. This is because all muon tracks are downgoing, and any tracks that reach the bottom of the detector are more likely to stop outside it. At the top of DeepCore right below the dust layer, where the ice is clearest, we expect the strongest separation. Because the group analysis did not show any large effects on switching to the TIL groups, full detector plots are only shown with the z-bin groups.

The total amount of DOM hits in the analysis for the 2012 and 2016 datasets

were $1.93 \cdot 10^8$ and $8.01 \cdot 10^8$ respectively. In the 2012 dataset (Figure 23 (a)), we see the expected distribution of NQE DOMs around the central values of 1.0, with a very high spread at $z < -400$ m and at the dust layer near $z = -100$ m. We also see a small increase in spread near the top of IceCube. We do see a separation of NQEs and HQEs, but also that the DeepCore HQEs exhibit an upwards slope at decreasing depths. This is unexpected both based on prior theory and the individual group analysis. The three groups in the middle of DeepCore that constitute the group analysis do not vary much with the depth as we saw previously, but in the upper and lower parts of DeepCore the RIDE varies from 1.25 to 1.4. These results show a flaw in only including three groups in the analysis, as a group analysis on the entirety of DeepCore could have shown us that the bottom and top of DeepCore would not have agreed on a single DOM-distance or zenith angle cut. The large spread on the HQEs above the dust layer and the HQEs below -400 m makes it difficult to get useful information about the HQE distribution.

To better quantify the upwards trend in DeepCore, we do a linear fit,

$$RIDE(z) = a \cdot z + b, \tag{14}$$

to only the region of DeepCore with small spread at $z = [-350, -150]$ m, see Figure 23 (b). Here we see an HQE slope of $a = 0.0006 \pm 0.0001$, predicting an increase in RIDE of 0.06 after ascending 100 meters. In the 200 meter fitting region, that amounts to an increase of over 0.1. Since the HQEs have a slope a that is not consistent with 0, the constant b loses meaning as a central RIDE value. The NQEs however are consistent with a slope of 0 and we can calculate a simple constant fit,

$$RIDE(z) = c. \tag{15}$$

We get the fitting constant $c = 1.004 \pm 0.003$, which is consistent with 1.000 within 2σ. It should be noted that the uncertainties on the fits are calculated by assuming that the uncertainty of the RIDE value of each DOM is $q/\sqrt{N}$ where q is the mean charge of the DOM and N is the amount of events the DOM is part of, which is likely an underestimation of the true uncertainties. While the NQE DOMs show a distribution centred at 1.0, the HQE RIDE currently has no single centred flat value for the 2012 dataset despite agreeable results from the group analysis.

Moving on to the 2016 datasets, see Figure 24, the NQEs are once again distributed around a central RIDE of 1.0, with a constant fit constant $c = 0.999 \pm 0.003$ consistent with 1.000 within 1σ. The HQEs have a slope of $a = 0.0001 \pm 0.0001$ which is consistent with a slope of 0. However, due to the RIDE value predictably being centred around 1.25, the separation between HQEs and NQEs is much less pronounced compared to the 2012 dataset. We see a much smaller HQE and NQE spread at
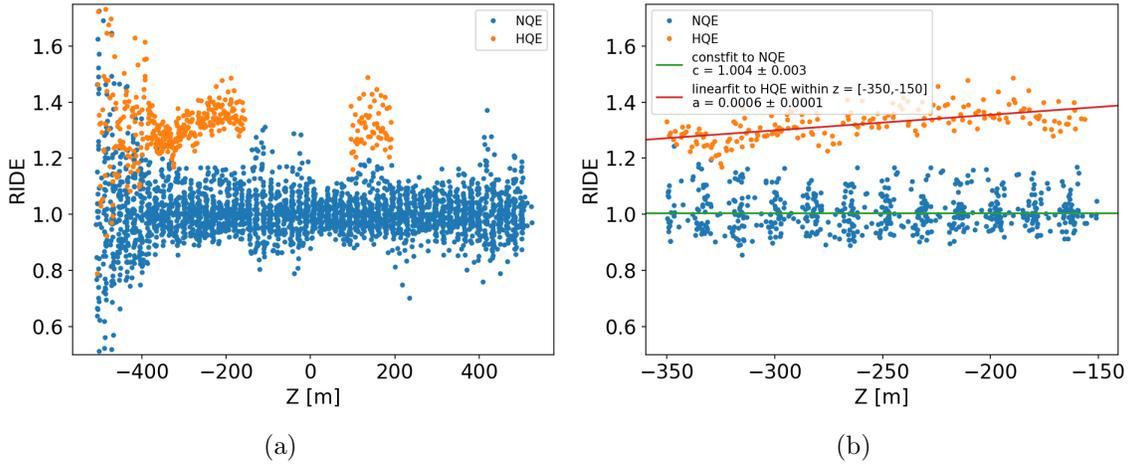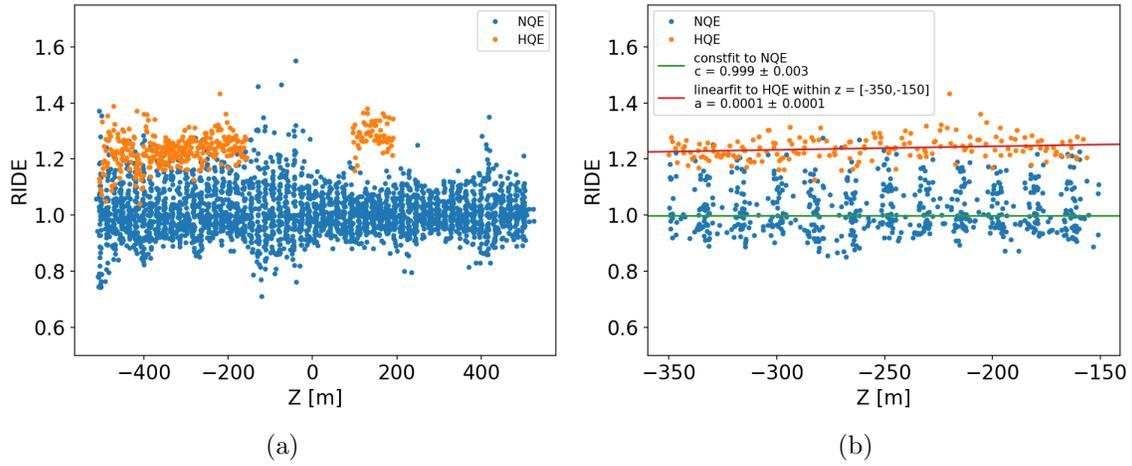
39

Figure 23: The 2012 dataset RIDE values for the NQEs and HQEs as a function of depth with no zenith angle cuts and DOM distance to track cuts of $[0, 130]$ m. (a) RIDE values across the entire detector. (b) a constant fit (Equation 15) to all NQEs and a linear fit (Equation 14) to the DeepCore HQEs. Displayed in the legend is the fit constant of the NQEs and the slope of the HQEs.

$z < -400$ m which indicates that the 2016 dataset contains many more stopped muons with tracks at the bottom of the detector.

In general, the 2016 dataset full detector plot agrees with what we expected from the group analysis and the 2012 group analysis having a large variety with change in DOM-distance is likely related to its depth-dependent HQE slope, as the 2016 dataset saw little variety with DOM-distance and no depth-dependent HQE slope. The exact nature of and reason for both the DOM-distance variance and the slope is currently unknown.

Due to the HQE slope in the 2012 dataset and a spread of the DOMs much larger than their individual uncertainties, the fit constants in Figures 23 and 24 are not very useful except confirmations of the presence of a slope or an NQE distribution around 1.0 as a function of depth. It is also difficult to use these plots to make direct comparisons between the central RIDE values of the 2012 and 2016 datasets difficult. This is better achieved by a histogram, see Figure 25, which better reflects the RIDE spread and shows that both datasets have an NQE RIDE consistent with 1.0 with a spread of 0.10 and 0.08 for the 2012 and 2016 datasets respectively. The 2012 HQE RIDE values (Figure 25 (a)) seem consistent with a mean of 1.35 from the histogram, while the 2016 HQE RIDE values (Figure 25 (a)) are inconsistent with a mean of 1.35 within 1σ. We also see a weaker separation between NQEs and HQEs.

40

Figure 24: The 2016 dataset RIDE values for the NQEs and HQEs as a function of depth with no zenith angle cuts and DOM distance to track cuts of $[0, 130]$ m. (a) RIDE values across the entire detector. (b) a constant fit (Equation 15) to all NQEs and a linear fit (Equation 14) to the DeepCore HQEs. Displayed in the legend is the fit constant of the NQEs and the slope of the HQEs.
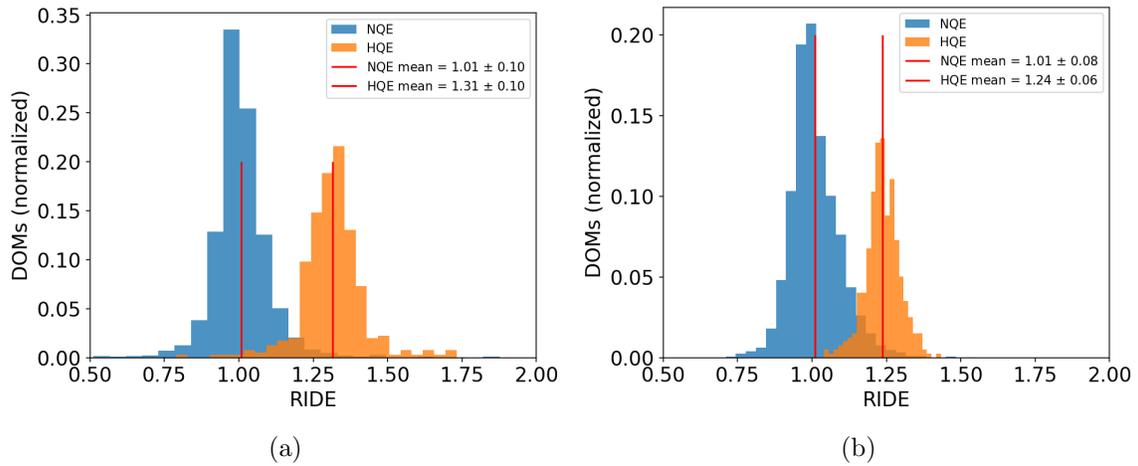
Figure 25: Histogram of the RIDE values. Red lines show the mean HQE and NQE value with accompanying standard deviation in the legend. (a) the 2012 dataset. (b) the 2016 dataset.

Returning the input DOM efficiencies of 1.35 HQE DOMs using the RIDE calculation was for in simulation was partially successful. The mean HQE RIDE value for the 2012 dataset, that uses simpler hadronic interaction, ice, and noise models, was found to be consistent with 1.35 but also contains a depth dependent slope of 0.06 per 100 meters. The 2016 dataset that contains more complex models and a wavelength dependent DOM efficiency has no depth dependent slope but has a mean HQE RIDE value that is inconsistent with 1.35. Both datasets had NQE RIDE values consistent with 1.0.

Because we are not able to return a flat consistent HQE RIDE value of 1.35 across either dataset, we do not expect reconstructions to return them either. Nevertheless, moving onto reconstruction could help provide better insight into the two datasets and current efforts should go towards improving the true simulation RIDE while also moving on to reconstructions. If reconstructions are consistent with the RIDE values returned in this thesis, then improvements to the RIDE will be of higher benefit. This will also allow us to move on to data more efficiently. When calculating the RIDE using real data, a large amount of stopped muons as calibration sources is necessary. Study into faster reconstruction methods, specifically the tagging of stopped muons, would be paramount to future work on the RIDE analysis.

3 Neural Network Reconstruction of Stopped Muon Parameters

One of the primary difficulties when moving from simulated truth to real data is reconstruction. Muon energies, positions, zenith angles, and other parameters need to be inferred directly from the DOM responses, and this can be both be inaccurate and computationally expensive. Traditional methods such as MPEFit (multiple photon electron)[49] use likelihood maximisation to find the best-fit track to explain the observed DOM responses. These methods are effective but also computationally expensive, as they need to maximise likelihood functions for every guessed parameter and every event.

Alternative methods include machine learning approaches that boast magnitudes faster runtime per event[50]. As the stopped muon criteria is a factor of primary importance for the RIDE analysis, we here present a neural network approach to classify stopped muons within the detector. A network such as this is also known as a classifier. Additionally, we also present a network to predict the endpoint positions of muons known to be stopped, known as a regressor.

3.1 Neural Networks

At its base, a neural network consists of a collection of nodes built into layers. There are three basic layer types. The first layer is known as the input layer and consists of the parameters fed into the network. In IceCube terms, this would be the DOM charge, timing, positions, etc. The last layer is the output layer containing the prediction. This would be whether a muon has stopped (classification) or the muon's endpoint (regression). In-between these two layers are the hidden layers containing the bulk of the neural network. In each layer the individual nodes take inputs from the nodes of the previous layer, processes those inputs, and gives an output to be fed into nodes of the next layer.

Each node calculates the weighted sum of its input,

$$y = \sum_{i=inputs} (w_i \cdot x_i) + b_i, \tag{16}$$

where y is the output, x_i is the input, w_i is the weight, and b_i is the bias. The weights and biases control how the inputs are calculated in the node. From the weighted sum, the neuron then calculates the output. This calculation is known as the activation function. The most simple of all is the linear activation function

$$A(y) = a \cdot y, \tag{17}$$

which simply return a straight line proportional to the input. Another simple activation function is the Heaviside step function

$$A(y) = \begin{cases} 0 & y < 0 \\ 1 & y > 0, \end{cases} \tag{18}$$

Which returns 0 if the weighted sum y is below 0 and 1 if the weighted sum y is above 0. This then gets fed into the next neuron. An illustration of the weighted sum and activation function can be seen in Figure 26. A network of a single neuron that takes an input and gives the final output is known as a perceptron. A neural network is the combination of multiple neurons in a sequence of hidden layers. Each neuron connects to multiple neurons of the previous and next layer. An illustration of a fully connected neural network can be seen in Figure 27. The output of the neural network is then compared to a series of labels with the same dimensionality as the output. In our examples, this would be a list of 0s and 1s for whether a muon is stopped or not or the x y, and z coordinates of the muon endpoint. The network then trains on these labels by changing the weights w and biases b in order to convert the input parameters into outputs that most closely match the labels. A network trains
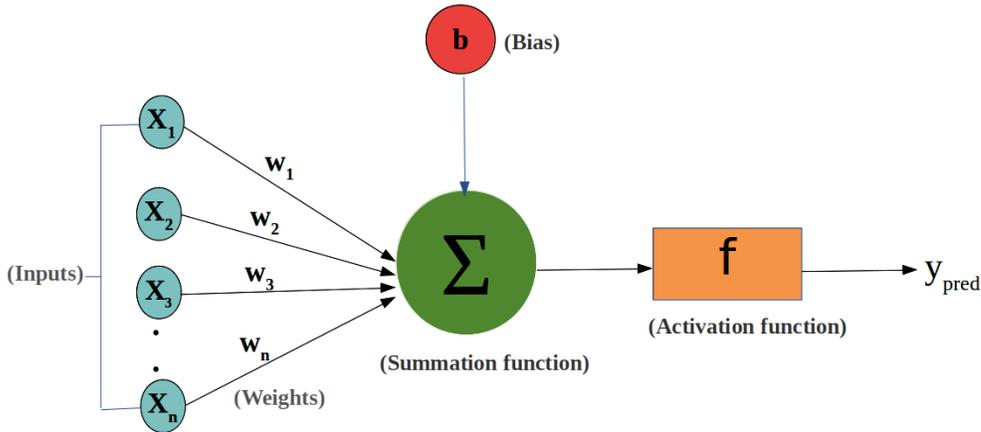
Figure 26: Illustration of a single perceptron network. The input parameters are calculated in a weighted sum. Then a bias is added and the result is fed into an activation function, producing the output prediction. Taken from Ref. [51].

using gradient descent on the activation functions. As such they both need to be differentiable and have meaningful derivatives. This is an issue if the hidden layers all have activation functions like Equations 17 and 18. The former has a constant derivative and the latter is not differentiable. A key point of neural networks is the introduction of nonlinearity. A linear neural network will just be a linear combination of the input layer and the benefits of a series of hidden layers is lost. Therefore, the hidden layer nodes should contain nonlinear differentiable activation functions. However, some nonlinear functions have derivatives that can get vanishingly small in very deep networks. This is known as the vanishing gradient problem. In most cases, every activation function has its advantages and disadvantages. There are a myriad of usable activation functions[53], but a common one is the rectified linear unit (ReLU) function

$$A(y) = max(0, y). \tag{19}$$

ReLU reduces the vanishing gradient problem and is nonlinear. It is non-differentiable at zero but the derivative can be chosen to be 0 or 1 and it is differentiable everywhere else. It is also very computationally inexpensive.

The network learns what a good prediction and a bad prediction is using a loss function. The loss function evaluates how well the network models the labels. The
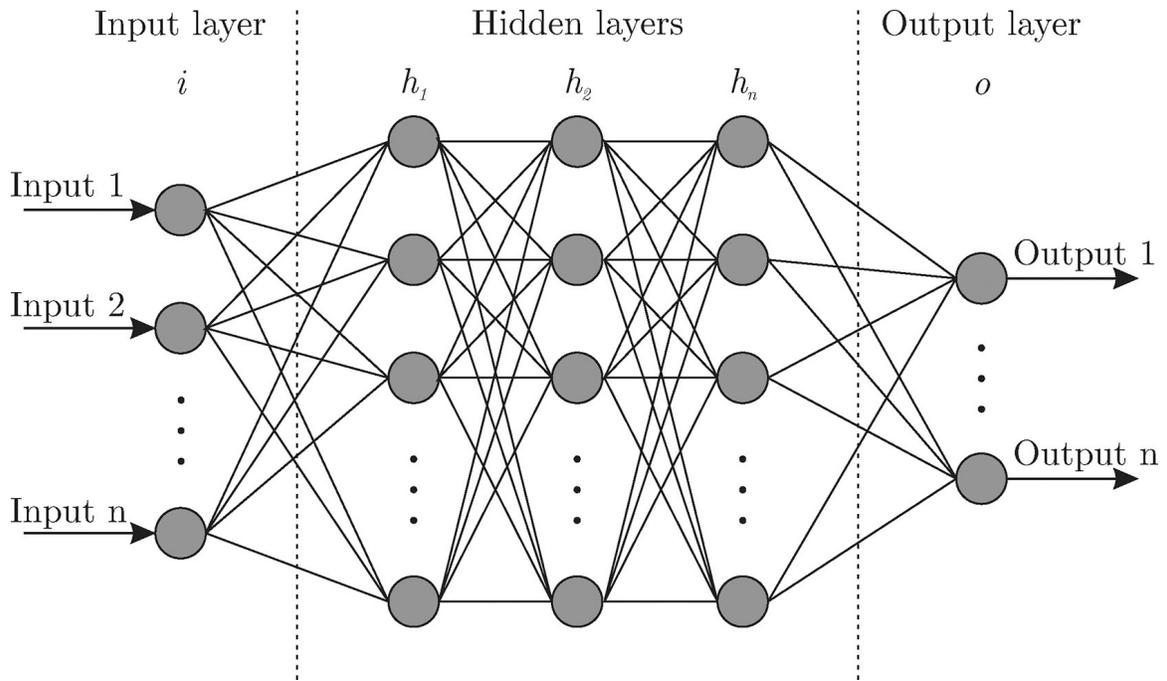
Figure 27: An illustration of a multi-layer perceptron network. Each node functions like Figure 26 and connects to nodes in the previous and next layers. The amount of nodes per layer can change to control the complexity of the network. Taken from Ref. [52].

chosen loss functions for our implementations can be seen in Sections 3.2.2 and 3.2.3.

There are many different types of neural networks. The one described generally above is typical of multi-layer perceptrons. Other networks include convolutions neural networks (CNN), inspired by the visual cortex. These networks respond to stimuli over a limited field known as the receptive field. The individual node operations are directly related to mathematical convolutions and this allows the network to effectively focus on individual parts of an input space like an image. An illustration of convolutional networks can be seen in Figure 28. The actual convolutions in CNNs are handled by the introduction of filters or kernels that are weighted dot products. The filter has a size partially based on the input dimensionality that scans over the input space to produce a map of dot products. The steps the scan takes is referred to as the stride. Both the size of the filter and the stride value affects the output shape. To retain an output shape similar to the input shape, zero padding can be done after applying the filters.

The neural networks described above base their decisions on current inputs. They assume that all inputs are independent and do not make use of sequential information to inform their decisions. This is an issue if the sequence of inputs is important such as when predicting the next word in a sentence. Recurrent neural networks (RNNs) are able to make predictions based on sequential inputs. They are called recurrent because their output is based on previous computations and they effectively have a memory on what is previously known about an input. This allows them to perform machine translations[55], generating image descriptions[56], and generating text[57], with good performance compared to traditional neural networks[58].

Convolutional networks that need to incorporate sequencing often use a combination of convolutional and recurrent neural networks to model sequential data. A recent neural network architecture known as a Temporal Convolutional Network (TCN) is a model that combines convolutional networks with the sequencing aspects usually seen in recurrent networks, and will be the neural network of focus in this thesis.

3.2   Temporal Convolutional Networks

TCN takes an input tensor of the form [batch_size,input_length,input_channels], referring to the size of the batch being processed by the network at a time, the size of each element in a batch (number of relevant DOMs per event), and the amount of input parameters (DOM position, charge deposited, etc.).

The TCN architecture will be briefly outlined in this thesis. More information can be found in Ref. [59]. Similar to CNNs, TCN convolves the input space using a filter with dimensions [filter_size,input_channels]. We convolve the input tensor by calculating the dot product using the filter to get an output tensor. If we want to maintain the same output tensor length as input tensor length we need to zero-pad
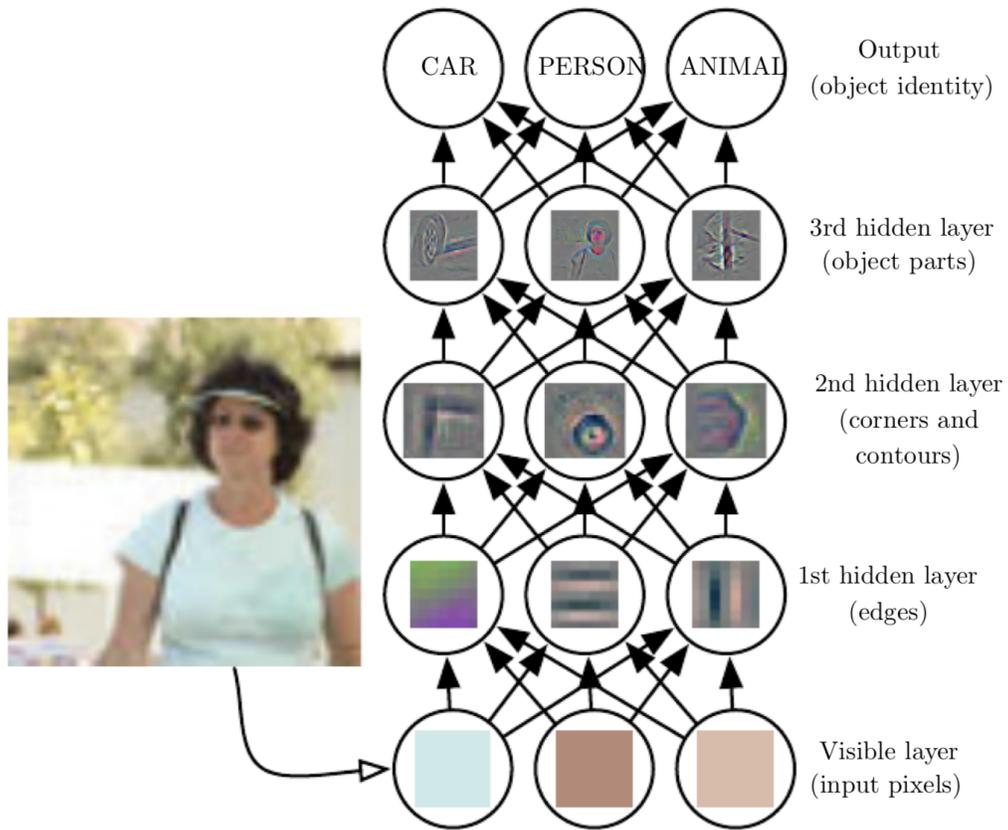
Figure 28: Conceptualisation of a convolutional neural network. The input image is split into pixels and the different hidden layers further split it into singular components. The final output layer uses these components to predict the object identity. Taken from Ref. [54].

the input tensor. In general, the number of zero-pads is equal to the filter size minus one.

A key aspect to TCNs is the temporal aspect or what is known as causal convolution. This means that an output element i in $\{0, ..., \text{input\_length} - 1\}$ can only depend on the input elements $\{0, ..., i\}$ i.e. the output element can only depend on input elements that came before it. This can be done by only applying zero-padding to the left, see Figure 29 (a). Ideally we want any output element to depend on all input elements in the sequence $\{0, ..., i\}$, which means a receptive field size of input_length. This is done by chaining multiple layers together, creating a receptive field of size

$$r = 1 + n(k - 1), \tag{20}$$

where r is the receptive field size, n is the number of layers, and k is the filter size. This solution, however, yields a huge number of layers, and thus nodes, weights, and biases, that need to be trained as the input_length increases. More specifically, if we want any output element to depend on all the input elements before it, the number of layers would be

$$n = \frac{l - 1}{k - 1}, \tag{21}$$

where l is the input_length. In order to keep a full receptive field while not letting the number of layers follow Equation 21, we introduce dilation. Dilation refers to the distance between input elements used to calculate an output element. In Figure 29 (a) the dilation is 1 because the input elements are adjacent. Figure 29 (b) shows a layer with a dilation of 2 where an output element is calculated from input elements separated by 1. This allows an increase in the receptive field without adding new layers. Combined with the chaining of layers, we can achieve a full receptive field while reducing the number of layers. However, if the number of dilations is kept constant then we still get a rapidly increasing number of layers as the input_length increases. This can be solved by allowing the number of dilations d to increase exponentially with each layer, $d = b^i$ where b is the dilation base and i is the number of layers below the current layer. The receptive field size is then

$$r = 1 + \sum_{i}^{n-1} b^i (k - 1), \tag{22}$$

where n is the total number of layers. In order to make sure no input elements are skipped, the filter size k should be at least as large as the dilation base b.

With the causal convolutions, the chaining of layers, and the introduction of exponential dilation, we can build the basic TCN network shown in Figure 30.
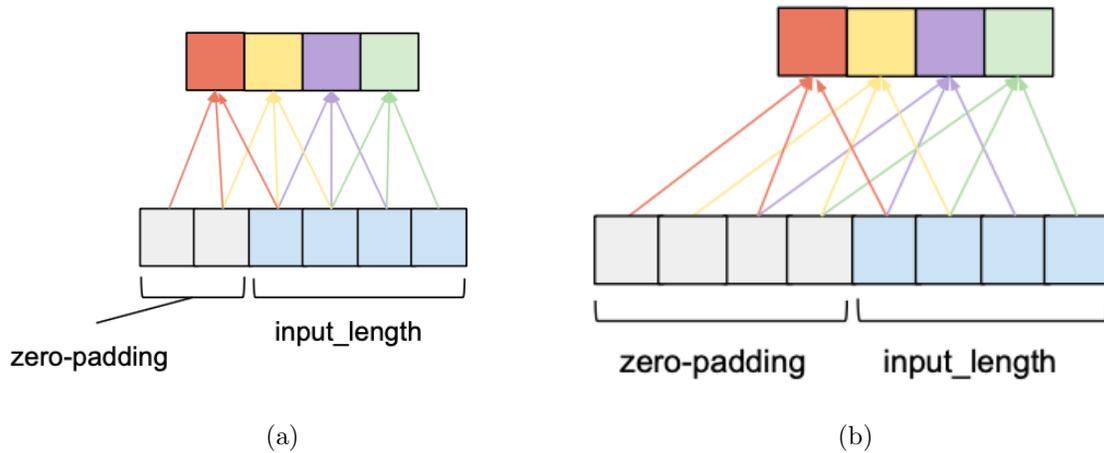
48

Figure 29: (a) Zero-padding to the left of a 1-dilated layer. (b) Zero-padding to the left of a 2-dilated layer. Taken from Ref. [60].
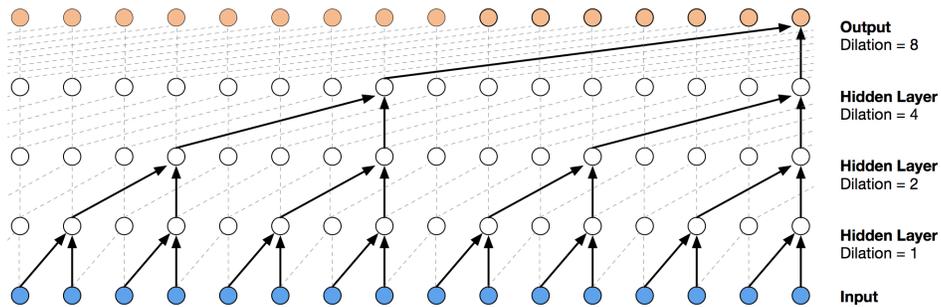

Figure 30: A basic TCN network with dilation factors of [1,2,4,8] and a filter size of 2. Sixteen input nodes are used to calculate one final output node. The arrows show only the node connections relevant to the right-most output node. Striped lines show all connections. Taken from Ref. [61].

Further improvements on the basic network can be made. The biggest one is the addition of residual blocks which replace the standard dilated convolution with two convolution layers that share the same dilation. Each of these layers is followed optionally by a weight normalisation that helps reduce exploding gradients. Then a ReLU activation function is applied to introduce nonlinearity to the network. And finally an optional dropout layer, which randomly ignores some nodes to prevent overfitting, is added. These blocks of two convolutional layers, normalisation, ReLU, and dropout replace the traditional single convolutional layer. In order to make the width of the output and input tensors match, an optional 1x1 convolution is added at the end of the block. The addition of the residual block has been shown to greatly improve performance in very deep networks[59].

Temporal Convolutional Networks are extremely powerful and versatile neural networks that boast a high number of advantages over traditional neural networks such as recurrent neural networks. These include the ability to run convolutions in parallel, avoiding the exploding/vanishing gradient problem, low training memory requirement, and variable input length. Coupled with better accuracy and performance across a wide range of problems[59], TCN makes for a very enticing network to use in IceCube reconstruction.

One of the attractions of current TCN implementations is the ease of use with Keras TCN[61]. This is a TCN implementation built using the Keras[62] framework, a deep learning python API running on the machine learning platform TensorFlow[63]. This implementation enables an out-of-the-box usage of TCN without intricate knowledge of neural network computing. The steps to run the TCN algorithm is defining the input values, adding one (or multiple) TCN layers, and defining the output layer.

3.2.1 Implementation

The first step to run the neural network is producing the input data. We run the cuts described in Section 2.2.1 to ensure that we only train on single-multiplicity atmospheric muon events. In addition, we also remove any muon bundles from the sample. We do not filter the events based on their true endpoint or angle of incident. We do not filter away DOMs in the outermost strings or based on their DOM-distance, but we do remove DOMs without any charge response for the event.

To include our events in a single trainable array, we need a consistent input length. This is done by picking a maximum amount of DOMs per event and zero-padding the input tensor if the amount of DOMs with charge is less than this maximum. A simple choice would be to include all 5160 DOMs. But the amount of DOMs that receive charge in any single-multiplicity non-bundled muon event is at absolute most a few hundred. Furthermore, we would require an extremely large amount of computation time and memory usage to include all DOMs with a sizeable number of events. In
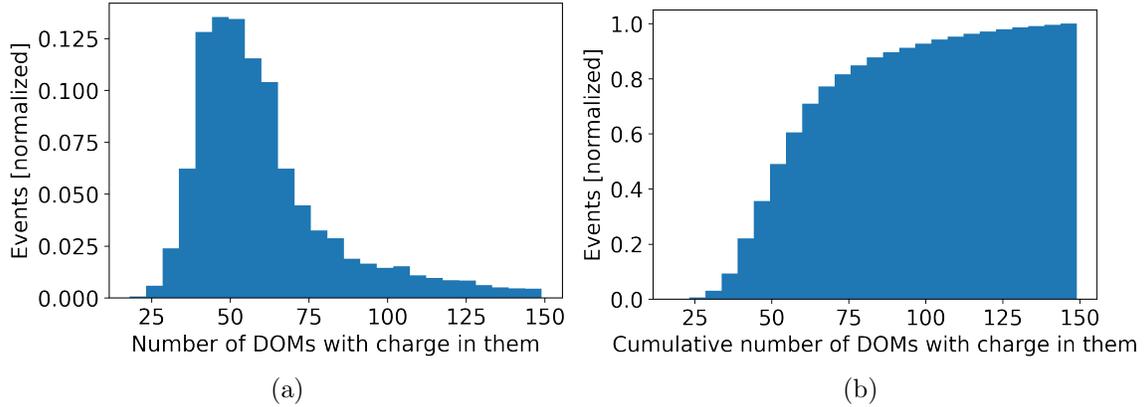
Figure 31: The distribution of charge-recording DOMs per event with a cutoff of 150 DOMs. (a) histogram over the number of DOMs with a charge response. (b) cumulative number of DOMs with charge in them.

effect, the smaller the number of maximum DOMs per event allows a larger number of events in the neural network. But if the maximum amount of DOMs is too small then there would be a significant amount of events that the network can not predict. The key is to find a balance between including the least amount of DOMs per event to improve memory and computation time, while making sure the vast majority of event types are represented by the network. Figure 31 presents a histogram of the number of charge-carrying DOMs per event. In our analysis, we put a threshold on the maximum number of DOMs per event at 150. This ensures that more than 99% of events are represented by our neural network while keeping the amount of DOMs per event manageable.

TCN can be chained into itself, allowing us to introduce multiple TCN layers which makes the network more complex and sophisticated at the cost of computation time. We found no difference in accuracy when using more than two TCN layers. In our implementation we used the default recommended settings in the Keras TCN package. This includes 64 filters with a filter size of 3, 1 stack of residual blocks, no normalisation or dropout layers, a dilation factor of [1,2,4,8,16,32], and a ReLU activation function. Changes from the default settings were not found to improve the network. The optimisation algorithm used is the Adam optimiser[64] with default options. We used a validation split of 0.1. In our input datasets, there is roughly 75% more non-stopped muons than stopped muons. Our implementation includes weights to ensure that the network is not biased towards non-stopped muons.

51

input_40: InputLayer	input:	[(None, 150, 5)]
	output:	[(None, 150, 5)]

tcn_57: TCN	input:	(None, 150, 5)
	output:	(None, 150, 64)

tcn_58: TCN	input:	(None, 150, 64)
	output:	(None, 64)

dense_34: Dense	input:	(None, 64)
	output:	(None, 1)
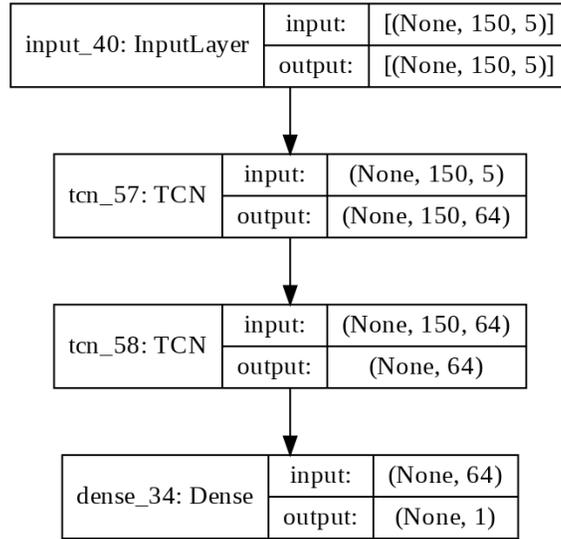
Figure 32: An illustration of the neural network in our implementation. We input a tensor with at most 150 DOMs with 5 parameters and an arbitrary batch size. This is fed into two TCN layers and finally a dense layer with a sigmoid activation function that outputs a single value between 0 and 1. The individual TCN layers inside each residual block are visualised in Figure 30.

3.2.2 TCN Classification of Stopped Muons

The initial implementation of TCN will be the classification of stopped muons. The final output layer is a dense layer which is a deeply connected layer that receives information from all nodes in the previous layer. The output from this layer needs to be a single value between 0 and 1 for each event. This is insured by using the sigmoid function defined as

$$A(y) = \frac{1}{1 + e^y}. \tag{23}$$

Any y mapped onto the sigmoid function will return a value between 0 and 1, where 1 defines a muon that has stopped within the detector, while 0 defines one that has not. An illustration of the neural network architecture can be seen in Figure 32. The input data will include a list of values at strictly 0 and 1 based on Section 2.2.2. The network will then return values representing its confidence of whether or not the muon is stopped. When picking whether a muon is stopped or not, we define the threshold as the network's confidence level above which a muon will be considered stopped. As we are running a classification network, the loss-function we are optimising is the binary cross-entropy (BCE) function defined as

$$L_{BCE} = -(q \ln(p) + (1 - q) \ln(1 - p)), \tag{24}$$

where q is the label (0 or 1) and p is the classifier prediction (between 0 and 1). We take the logarithm as $-\ln(p)$ approaches infinity as p approaches 0 and correspondingly for $-\ln(1 - p)$ as p approaches 1. The input tensor includes roughly a million events and five input parameters: The DOM positions (x, y, z), their charge, and their timing. We previously experimented with more parameter inputs including the pulse-width, DOM indexes, and TIL groups. These inputs were not found to provide better performance and were scrapped.

In the TCN implementation, we still work with our two datasets. Seeing the difference between the predictability of the stopped muon criteria for the 2012 and 2016 dataset could provide insight into why the group analysis predicts different optimisations. Figure 33 shows histograms of the five parameters for events labelled stopped and non-stopped. We generally see, as expected, that DOMs in the centre of the detector are triggered more often by stopped muons and that DOMs at the bottom of the detector, closer to the -400 m depth restriction, are triggered less by stopped muons. Since stopped muons are expected to last a shorter period of time within the detector, we also see more stopped muons at shorter timescales. As expected, the amount of charge deposited in the DOM is nearly independent of the stopped muon criteria in both datasets, although there is a small bias towards higher charge for

stopped muons. The reason for this is likely that non-stopped muons have a higher energy and are more likely to produce more light in the detector. Although the plots show the inputs in units of meters, photoelectrons, and nanoseconds, the input data fed into TCN is first scaled and normalised to have a mean of 0 and a width of 1.

3.2.3 TCN Regression on Muon Endpoint

Current reconstruction methods assume through-going tracks and do not tell us where the muon stopped. Together with the stopped muon classification, we started preliminary development of predicting the endpoint position of the muon. If such predictions are successful, then the muon endpoint could be found directly. Additionally, regression performance would inform us about what coordinates have the potentially largest effect on the classifier's ability to predict if a muon is stopped. To maximise the effectiveness of the endpoint prediction, we only train and test on stopped muons, with the future hope that one neural network can find stopped muons while another pinpoints the endpoint position of those stopped muons.

The same configuration and input tensor is used as with the stopped muon classification network, but the final output layer uses a linear activation function

$$A(y) = cy, \tag{25}$$

and contains 3 nodes. The loss function used is the mean absolute error function

$$L_{MAE} = \frac{1}{3} \left(|x_{true} - x_{pred}| + |y_{true} - y_{pred}| + |z_{true} - z_{pred}| \right), \tag{26}$$

where x_{true}, y_{true}, and z_{true} are the true coordinates, and x_{pred}, y_{pred}, and z_{pred} are the predicted coordinates. The overall loss is then the mean loss across all samples. Experiments were done with the mean squared error function

$$L_{MSE} = \frac{1}{3} \left((x_{true} - x_{pred})^2 + (y_{true} - y_{pred})^2 + (z_{true} - z_{pred})^2 \right), \tag{27}$$

but no performance improvement was found.

3.3 Neural Network Performance

Both the classification and regression networks ran for 20 epochs or until the loss did not decrease for 3 epochs. After 3 epochs, the loss stayed relatively constant and gained minimal improvement with further training, see Figure 34. After the second epoch, the validation loss began fluctuating and saw no improvement either. This structure is similar for all networks.
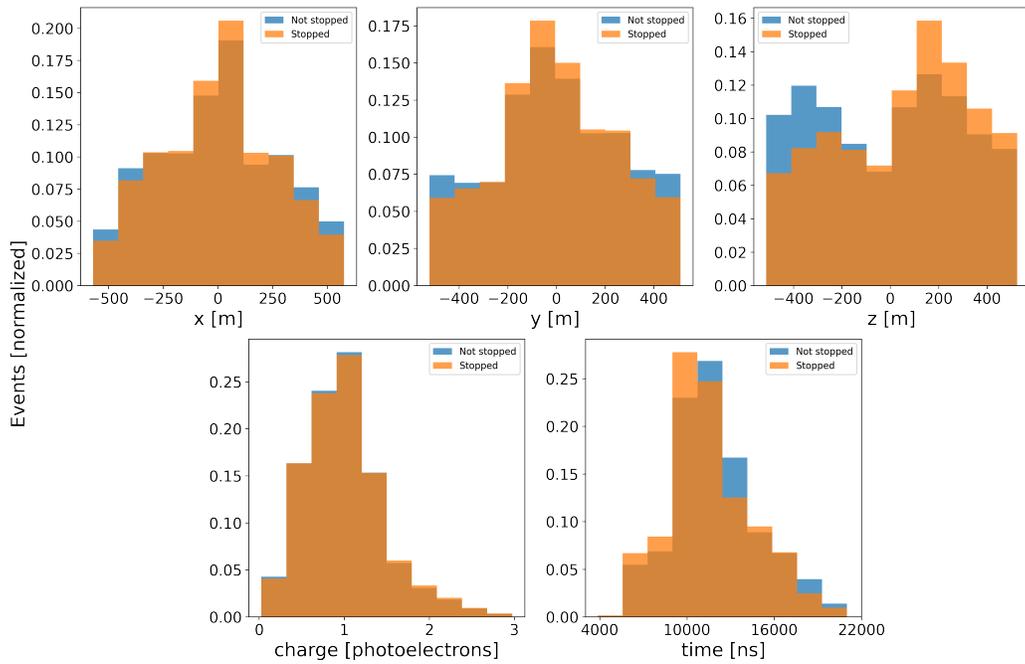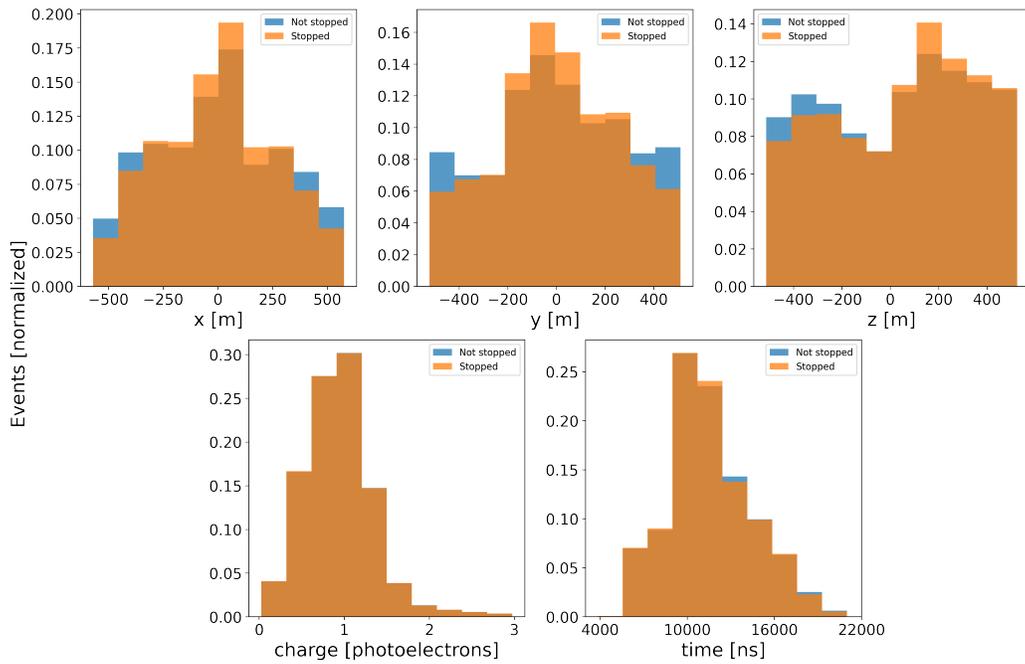
Figure 33: The five DOM input parameters of the networks. Shown are the distribution of the parameters for true stopped and non-stopped muons separately. (a) the 2012 dataset inputs. (b) the 2016 dataset inputs.
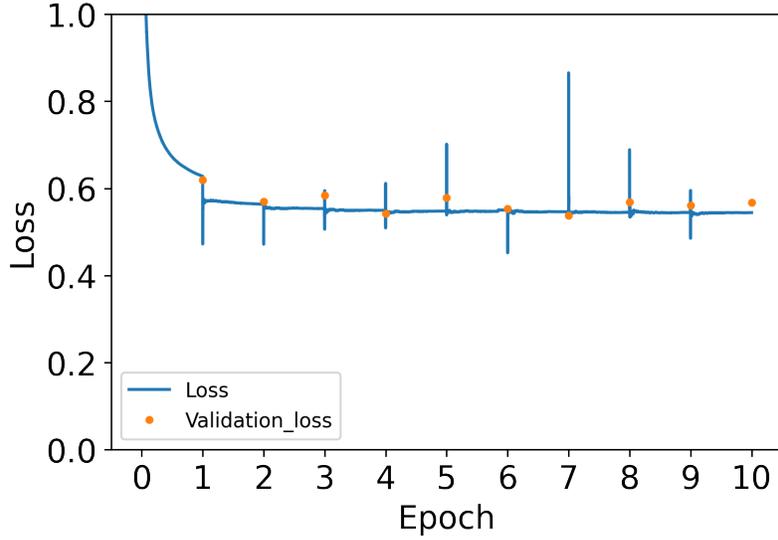
Figure 34: The training and validation loss of the 2012-TCN classifier network. The loss decreases rapidly in the first few epochs, but stagnates above ∼3 epochs with an unstable validation loss. At the start of each epoch the training loss is unstable as the network is starting over on the shuffled dataset. All other networks behave similarly.

3.3.1 Classification

The initial classification results are presented in terms of the TCN predictions compared to the labels. The prediction confidence for the network goes from 0 to 1. 0 and 1 means high confidence of non-stopped and stopped muons respectively and 0.5 means very little confidence of either. All true labels are placed in the 0- and 1-bins and also represent what a perfect network would predict. A good network would have very few events with low confidence (around 0.5) and predict a high amount of events accurately. We start by looking at the 2012 dataset, see Figure 35 (a). Every stopped muon has a confidence of at least 0.7 with the vast majority landing in the 1-bin. The majority of non-stopped muons have confidences below 0.2, but a few are between 0.4 and 0.6, with even fewer landing in the bins signifying high confidence of a stopped muon. Overall, there is a very strong separation between the event labels and very few events with low or wrong confidence.

Initially, training was done on a 2016 dataset that did not have coincidence events removed (see Section 2.2.1). The prediction confidence of this dataset is seen in Figure 35 (b). No event is above 0.8 or below 0.2 in confidence, indicating that the network has very little confidence on the both stopped and non-stopped muons. Even within the low-confidence region between 0.4 and 0.6, the labels are not separated at all.
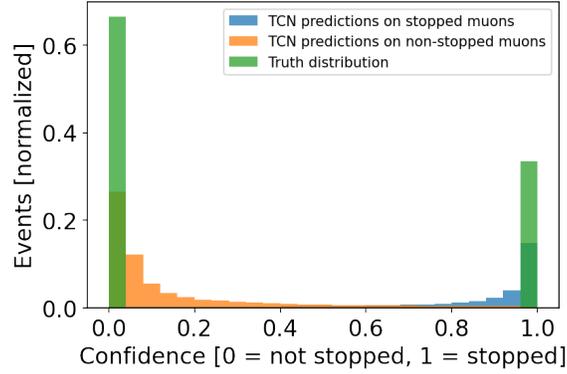
(a)
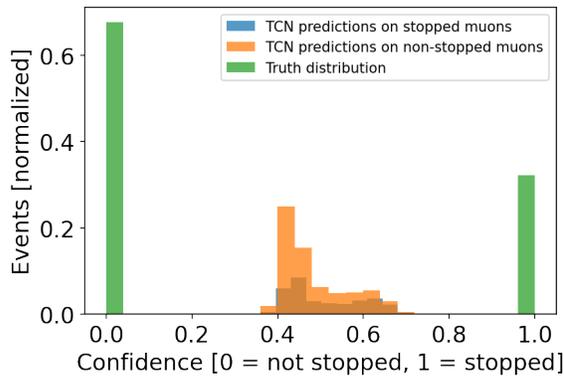
(b)
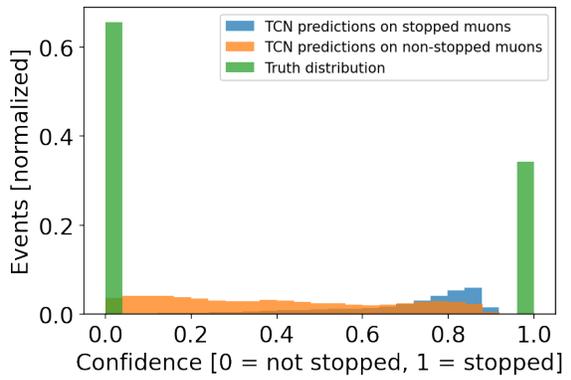
(c)

Figure 35: The neural network confidence on stopped and non-stopped muons respectively alongside the true labels. (a) the 2012-TCN predictions. (b) the 2016-TCN predictions with coincident events. (c) the 2016-TCN predictions without coincident events.

With the coincident events removed, the network performs much better, see Figure 35 (c). The majority of stopped muons have a confidence over 0.7, although a few are still below 0.6 and none are placed in the 1-bin. The non-stopped muons populate almost the entire confidence space. This indicates that the network performs well at correctly identifying stopped muons but does not perform well at identifying non-stopped muons. Even placing the threshold at a high confidence results in a large amount of false positives. Nevertheless, the attempt at stopped muon classification shows promise at accurately classifying stopped muons with further testing and analysis. However, it is already apparent that the 2016 dataset is not able to accurately predict stopped muons with the wanted confidence. We continue to analyse this dataset to quantify its performance compared to the 2012 dataset.

To better quantify the confidence plots in Figure 35, we can use different metrics to gauge the predictive ability of the networks as well as find the optimal threshold for when to mark an event as a stopped muon. To consider this, we first have to define the metrics. We define four values: recall/sensitivity or true positive rate (TPR), the number of correctly identified stopped muons out of total stopped muons; specificity or true negative rate (TNR), the number of correctly identified non-stopped muons out of total non-stopped muons; type 1 error or false positive rate (FPR), the number of wrongly identified non-stopped muons out of total non-stopped muons; and type 2 error or false negative rate (FNR), the number of wrongly identified stopped muons out of total stopped muons. In addition, we define precision as the number of correctly identified stopped muons out of all the muons identified as stopped.

Based on these metrics, we can define the receiver operating characteristic (ROC) curve[65] as the true positive rate versus the false positive rate, see Figure 36. This is a general tool to evaluate the performance of a neural network and tells us how many true positives we would get given a rate of false positives. The ideal neural network would have a true positive rate of 1 and a false positive rate of 0. How much the network curves towards this ideal gives a good representation of its performance. In Figure 36, a comparison of the ROC curves for the 2012- and the two 2016-TCN networks is shown. As expected, the 2012-TCN has a much better performance, with a higher true positive rate at a given false positive rate, and that the 2016-TCN improves measurably with coincident events removed. In particular, the area under the ROC curve (AUC) represents the probability that a randomly chosen stopped muon is correctly marked with higher confidence than a randomly chosen non-stopped muon[66]. Note that this is not the same as the probability of correctly identifying a stopped muon. An AUC of 0.5 means that the classifier is just as likely to incorrectly or correctly put a stopped muon at a higher confidence than a non-stopped muon. In our analysis, the base 2016-TCN, the 2016-TCN without coincident events, and the 2012-TCN have an AUC of 59%, 80%, and 95% respectively. Further analysis will only be done on the 2012-TCN and the 2016-TCN without coincident events.
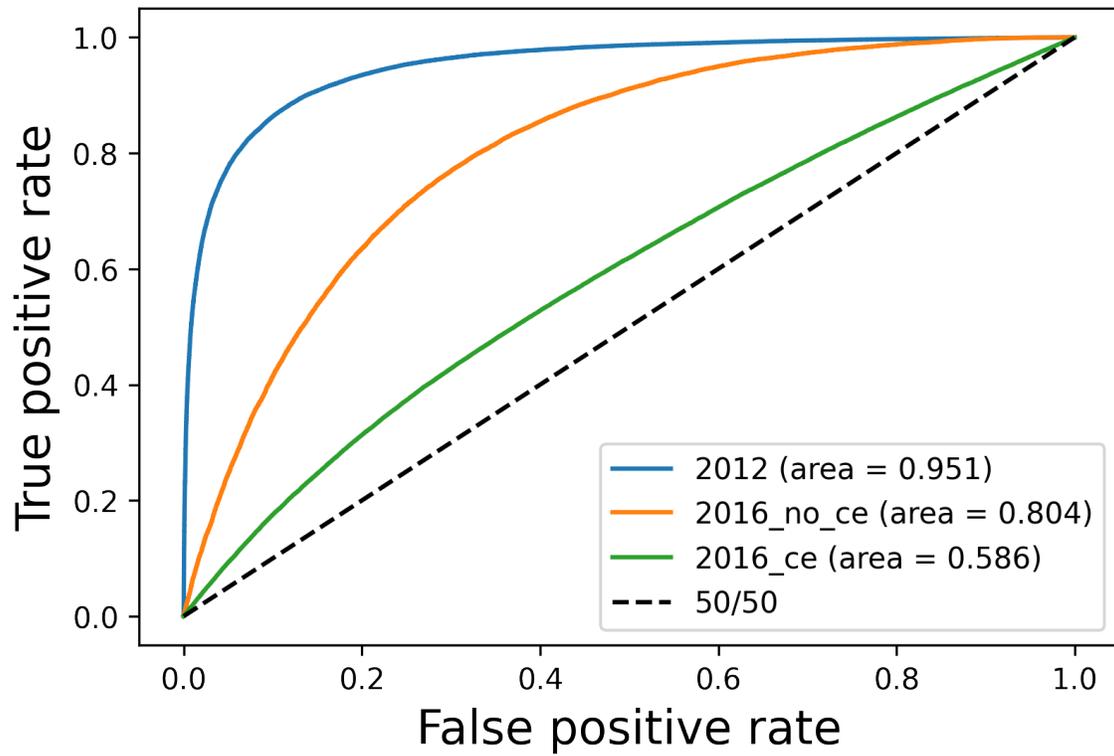
Figure 36: ROC curve of the 2012-TCN, the 2016-TCN without coincident events (no_{ce}), and the 2016-TCN with coincident events (ce), alongside their respective AUC.

While the ROC curve is a useful metric to evaluate the general performance of different networks, it does not tell us the optimal threshold at which the networks should predict a muon to be stopped. This depends on what the network should be optimised for. Accuracy (ACC) is a conventional metric designed to maximise the percentage of correctly identified muons, defined as

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{28}$$

where TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives respectively. However, this metric is not useful when correctly identifying true negatives is more important than correctly identifying true positives. At IceCube, we see about 100 billion atmospheric muon events per year[67]. With such a large amount of data, it is more important that we correctly discard non-stopped muons than avoid incorrectly discarding stopped muons. That is, even with a huge amount of false negatives, IceCube sees more than enough atmospheric muons to get relevant statistics for the RIDE analysis. The optimal threshold will then aim to minimise the false positive rate and/or maximise the precision. Simply picking the threshold that reduces the FPR to 0 would classify all muons as not non-stopped. So it is still important to have a threshold that maintains a decent number of stopped muons. A useful metric to control the desired precision while retaining a significant number of stopped muons is the F-score[68], defined as

$$F_\beta = (1 + \beta^2) \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}, \tag{29}$$

where the β factor is picked such that recall is considered β times as important as precision. A β of 1 returns the balanced F-score or the F_1 score which is the harmonic mean of precision and recall. We place high priority on the precision and our picked β value is 0.1, indicating that precision is 10 times more important than recall. See Appendix B for different possible values of β. Figure 37 puts the optimal threshold values at 0.97 and 0.84 for the 2012-TCN and 2016-TCN respectively. This provides the confusion matrices shown in Table 3. With this threshold, we have a 2012-TCN precision of 97.9% while retaining a sensitivity of 39.1%. Meanwhile, the 2016-TCN has a precision of 72.6% with a sensitivity of 20.7%. This is a direct quantifiable comparison between the two models and shows that the 2016-TCN does not perform at acceptable levels for the current analysis. One method of analysing why the 2016-TCN classifier performs as poorly as it does compared to the 2012-TCN classifier is comparing the performance of the two datasets on regression networks. Knowing exactly how well each regressor predicts the three coordinates on the two datasets could yield valuable insight into why the stopped muons are classified with such
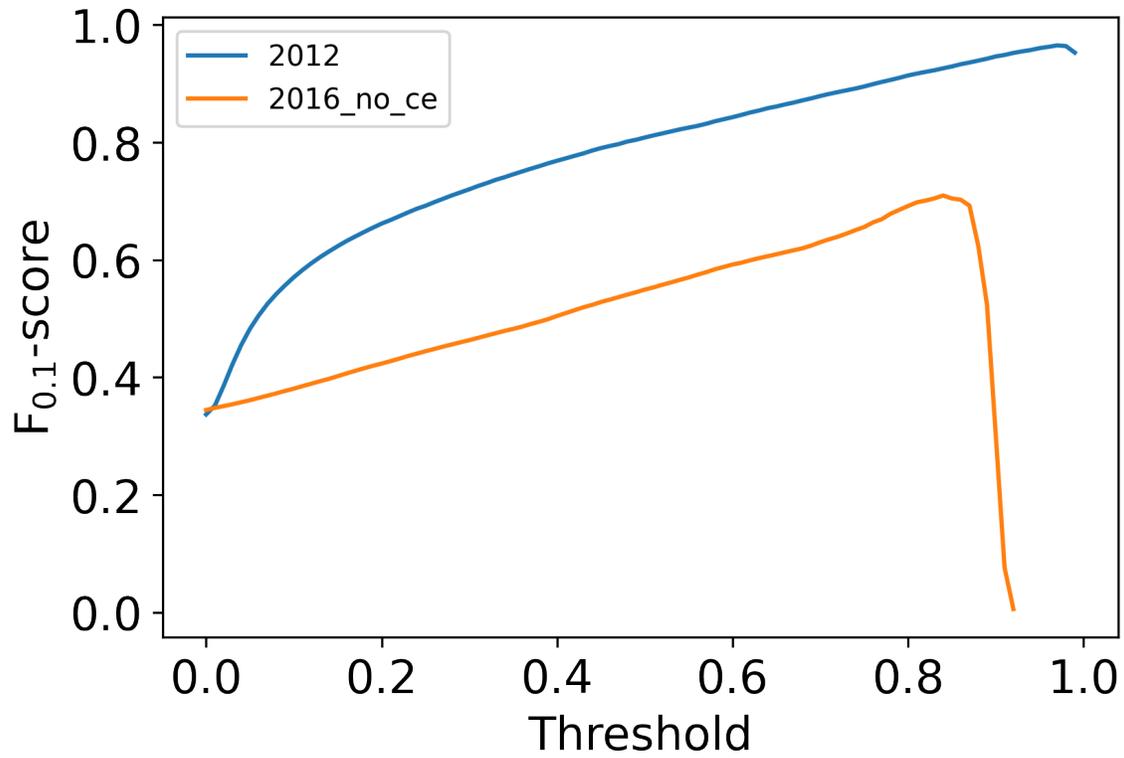
Figure 37: The $F_{0.1}$-score (see Equation 29) as a function of the threshold. The threshold with the maximum $F_{0.1}$-score is 0.97 and 0.84 for the 2012-TCN and 2016-TCN respectively.

| | 2012 Predicted | |
	Not stopped	Stopped
Actual Not stopped	104665	439
Actual Stopped	32265	20690

| | 2016 Predicted | |
	Not stopped	Stopped
Actual Not stopped	60735	2575
Actual Stopped	26225	6830

Table 3: Confusion matrix of the 2012 (top) and 2016 dataset (bottom). The size of the two datasets differ and the total amount of muons in the matrix are different.

different performance.

3.3.2 Regression

The regression network results are represented in terms of the true coordinates versus the predicted coordinates. We plot the prediction coordinates as a function of the true coordinates with the goal of every value sitting as close to the 1:1 line as possible. To better quantify the distribution of events, we plot it as a 2D normalised histogram. Here we should also see a distribution with more events near the centre of the detector in x and y coordinates and a sharp decrease where the outer string cuts are applied. The z coordinate distribution should decrease sharply at $z < -400$ m, and softly as we approach the top of the detector.

Figure 38 compares the 2012-TCN (a) and the 2016-TCN (b). The x coordinate shows similar performance across the two networks, with slightly more scatter for the 2016-TCN. The y coordinate is also more scattered for the 2016-TCN and the population density around the 1:1 line is smaller. The z coordinate in the 2012-TCN shows pockets of increased density around $z = 0$ m, 200 m $< z <$ 300 m, and a very high-density pocket at $z = -400$ m. The regions around these pockets have a relative density of about one third to two thirds. This distribution is well predicted by the 2012-TCN excepting a skew at the high-density pocket near the bottom of the detector. It follows from expectations (Figure 33) that there is a higher distribution of DOM responses at around $z = 200$ m and $z = -400$ m. It is likely that muon endpoints close to regions of more DOMs are better modelled and therefore have

higher a density at the 1:1 line, but this does not explain the pocket at $z = 0$ m where the dust layer is. The 2016-TCN is generally not able to predict the muon z coordinate at all. The true distribution of muons is closer to the centre of the detector (x-axis in Figure 38 bottom), but the predicted values are highly scattered. The pockets seen in the 2012 distribution are also not apparent. The inability to properly predict the 2016 dataset z coordinate could be why the 2016 classifier performed poorly and the low precision as opposed to just a low accuracy might be why the 2016 classifier non-stopped muon confidence is low. As with the loss functions, there are multiple ways of gauging the network performance for regression. We mentioned the mean absolute error (MAE) and mean squared error (MSE). There is also the square root of the mean squared error (RMSE), and the coefficient of determination or the R^2 score

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(q_i - p_i)^2}{\sum_{i=1}^{N}(q_i - \overline{q})^2}, \tag{30}$$

where q_i is the true value, p_i is the predicted value and $\overline{q}$ is the mean true value. The R^2 score is the relative variance explained by the model. If the residual is 0 then R^2 is 1. If the model is as good as simply predicting $\overline{y}$ for every value then R^2 is 0. If the model is worse than simply predicting the mean truth then R^2 is negative.

Both MAE and MSE/RMSE provide absolute metrics to the model performance. We report the MAE here since it treats all errors proportionally and allows evaluation on the direct mean error. The R^2 is a relative metric and tells us at a glance how each model performs regardless of the spread of the input data. The MAE and R^2 are reported in Table 4.

A couple of things can be learnt from the table. The MAE reduction in the y coordinate is due to the true y coorindates being less scattered and not necessarily because the model predicts it better than the x coordinates, as shown by a similar R^2. Both networks have trouble with predicting the z coordinate, with the 2016-TCN having significant more trouble at more than twice the MAE of the 2012-TCN and half the R^2. The prediction of the x and y coordinates in the 2016-TCN is also slightly worse with the MAE metric but performs similarly on the R^2. This suggests a slightly larger scattering of true values for the 2016-TCN x and y coordinates. The magnified z coordinate errors leads to the 2012-TCN having an MAE of around 30 m less and an R^2 around 0.1 more than the 2016-TCN.

Overall, the 2012-TCN regressor performs similar to the 2016-TCN regressor on x and y coordinate predictions, but more than twice as well on the z coordinate prediction, as expected from the 2D histogram plots. Both networks have a mean absolute prediction over 50 m for all three coordinates and, at current levels, the regressor network is not able to predict the muon endpoint positions to a satisfying
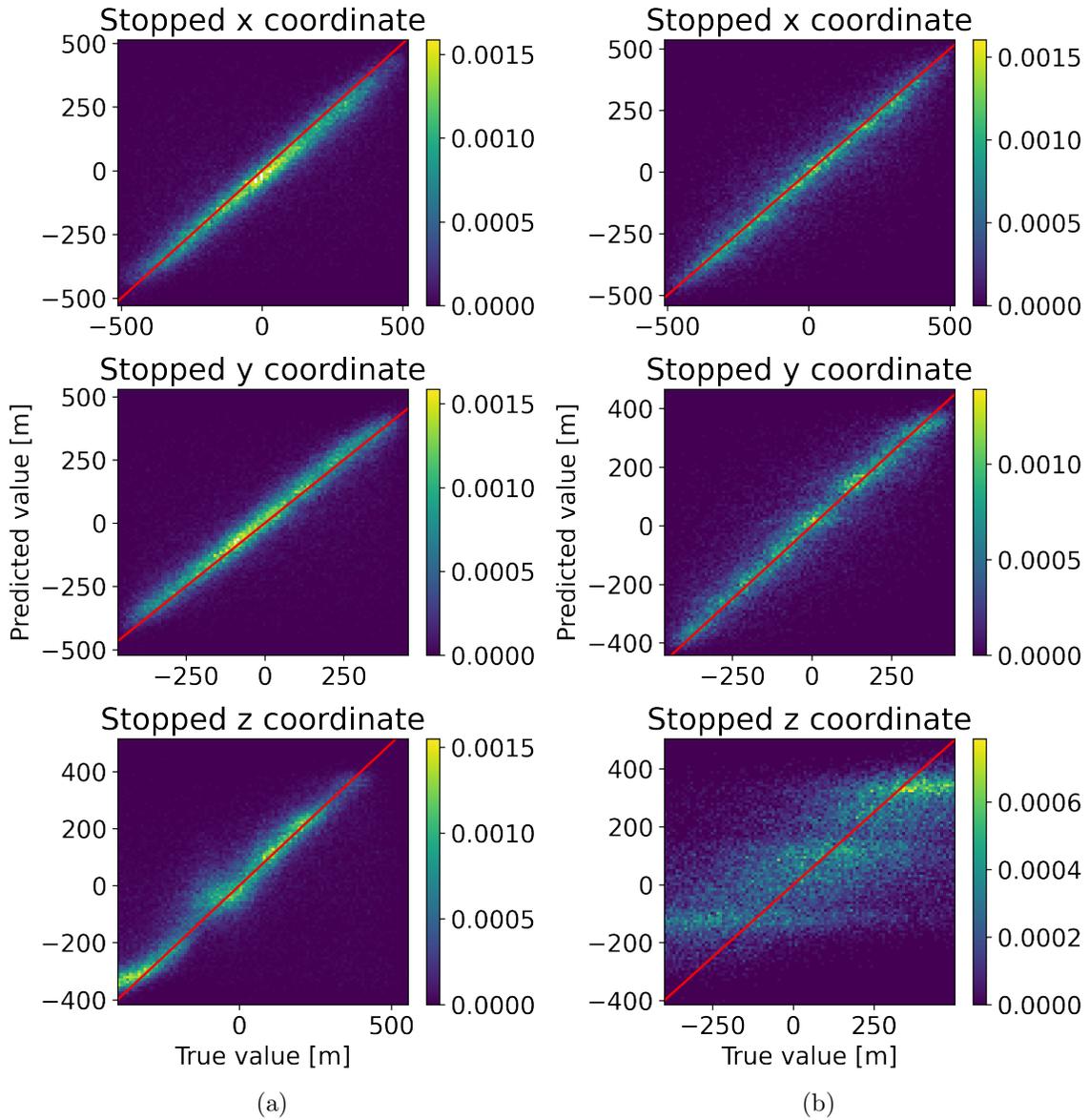
Figure 38: Predicted stopped muon position as a function of true muon position across the x, y, and z coordinates. Also shown is a 1:1 line representing perfect agreement between true and predicted value. (a) the 2012-TCN regressor. (b) the 2016-TCN regressor.

	Metric	x	y	z	Mean
2012	MAE [m]	62.1	56.6	66.4	61.7
	R^2	0.82	0.82	0.74	0.79
2016	MAE [m]	67.3	66.8	150.5	94.9
	R^2	0.83	0.81	0.36	0.67

Table 4: The mean absolute error (MAE) and coefficient of determination (R^2) for the 2012- and 2016-TCN across the three individual coordinates and the mean of those. The best R^2 is 1 and the best MAE is as low as possible.

accuracy. The networks used in the regression and classification problems were of similar size and complexity, and a prediction of the exact endpoint as opposed to just a classification on whether the endpoint is within the detector or not differ greatly in difficulty. However the usage of similar neural networks provides information into the distribution of the endpoints in the two datasets, since the difference in performance must come from the input data. It is apparent that the z coordinate endpoint distribution of the 2016 dataset is the primary cause of error in both the regressor and classifier. The relatively flatter distribution of the DOM z position in Figure 33 (b) top right could point towards a larger amount of noise or light from muons hitting DOMs further away. However, this did not impact the x and y positions of the DOMs nor the endpoint predictions. It is currently not definitely known why the 2016 dataset z coordinate behaves like it does.

Due to the regressors not predicting the muon endpoint to a satisfying degree, current focus should instead be on optimising the classifier network to select stopped muons and using traditional reconstruction methods to find the endpoint. This approach will drastically reduce the amount of reconstruction necessary for statistically relevant analyses.

4 Conclusion

In this thesis, we have analysed two different simulation datasets in order to gauge the relative individual DOM efficiency (RIDE) in the IceCube detector. Input values for each DOM are set at 1.00 for NQEs and 1.35 for HQEs. The goal of this thesis was to reproduce these numbers from simulated DOM responses. Previous studies used uniformly produced muons and a uniform ice model to calculate the RIDE. This thesis has moved onto two dCORSIKA models that simulate primary and secondary particle interactions in the atmosphere and their propagation in the ice. These models also contain tilted anisotropic ice layers to more accurately model observed ice con-

ditions. Muon depth penetration and ice properties necessitated a DOM efficiency that depends on its depth. We have shown through analysis of DOMs binned into three depth-based groups in DeepCore, that cuts on the muon zenith angle do not meaningfully alter the RIDE value. We have also shown that cuts based on the distance of DOMs from the muon track of each event should be within 0 and 125 meters in the 2012 dataset through the optimisation of an HQE RIDE value of 1.35. The 2016 dataset was optimised for inter-group and inter-DOM agreement, yielding similar conclusions about the zenith angle and an optimised DOM-distance of 0 to 130 meters. We have shown that switching between a group model based on a uniform 10 meter depth spacing and one based directly on tilted ice layer properties does not substantially affect the RIDE value.

Full detector DOM analysis based on the parameters above were performed, and we showed an NQE consistency with 1.00. The 2012 RIDE values were consistent with an input of 1.35 and showed high separation between NQEs and HQEs, but exhibited an HQE slope of 0.06 per 100 meters with decreasing depth in the DeepCore region. The 2016 RIDE values were inconsistent with the input but showed flat and nearly separated HQE and NQE RIDE values. The existence of a wavelength dependent DOM efficiency that is not currently accounted for in the analysis is discussed, with sights set on moving onto reconstruction and real data once pure simulation results can return the input 1.35 values as expected.

Additionally, we have produced neural network models that can effectively tag stopped muons in the detector with a purity of 97.9% and 72.6% for the 2012 and 2016 datasets respectively, while retaining a sensitivity of over 20% for both sets. More sophisticated models and larger datasets may be able to push these numbers even further. With the computational speed of neural networks, once trained, this method can be a competitive classifier of stopped muons over traditional reconstruction methods. Neural networks were also trained to predict the endpoint of muons that have been labelled as stopped. Prediction results are summarised in Table 4, and we conclude that both datasets performed similarly on predicting the x and y coordinates of stopped muon endpoints. However, the 2016 dataset had an MAE twice that of the 2012 dataset and half the R^2 on the z coordinate. The mean of the three coordinate metrics was an MAE of 61.7 m and R^2 of 0.79 for the 2012 dataset and an MAE of 94.9 m and an R^2 of 0.67 for the 2016 dataset. This highlighted the z coordinate prediction difficulties with the 2016 dataset and allowed us to gain insight into why the 2016 classifier had a worse performance than the 2012 classifier. Neither regression networks currently produce predictions with acceptable accuracy and we conclude that the classification network can be used to tag stopped muons before using traditional reconstruction methods.

This thesis has calculated the RIDE value of the entire detector with more accurate atmospheric muon models and developed metrics for optimising the zenith angle and

distance from DOM to track directly. We have shown the effect on the RIDE value of two different simulated datasets. Machine learning models for tagging stopped muons and predicting their endpoint was also developed. Future work would include stronger analysis into machine learning models with attempted reconstruction of RIDE values based on machine learning predictions instead of true labels. Group analysis based on more groups such as the entirety of DeepCore could better predict optimal zenith angle and DOM-distance values. A better understanding of the 2016 dataset especially regarding the muon endpoint positions would be beneficiary. Work is still required to return input DOM efficiencies in simulation files. After RIDE values are found to be consistent with inputs in simulation, we would move on to reconstructed variables. Finally, once reconstructions are found to be consistent, we can compare with data to produce a relative individual DOM efficiency factor.

Appendices

A   Additional Group Analysis Plots

Figures 39 through 54 show the complete group analysis plots, including the RIDE value and standard deviation for the 2012 and 2016 datasets with both the z-bin and TIL groups. The Figure captions will contain brief information about the year, group definition, parameter, and RIDE or standard deviation.
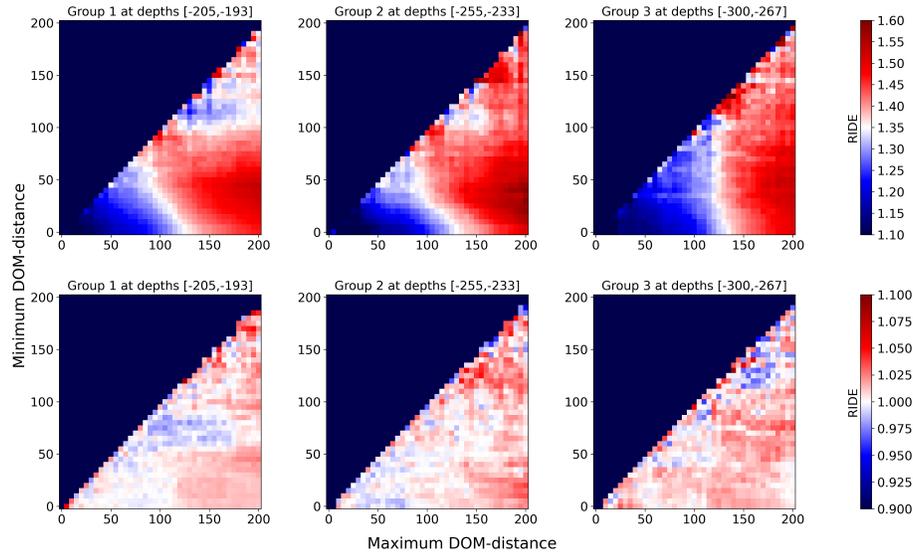
Figure 39: 2012, TIL groups, DOM-distance, mean RIDE.
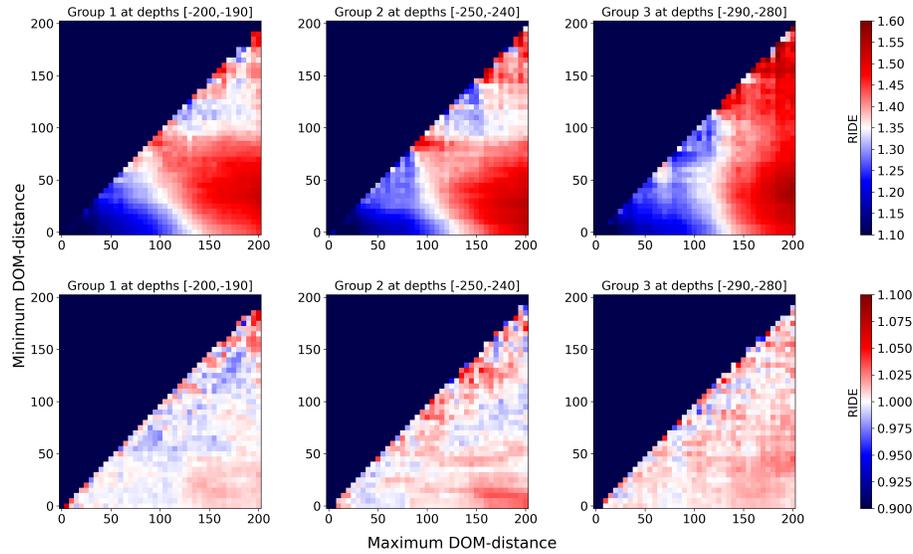

Figure 40: 2012, z-bin groups, DOM-distance, mean RIDE.
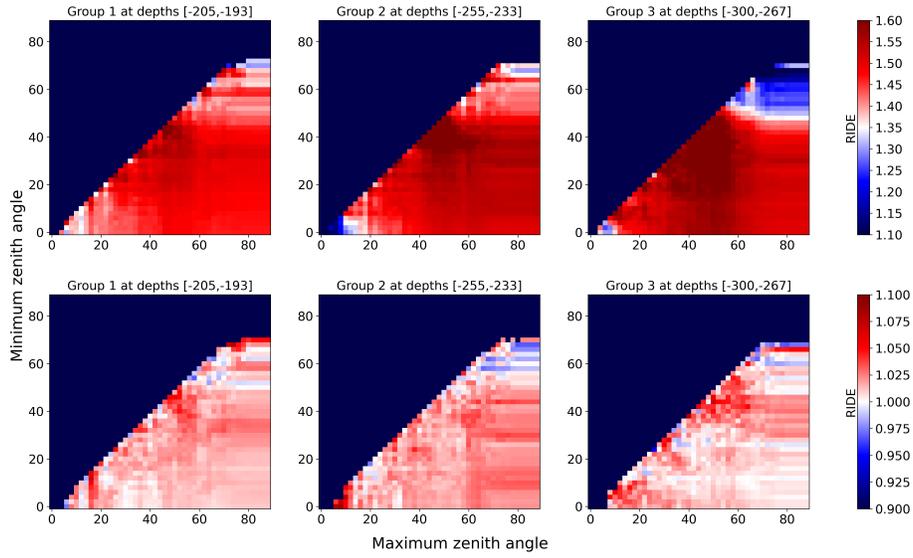
Figure 41: 2012, TIL groups, zenith-angle, mean RIDE.
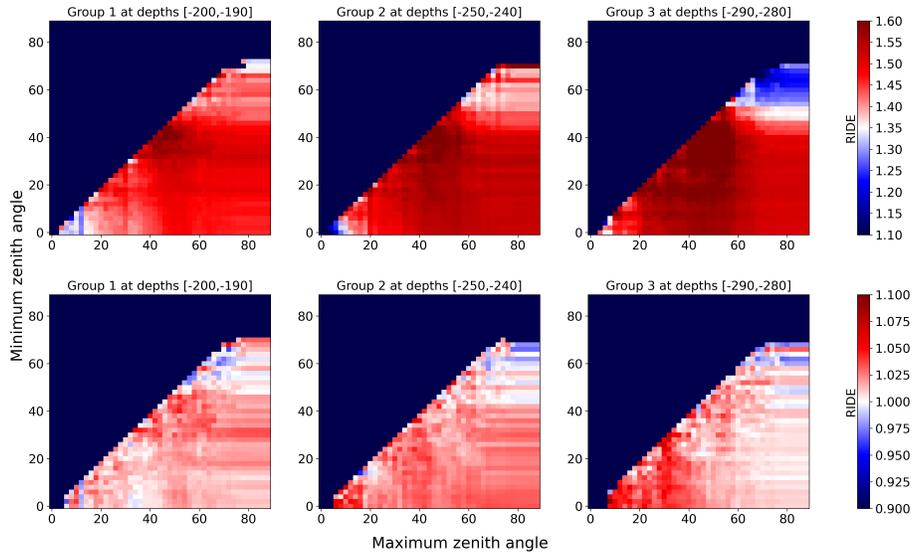
Figure 42: 2012, z-bin groups, zenith-angle, mean RIDE.
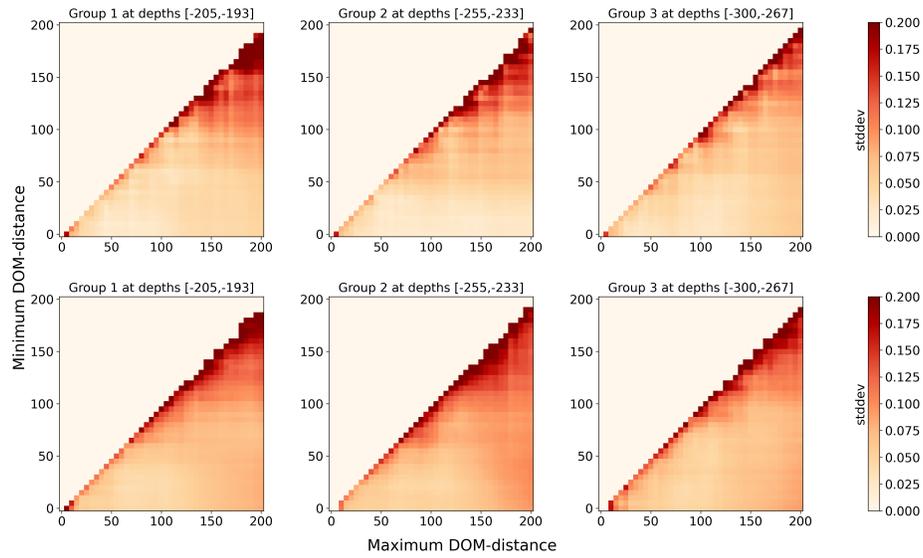
Figure 43: 2012, TIL groups, DOM-distance, standard deviation.
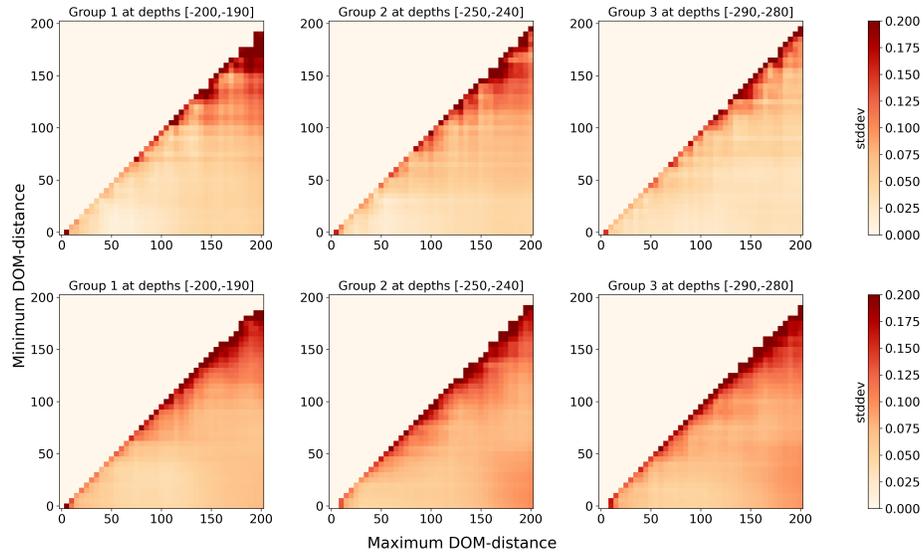
Figure 44: 2012, z-bin groups, DOM-distance, standard deviation.
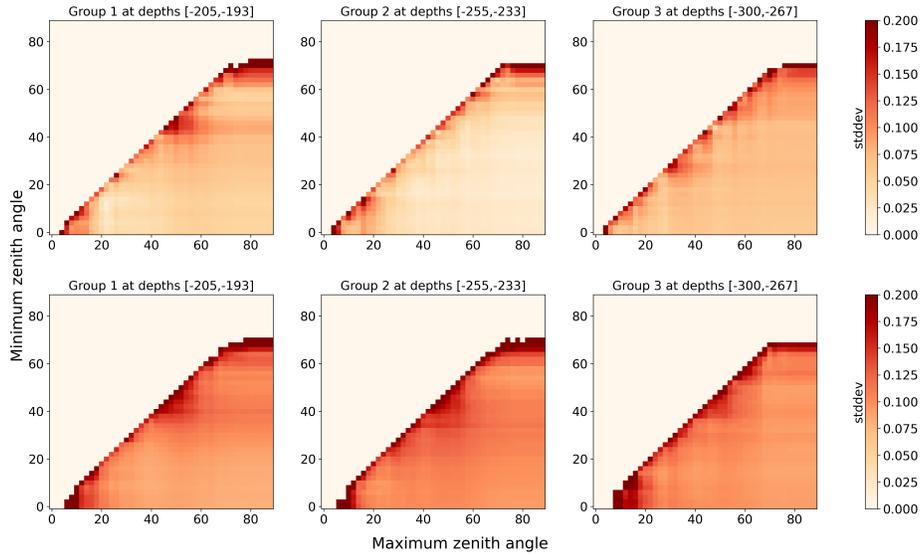
Figure 45: 2012, TIL groups, zenith-angle, standard deviation.
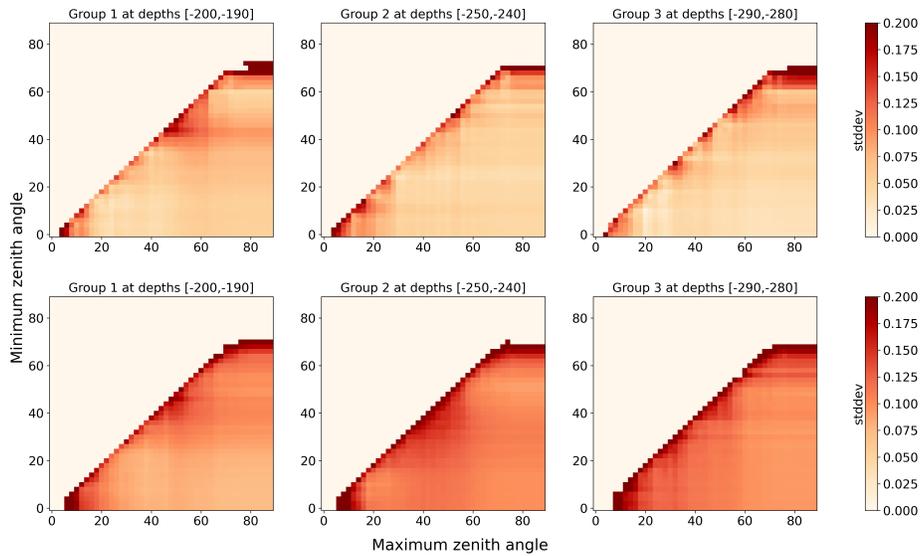
Figure 46: 2012, z-bin groups, zenith-angle, standard deviation.
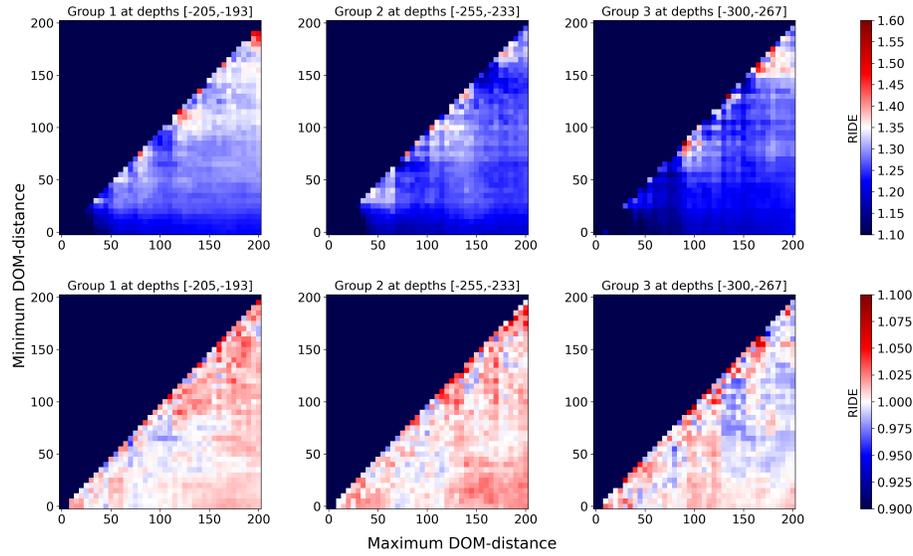
Figure 47: 2016, TIL groups, DOM-distance, mean RIDE.
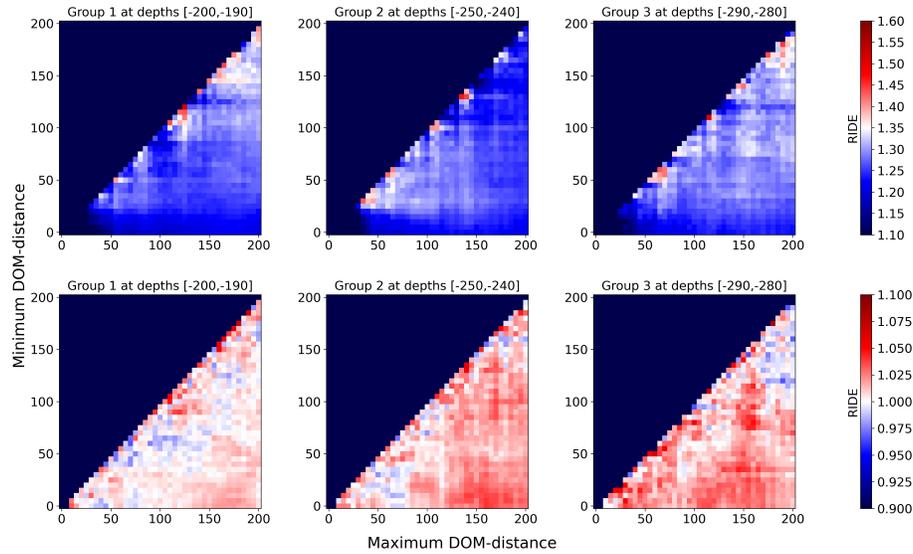
Figure 48: 2016, z-bin groups, DOM-distance, mean RIDE.
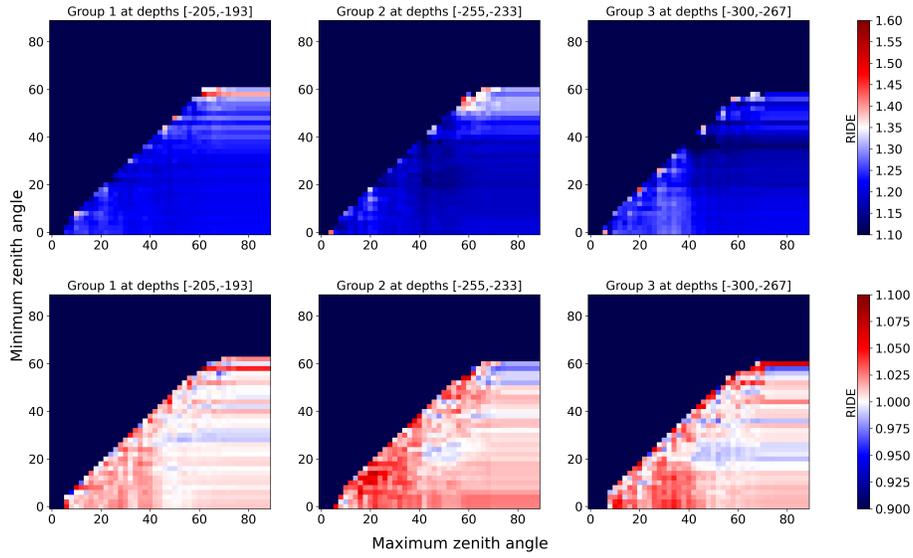
Figure 49: 2016, TIL groups, zenith-angle, mean RIDE.
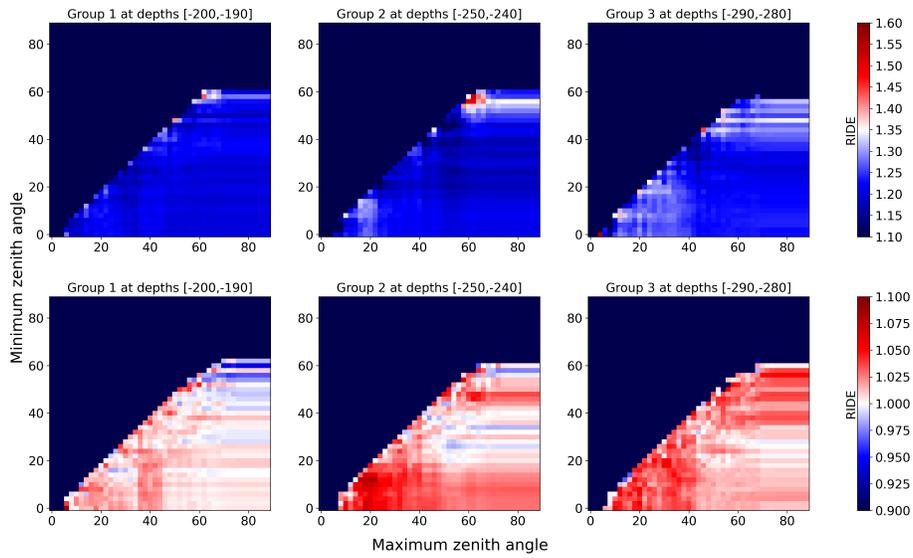
Figure 50: 2016, z-bin groups, zenith-angle, mean RIDE.

74
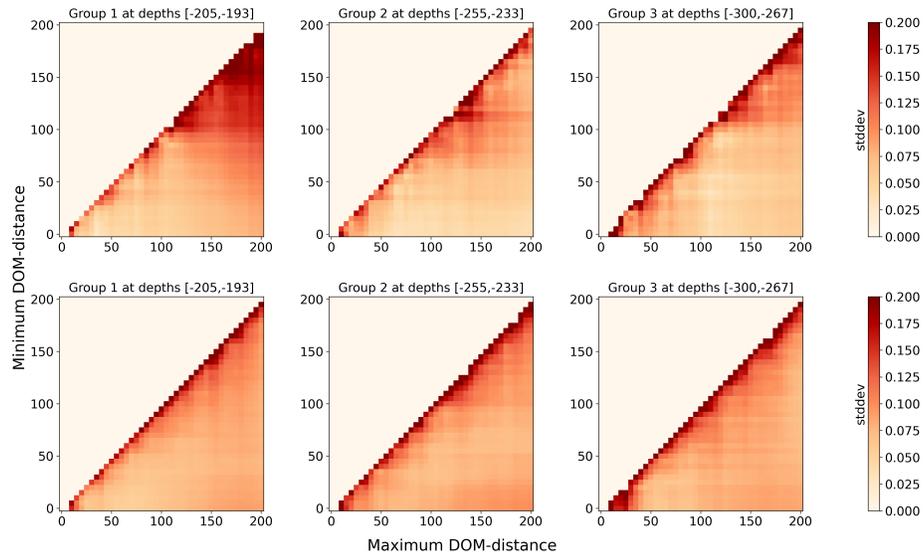
Figure 51: 2016, TIL groups, DOM-distance, standard deviation.
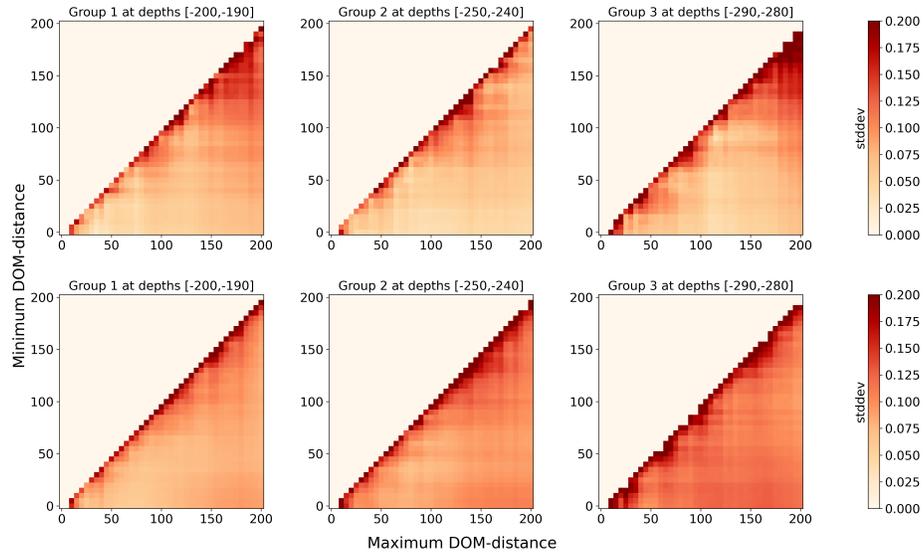
Figure 52: 2016, z-bin groups, DOM-distance, standard deviation.
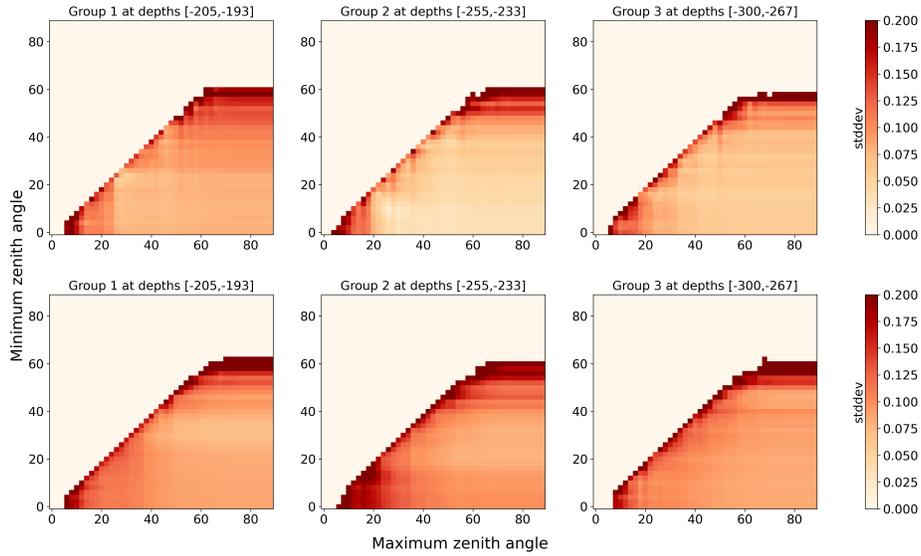
Figure 53: 2016, TIL groups, zenith-angle, standard deviation.
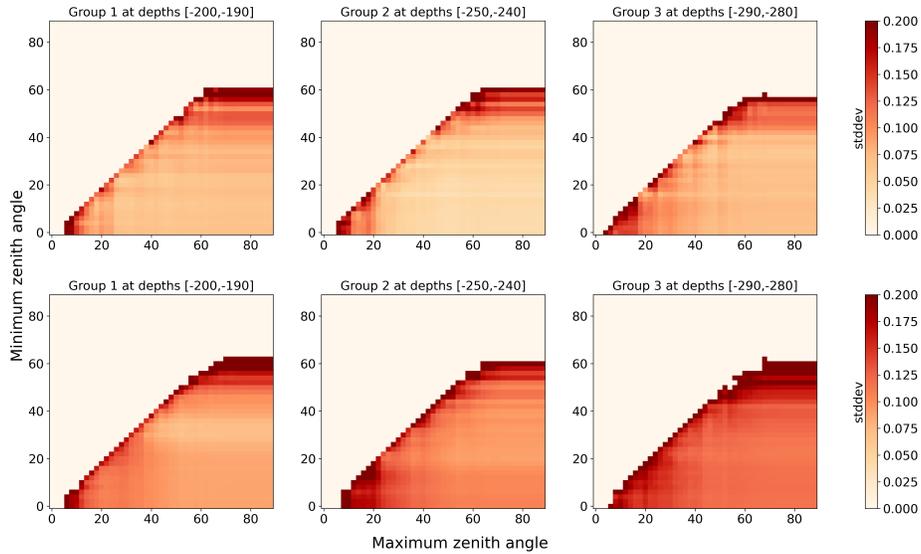
Figure 54: 2016, z-bin groups, zenith-angle, standard deviation.

76

From the figures above, we see slight differences between the two group models, but not anything that substantially change the conclusions of the thesis.

We also plot the two metrics σ_μ and σ_{max} for both datasets including the TIL groups. The Figure captions will contain brief information about the year, group definition, and parameter.
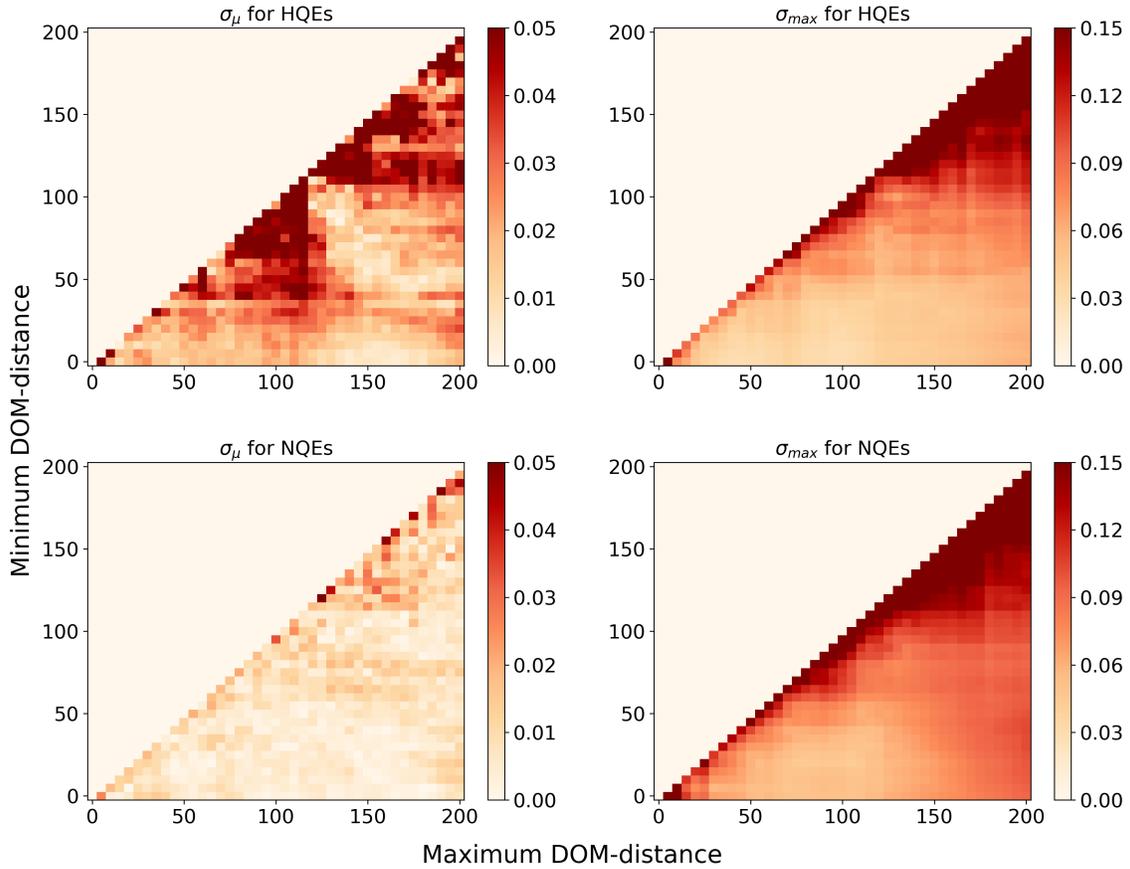
Figure 55: 2012, TIL groups, DOM-distance.

The largest difference between the two models is the general decrease in the NQE σ_{max} with the TIL groups which could point towards a potential decrease in scatter on the DOMs. However this decrease is not consistent across all plots and less than 0.05. The z-bin groups also predict a lowest HQE σ_μ in the 2012 dataset around the concluded best RIDE value of $[0, 130]$.
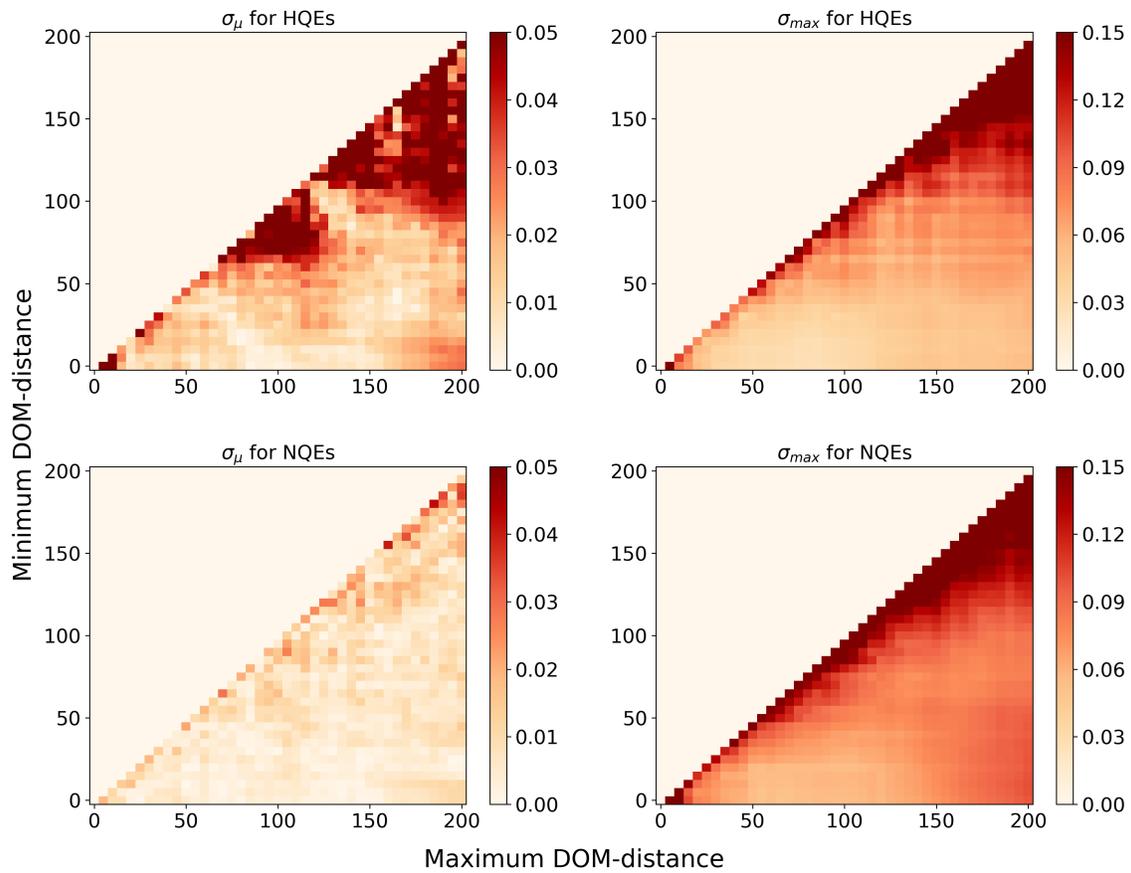
Figure 56: 2012, z-bin groups, DOM-distance.
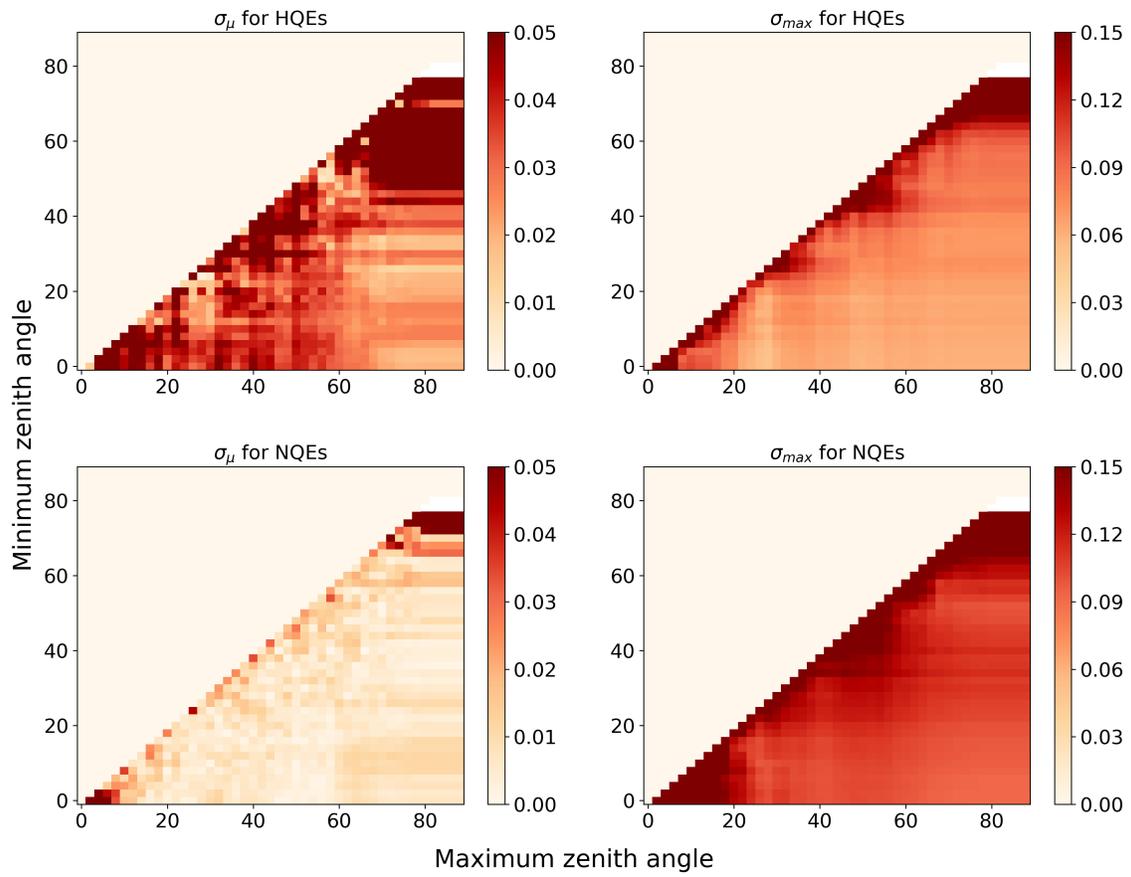
Figure 57: 2012, TIL groups, zenith-angle.
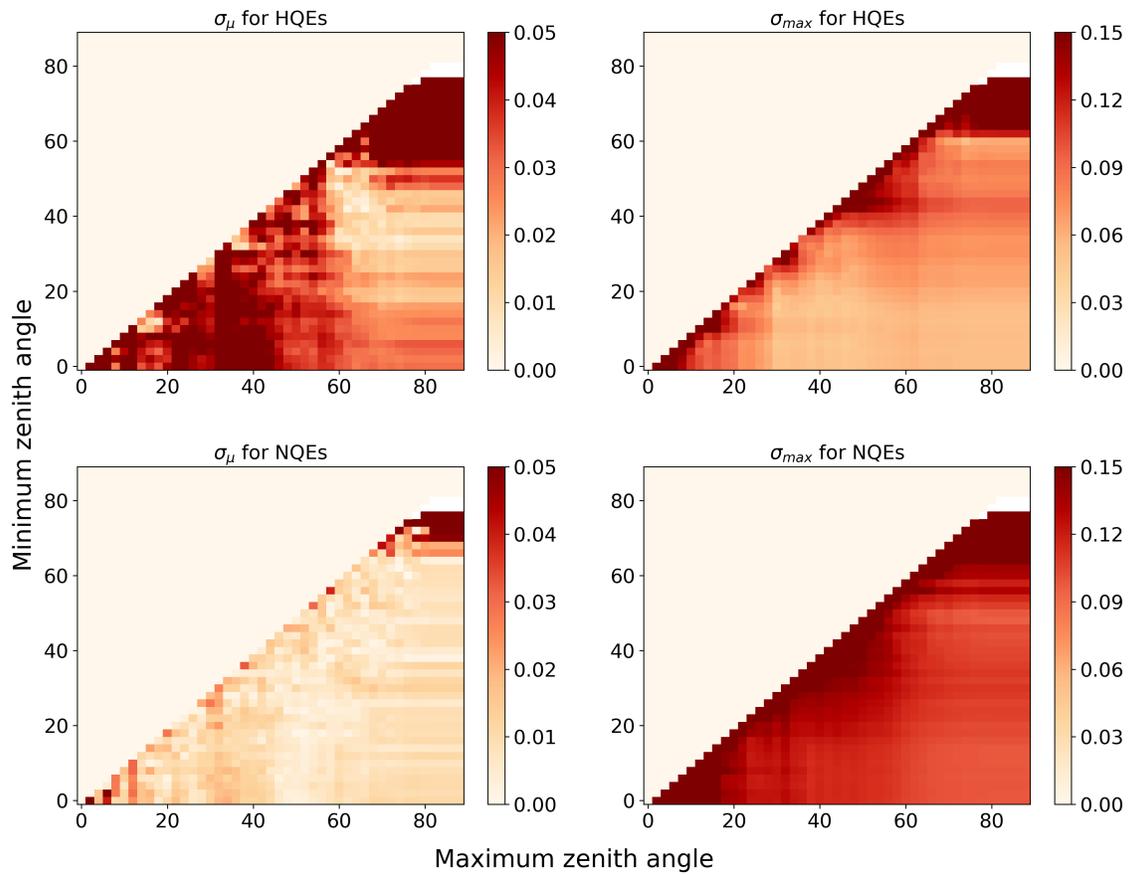
Figure 58: 2012, z-bin groups, DOM-distance.
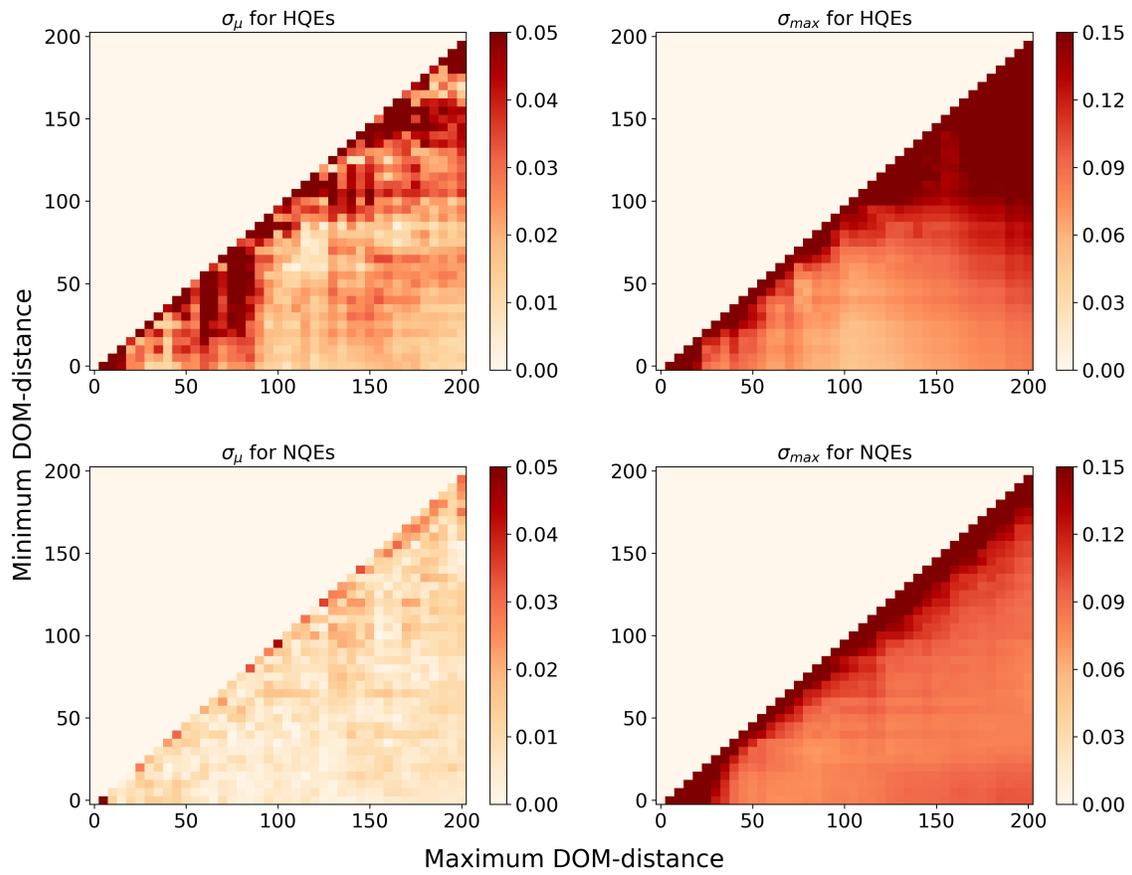
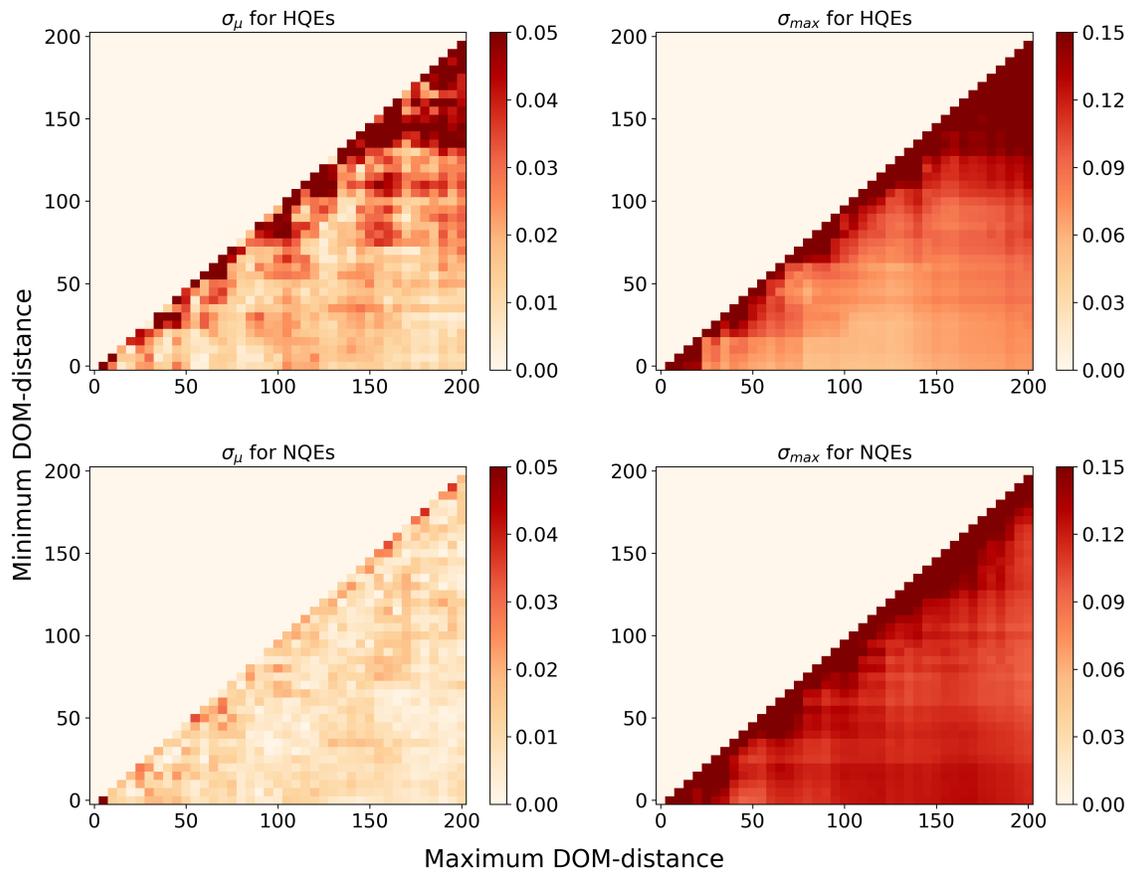Figure 59: 2016, TIL groups, DOM-distance.

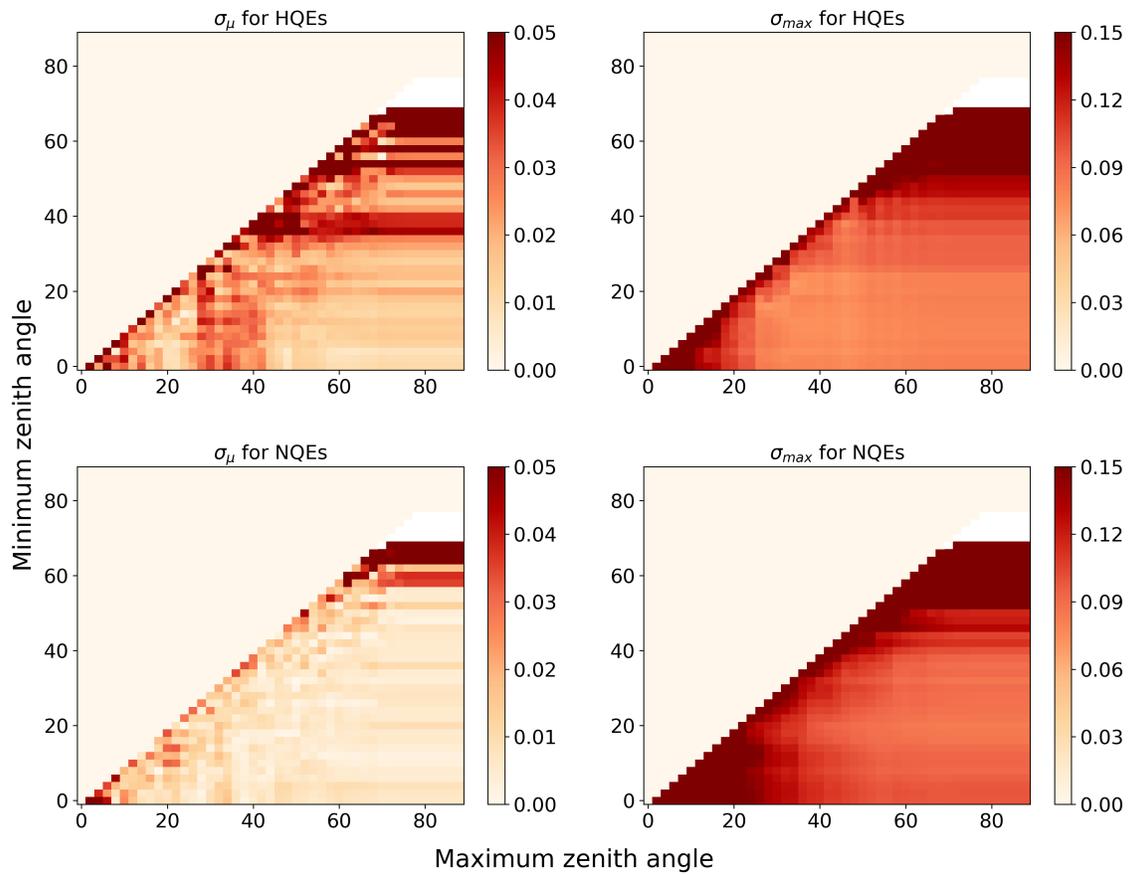Figure 60: 2016, z-bin groups, DOM-distance.

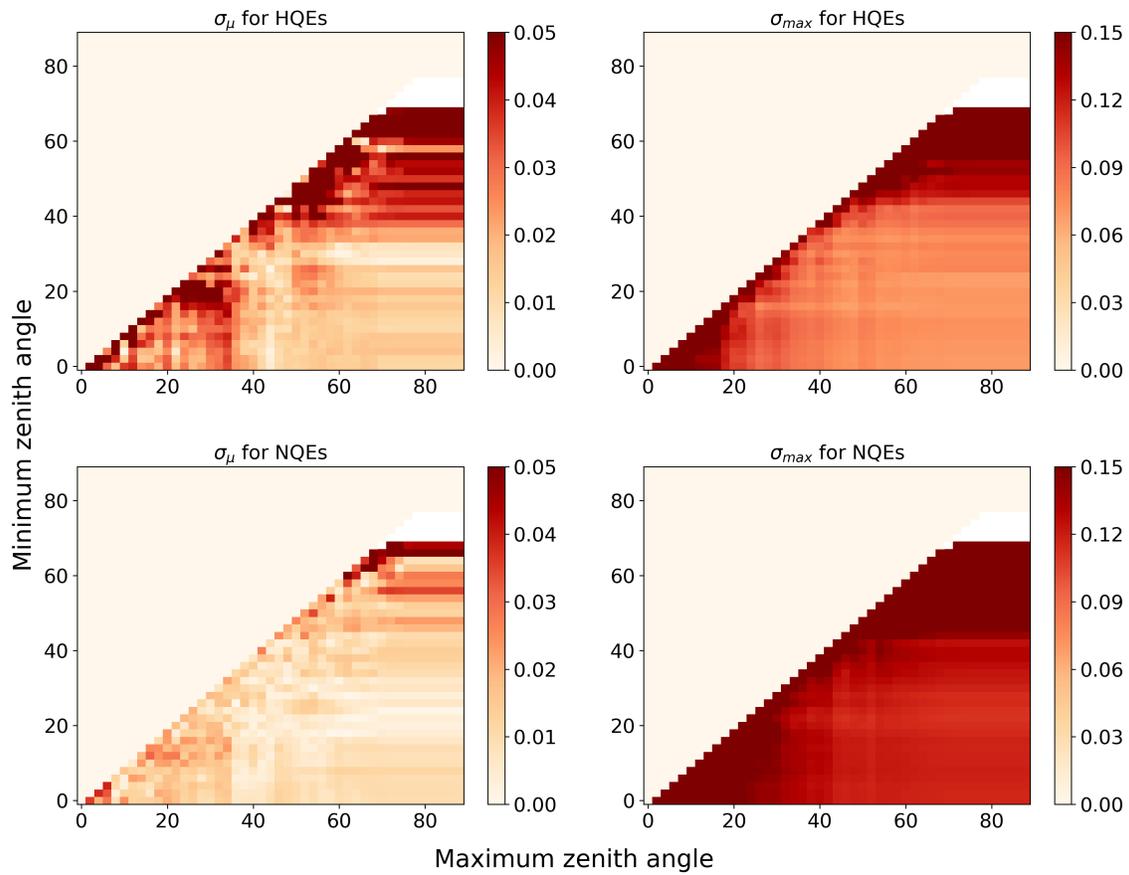Figure 61: 2016, TIL groups, zenith-angle.

Figure 62: 2016, z-bin groups, DOM-distance.

B   F-scores

Below is a collection of plots of different F-scores to showcase how the optimal threshold varies with the value of β. The confusion matrices based on each F-score is however not shown.
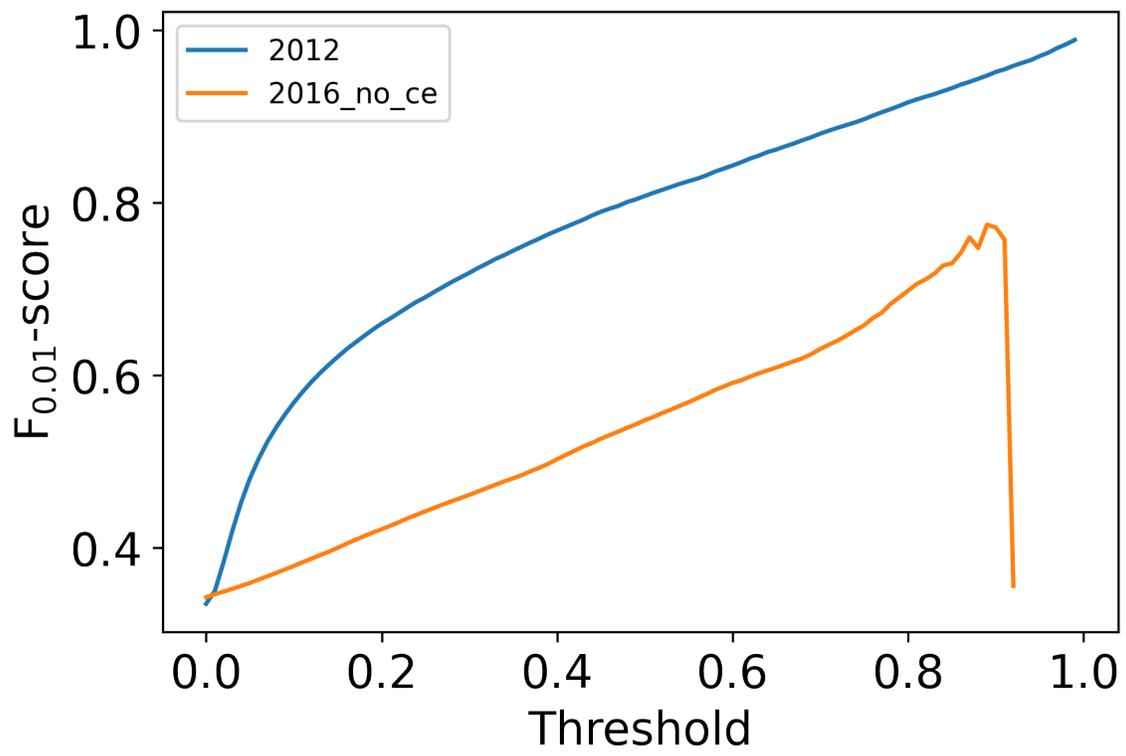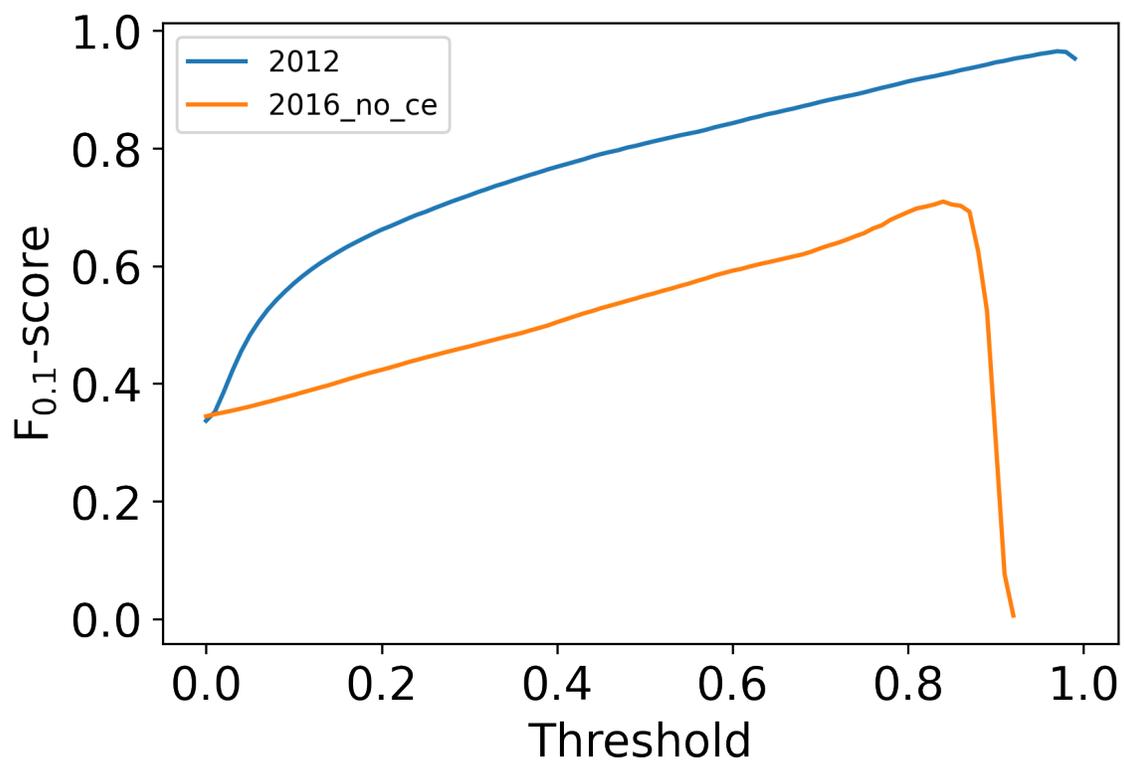
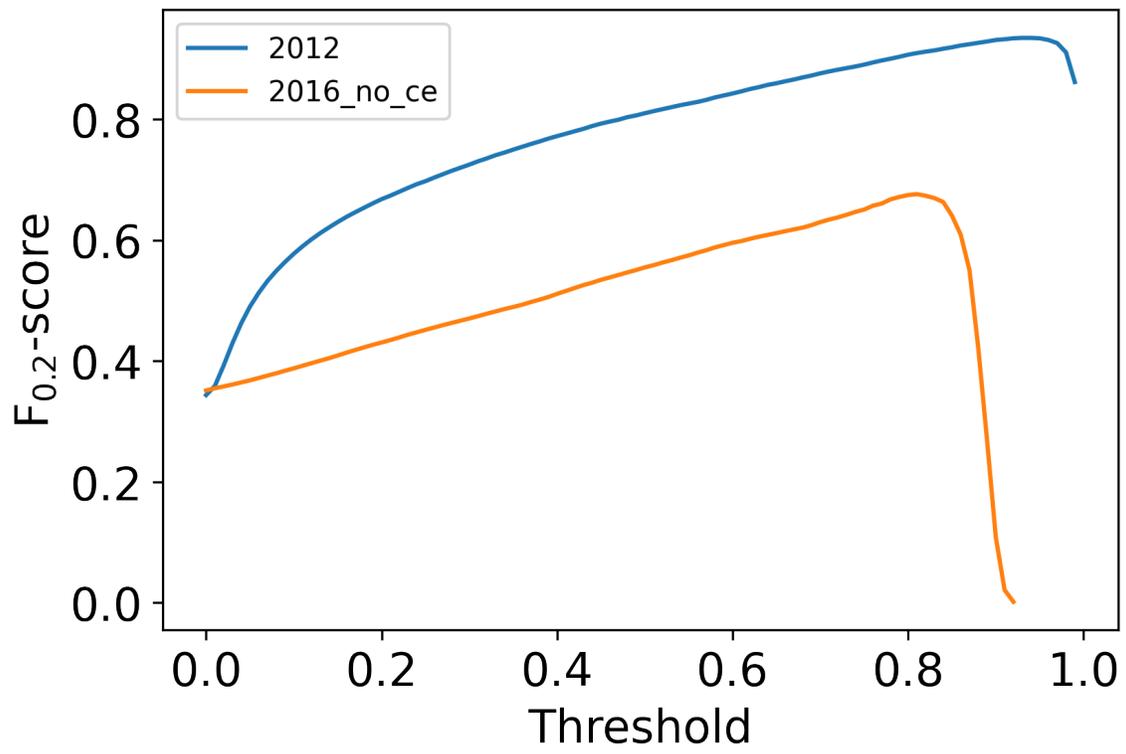Figure 63: $F_{0.01}$ score.

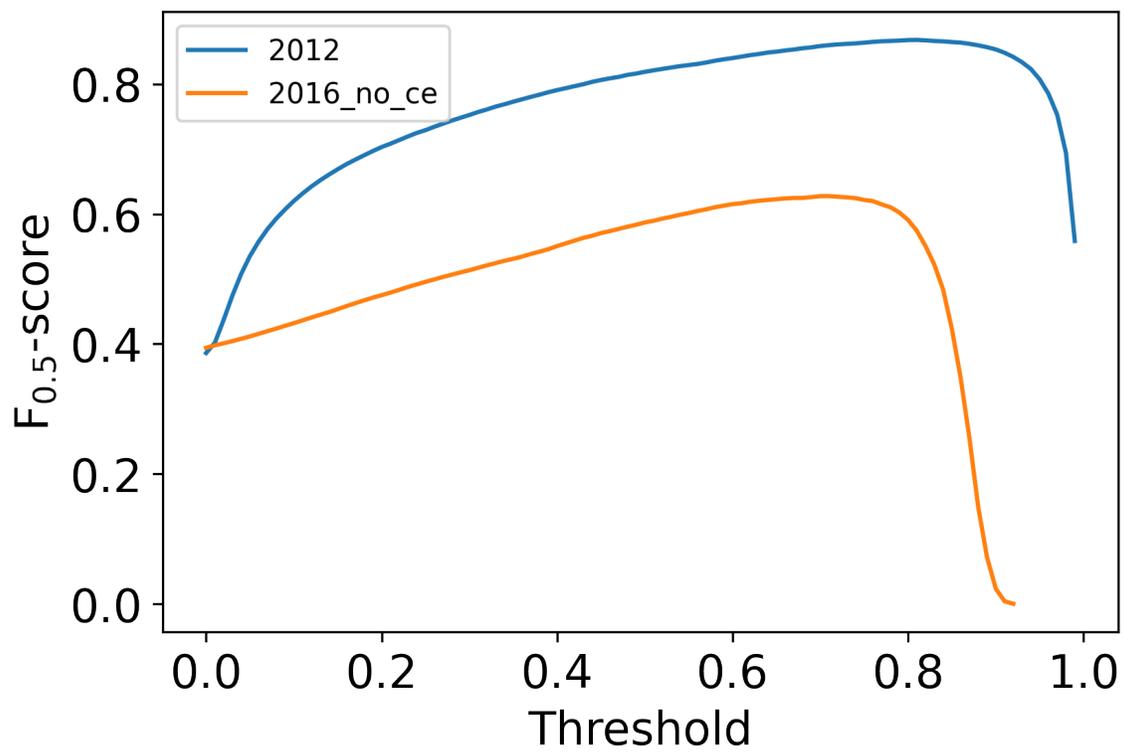Figure 64: $F_{0.1}$ score.

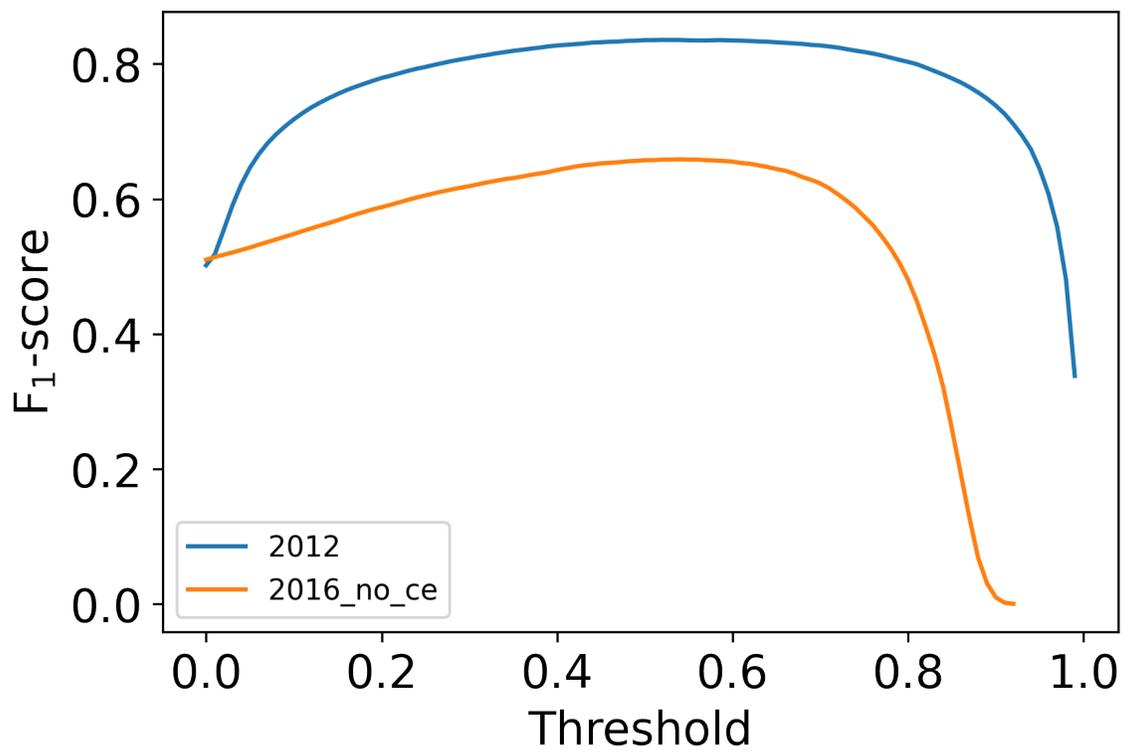Figure 65: $F_{0.2}$ score.

Figure 66: $F_{0.5}$ score.

Figure 67: F_1 score.
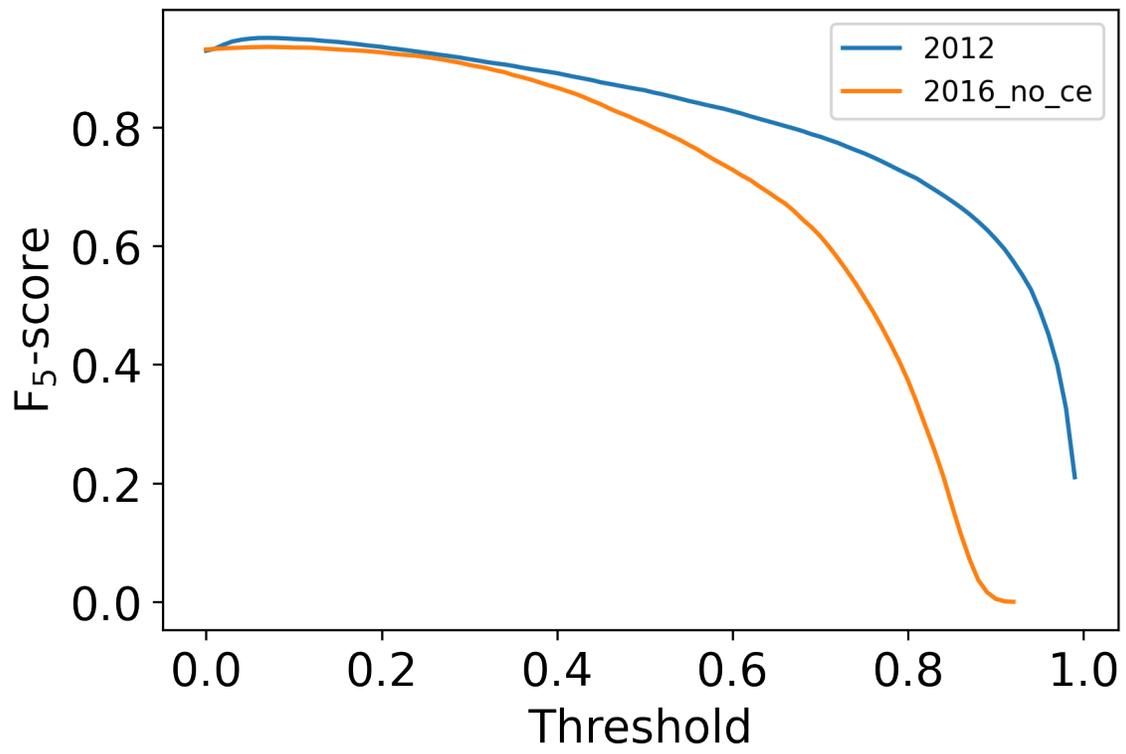
Figure 68: F_5 score.
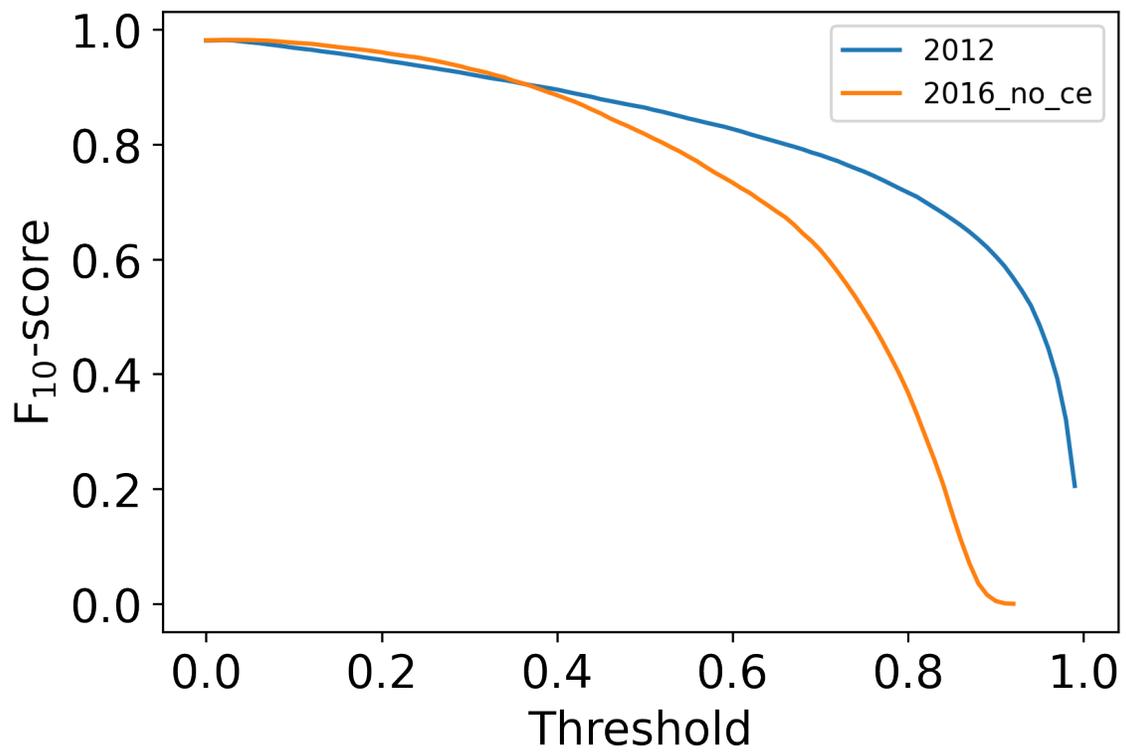
Figure 69: F_{10} score.

We see from Figures 63 through 69 that the $F_{0.01}$ score does not plateau at any threshold for the 2012 dataset and would predict the optimum value as high as possible, and that the 2016 dataset is unstable near its maximum. The $F_{0.1}$ and $F_{0.2}$ are very similar, while the $F_{0.5}$ and harmonic mean F_1 scores simply push the optimal threshold lower for both datasets. The F_5 and F_{10} scores are unreliable and pick an optimal threshold near zero. It is clear from all these figures that an F-score from 0.1 to 0.2 is ideal for our purposes, while pushing it up to 1 is possible but lowers the threshold considerably.

References

[1] P. F. de Salas, D. V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena, C. A. Ternes, M. Tórtola, and J. W. F. Valle, "2020 global reassessment of the neutrino oscillation picture", Journal of High Energy Physics **2021**, 10.1007/jhep02(2021)071 (2021).

[2] S. Mertens, "Direct Neutrino Mass Experiments", J. Phys. Conf. Ser. **718**, 022013 (2016).

[3] J. Christman, *The weak interaction* (Physnet, 2011).

[4] J. Rohlf, *Modern physics from alpha to z0* (Wiley, 1994).

[5] C. G. Tully and G. Zhang, *Multi-messenger astrophysics with the cosmic neutrino background*, 2021.

[6] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge", Zh. Eksp. Teor. Fiz. **34**, 247 (1957).

[7] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles", Prog. Theor. Phys. **28**, 870–880 (1962).

[8] G. Fantini, A. G. Rosso, F. Vissani, and V. Zema, *The formalism of neutrino oscillations: an introduction*, 2020.

[9] Y. Fukuda et al., "Evidence for oscillation of atmospheric neutrinos", Physical Review Letters **81**, 1562–1567 (1998).

[10] I. Esteban, M. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou, "The fate of hints: updated global analysis of three-flavor neutrino oscillations", Journal of High Energy Physics **2020**, 10.1007/jhep09(2020)178 (2020).

[11] F. Halzen and U. Katz, "The era of kilometer-scale neutrino detectors", Advances in High Energy Physics **2013**, 10.1155/2013/680584 (2013).

[12] K. Abe et al., "Real-time supernova neutrino burst monitor at super-kamiokande", Astroparticle Physics **81**, 39–48 (2016).

[13] S. Adrián-Martínez et al., "Searches for point-like and extended neutrino sources close to the galactic center using the antares neutrino telescope", The Astrophysical Journal **786**, L5 (2014).

[14] A. Suzuki (KamLAND), "Results from KamLAND reactor neutrino detection", Phys. Scripta T **121**, edited by L. Bergström, O. Botner, P. Carlson, P. O. Hulth, and T. Ohlsson, 33–38 (2005).

[15] H. H. Chen, "Direct approach to resolve the solar-neutrino problem", Phys. Rev. Lett. **55**, 1534–1536 (1985).

[16] R. Abbasi et al., "The icecube data acquisition system: signal capture, digitization, and timestamping", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **601**, 294–316 (2009).

[17] D. F. Cowen and the IceCube Collaboration, "Tau neutrinos in icecube", Journal of Physics: Conference Series **60**, 227–230 (2007).

[18] J. Madsen, *Ultra-high energy neutrinos*, 2019.

[19] R. Abbasi et al., "The design and performance of icecube deepcore", Astroparticle Physics **35**, 615–624 (2012).

[20] M. G. Aartsen et al., "First observation of pev-energy neutrinos with icecube", Physical Review Letters **111**, 10.1103/physrevlett.111.021103 (2013).

[21] M. Aartsen et al., "The icecube neutrino observatory: instrumentation and online systems", Journal of Instrumentation **12**, P03012–P03012 (2017).

[22] *The icecube collaboration*, https://icecube.wisc.edu/science/icecube/, Accessed: 2021-04-29.

[23] H. Corp, "HAMAMATSU PHOTOMULTIPLIER TUBE DATA SHEET R7081-02 for ICECUBE Experiment",

[24] R. Abbasi et al., "Calibration and characterization of the icecube photomultiplier tube", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **618**, 139–152 (2010).

[25] D. Chirkin, "Study of South Pole ice transparency with IceCube flashers", in 32nd International Cosmic Ray Conference (2011).

[26] M. Aartsen, M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers, M. Ahrens, I. Al Samarai, D. Altmann, K. Andeen, and et al., "Neutrino emission from the direction of the blazar txs 0506+056 prior to the icecube-170922a alert", Science **361**, 147–151 (2018).

[27] M. G. Aartsen et al., "Measurement of atmospheric neutrino oscillations at 6–56 gev with icecube deepcore", Physical Review Letters **120**, 10.1103/physrevlett.120.071801 (2018).

[28] M. G. Aartsen et al., "Measurement of atmospheric tau neutrino appearance with icecube deepcore", Physical Review D **99**, 10.1103/physrevd.99.032007 (2019).

[29] É. Bourbeau, "Phd. thesis [embargoed]", PhD thesis (University of Copenhagen, 2021).

[30] R. Stokstad, "Design and performance of the icecube electronics", Proceedings - 11th Workshop on Electronics for LHC and Future Experiments, LECC 2005 (2005).

[31]M. Aartsen et al., "Measurement of south pole ice transparency with the icecube led calibration system", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **711**, 73–89 (2013).

[32]M. Aartsen et al., "Icecube-gen2: a vision for the future of neutrino astronomy in antarctica", arXiv:1412.5106 (2014).

[33]É. Bourbeau, `https://events.icecube.wisc.edu/event/100/contributions/522/attachments/361/383/RIDE_update_VIII_AtlantaMeeting2018.pdf`, Private communication, presentation on non-public IceCube wiki., Spring 2018.

[34]D. Groom and S. Klein, "Passage of particles through matter", European Physical Journal C - EUR PHYS J C **15**, 163–173 (2000).

[35]M. Aartsen et al., "Characterization of the atmospheric muon flux in icecube", Astroparticle Physics **78**, 1–27 (2016).

[36]T. DeYoung, "IceTray: A software framework for IceCube", in 14th International Conference on Computing in High-Energy and Nuclear Physics (2005).

[37]D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, "CORSIKA: A Monte Carlo code to simulate extensive air showers", (1998).

[38]D. Chirkin, "dCORSIKA update",

[39]J. R. Hörandel, "On the knee in the energy spectrum of cosmic rays", Astroparticle Physics **19**, 193–220 (2003).

[40]C. Kopper, *Clsim*, `https://github.com/claudiok/clsim`, Accessed: 2021-04-13.

[41]E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, "Cosmic ray interaction event generator sibyll 2.1", Physical Review D **80**, `10.1103/physrevd.80.094003` (2009).

[42]F. Riehn, H. P. Dembinski, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, *The hadronic interaction model sibyll 2.3c and feynman scaling*, 2017.

[43]I. Collaboration et al., *The icecube neutrino observatory part vi: ice properties, reconstruction and future developments*, 2013.

[44]Dmitry Chirkin, *Ice model update*, `https://events.icecube.wisc.edu/event/75/contributions/4869/attachments/3947/4325/icetalk.pdf`, UW-Madison, Apr. 2016.

[45]S. Blot, *[clsim] implement correct wavelength acceptance for hqe doms*, `https://code.icecube.wisc.edu/projects/icecube/ticket/2162`, Accessed: 2021-05-04.

[46]M. J. Larson, "A search for tau neutrino appearance with icecube-deepcore", PhD thesis (University of Copenhagen, 2018).

[47]N. Kulacz, "In situ measurement of the icecube dom efficiency factor using atmospheric minimum ionizing muons", MA thesis (University of Alberta, 2019).

[48]T. Feusels, "Extensive air showers measured by the 79-string icecube observatory at south pole", 283– (2011).

[49]J. Ahrens et al., "Muon track reconstruction and data selection techniques in amanda", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **524**, 169–194 (2004).

[50]R. Abbasi et al., *A convolutional neural network based cascade reconstruction for the icecube neutrino observatory*, Submitted to Journal of Instrumentation, 2021.

[51]V. Jain, *What's the role of weights and bias in a neural network?*, `https://towardsdatascience.com/whats-the-role-of-weights-and-bias-in-a-neural-network-4cf7e9888a0f`, Accessed: 2021-04-13.

[52]F. Bre, J. M. Gimenez, and V. D. Fachinotti, "Prediction of wind pressure coefficients on building surfaces using artificial neural networks", Energy and Buildings **158**, 1429–1441 (2018).

[53]C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, *Activation functions: comparison of trends in practice and research for deep learning*, 2018.

[54]I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, `http://www.deeplearningbook.org` (MIT Press, 2016).

[55]S. Liu, N. Yang, M. Li, and M. Zhou, "A recursive recurrent neural network for statistical machine translation", in Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: long papers) (June 2014), pp. 1491–1500.

[56]A. Karpathy and L. Fei-Fei, *Deep visual-semantic alignments for generating image descriptions*, 2015.

[57]I. Sutskever, J. Martens, and G. Hinton, "Generating text with recurrent neural networks", in (Jan. 2011), pp. 1017–1024.

[58]V. Ravi, S. Kp, and P. Poornachandran, "Evaluation of recurrent neural network and its variants for intrusion detection system (ids)", International Journal of Information System Modeling and Design **8**, 43–63 (2017).

[59]S. Bai, J. Z. Kolter, and V. Koltun, *An empirical evaluation of generic convolutional and recurrent networks for sequence modeling*, 2018.

[60]F. Lässig, *Temporal convolutional networks and forecasting*, `https://medium.com/unit8-machine-learning-publication/temporal-convolutional-networks-and-forecasting-5ce1b6e97ce4`, Accessed: 2021-04-13.

[61] P. Remy, *Temporal convolutional networks for keras*, `https://github.com/philipperemy/keras-tcn`, 2020.

[62] F. Chollet et al., *Keras*, `https://keras.io`, 2015.

[63] M. Abadi et al., *Tensorflow: large-scale machine learning on heterogeneous distributed systems*, 2016.

[64] D. P. Kingma and J. Ba, *Adam: a method for stochastic optimization*, 2017.

[65] K. Hajian-Tilaki, "Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation", Caspian journal of internal medicine **4**, 627–635 (2013).

[66] J. Hanley and B. Mcneil, "The meaning and use of the area under a receiver operating characteristic (roc) curve", Radiology **143**, 29–36 (1982).

[67] M. Ahlers, K. Helbing, and C. Pérez de los Heros, "Probing particle physics with icecube", The European Physical Journal C **78**, `10.1140/epjc/s10052-018-6369-9` (2018).

[68] C. VAN RIJSBERGEN, "Foundation of evaluation", Journal of Documentation **Volume 30**, 365–373 (1974).