WWU
MÜNSTER

Experimentelle Physik

Development of detector calibration and graph neural network-based selection and reconstruction algorithms for the measurement of oscillation parameters with KM3NeT/ORCA

Entwicklung von Detektorkalibration und auf Graph Neural Networks basierten Rekonstruktions- und Selektionsalgorithmen zur Bestimmung von Oszillationsparametern in KM3NeT/ORCA

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich Physik
der Mathematisch-Naturwissenschaftlichen
Fakultät der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Daniel Guderian
aus Ahlen

— 2022 —
(leicht überarbeitete Version, Juli 2022)

Abstract

NEUTRINOS are elementary particles of high interest for many different fields of scientific research. They serve as messengers from the high-energy universe, reveal the inner workings of our Sun and, maybe most prominently, are involved in particle decays. The decays were not only historically the first observations whose explanation required these "ghost-like" particles to exist, but also cause countless neutrinos to rain down on the surface of the Earth from processes higher up in the atmosphere.

Yet, one challenge is common to all experiments setting out to measure the properties of these particles: Since they only interact via the weak nuclear force, they are extremely complicated to detect. Additionally, their mass is so tiny that it is assumed to be zero in the scope of the Standard Model of particle physics. However, evidence for a non-zero mass was found in the form of *neutrino oscillations*, requesting extensions to their theoretical description. For this discovery, T. Kajita and A. B. McDonald were awarded the Nobel Prize in physics in 2015. Oscillation in this context refers to the fact that the flavor of a neutrino can change over the distance it travels. The probability for the transition follows an oscillatory pattern, hence the name.

One experiment aiming to detect atmospheric neutrinos to determine the parameters that govern the flavor transitions is KM3NeT/ORCA. It is a multi-megaton water Cherenkov neutrino telescope in the Mediterranean Sea, currently under construction. It utilizes multi-PMT optical sensors to capture the light created by neutrino interactions in the vicinity of the detector.

In order to extract any physical information from the interactions, their properties, namely energy and direction, have to be reconstructed from the recorded light. To this end, novel deep learning techniques have been employed in the scope of this work to improve the resolution of the reconstruction and thus the sensitivity to the physics goals. Additionally, classifier neural networks are designed to help with particle identification, which is fundamental to any physics analysis.

As per usual for supervised learning approaches, and no different for the graph neural networks utilized in this particular case, sufficient labeled

data to train on have to be produced that model the reality as close as possible. One prerequisite to generating these is a accurate time and position calibration of the detector. Furthermore, the calibrations are required to reach the maximum physics potential of the experiment.

Both of these topics, ensuring a reliable calibration and realizing the deep learning methods by applying them to simulations and real data, are tackled in this thesis. In the introduction, the fundamental properties of neutrinos are summarized, with particular emphasis on their ability to oscillate. The next chapter demonstrates how neutrinos can be detected and how their oscillation signature reveals itself.

The first main part describes the calibration studies, for which first an overview of currently available calibration techniques in KM3NeT is given. Then, the development of an atmospheric muon-based detector calibration is detailed, and consistency studies and comparisons in simulations and real data are presented.

The second main part encompasses all studies related to the deep learning approach. It starts with an introduction to the topic in general and in particular to the graph neural networks employed. Next, training and performance are evaluated, after which the outputs are adopted for a neutrino selection. Based on this, exemplary oscillation analyses can demonstrate the proficiency of the selected sets and the reconstruction. Moreover, the developed tools are applied to real data, allowing to probe the neutrino simulations in detail and infer the best fitting oscillation parameters.

In a final consideration, the two main topics are united, as the influences of the calibration on reconstruction, selection and the final physics results are evaluated.

Zusammenfassung

NEUTRINOS sind Elementarteilchen und für viele verschiedene Bereiche der wissenschaftlichen Forschung von großem Interesse. Sie dienen als Boten des hochenergetischen Universums, geben Aufschluss über das Innenleben unserer Sonne und sind an den Teilchenzerfällen beteiligt. Die Zerfälle waren nicht nur historisch die ersten Beobachtungen, deren Erklärung die Existenz dieser "geisterhaften" Teilchen benötigte, sondern sorgen auch dafür, dass zahllose Neutrinos aus Prozssen in der Atmosphäre auf die Erdoberfläche niederregnen.

Eine Herausforderung ist jedoch allen Experimenten gemeinsam, die die Eigenschaften dieser Teilchen messen wollen: Da sie nur über die schwache Kernkraft wechselwirken, ist ihr Nachweis äußerst kompliziert. Außerdem ist ihre Masse so gering, dass sie im Rahmen des Standardmodells der Teilchenphysik als Null angenommen wird. Allerdings wurden Beweise für eine Masse ungleich Null in Form von Neutrino-Oszillationen gefunden, die eine Erweiterung ihrer theoretischen Beschreibung erforderlich machten. Für diese Entdeckung wurden T. Kajita und A. B. McDonald 2015 mit dem Nobelpreis für Physik ausgezeichnet. Oszillation bezieht sich in diesem Zusammenhang auf die Tatsache, dass sich der Flavor eines Neutrinos über den Weg, den es zurücklegt, ändern kann. Die Wahrscheinlichkeit für diesen Übergang folgt einem oszillatorischen Muster, daher der Name.

Ein Experiment, das darauf abzielt, atmosphärische Neutrinos nachzuweisen, um die Parameter zu bestimmen, die die Flavor-Übergänge beschreiben, ist KM3NeT/ORCA. Es handelt sich um ein megatonnen Wasser-Cherenkov-Neutrino-Teleskop im Mittelmeer, das sich derzeit im Bau befindet. Es nutzt optische mehrfach-PMT-Sensoren, um das durch Neutrinowechselwirkungen in der Nähe des Detektors erzeugte Licht einzufangen.

Um physikalische Informationen aus den Wechselwirkungen zu gewinnen, müssen deren Eigenschaften, wie Energie und Richtung, aus dem aufgezeichneten Licht rekonstruiert werden. Zu diesem Zweck wurden im Rahmen dieser Arbeit neuartige Deep-Learning-Techniken eingesetzt, um die Auflösung der Rekonstruktion und damit die Empfindlichkeit für die physikalischen Ziele zu verbessern. Darüber hinaus werden neuronale Klassifizierungsnetze entwickelt, die bei der Identifizierung von Teilchen helfen,

die für jede physikalische Analyse von grundlegender Bedeutung ist.

Wie bei Supervised-Learning üblich, und nicht anders bei denen in diesem speziellen Fall verwendeten Graph Neural Networks, müssen ausreichend gelabelte Daten zum Trainieren erzeugt werden, die die Realität so gut wie möglich abbilden. Eine Voraussetzung für die Erzeugung solcher Daten ist eine präzise Zeit- und Positionskalibrierung des Detektors. Außerdem sind Kalibrierungen erforderlich, um das maximale physikalische Potenzial des Experiments zu realisieren.

Beide Themen, die Sicherstellung einer zuverlässigen Kalibrierung und die Umsetzung der Deep-Learning-Methoden durch Anwendung auf Simulationen und reale Daten, werden in dieser Arbeit angegangen. In der Einleitung werden die grundlegenden Eigenschaften von Neutrinos zusammengefasst, mit besonderem Augenmerk auf ihre Fähigkeit zu oszillieren. Das nächste Kapitel zeigt, wie Neutrinos nachgewiesen werden können und wie sich ihre Oszillationssignatur beobachten lässt.

Der erste Hauptteil beschreibt die Kalibrierungsstudien, für die zunächst ein Überblick über die derzeit verfügbaren Kalibrierungsverfahren in KM3NeT gegeben wird. Dann wird die Entwicklung einer auf atmosphärischen Myonen basierenden Detektorkalibrierung beschrieben. Darüber hinaus werden Konsistenzstudien und Vergleiche in Simulationen und realen Daten vorgestellt.

Der zweite Hauptteil umfasst alle Studien im Zusammenhang mit dem Deep-Learning-Ansatz. Er beginnt mit einer Einführung in das Thema im Allgemeinen und insbesondere in die verwendeten Graph Neural Networks. Anschließend werden Training und Performanz ausgewertet und die Ergebnisse für eine Neutrinoselektion übernommen. Darauf aufbauend können beispielhafte Oszillationsanalysen die Leistungsfähigkeit der Selektionen und der Rekonstruktion demonstrieren. Des Weiteren werden die entwickelten Werkzeuge auf reale Daten angewandt, um die Neutrino-Simulationen im Detail zu untersuchen und die am besten die Daten beschreibenden Oszillationsparameter zu ermitteln.

In einer abschließenden Betrachtung werden die beiden Hauptthemen zusammengeführt, indem die Einflüsse der Kalibrierung auf die Rekonstruktion, die Selektion und die physikalischen Ergebnisse bewertet werden.

Contents

1

Fundamental neutrino properties

Right from its postulation by Wolfgang Pauli in 1930, the neutrino attained a special role in the then-emerging field of particle physics. Rather desperately admitting "I have done a terrible thing, I have postulated a particle that cannot be detected", he saw no other way of saving the momentum and energy conservation laws than by introducing a new, uncharged particle that would be emitted in the beta decay, along with an electron [1]. Given the fact that the neutrino neither interacts via the electromagnetic force nor via the strong force and, additionally, has a minimal mass compared to other elementary particles [2], Pauli was right about the difficulties involved in detecting it. This is why it took 30 years of theoretical and technological advances to achieve the first detection of neutrinos in Reines and Cowan's experiment [3]. Since then, the neutrino and its discovery potential have come a long way.

Further proving its unique standing, many fundamental properties are still unknown or flawed with large uncertainties, including something as elementary as its mass. Consequently, numerous experiments strive to determine these properties in high precision measurements. At the same time, neutrinos can be utilized as messengers from extra-galactic sources in extensive volume experiments.

This chapter will introduce the fundamental physics of neutrinos from the creation over the propagation to the interactions relevant for neutrino experiments.

1.1 Neutrinos and the Standard Model of particle physics

Once established as an elementary particle, the neutrino is naturally a constituent of the Standard Model of particle physics [5]. Primarily developed in the second half of the 20th century, the Standard Model

generations of fermions
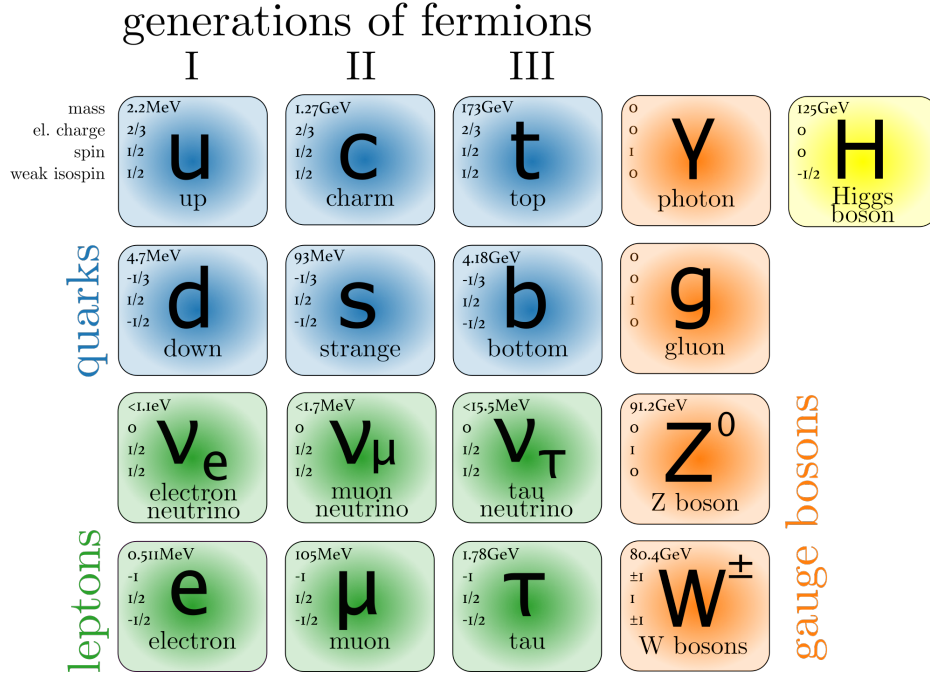
Figure 1.1: The Standard Model of particle physics. Quarks (blue) and leptons (green) are fermions and together form the known matter. The gauge bosons (orange) mediate the fundamental forces. Completing the model is the Higgs particle. Values taken from [4].

is one of physics most well-tested theories. As a quantum field theory, the observable particles are described as excitations of their underlying fields. Despite finding huge success in the prediction of missing particles, like the top quark [6] and the Higgs boson [7], there are still open questions that require extensions to the model, like the prediction of dark matter particles, or, notably a non-zero mass of the neutrino.

The Standard Model is organized into fermions, holding a half-integer spin, and bosons with spin 1 [2]. Additionally, there is the Higgs boson with spin 0 that creates a field to which fermions couple, giving them their mass. Massive gauge bosons gain their mass via symmetry breaking of this field in the so called Higgs mechanism [8]. The bosons mediate the three fundamental forces covered in the Standard Model: electromagnetic (γ), strong (g) and weak (Z and W). The fourth elementary force, the gravitation, is not described by this model. The fermions can be divided into quarks and leptons. The difference being that the former are subject to the electromagnetic, strong and weak force while the latter are only involved in weak and electromagnetic (if charged) interactions.

Within the fermions, there are three different generations, which differ by their so-called *flavor* and, most notably, by their mass. The first generation contains the three elementary particles ordinary matter is composed of: up (u) and down (d) quark build up protons (uud) and neutrons (udd) that
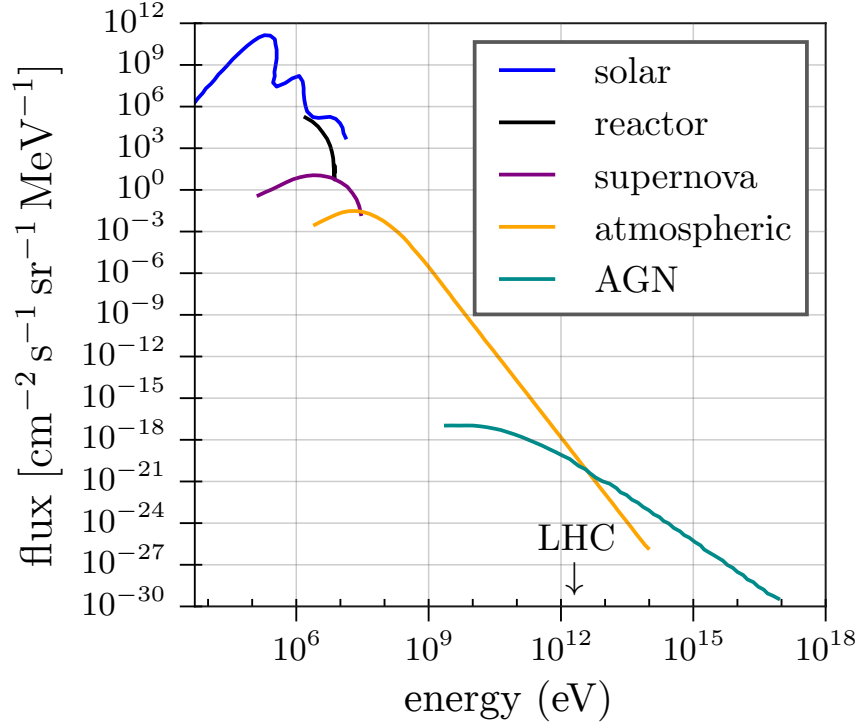
Figure 1.2: Neutrino flux per energy from different sources at Earth. Modified from [11].

form the positively charged nucleus of an atom, held together by the strong force. Negatively charged electrons can populate different energy levels of the atom's shell. Contrary to the electron, the electron neutrino is no stable matter constituent but is involved in particle decays and fusion processes. The leptons introduced in the other generations are the muon neutrino, first detected in 1962 [9], along with the muon and the tau neutrino, rather recently discovered in 2000 [10], with the tau. In experiments searching for neutrinos, the fact that different flavors generally have different production probabilities, interaction cross sections and event signatures can be exploited to distinguish them.

1.2 Sources of neutrinos

Neutrinos are created in a variety of different scenarios of which some will be discussed in the following, going from higher to lower energies. An overview of the fluxes for different neutrino energies spanning sub-MeV to several PeV is shown in Fig. 1.2. Objects capable of accelerating particles to such extreme energies as PeV (three orders of magnitude higher than the energies reached at the world's largest particle accelerator LHC [12]) are believed to be active galactic nuclei (AGN) [13]. The model behind this proposes a jet emerging from a supermassive black hole, such as in the

center of a galaxy, in which charged particles are accelerated in so-called Fermi acceleration [14]. Here, charged particles are repeatedly accelerated by magnetic interaction with traversing shock waves radiating from the objects themselves. With the accelerated primary particles interacting by producing light mesons, neutrinos are created in the subsequent decays, yielding the flux of the highest energetic neutrinos. Further candidates for sources of high-energy neutrinos are gamma ray bursts and supernova remnants. [15] provides a comprehensive overview. So far, the IceCube experiment has successfully measured the flux of astrophysical neutrinos [16] with indications of individual sources in the form of a flaring blazar [17] and a tidal eruption event [18]. Detecting *cosmic neutrinos* has established itself as an integral part of multi-messenger astronomy, next to observations in various frequencies of the electromagnetic spectrum, cosmic rays and its most recent addition of the first detection of gravitational waves in 2015 [19] and, in particular, the observation of merging neutron stars in 2017 [20].

Accelerated primary particles (protons, lighter and heavier nuclei) can also reach Earth directly. These cosmic rays will interact in the atmosphere with nuclei (N) and create particle showers [21]. Hadronic interactions produce light mesons such as pions (π) and kaons (K) and other hadronic components (X), which decay into leptons (e, μ), including neutrinos (ν):

$$CR + N \rightarrow X + \pi^{\pm}, K^{\pm}$$
$$\pi^{\pm}, K^{\pm} \rightarrow \mu^{\pm} + \overset{(-)}{\nu}_{\mu} \tag{1.1}$$
$$\mu^{\pm} \rightarrow e^{\pm} + \overset{(-)}{\nu}_{e} + \overset{(-)}{\nu}_{\mu}.$$

Thus, cosmic rays induce *atmospheric neutrinos*. The neutrino energies range from up to several TeV down to sub-GeV.

Their flux has been studied by various groups, most notably Honda et al. [22], whose data are also used in the KM3NeT experiment. The neutrino flux from all flavors possible at creation is shown as a function of the energy in Fig. 1.3, left. The y-axis in this case is multiplied by E^3, indicating that in the shown energy region the spectrum follows a power law with spectral index of roughly three. As expected from Eq. 1.1, the flux of $\overset{(-)}{\nu}_{\mu}$ is larger than $\overset{(-)}{\nu}_{e}$ and exhibits a slightly harder spectrum. In particular, the $\overset{(-)}{\nu}_{e}$ flux softens for higher energies, as the muons they are created from have less time to decay before reaching the surface due to time dilation. Additionally, for small muon energies, energy loss processes (bremsstrahlung, pair production) become more prominent, creating more electron (anti)neutrinos. This is why the fraction of

$$\frac{\nu_{\mu} + \bar{\nu}_{\mu}}{\nu_{e} + \bar{\nu}_{e}}, \tag{1.2}$$

which is plotted in Fig. 1.3, right, becomes smaller towards lower energies.
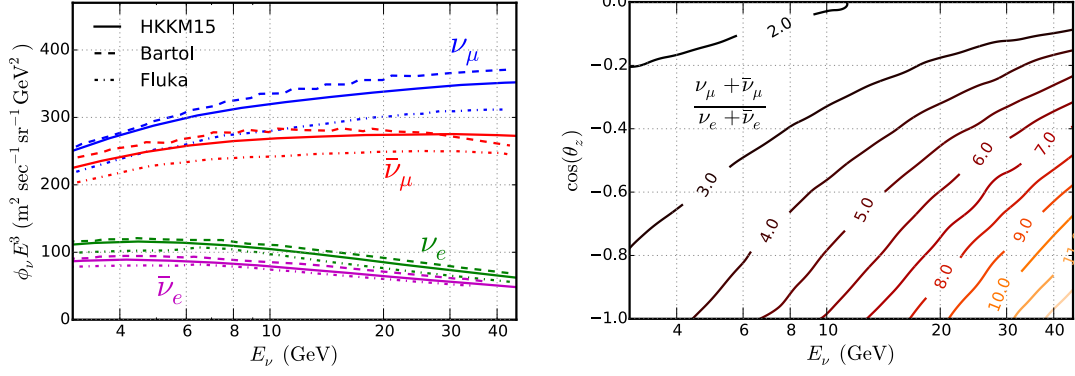
Figure 1.3: Left: Flux of atmospheric neutrinos for different flavors (colors) versus the energy as predicted by different models. Right: Contours of the same fraction of muon and electron neutrino flavors for different zenith angles ($\cos\theta = 0$ means horizontal, $\cos\theta = -1$ is vertically downgoing) and energies. Both from [23].
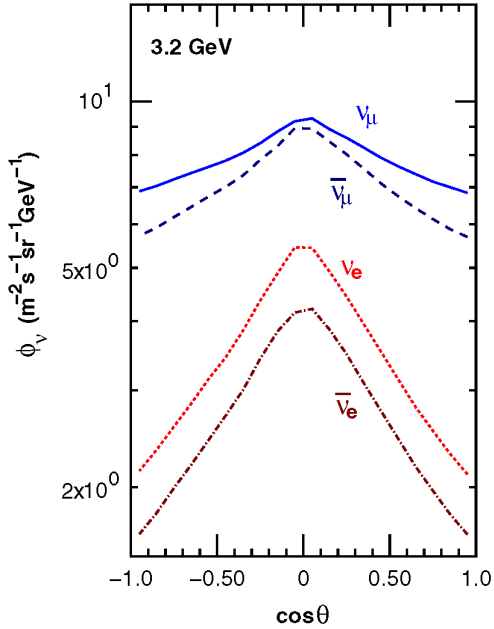
Figure 1.4: Flux of atmospheric neutrinos per flavor as a function of the zenith angle. This is assuming the Honda flux model and specific for the Kamioka mine, yet very similar to what is expected for the KM3NeT/ORCA site where the data for this thesis are taken (differences due to atmospheric density and magnetic field). From [24].

At the same time, a strong dependence on the zenith angle can be observed in the ratio plot. The zenith dependence alone is drawn in Fig. 1.4 for neutrinos with $3.2\,\text{GeV}$. A distinct peak for horizontal directions ($\cos\theta = 0$) for both flavors can be observed. This can be explained by the longer distance the muons travel through the atmosphere under these angles, which gives them more time to decay. Contrary to that, for the edges of the spectrum, the flux is lower, as for vertical events the path through the atmosphere is minimal. The muon flavor is less affected by the decrease, as a portion of muon (anti)neutrinos is in any case created in the pion and kaon decay in the upper atmosphere. The shorter path through the atmosphere thus causes the ratio of muon to electron neutrino flavors towards $\cos\theta = 0$.

On average, antineutrinos are created in smaller quantities because of the

large fraction of positively charged protons in the cosmic ray composition. They generate more π^+ which result in more ν_μ and ν_e.

It should be noted that, as can be seen in Fig. 1.3, left, different models differ in absolute normalization by up to 15%, while the shapes of the distributions are generally similar [25]. This makes it one of the major contributions to the uncertainty of oscillation research.

In addition to natural sources, neutrinos are also created in accelerators to study their properties in a more controlled environment. Typical energies range from hundred MeV to a few GeV [26, 27]. To this end, light mesons are created in hadronic collisions that rapidly decay into muons and neutrinos. While the muons are stopped, the neutrinos can be measured in a far detector. There exist, for example, plans to install such a setup consisting of an accelerator located at Protvino, near Moscow, and KM3NeT/ORCA as the far detector [28].

Neutrinos with energies in the MeV region are believed to play an important role in core collapse supernovae [29]. They are created during the formation of neutrons via

$$p + e^- \rightarrow n + \nu_e \tag{1.3}$$

in the core and carry away a significant fraction of the energy released in the process. For large-scale water Cherenkov detectors like KM3NeT, the energies of *supernova neutrinos* are too low to be detected as reconstructable events. Instead, these neutrinos can be identified by an increase of local coincidences in the background rate [30].

Other contributions to the MeV neutrino flux come from reactor and solar neutrinos. *Reactor neutrinos* are predominantly electron antineutrinos created in the β^- decay of nuclear fission products [31]:

$$n \rightarrow p + e^- + \bar{\nu}_e. \tag{1.4}$$

Solar neutrinos, on the other hand, are electron neutrinos mainly from hydrogen burning in the Sun [32]:

$$4\,{}^1_1\mathrm{H} \rightarrow\ {}^4_2\mathrm{He}\ +\ 2\,e^+\ +\ 2\,\nu_e\ +\ 26{,}7\,\mathrm{MeV} \tag{1.5}$$

Analyzing the data from measurements of atmospheric, reactor and solar neutrinos, it became apparent that a discrepancy between the expected and observed neutrino flux cannot be fully explained by Standard Model processes alone (the "solar neutrino problem") [33].

1.3 Neutrino oscillations

Delivering an elegant solution to the solar neutrino problem is considering the possibility that neutrinos can change their flavor through propagation

in space. This phenomenon is called neutrino oscillation and is practically unique to neutrinos due to their particular properties: The introduced flavor eigenstates ν_e, ν_μ and ν_τ are composed of mass eigenstates ν_1, ν_2 and ν_3 with different masses. The following considerations and calculations (until Sec. 1.3.2) are based on Giunti's exhaustive book "Fundamentals of Neutrino Physics and Astrophysics" [34] and some of the chosen structure is inspired by [35] and [36]. In particular, this means that neutrinos do have masses, thus require extensions to the Standard Model. The phenomenological explanation of a periodically changing probability to observe a certain flavor can be directly derived from the unevenness of the mass eigenstates. Assuming the quantum mechanical system of superimposing mass eigenstates is created in coherence and keeps the phase relation during propagation, different masses advance differently in space, causing changes to the mixing that constitutes the flavor over time. In this picture, each mass eigenstate can be described by plane wave solutions. Measuring the flavor over macroscopic distances allows making this quantum mechanical effect visible. This, however, is only possible due to the large coherence length neutrinos exhibit because of their small mass and the resulting small mass differences of the eigenstates. In conclusion, the flavor eigenstate is involved in production (decay) and interaction, adhering to the lepton family conservation law (at least in Standard Model processes), while the mass eigenstate is involved in propagation, allowing for oscillation.

Expressing this in a mathematical formalism, the flavor eigenstates $|\nu_\alpha\rangle$ can be represented by the mass eigenstates $|\nu_k\rangle$ via the transformation by a unitary matrix U, often called PMNS matrix (after Pontecorvo, Maki, Nakagawa and Sakata) [37],

$$|\nu_\alpha\rangle = \sum_i U_{\alpha k}^* |\nu_k\rangle, \tag{1.6}$$

$$|\nu_k\rangle = \sum_\alpha U_{\alpha k} |\nu_\alpha\rangle. \tag{1.7}$$

The matrix U contains the mixing angles $\theta_{i,j}$ and a complex phase δ in the case of CP-violation for neutrinos [37]:

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{1.8}$$

where $c_{ij} = \cos(\theta_{ij})$ and $s_{ij} = \sin(\theta_{ij})$. If neutrinos are their own anti particle (Majorana, rather than Dirac), additional phases enter the matrix. In any case, they do not have an effect on the oscillation properties.

1.3.1 Neutrino oscillation in vacuum

Since the mass eigenstates can be interpreted as a free particle moving through space, their time evolution can be described by the Schrödinger equation

$$i\frac{d}{dt}\left|\nu_k(t)\right\rangle = H\left|\nu_k\right\rangle, \quad \text{with} \quad \left|\nu_k(0)\right\rangle = \left|\nu_k\right\rangle, \tag{1.9}$$

and the system's Hamiltonian H. More specifically, the equation has solutions in the form of plane waves:

$$\left|\nu_k(t)\right\rangle = e^{-i(E_k t - \vec{p}_k \cdot \vec{x})}\left|\nu_k(0)\right\rangle. \tag{1.10}$$

Here, E_k is the energy of the mass eigenstate, $\vec{p}_k$ the three-dimensional momentum vector and $\vec{x}$ the neutrinos position. As usual in particle physics, the constants c and $\hbar$ are omitted in equations except where noted. In the ultrarelativistic limit ($|\vec{p}_i| = p_i \gg m_i$) the energy can be approximated by

$$E_k = \sqrt{p_k^2 + m_k^2} \simeq p_k + \frac{m_k^2}{2p_k} \approx E_\nu + \frac{m_k^2}{2E_\nu}, \tag{1.11}$$

with E_ν being the total energy of the particle, as the mass contribution to the overall energy becomes small, $E_k \to E_\nu$. This limit can be safely assumed as the masses of the neutrinos are of $\mathcal{O}(\text{eV})$ and the energies studied are larger than $\mathcal{O}(\text{MeV})$, already yielding a large Lorentz factor. For the same reason, $t \approx L$, the path length the particle has traveled in natural units, causing the phase of Eq. 1.11 to simplify to

$$\left|\nu_k(L)\right\rangle = e^{-i\frac{m_k^2 L}{2E_\nu}}\left|\nu_k(0)\right\rangle. \tag{1.12}$$

From this equation, the statement about different mass eigenstates propagating through space differently becomes clear, as the frequency of the solution depends on the mass of the eigenstate m_k. Oscillation of the flavor eigenstate directly follows from the fact that a mixture of the mass eigenstates, each with their own frequency (mass), interferes with each other.

The probability of observing a neutrino that is produced with flavor ν_α exhibiting the flavor ν_β at a distance of L can be computed with the square of the transition amplitude [34]:

$$P_{\alpha\to\beta}(L) = \left|\langle\nu_\beta(L)|\nu_\alpha\rangle\right|^2 = \left|\sum_i U_{\alpha k}^* U_{\beta k} e^{-i\frac{m_k^2 L}{2E_\nu}}\right|^2 \tag{1.13}$$

which becomes

$$P_{\alpha\to\beta}(L) = \delta_{\alpha\beta} - 4\sum_{k>j}\mathrm{Re}\left(U_{\alpha k}^* U_{\beta k} U_{\alpha j} U_{\beta j}^*\right)\sin^2\left(\frac{\Delta m_{kj}^2 L}{4E_\nu}\right)$$

$$+ 2\sum_{k>j}\mathrm{Im}\left(U_{\alpha k}^* U_{\beta k} U_{\alpha j} U_{\beta j}^*\right)\sin\left(\frac{\Delta m_{kj}^2 L}{2E_\nu}\right) \tag{1.14}$$
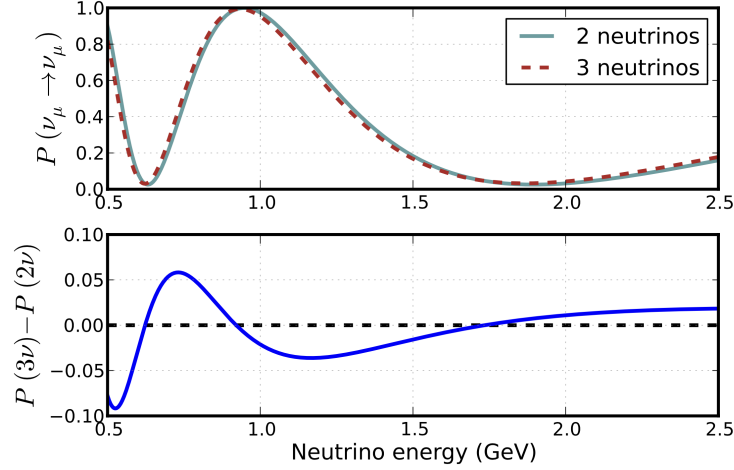
Figure 1.5: Survival probability for ν_{mu} after $L = 1000\,\mathrm{km}$ in vacuum. Calculated using two- and three-neutrino formalisms and best fit values (top) with the difference between the two cases on the bottom. From [36], p 25.

with $\Delta m_{kj}^2 = m_k^2 - m_j^2$ when writing out real and imaginary parts. The sign in front of the imaginary part distinguishes between neutrinos (+) and antineutrinos (-).

From this representation, the influences of the variables on the transition probability become clear: The mixing matrix U in front of the sine determines the amplitude of the oscillation and the squared mass difference, the energy of the neutrino and its traveled distance inside the sine together determine the oscillation frequency.

Considering this theory and the fact that neutrino oscillations have been observed [38, 39], directly proves that neutrinos have to have mass, as two mass-squared differences are measured. With the absolute values still being scrutinized by experiments such as KATRIN [40], at least two masses must be different from zero.

To be able to give the mass differences in convenient units of eV^2, the phase of Eq. 1.14 can re rearranged as

$$\frac{\Delta m^2\,c^3\,L}{4\hbar E} = \frac{\mathrm{GeV\,fm}}{4\hbar c}\frac{\Delta m^2}{\mathrm{eV}^2}\frac{L}{\mathrm{km}}\frac{\mathrm{GeV}}{E} \approx 1.27\frac{\Delta m^2}{\mathrm{eV}^2}\frac{L}{\mathrm{km}}\frac{\mathrm{GeV}}{E}, \qquad (1.15)$$

considering c and $\hbar$ again. In the same way, the distance L can now be given in km and the energy in GeV.

Two neutrino flavor description

In practice, many processes can be sufficiently described by assuming two neutrino flavors that govern transition probabilities, as is exemplified by the muon neutrino survivability in Fig. 1.5. In this case, there is only one
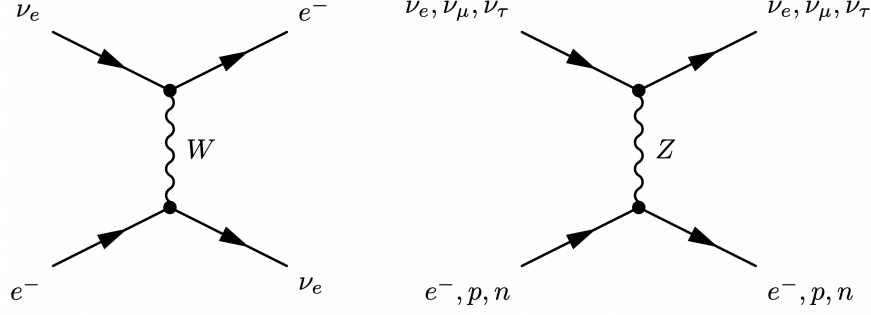
Figure 1.6: Feynman diagrams of all possible coherent forward elastic scattering processes from neutrinos in normal matter. From [34], p. 324.

mass splitting $\Delta m^2 = m_2^2 - m_1^2$ and the mixing matrix simplifies to a 2×2 matrix with a single mixing angle θ and no imaginary phase [34]:

$$U_{2\nu} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{1.16}$$

With this, the flavor transition probability from Eq. 1.13 turns into

$$P_{\alpha \to \beta, \alpha \neq \beta}^{2\nu} = \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \text{ (natural units).} \tag{1.17}$$

From this, the dependence of the oscillation frequency on the mixing angle is directly visible. In particular, $\theta = \pi/4 = 45°$ is called "maximal mixing" in which full transitions between different neutrino flavors can be realized.

1.3.2 Neutrino oscillation in matter

So far, neutrino oscillations in vacuum were discussed where the particles travel unperturbed through space. However, for neutrino experiments on Earth, matter effects play an important role in the transition probabilities. Earth's matter presents a large number of potential targets to interact with, which creates an effective potential that neutrinos can scatter off coherently [41]. For this potential, only the charged current interactions of electron neutrinos with electrons need to be considered. The other possible interactions are neutral current contributions from all neutrino flavors scattering off any of the atom's constituents (see Fig. 1.6). However, the effects from electron and proton scattering cancel out, as the coupling strength is the same, but the sign is different for positively and negatively charged targets. The total amount of electrons and protons is the same in uncharged matter. In the same way, neutral coupling to neutrons is equivalent for all neutrino flavors, causing no neutral current contributions to the oscillation phase [42]. The effective potential V_{CC} in this case (as

derived in [34], p. 324) is thus dependent on the number of electrons encountered, n_e,

$$V_{CC}(n_e) = \pm\sqrt{2}G_F n_e. \tag{1.18}$$

G_F is the Fermi constant describing the strength of the weak coupling. The sign is positive for neutrinos and negative for antineutrinos, respectively.

For describing the complete evolution of the neutrino state, the Hamiltonian in Eq. 1.9 needs to be expanded by a matter contribution H_M:

$$H = H_0 + H_M, \tag{1.19}$$

where H_0 describes the propagation in vacuum. The eigenvalue of H_M is the effective potential introduced in 1.18, as only electron neutrinos are affected:

$$H_M \left|\nu_e\right\rangle = V_{CC} \left|\nu_e\right\rangle. \tag{1.20}$$

From this, an effective mass-squared difference Δm_M^2 and an expression for the effective mixing angle θ_M in the two flavor approximation can be calculated [34]:

$$\Delta m_M^2 = \sqrt{(\Delta m^2 \cos(2\theta) - A)^2 + (\Delta m^2 \sin(2\theta))^2} \tag{1.21}$$

with $A = 2E_\nu V_{CC}$ and

$$\tan(2\theta_M) = \frac{\tan(2\theta)}{1 - \frac{A}{\Delta m^2 \cos(2\theta)}}. \tag{1.22}$$

The complete three-flavor case follows the same principle, yielding two effective mass splittings and modified mixing angles, see for example [43].

Special cases for the potential can be considered that allow to isolate effects and scrutinize certain parameters:

- Vacuum oscillations should be recovered for disappearing electron densities, $n_e \to 0$. This is indeed the case, as $\theta_M \to \theta$ in Eq. 1.22.
- For $A = \Delta m^2 \cos 2\theta$ a resonance is found. Here, the mixing approaches $\pi/4$ and is maximal. This effect is called *resonance enhancement* and more specifically for media with slowly changing densities, it is known as the *MSW effect* (after Mikheyev, Smirnov and Wolfenstein) [44]. The enhancement implies that even for small mixing in vacuum oscillation in matter can become maximal. For the resonance condition, the signs of A and $\Delta m^2 \cos 2\theta$ need to be equal. In particular, the sign of A must be positive for neutrinos and negative for antineutrinos to resonate. This fact offers the opportunity to test neutrinos and antineutrinos separately for resonance and thus scrutinize the sign of Δm^2. From observations of solar neutrinos in the SNO experiment where the ν_e and total neutrino flux were measured, the sign of Δm_{21}^2 was determined to be positive [45].
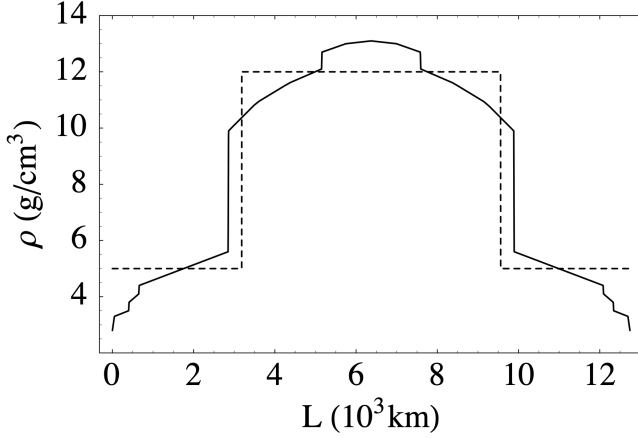
Figure 1.7: Density profile of Earth (solid line). A rough parameterization (dashed line), also referred to as "caste wall", can be used to approximate the changes encountered by traversing neutrinos through the core. From [47].

- For $A \gg \Delta m^2 \cos 2\theta$ matter effects take over and suppress oscillations. The mixing angle approaches $\pi/2$ in a saturation effect.

Finally, there exists yet another way flavor transitions in matter are enhanced compared to vacuum. In the *parametric enhancement*, periodically changing densities can increase oscillation probabilities for neutrinos with certain energies [46]. The idea is that for different media the matter potential, which influences the effective mixing angles and mass-squared differences, changes at just the right frequency to steadily increase the transition probability. In the specific case of Earth, neutrinos experience different densities while traversing the mantle, the core and the mantle again (compare Fig. 1.7). Each time the oscillation probability increases until reaching the transition of densities. Instead of decreasing in the former medium, the now modified oscillation properties cause further enhancement in the next medium. This effect is maximal for distances of half of the oscillation length (see Eq. 1.15, but with the effective mass-squared difference Δm_M). Again, the effect is sensitive to the sign of Δm^2. Peculiarly, this condition is met for neutrinos passing through Earth on the GeV scale, allowing experiments to scrutinize oscillation parameters better.

1.3.3 Neutrino mass ordering

As briefly mentioned, the sign of the mass-squared difference Δm_{21}^2 has already been measured to be positive by exploiting the matter effects in the sun. In the case of three mass eigenstates only one further difference needs to be fixed to determine the complete system. However, so far only the absolute value is known from the study of atmospheric neutrinos [38] leaving two options for the sign that are called *normal* (NO) and *inverted ordering* (IO). In Fig. 1.8 the scenarios are illustrated; Mass eigenstates m_1 and m_2 are chosen to exhibit the small Δm^2 and m_3 has either a larger mass in NO or a smaller mass in IO.

Clarifying the neutrino mass ordering is essential for modeling the origin
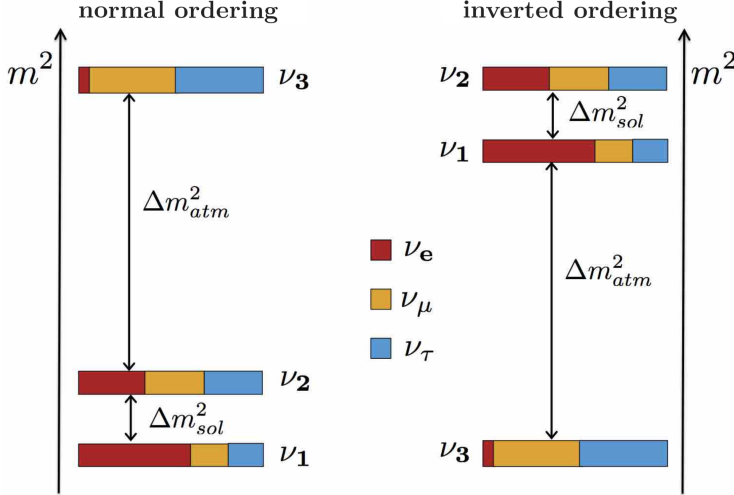
Figure 1.8: Schematic drawing of mass-squared differences between the neutrino mass eigenstates ν_1, ν_2 and ν_3 in normal and inverted ordering. The colors indicate the fractions of flavor contributions to each mass eigenstate. Modified from [48], p. 50.

of mass for leptons as well as the understanding of the mass spectrum observed in charged quarks and leptons [49]. Particularly in the neutrino sector, it can help constrain the models of experiments searching for CP-violating processes, the absolute neutrino masses and the Dirac or Majorana nature of neutrinos. It additionally has implications for astrophysics, for example for the neutrino flavor conversion in supernovae [50]. Lastly, it affects the oscillation parameters themselves, as will be seen in Sec. 1.3.4. Nevertheless, the impact is small, so that it is challenging to infer the mass ordering from a precise measurement of the mixing angles alone. Instead, a separation of neutrino and antineutrino events can help with the determination, as the matter resonance effects are sensitive to the particle type as discussed in Sec. 1.3.2. However, for large-scale neutrino experiments such as KM3NeT/ORCA, the signature of particles and antiparticles are indistinguishable. The lever arm on the neutrino mass ordering comes from the fact that the interaction cross section for antineutrinos is roughly half as large as for neutrinos. This way, a total expected flux that contains uneven contributions from neutrinos and antineutrinos for normal and inverted ordering can be probed. To this end, the event counts N in bins of energy and zenith angle comparing both hypotheses can be evaluated:

$$A = \frac{N_{IO} - N_{NO}}{N_{NO}}. \tag{1.23}$$

This asymmetry A is shown for electron and muon neutrinos in Fig. 1.9.

In this plot, realistic detector resolutions in θ and energy have been considered. It illustrates that accurate reconstruction of these quantities is needed to resolve the structures and thus achieve the desired sensitivity to the mass ordering. The pattern for electron neutrinos on the left side defines large continuous regions to distinguish NO and IO up to 15 GeV, particularly for upgoing events. Finer structures are found for muon neutrinos with
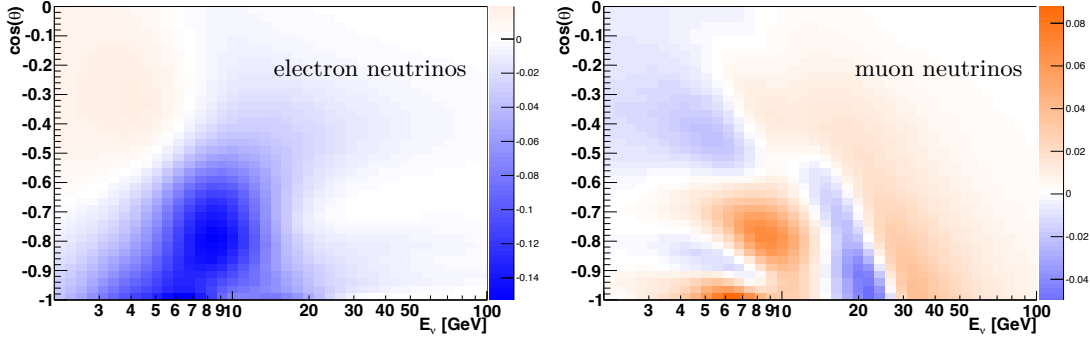
Figure 1.9: Asymmetry between expected number of neutrino events ($\nu + \bar{\nu}$) between NO and IO for bins in the zenith angle and energy of the neutrino. Left side for electron, right side for muon neutrinos. The angle and energy are smeared with a realistic detector resolution. From [48], p. 56.

similar overall discrimination strength. Larger flux and higher angular resolution give this channel the higher discovery potential [51].

1.3.4 Current status of neutrino oscillation experiments

Numerous experiments over the past decades have been involved in measuring neutrino oscillation using various sources and techniques. The results are well described within the three-flavor scheme and the world best fit values are listed in Tab. 1.1 for normal and inverted ordering [53]. The mixing angles θ_{12} and θ_{23} are significantly larger than θ_{13}, while θ_{23} is in the vicinity of maximal mixing. The small θ_{13}, even though non-zero, justifies the often used two-flavor approximation. The mass-squared difference Δm_{21}^2 is small in comparison to any of the other possible combinations Δm_{31}^2 and Δm_{32}^2, which are, in comparison, almost identical. Because of that, the latter are often referred to simply as Δm_{3l}^2 for "large".

The contributions from individual experiments to the determination of these values are listed in Tab. 1.2. The large variety of experiments illustrates the complexity of creating a complete picture of neutrino oscillations as well as the need for global cooperation in order to mount the challenge of different parameters only revealing themselves in very particular scenarios. Mixing angle θ_{12} and Δm_{12}^2, for instance, can only be studied in long baseline (LBL) experiments for which neutrinos are produced either by the Sun or reactors. Typical values for L/E are in the order of $16 \, \text{km/MeV}$ [36]. Atmospheric and also LBL accelerator neutrinos allow for investigating the large mass-splitting and θ_{23} where most significance comes from the disappearance of muon (anti)neutrinos, operating at $L/E \sim 500 \, \text{km/GeV}$. The small mixing angle θ_{13} can be best probed by medium baseline (MBL) reactor neutrinos at $L/E \sim 1 \, \text{km/MeV}$.

The CP-violating phase is barely constrained, still allowing a large parameter space in the $3\,\sigma$ interval. In the same way, no decisive evidence

Table 1.1: Oscillation parameters for the three-flavor scheme as fitted to global data for normal and inverted ordering. The values are from the NuFIT web page [52] and updated in October 2021 from their last publication [53].

	Normal Ordering (best fit)		Inverted Ordering ($\Delta\chi^2 = 7.0$)	
	bfp $\pm 1\sigma$	3σ range	bfp $\pm 1\sigma$	3σ range
$\sin^2\theta_{12}$	$0.304^{+0.012}_{-0.012}$	$0.269 \to 0.343$	$0.304^{+0.013}_{-0.012}$	$0.269 \to 0.343$
$\theta_{12}/°$	$33.45^{+0.77}_{-0.75}$	$31.27 \to 35.87$	$33.45^{+0.78}_{-0.75}$	$31.27 \to 35.87$
$\sin^2\theta_{23}$	$0.450^{+0.019}_{-0.016}$	$0.408 \to 0.603$	$0.570^{+0.016}_{-0.022}$	$0.410 \to 0.613$
$\theta_{23}/°$	$42.1^{+1.1}_{-0.9}$	$39.7 \to 50.9$	$49.0^{+0.9}_{-1.3}$	$39.8 \to 51.6$
$\sin^2\theta_{13}$	$0.02246^{+0.00062}_{-0.00062}$	$0.02060 \to 0.02435$	$0.02241^{+0.00074}_{-0.00062}$	$0.02055 \to 0.02457$
$\theta_{13}/°$	$8.62^{+0.12}_{-0.12}$	$8.25 \to 8.98$	$8.61^{+0.14}_{-0.12}$	$8.24 \to 9.02$
$\delta_{\mathrm{CP}}/°$	230^{+36}_{-25}	$144 \to 350$	278^{+22}_{-30}	$194 \to 345$
$\dfrac{\Delta m^2_{21}}{10^{-5}\ \mathrm{eV}^2}$	$7.42^{+0.21}_{-0.20}$	$6.82 \to 8.04$	$7.42^{+0.21}_{-0.20}$	$6.82 \to 8.04$
$\dfrac{\Delta m^2_{3\ell}}{10^{-3}\ \mathrm{eV}^2}$	$+2.510^{+0.027}_{-0.027}$	$+2.430 \to +2.593$	$-2.490^{+0.026}_{-0.028}$	$-2.574 \to -2.410$

Table 1.2: Detection channels and experiments contributing to the determination of oscillation parameters. The dominant contribution is highlighted. Inspired by [54] and [36], p. 33.

source/channel	parameter	experiment
solar neutrinos $\nu_e \to \nu_\alpha$	$\boldsymbol{\theta_{12}}, \Delta m^2_{21}, \theta_{13}$	Homestake, Chlorine, GALLEX/GNO, SNO, SAGE, Borexino
reactor LBL $\bar{\nu}_e \to \bar{\nu}_e$	$\boldsymbol{\Delta m^2_{21}}, \theta_{12}, \theta_{13}$	KamLAND
reactor MBL $\bar{\nu}_e \to \bar{\nu}_e$	$\boldsymbol{\theta_{13}}, \Delta m^2_{3l}$	Daya Bay, RENO, Double Chooz
atm. neutrinos $\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_\mu$	$\boldsymbol{\theta_{23}}, \Delta m^2_{3l}, \theta_{13}, \delta_{CP}$	Super-Kamiokande, IceCube/DeepCore
accelerator LBL $\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_\mu$ $\nu_\mu \to \nu_e$	$|\boldsymbol{\Delta m^2_{3l}}|, \boldsymbol{\theta_{23}}$ $\boldsymbol{\delta_{CP}}, \boldsymbol{\theta_{13}}, \theta_{23}, \mathrm{sign}(|\Delta m^2_{3l}|)$	Minos, NOνA, T2K
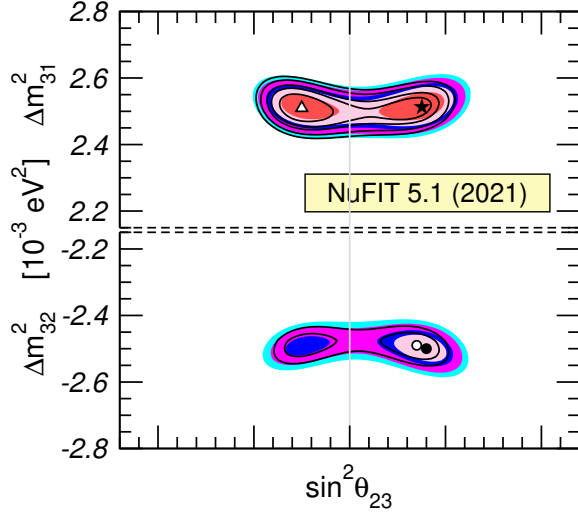
Figure 1.10: Results from fit to global data of Δm_{3l}^2 and $\sin^2 \theta_{23}$ with contours of confidence levels $1\,\sigma$, 90%, $2\,\sigma$, 99% and $3\,\sigma$. For Δm_{31}^2 on the top, normal ordering is considered, while inverted ordering is assumed for Δm_{32}^2 on the bottom. Modified from [52].

is found favoring one of the two possible mass orderings. Especially with recent updates from the T2K and NOνA experiments [55],[56], that prefer IO over the so far favored NO, a $2.7\,\sigma$ preference for NO remains. Yet, this being an active field of contemporary research, results soon to be made available from Super-Kamiokande indicate a lower significance for NO [57], which will then also affect the global fit.

Summarizing the uncertainties with which oscillation parameters are determined currently from Tab. 1.1, the following overview indicates the fields of research with the highest interest:

$$\begin{array}{lll} \theta_{12} : 14\% & \theta_{13} \quad : 8.5\%, & \theta_{23} \;\; : 25\% \\ \Delta m_{21}^2 : 16\% & |\Delta m_{3l}^2| : 6.5\%, & \delta_{CP} : 83\%. \end{array}$$

The relative precision is calculated by $2(x_{\text{up}} - x_{\text{low}})/(x_{\text{up}} + x_{\text{low}})$, where x_{up} (x_{low}) are the upper (lower) 3σ confidence intervals, like done in [53]. As can be seen, next to the mass ordering and δ_{CP}, θ_{23} is currently the least well known mixing angle. When having a closer look at the parameter's allowed regions, like done in Fig. 1.10, it can be observed that two different minima in the phase space emerge. With the similar significance on both sides, the octant for θ_{23} is still unknown. There is only a mild preference for the second octant ($\sin^2 \theta_{23} < 0.5$) in the NuFIT 5.1 data, but maximal mixing ($\sin^2 \theta_{23} = 0.5$) is disfavored with $3.9\,\sigma$. This open question provides the motivation to study θ_{23} in KM3NeT/ORCA data, as will be discussed in Chap. 2.5.1 and results will be presented in Chap. 8.2. Next to Δm_{3l}^2, ORCA will be able to determine the neutrino mass ordering with $4.4\,\sigma$ ($2.3\,\sigma$) after three years of data taking for normal (inverted) ordering with the full detector [58].

Furthermore, sterile neutrinos are an extension to the three-flavor oscillation and an attempt to explain tensions observed in some experiments [59]. They would, however, not directly be created in particle interactions

but merely leave an impact on the oscillation for the three known flavors. The sensitivity of ORCA to sterile neutrinos has been explored in [60].

1.4 Interactions of neutrinos with matter

As briefly mentioned in Sec. 1.1, neutrinos can only interact via the weak nuclear force. This results in an intrinsically small total cross section compared to all other Standard Model particles. For the neutrino energy region of interest for KM3NeT/ORCA and this thesis (GeV scale), the dominant interaction mechanism is the scattering off nucleons. For a detailed discussion see [35] and find the complete, energy-dependent description in [61].

In general, neutrinos, as well as antineutrinos, can undergo a neutral current (NC) or a charged current (CC) interactions:

$$\text{NC}: \overset{\scriptscriptstyle(-)}{\nu}_l + N \xrightarrow{\text{Z}} \overset{\scriptscriptstyle(-)}{\nu}_l + \text{hadr.} \tag{1.24}$$

$$\text{CC}: \overset{\scriptscriptstyle(-)}{\nu}_l + N \xrightarrow{\text{W}} l^{\pm} + \text{hadr.} \tag{1.25}$$

Here, N stands for a nucleon (neutron or proton) and "hadr" for a hadronic shower.

In the NC interaction, the mediator is the electromagnetically neutral Z^0 boson, which does not allow the exchange of weak hypercharge Y. Analogously to the electric charge for their corresponding interaction, the weak hypercharge denotes how the weak force acts on a particle. It is defined as

$$Y = 2(Q - T_3), \tag{1.26}$$

where Q is the electric charge and T_3 the third component of the weak isospin, a fundamental property assigned to every elementary particle (see Fig. 1.1).

In the case of a CC interaction, the $W^{\pm}$ does indeed transfer hypercharge, enabling a change in the weak isospin and thus in the observable flavor of the particle.

Due to helicity (the projection of the spin of a particle to its direction) constraints, the cross section for CC interactions of antineutrinos is three times smaller than for neutrinos [61]. This is manifested in Fig. 1.11, where the total cross section for (anti)neutrino scattering off a nucleon is shown. Also illustrated are the single contributions from different interaction mechanisms, namely *quasi-elastic scattering* (QE), *resonance production* (RES) and *deep inelastic scattering* (DIS).

- Starting at energies below $1\,\text{GeV}$, quasi-elastic scattering refers to the process of the neutrino elastically scattering off an entire nucleon, ejecting it from the target atom. In the case of NC interactions this is called elastic scattering.
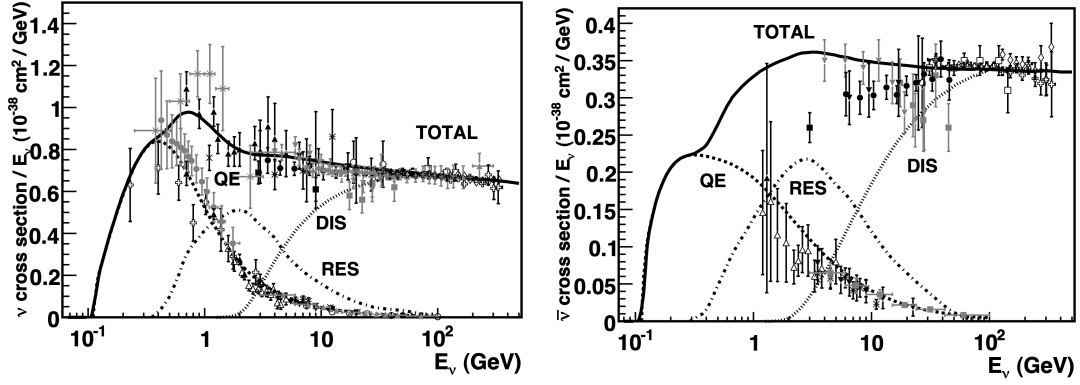
Figure 1.11: Total cross section divided by the energy (solid line) for neutrinos (left) and antineutrinos (right) scattering off nucleons versus the energy. Shown are also the different contributions (dashed lines). From [61].

- For energies between $1\,\text{GeV}$ and $10\,\text{GeV}$, the dominant contribution comes from resonance production, meaning that the target nucleon is excited to a higher energy state, such as Δ or N^*, that will later decay back emitting light mesons.
- For larger energies, deep inelastic scattering takes over. Here, the inner structure of the nucleon, the quarks, can be resolved. By this process the nucleon is destroyed and an hadronic shower is created.

As can be concluded from the description of the processes, with increasing energy the inelasticity of the interaction y, defined as

$$y = \frac{E_{\text{hadr}}}{E_\nu}, \tag{1.27}$$

increases. It is also referred to as the *bjorken y* and describes the fraction of the neutrino energy E_ν transferred to the target system E_{hadr}.

1.5 Neutrino detection principles

As discussed so far, the physical nature of the neutrino leads to considerable challenges to register neutrinos or their interactions with significant statistics. Nevertheless, there are several different techniques that have been realized in experiments over the past 55 years.

One possibility is to make use of the fact that neutrinos can induce the inverse beta decay on specific elements yielding daughter nuclides that can be identified by the signature they produce as unstable isotopes. One example is the Homestake experiment [62], which used chlorine-37 as target material, i.e.

$$\nu_e + {}^{37}\text{Cl} \rightarrow \nu_e + {}^{37}\text{Ar} + e^-, \tag{1.28}$$
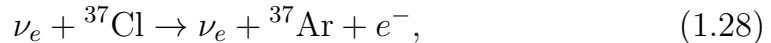
and subsequently looked for radioactivity from the argon-37 created in the neutrino capture. This way, MeV-neutrinos from the Sun were detected
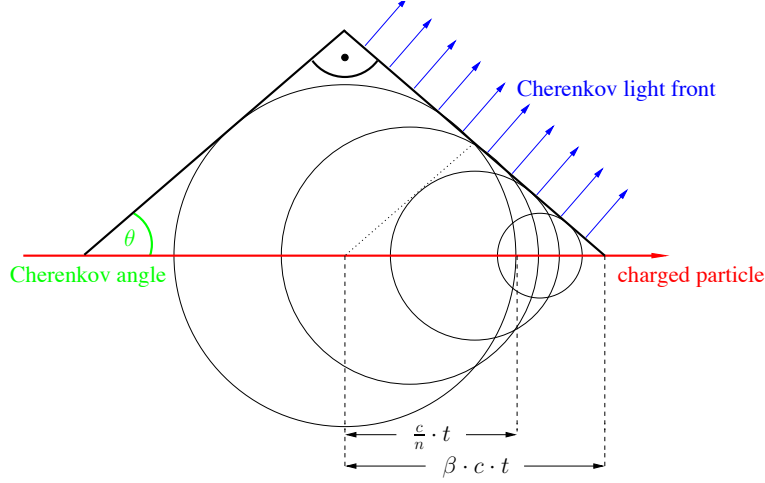
Figure 1.12: Geometry of the propagation of Cherenkov light due to a relativistic charged particle in a dielectric medium. From [67], p. 28.

as early as 1970. The same principle was later realized with gallium-71 in experiments such as SAGE [63], GALLEX [64] and its successor GNO [65] taking advantage of the lower threshold energy needed for the reaction of 200 KeV, compared to the 800 KeV for the chloride, allowing to probe further regions of the sun's neutrino spectrum.

1.5.1 Cherenkov neutrino detection

The other possibility to achieve neutrino detection is to look for the signatures of secondary particles from neutrinos scattering off matter (preceding Sec. 1.4). If the secondary particles are charged and travel through a dielectric medium with a phase velocity greater than the speed of light in that medium they will emit so-called Cherenkov radiation [66]. The light radiates spherically from along the particle's path generating a conical wavefront under a characteristic angle θ_C, depending on the refraction index of the medium n (see Fig. 1.12). For a relativistic velocity $\beta = v/c$ of the particle, the angle reads as

$$\cos(\theta_C) = \frac{1}{\beta n}. \tag{1.29}$$

In the case of relativistic particles ($\beta \approx 1$) and seawater as medium ($n \approx 1.35$), the Cherenkov angle is about 42°. The detector's photo sensors then capture the light and from the hit distributions the properties of the interacting particle can be reconstructed. This technique is utilized in large-scale neutrino telescopes like KM3NeT, whose photo sensors are a few meters to tens of meters apart to cover volumes up to the km^3 scale.

The spectrum of the emission peaks in the ultraviolet and blue wavelengths (see also some discussion about the modeling of this light in Chap. 4.1), producing the characteristic color visible in the cooling water
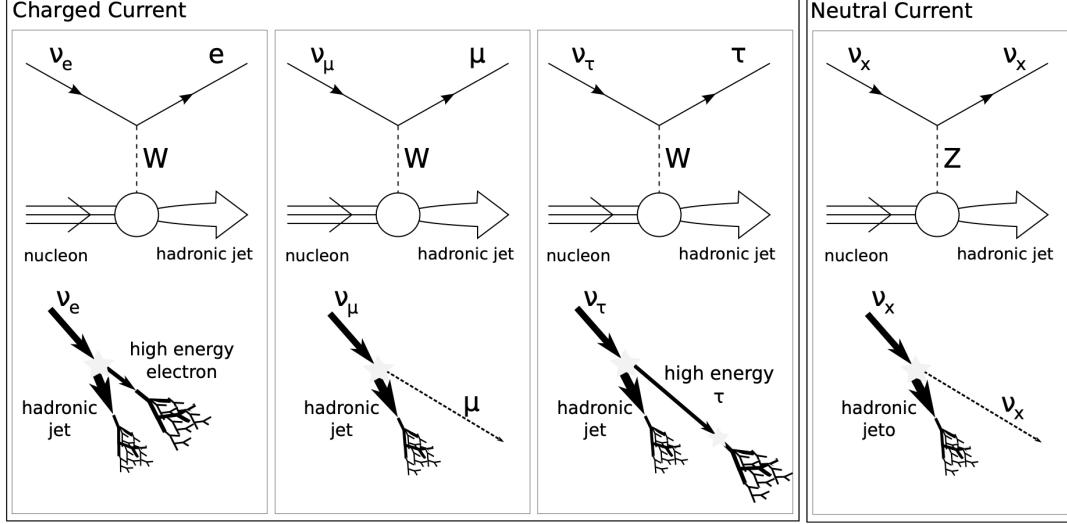
Figure 1.13: Overview of signatures produced in each flavor's charged current and the neutral current interaction. From [70].

of nuclear power plants. There is, however, a lower energy threshold E_{th} for the production of Cherenkov light, as the particle of rest mass m_0 must be faster than $\beta > 1/n$:

$$E_{th} = \frac{m_0}{\sqrt{1 - 1/n^2}}. \tag{1.30}$$

This yields $T_{th}^e \approx 0.25\,\text{MeV}$ for electrons and $T_{th}^\mu \approx 53\,\text{MeV}$ for muons as minimum kinetic energies ($T = E - m_0 c^2$).

For exploring the region below these energies, experiments like KAMland [68] and BOREXINO [69] utilize the inverse beta decay and liquid scintillators to measure the energy of the created positron.

In equations 1.24 and 1.25, it was stated that the neutrino interaction can produce either exclusively a hadronic shower in the detector for the case of a neutral current interaction with the neutrino escaping undetected or have a hadronic part accompanied by a lepton for the charged current, whose flavor depends on the flavor of the incident neutrino. Since electron, muon and tau differ in their mass and, in turn, their stability, the signatures of the Cherenkov light emission are also distinct for each lepton flavor and are thus further addressed in the next section.

1.5.2 Interaction signatures in water Cherenkov detectors

Based on the topology of the light production profile, which, in the end, is the feature that is measured in Cherenkov detectors, two main categories of event signatures can be defined: *showers* and *tracks*. The contributions from each flavor to these classes are illustrated in Fig. 1.13 and will be discussed in the following dedicated sections.

Electromagnetic showers

Electrons with energies in the GeV scale lose the dominant part of their kinetic energy via the process of bremsstrahlung [4], causing the emission of photons while slowing down. These photons react with that medium themselves in the form of pair production and Compton scattering, creating electrons and positrons again. The cascading event structure is called an *electromagnetic shower*. During that process, all electrons and positrons with kinetic energies above their Cherenkov threshold (0.25 MeV) will create visible light. Due to the attenuation length of 36 cm for electrons in water [4], the overall signature appears point-like in a detector with dimensions such as KM3NeT/ORCA. The amount of all generated Cherenkov photons is proportional to the energy of the shower, with fluctuations less than 1% [35]. Thus, the energy can be reconstructed relatively precisely for events fully contained in the detector that effectively functions as a calorimeter.

Due to the spherical propagation of light from the center of the shower, direction reconstruction is more challenging. Yet, the fact that the maximum of the photon emission, even though rather broad, peaks in the direction of the incident neutrino allows exploiting this information when reconstructing the direction.

Hadronic showers

Hadrons, as released in every of the listed neutrino interactions, can form *hadronic showers*. From the initial energy transfer to the nucleon, further hadrons (often light mesons) are created through the strong nuclear force abiding by the law of confinement. Despite having a longer attenuation length of 83 cm in water [35], the dimensions of the hadronic showers are comparable to the electromagnetic showers, as the energy in the interaction is usually transferred to two or more hadrons. Again, all participating hadrons with energies above the Cherenkov threshold emit light. However, the overall light yield is smaller than in electromagnetic cascades for the same energy because the masses of the particles are higher and so are their respective thresholds.

Naturally, the lightest meson, the uncharged pion, is created in large abundances in these processes and decays predominantly into two photons. This adds an electromagnetic component to the shower that increases with the shower's energy. For energies of TeV and PeV, both kinds of showers look similar. For lower energies, however, there is a considerable variation on an event-by-event basis. For energies of 1-20 GeV, as studied in [35], this can be explained by the fact that the number, type and propagation of hadrons in the final state varies. In a Cherenkov detector, only the charged particles generate observable light. Additionally, the deviation of the hadron's emission direction from the neutrino's incident direction is larger ($\mathcal{O}(10°)$).

Muon tracks

Muons, as created directly in the atmosphere, secondary to cosmic rays, or during a muon neutrino charged current interaction, can travel up to kilometers in water, depending on their energy. The dominant process for the energy loss is ionization, while Cherenkov light is created but does not provide significant energy transfer. For energies above $1\,\text{TeV}$, stochastic processes (bremsstrahlung, pair production and photonuclear interactions) play a more important role causing *catastrophic energy losses*, which comprises electromagnetic showers along the track. Towards the few GeV region, muons behave like *minimum ionizing particles*, which exhibit an almost constant energy loss over a straight path until they are stopped or decay. In water, the length of a muon track is about $4\,\text{m/GeV}$ [4].

A detailed description of how to classically reconstruct the direction and energy of such events from these first physics principles will be given in Chap. 4.1.

Features of tau interactions

Tau particles decay much faster than muons ($29\,\text{ns}$ and $2.2\,\text{µs}$, respectively [4]), thus do not travel significantly ($5\,\text{cm/TeV}$) before decaying in large scale Cherenkov detectors for energies in the GeV region. With the branching rations of the weak decays of

$$\tau^- \to \text{hadronic shower} : 64.8\%$$
$$\tau^- \to \nu_\tau + e^- + \bar{\nu}_e : 17.8\%$$
$$\tau^- \to \nu_\tau + \mu^- + \bar{\nu}_\mu : 17.4\%,$$

in most cases two overlapping showers, from the interaction generating the tau and the decaying particles, are expected. Only in roughly 20% a muon is produced, which would additionally create a track emerging from the shower vertex. Because of this, the only way to identify tau neutrino interactions in experiments such as KM3NeT/ORCA is via an excess of observed showers.

2 Detecting neutrinos with the KM3NeT neutrino telescope

K M3NeT is a European research infrastructure designed for the detection of neutrinos in the deep Mediterranean Sea [48]. With that, it is already the second generation of deep-sea Cherenkov detectors, succeeding the ANTARES experiment [71] and heavily benefiting from its experience. Currently, the detector is under construction at two different building sites, both using the same technology but distinct layouts optimized for their respective physics goals.

On the one hand, there is the KM3NeT/ARCA detector (**A**stroparticle **R**esearch with **C**osmics in the **A**byss) located 100 kilometers south-east of Sicily, Italy, in a depth of 3500 m. Covering the high-energy spectrum from a few tens of GeV up to PeV, it is designed to detect neutrinos of cosmic origin, much like the IceCube neutrino telescope at South Pole [72]. To this end, the instrumentation of photo detection units is relatively sparse to cover a large effective volume, optimized for detecting astrophysical neutrinos. Of particular interest is the identification of point sources from the measured flux.

On the other hand, KM3NeT/ORCA (**O**scillation **R**esearch with **C**osmics in the **A**byssis) is optimized for the GeV neutrino energy range with a few GeV as lower detection threshold. This allows for detecting atmospheric neutrinos and studying fundamental neutrino properties, such as the neutrino mass ordering and oscillation parameters. Lower target energies require much denser instrumentation compared to ARCA. The site is located 40 kilometers off-shore of Toulon, south France, at a depth of 2450 m under sea level, in the vicinity of the ANTARES detector.

Inside the experiment and also later for the data analysis related topics, a variety of specific terms and abbreviations is used. To help with that for reading this dissertation, a glossary can be found at the end.
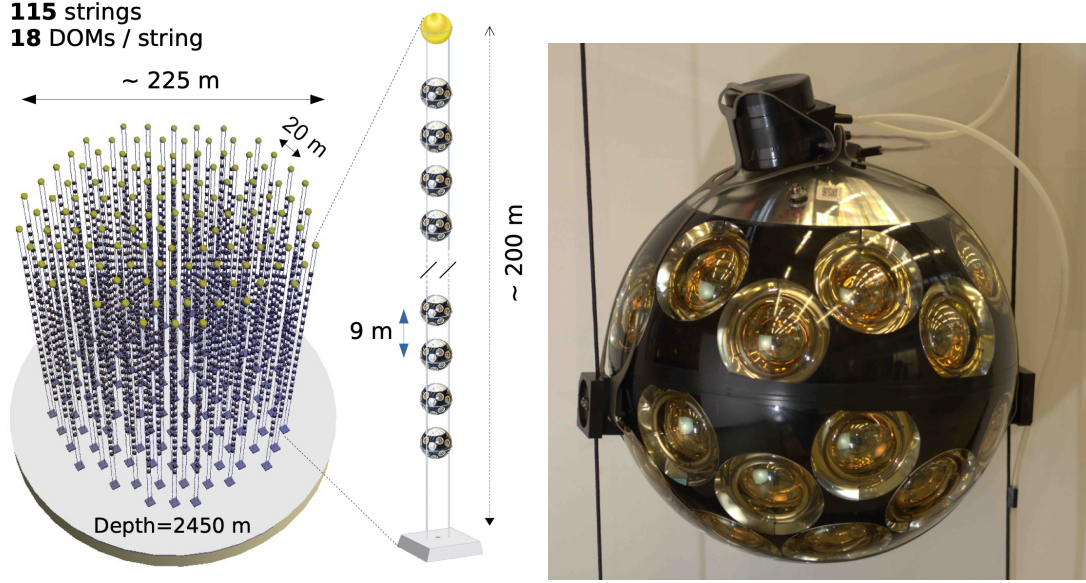
Figure 2.1: Left: The ORCA building block with its final dimensions. Next to it, an illustration of a DU. From [28]. Right: Picture of a DOM as mounted on Dyneema© ropes. From [48].

2.1 The KM3NeT detector sites

Both sites of the KM3NeT detector [48] share the same basic structure and detection principle for recording Cherenkov light produced by secondary particles in neutrino interactions. They are arranged in building blocks of 115 detection units (DUs), which hold the optical sensors in place, provide them with power, as well as manage the data transfer (see Fig. 2.1, left).

For that, two Dyneema© ropes provide the backbone of a DU, between which the optical sensors and additional spacers to maintain shape are mounted. Along the ropes, an electro-optical cable, covered in plastic tubes and filled with oil to balance the pressure, connects each sensor to the power grid and uses optical fibers to transmit the signals to shore. The base of the DU keeps the DU at its destined position and allows for interfacing with the sea infrastructure network, while a buoy at the top provides additional lift to keep the DU straight.

Each DU consists of 18 photo sensors, called Digital Optical Modules (DOMs). The position of the DOM on a DU is also referred to as "floor". The lowest DOM, for example, would be the first floor. A picture of a DOM can be seen in Fig. 2.1, right. In it, 31 3-inch Hamamatsu photomultiplier tubes (PMTs) cover an almost 4π angular acceptance for photon detection. They are housed in a transparent 17 inch glass vessel comprised of 2 hemispheres to withstand the hydrostatic pressure present at the operating depths. The PMTs are mounted on a 3D printed holding structure with an optical gel filling up the space to the glass to improve optical coupling to the PMT's photocathode. The signal at the anode of the PMT is digitized with the

Figure 2.2: Footprint of the *xy*-positions of the ORCA6 (orange) and ORCA4 (green) detectors. Different nomenclature for the position of a DU is used for the two detector configurations.

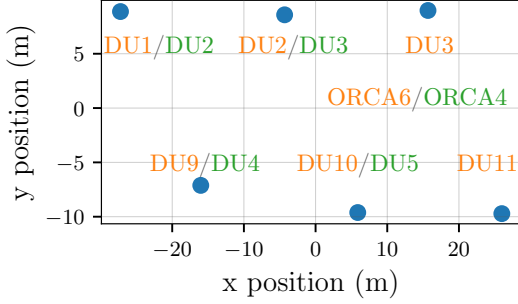main logic board by saving the time of arrival of the first photon, which is defined as the moment the amplitude of the voltage signal surpasses a set threshold (0.3 photoelectrons), and the time-over-threshold (ToT), marking the time until the amplitude falls below the threshold again. For every photon detection, the time and ToT information along with an identifier for the PMT are sent to the shore station, where the triggering of events takes place in real-time.

The main difference between ARCA and ORCA is their respective dimensions. In its final form, ARCA will consist of 2 building blocks with a radius of about $500\,\mathrm{m}$ while maintaining a mean distance between DUs of $90\,\mathrm{m}$, which overall yields about one km^3 of instrumented volume. Here, the DOMs on the DU have a $36\,\mathrm{m}$ vertical spacing giving the detector a height of about $700\,\mathrm{m}$. Currently, 8 DUs are deployed and continuous enlargement is planned in the coming years.

The low-energy counterpart ORCA features one building block with a radius of about $107\,\mathrm{m}$ ($20\,\mathrm{m}$ average distance) and a vertical spacing of $9\,\mathrm{m}$. With a total height of $180\,\mathrm{m}$ this provides 6 megatons of seawater as effective volume. At the moment, 10 DUs are deployed, while also here, steady expansion is scheduled for the next months and years.

The data used for the analysis in this thesis were recorded with the ORCA6 detector, of which a footprint is shown in Fig. 2.2. Some studies on the calibration chapter are based on the predecessor ORCA4, while a change in nomenclature took place at that time, as is indicated in the figure.

2.2 Data acquisition and triggering

With the aforementioned realization of sending all PMT hit data (PMT id, arrival time and ToT) to shore comes a vast amount of information to be processed [48]. This would amount to 25 GB/s for a completed building block when maintaining a 1 ns precision for the time information. For data reduction, filter algorithms scan the data stream for event-like signals online in a dedicated computing facility, rather than saving everything to disk. To this end, three levels of filtered data are defined: L0, L1 and L2. The

two latter ones make use of causal time-position correlations. If a trigger is activated, an *event* is saved.

The L0 trigger level refers to the unfiltered data stream as detected by the PMTs whose amplitude surpasses the threshold value. This type of data is recorded for short amounts of time in order to collect information for calibration and reconstruction purposes: the homogeneous, high-statistics data set provides the basis for the intra-DOM time calibration (see Chap. 3.1) and yields accurate background rates that can be inserted in the event reconstruction to infer the background hypothesis.

L1 triggered data are required to have hits in at least two different PMTs of one DOM within a time window of typically 10 ns. Based on this, an event is defined, which also includes L0 hits 1 μs before and 1.5 μs after the L1 triggered cluster. In a data analysis context, all hits of an event are sometimes referred to as *snapshot hits* or simply *hits* while the subset flagged additionally as L1 is referred to as *triggered hits*.

Lastly, L2 filter algorithms allow for further specialization and identification already at the online data processing level. In principle, two kinds of filters can be distinguished; one for track-like and one for shower-like events. They are discussed in detail in [48] but are not yet relevant in part-size detectors such as those studied in this thesis. Both reconstruction methods, the standard likelihood fit and the later introduced deep learning approach, are based only on L0 and L1 trigger information.

2.3 Optical background sources

For the KM3NeT detectors, three types of background sources concealing neutrino-induced events can be distinguished: radioactive decays in the water (and to a lesser extent the glass vessel), bioluminescence and atmospheric muons. All of these have the potential to trigger at the L1 level. But there are also countermeasures to identify and thus reject them.

2.3.1 Radioactive decays in water

Seawater contains a small fraction of about 0.04% of radioactive Potassium-40 (^{40}K). It mainly decays into Calcium-40 via the beta decay emitting an electron, or in about 10 % via the electron capture into Argon-40 emitting a photon in the subsequent gamma decay. With its maximal energy of 1.31 MeV, the electron can produce Cherenkov radiation. Similarly, the excited ^{40}Ar state has an energy of 1.46 MeV and its gamma ray can scatter via the Compton effect to produce detectable photons from the subsequent electrons. Together, these ^{40}K decays form a constant, isotropic optical background for the detector, which can cause L1 triggers randomly to fire but in turn can also be utilized for calibration purposes. The mean traveling

distance of light from one decay of a few tens of centimeters [73] can induce coincident hits in PMTs of the same DOM, allowing for an intra-DOM time calibration (see 3.1) and PMT detection efficiency evaluations.

2.3.2 Bioluminescence

Bioluminescence is the ability of living organisms to emit visible light for luring prey or attracting mates. There exists an exceptionally distinct variety of these organisms in the deep sea, where no more surface light is present [74]. The event signature produced can be twofold: larger animals, like fish or jellies, can create bioluminescence bursts that persist up to a few seconds and thus emit a large number of photons that travel tens of meters. These bursts are typically induced by physical stress on the animals as they, for example, touch parts of the detector.

In contrast, bioluminescent bacteria are a weaker yet more constant light source. This way, they contribute to the overall background rate, increasing the chances of pure noise events firing a trigger. It can be directly observed in the trigger rates that higher sea current, applying forces on the bacteria, increases their bioluminescent activity. In the same way, season-based changes in the nutrition levels of the deep seawater amplify the activity.

In order to discriminate against contributions from bioluminescence, the so-called high-rate veto (HRV) is implemented in the data acquisition systems. To that end, the hit count of a single PMT per time is monitored and the PMT is excluded from triggering (and also later in reconstruction) if there are more than 2000 hits per $100\,\mu s$. For ORCA, this results in an HRV threshold of $20\,kHz$, while the average background rate is between $5\,kHz$ and $7\,kHz$.

2.3.3 Atmospheric muons

As stated in Eq. 1.1, muons are created in air showers induced by cosmic rays during decays in the upper atmosphere by the same processes as atmospheric neutrinos are. While traversing air, water or rock, they will emit Cherenkov radiation and lose energy until they are eventually stopped altogether. This is the reason why some experiments are located deep below the surface: often, atmospheric muons are a significant background source that is reduced by the shielding the overburden of a detector provides. Still, at a depth of $2450\,m$ the rate of atmospheric muons in ORCA is expected to be $\sim 10^5$ times larger than muons from GeV-scale neutrinos interacting in the detector volume. Since a muon track is detected in both cases, it is difficult to discriminate against the atmospheric muon background for the downgoing direction. In the future, with more extended detectors, it might be worthwhile studying the feasibility of defining veto regions that would shield an active volume to identify muon tracks starting inside the detector.
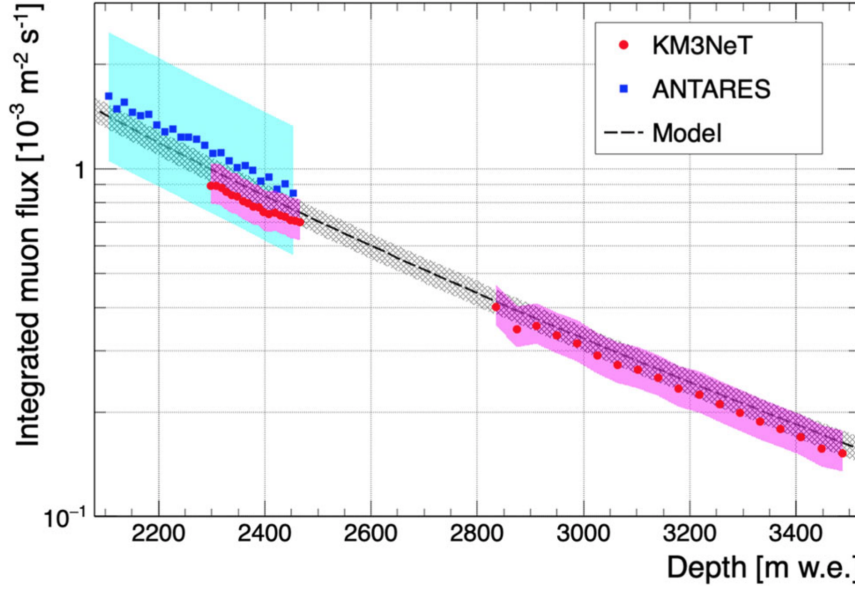
Figure 2.3: Muon flux versus the depth in the sea. ARCA and ORCA contribute with the red points while also ANTARES data in blue is considered. The shaded areas signal the uncertainties. From [75].

Instead, muons cannot traverse the Earth, which enables the search for upgoing neutrinos that interact in the vicinity of the detector. However, there is the possibility that some atmospheric muon events are falsely reconstructed as upgoing. So, to guarantee a robust neutrino identification, both calibration and reconstruction should be as accurate as possible. Improvements in both fields have been studied in the scope of this thesis.

The first physics result based on data from the partially built ARCA and ORCA site shows the depth dependence of the muon flux [75] that is well in agreement with the assumed model and previous data from ANTARES, as can be seen in Fig. 2.3. Due to the considerable height, especially for the ARCA detector, the decreasing number of detected muons is observable already with one and two DUs for ORCA and ARCA, respectively.

2.4 Monte Carlo simulations

Monte Carlo (MC) simulations play an essential part in multiple areas of a physics experiment: from the optimization of different parts and layouts in the planning phase over necessary consistency checks for validating real data to the detailed simulations of the expected physics processes considered in an analysis.

In KM3NeT/ORCA, the MC simulations are done in separate stages, covering the generation, propagation of particles and light, and triggering in the detector. The different components that together constitute the expected data are briefly presented in the following.

The random noise events are solely based on a rate deduced from real data, covering the ^{40}K and the bioluminescence contributions. The hits on the PMTs are directly generated randomly and the trigger algorithm is applied.

Atmospheric muons are directly generated with the program MUPAGE [76], which provides a parameterized description of the expected flux at a given depth, considering the interaction of the primary cosmic ray in the atmosphere. It also includes the possibility of multi-muon events. Starting from these particles (type, energy, direction), the propagation of the muons and the generation and propagation of the induced Cherenkov light are achieved in the subsequent program, which already holds the detector information. Here, the hits are generated taking into account the information about the water (optical properties), the DOM (transmission properties of the glass and gel and their coupling) and the PMTs (angular detection efficiency). For efficient large-scale muon MC productions, probability density functions (PDFs, see Chap. 4.1) for the light propagation are used to speed up the process. Next, the PMT response to the incident photon is simulated, yielding arrival times and ToTs. The random noise background is also added to represent a realistic muon event. Again, for the final detector response, the triggering is applied as well.

Atmospheric neutrinos are generated using the software package gSeaGen [77], based on the widely used GENIE [78] generator. The application developed inside the KM3NeT collaboration is designed to efficiently generate large statistics of neutrino interaction events. The generator propagates the neutrinos and simulates the interaction if it is located in the vicinity of the detector to create the secondary particles. For lower-energetic neutrinos, the particle and light propagations are fully simulated in Geant4-based [79] code. Equivalently to the atmospheric muons, the physics for higher-energetic muon tracks are modeled using PDFs. Based on the hits, events are created with the same trigger algorithms as above.

Due to their relatively low detection rate but high significance for the analysis, neutrino events are typically simulated in much larger statistics and then weighted accordingly to reduce statistical uncertainties.

2.5 Measuring oscillations of atmospheric neutrinos

In order to utilize the full potential of Cherenkov-based neutrino research, the detector has to fulfill a few requirements:

- In order for the Cherenkov light to reach the photo sensors, the medium must be transparent in addition to being dielectric.
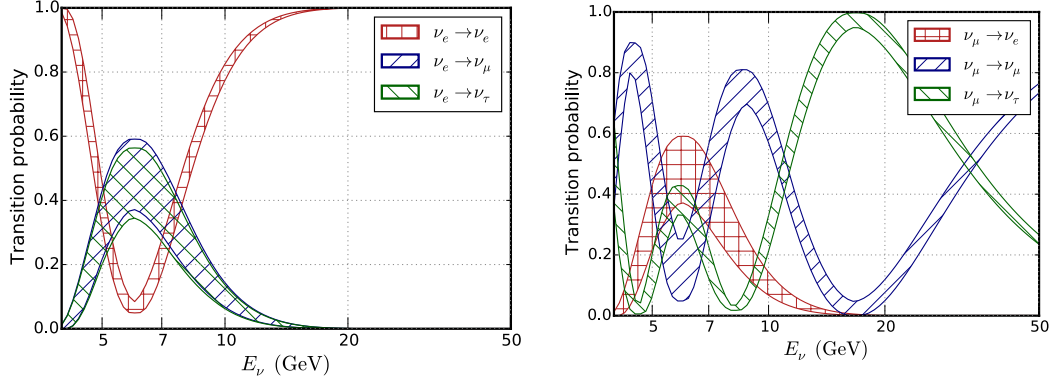- The detector's active volume should be extensive to accumulate a

Figure 2.4: Transition probabilities for ν_e (left) and ν_μ (right) for $\cos\theta = -0.7$ (traversing the Earth's mantle) versus the neutrino's energy. Normal mass ordering is assumed. The areas between the curves indicate the uncertainties on the oscillation parameters used in the calculation. From [23].

certain amount of statistics, as the interaction cross section for a single neutrino is small.

- Providing large interaction volumes should not be too cost-intensive; use naturally abundant media such as water, ice or air.
- There should be some form of shielding from atmospheric muons and exterior light.

All these points are met by the experiments KM3NeT (as mentioned), Super-Kamiokande and IceCube. The two latter are briefly introduced in the following section after the mechanism with which these experiments study oscillations is presented.

2.5.1 Oscillation signature of atmospheric neutrinos

As already indicated in Sec. 1.3.4, experiments studying atmospheric neutrinos operate at L/E of $10 - 10^3$ km/GeV and are mainly able to determine the large mass splitting and the mixing angle θ_{23}.

Following Eq. 1.17 and expanding it to the three flavor case (as for example done in [23], from which this section is inspired by), the transition probabilities for electron and muon neutrinos are plotted in Fig. 2.4. For this, the actual matter potential of the mantle with an electron number density of $N_e = 2.5\,\mathrm{cm}^{-3} N_A$ is assumed, where N_A is the Avogadro constant. From the discussion in Sec. 1.3.2 and in particular Eq. 1.22 it is found that neutrinos traversing the mantle experience a resonance at around 6 GeV. In the same way, saturation is found starting above 12 GeV, which is observed in the plots. The same is seen in Fig. 2.5, which contains only the survival probability as a dependence on the zenith angle and energy for directions larger than $\cos\theta = -0.85$, where the dashed line indicates the zenith angle used for Fig. 2.4. For electron neutrinos on the left, the flux
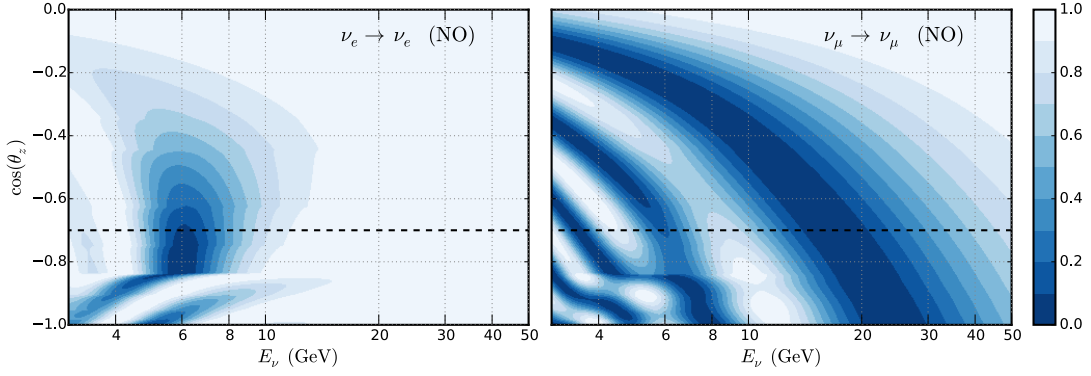
Figure 2.5: Survival probabilities for ν_e (left) and ν_μ (right) as a function of the neutrino's zenith angle and energy. Here, $\cos\theta = -1$ corresponds to a path through the center of the Earth, while $\cos\theta = 0$ is the horizontal direction, which exhibits almost no matter effects. Assumes NO. The dashed line marks the zenith angle for which the transition probabilities are shown in Fig. 2.4. From [23].

almost fully disappears for energies around $6\,\text{GeV}$ and directions between $\cos\theta = [-0.85, -0.6]$.

Furthermore, for $\cos\theta$ closer to 1 the neutrino passes through the core and parametric enhancement due to the periodic density changes modifies the survival probability. Superpositions of different matter effects cause the pattern to become more complex, especially in the case of ν_μ. Additionally, the uncertainties resulting from the current precision with which oscillation parameters are known are significant for small energies (Fig. 2.4, right).

For the saturated regime, the oscillations are dominated by transitions between ν_μ and ν_τ.

From these considerations, it should become clear how to probe for oscillations and infer oscillation parameters from it; the zenith angle and the energy of the neutrino have to be reconstructed and spectra of these quantities can then be compared to hypotheses generated assuming certain sets of oscillation parameters. The best-fitting model constitutes the result as will be carried out in Sec. 8.2 for the first ORCA data.

2.5.2 Further atmospheric neutrino experiments

This section focuses specifically on oscillation research experiments that utilize atmospheric neutrinos like KM3NeT/ORCA, namely Super-Kamiokande and IceCube/DeepCore. Further contributions to the determination of θ_{23} and Δm^2_{3l} come from the LBL accelerator experiments Minos [80], NOνA [26] and T2K [27].
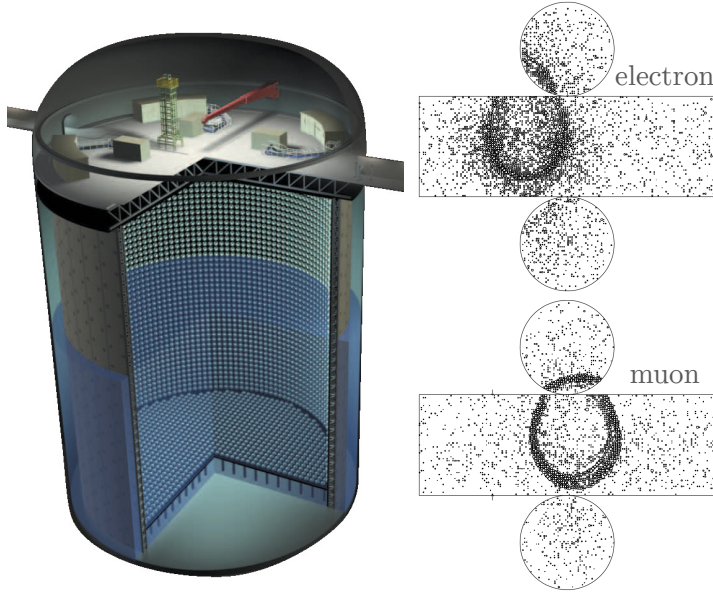
Figure 2.6: Left: Artist illustration of the Super-Kamiokande detector with its inner walls of a steel tank equipped with PMTs to register Cherenkov light in the water. From [82]. Right: Hit pattern on PMTs as created by a traversing electron (top) and muon (bottom) in the detector. Modified from [83].

Super-Kamiokande

The first experiment to add major contributions to the field of atmospheric neutrino studies was Super-Kamiokande [81]. Its researcher Takaaki Kajita received the Nobel prize in 2015 for the discovery of neutrino oscillations, for which they reported evidence for the first time in 1998 [57]. The prize was shared with Arthur B. McDonald from the Sudbury Neutrino Observatory who contributed to the same topic. The detection principle is similar to ORCA, using the Cherenkov light produced by the secondary leptons in neutrino interactions. The detector is located in a former mine in Kamioka, Japan, providing the shielding from atmospheric muons. It comprises a 50 kton water tank (42 m tall, 39 m in diameter) equipped with more than ten thousand PMTs with 50 cm diameter on the walls. A schematic drawing is shown on the left side of Fig. 2.6. The water is purified to lower the background, which helps to bring down the energy threshold to about 5 MeV.

Due to the high density of light sensors the Cherenkov cones can be detected as ring like projections on the detector walls. Here, electrons create more fuzzy rings due the higher probability of causing electromagnetic showers. Muons on the other hand form more clear-cut ring structures as can be seen in a comparison on the right side of Fig. 2.6.

Reconstructing the properties of the detected interactions, they are able to observe a clear oscillation signal when comparing the data to the non-oscillation hypothesis. One example is presented in Fig. 2.7, where the spectrum of the zenith angle is shown for a selected class of sub-GeV events with a momentum larger than 400 MeV/c. The data in crosses follow a best fit to a set of oscillation parameters rather than the boxes that contain the non-oscillation hypothesis.

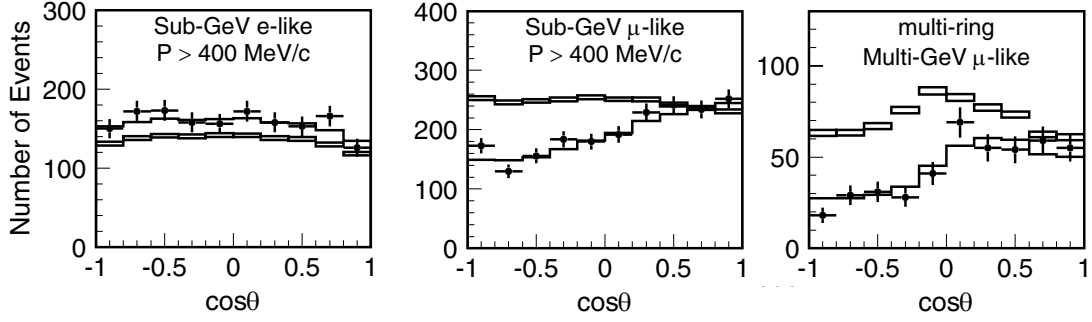In addition to atmospheric neutrinos, Super-Kamiokande also studies

Figure 2.7: Spectra of the reconstructed $\cos\theta$ for a selected class of events. The crosses mark the data and the line the best fit for a set of oscillation parameters. The MC predictions for no oscillation in the boxes (including statistical uncertainties) are compared and not in agreement with the data. Modified from [83].

neutrinos of solar origin. It is further looking for potential proton decay and able to detect neutrinos from supernovae.

IceCube/DeepCore

The IceCube Neutrino Observatory [72] is located at South Pole and utilizes the Antarctic ice as Cherenkov medium. Similar to KM3NeT, it is built up of strings (86) holding the photo sensors that are fixed by the ice after deployment. With its $1\,\mathrm{km}^3$ of instrumented volume and over 5000 optical modules, each housing a 10 inch PMT, it is primarily designed to look for astrophysical neutrinos in the TeV to PeV range. In a depth between 1760 m and 2450 m below the ice surface, the low-energy extension DeepCore [84] is located in a particularly clean region of ice. It features denser instrumentation comprising 8 additional strings added between existing ones to yield an average horizontal spacing of 72 m and vertical distances of mainly 7 m, as is illustrated in Fig. 2.8. The denser instrumentation allows for lowering the energy threshold to about 5 GeV and thus studying neutrino oscillations [85]. Evaluating the data from 3 years, IceCube yielded the contours drawn in Fig. 2.9, which shows the precision with which Δm_{3l}^2 and θ_{23} were determined in comparison to other experiments. To further improve the sensitivity to oscillation parameters and also the neutrino mass ordering, another extension is planned to increase the density of photo sensors even further. The IceCube Upgrade [87] will also feature multi-PMT optical modules following the example of the KM3NeT DOM.
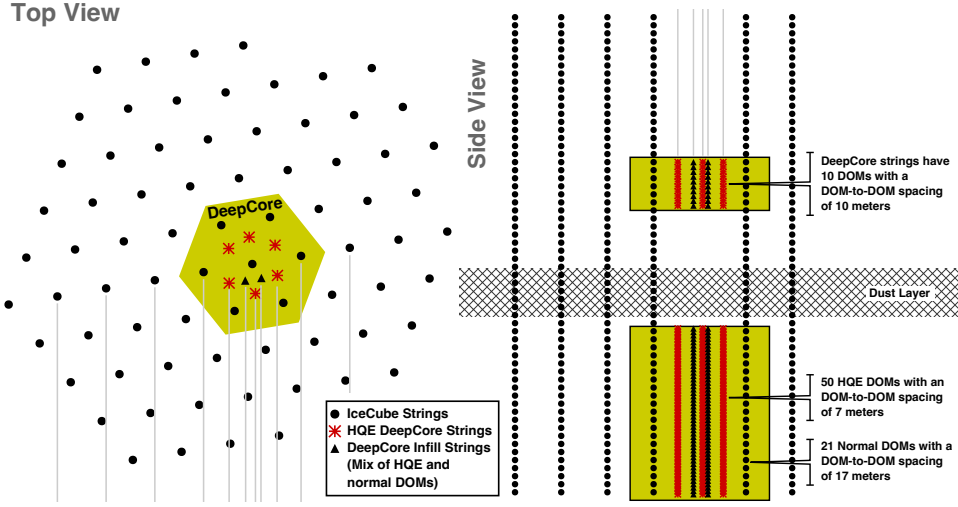
Figure 2.8: Schematic drawing highlighting DeepCore inside the IceCube detector. Top view on the left and side view on the right, indicating the positions of the additional strings with denser instrumentation. The later deployed DeepCore modules feature also higher quantum efficiency PMTs. Modified from [84].
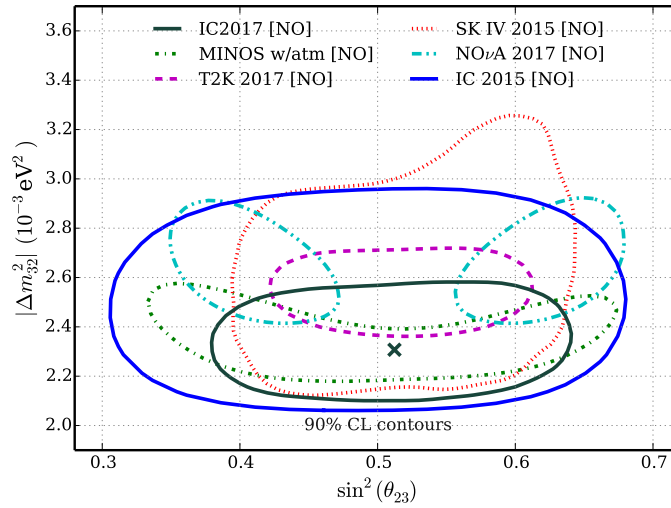
Figure 2.9: Results from the best fits to data for different experiments studying Δm^2_{3l} and θ_{23}. In black the 90% confidence level determined by IceCube/DeepCore. From [86].

Calibration methods in KM3NeT

THE goal of discovering neutrino point sources in the universe with ARCA and carrying out high-precision measurements of the fundamental neutrino properties with ORCA requires a shared, essential prerequisite to achieve the desired resolutions: an accurate time and position calibration of the DOMs and the PMTs within. Namely, the relative timing accuracy aimed for is 1 ns and the corresponding spacial deviation is about 20 cm considering the speed of light in water [48].

To this end, several calibration methods are in place to accomplish an accurate, robust and time-variant calibration used for online triggering of events in data acquisition, offline reconstruction stage and MC production. The conventional methods include measured and analyzed data from hardware devices, such as acoustic emitters and receivers, lasers and compasses and are presented in this chapter. Additionally, an atm. muon-based method has been developed in the scope of this thesis that utilizes the optical data of recorded events and aims at finding the underlying detector calibration by optimizing for the reconstruction quality. Its working principle is explained in the next chapter, together with the classical track reconstruction it is based on.

3.1 Conventional time calibration methods

In the layout of the KM3NeT detector, there are three different levels of time calibrations to be considered:

- inter-DU: offsets between detection units as parts of the detector,
- inter-DOM: offsets between the single optical modules on a detection unit and
- inter-PMT: offsets between the individual photomultiplier tubes inside an optical module.

In principle, the origin for each level's need for adjustment is the same: Different positions and thus different signal paths and transmission times

need to be determined and accounted for. Additionally, the signal travel time is influenced by the PMTs themselves and the electrical components it is transmitted through and processed by. These components, in turn, cannot be assumed to be identical, nor their influence to be stable for all times and conditions. This includes differences between testing onshore and operation in the water.

Inter-PMT time calibration

Starting from the smallest scale, timing differences between PMTs within a DOM must be determined. For this, the radioactive isotope Potassium-40, naturally abundant in seawater, is utilized. In its β^- decay into Calcium, electrons with up to $1.3\,\text{MeV}$ [88] are released, possessing enough energy to emit Cherenkov light. With the electrons' mean free path of a few centimeters, such decays can isotropically create coincidences in a few (mostly two) PMTs in the vicinity of the module. The difference in detection times of the signals for a pair of PMTs can then be evaluated. An example for two adjacent PMTs is displayed in Fig. 3.1, left. The position of the maximum is the time offset between the two PMTs. In a global fit that considers all PMT pairs of a DOM with sufficient statistics, the offsets are minimized [89]. Additionally, the detection efficiency of both PMTs combined and their (combined) transit time spread can be determined from the integral and the width of the distribution, respectively. These values are later used during MC production to simulate the PMT response accurately. The efficiency, for example, might change due to sedimentation and biofouling and recover at a later point as a result of high sea currents that clear some of the residue. The term transit time refers to the time the avalanche of electrons takes to traverse the entire system of dynodes in the PMT, starting from the incident light hitting the photocathode and ending at the anode. Since the accelerating of electrons in the electric fields of the PMT and the releasing of further electrons at the dynodes are statistical processes, the transit time exhibits a spread

In principle, it is also possible to use ^{40}K decays from the glass sphere in the darkroom at shore, though with longer measurement times to achieve the same statistics. This step is left out in the current calibration plan and is instead done shortly after deployment in the water. Directly after that a high voltage tuning is carried out. In this case and for all subsequent adjustments of the high voltage setting the PMT is operated with, the inter-PMT time calibration has to be updated as well. Higher (lower) voltage results in shorter (longer) transit times and thus potentially different offsets between PMTs, as each PMT is tuned individually. The aim of adjusting the high voltage is to reduce the gain spread from all PMTs, resulting in equal detection efficiencies.
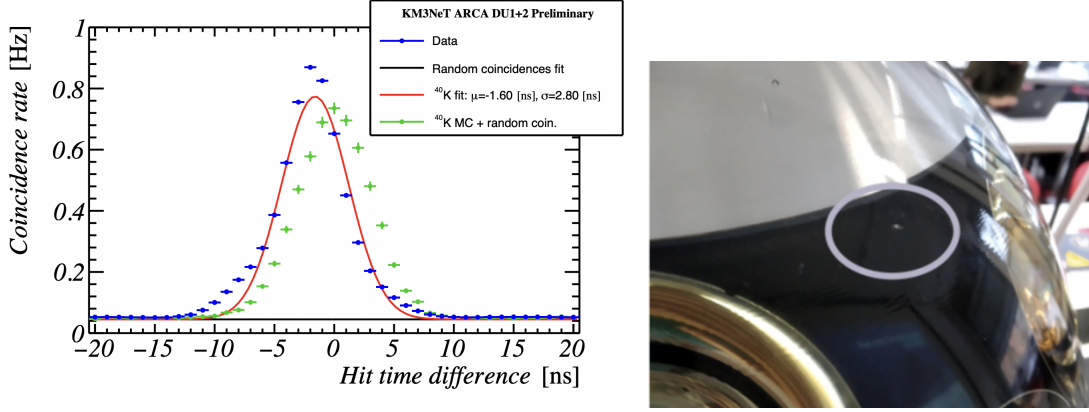
Figure 3.1: Left: Rate of coincidences due to ^{40}K decays in the water versus hit time differences for a pair of adjacent PMTs. The data (blue) are fitted with the red curve. From it, the position of the maximum yields the offset between these two PMTs. After correction, the green points are obtained. Additionally, efficiency (integral) and transit time spread (width) can be extracted. [89] Right: Picture of a DOM highlighting the nanobeacon. [90]

Inter-DOM time calibration

There exist two entirely different methods to obtain the timing difference between DOMs on a DU in the laboratory and *in situ*. In the first case, lasers directly expose the photocathode of the PMTs with light. Detection times are measured and differences between DOMs are adjusted.

After deploying the fully-functioning detection unit in the water, coincident light reaching several DOMs can be utilized to determine individual time offsets. This light can either come from atm. muons or be emitted by so-called *nanobeacons* [90]. The nanobeacon is a custom-made LED located at 45° off-axis (see Fig. 3.1, right) that can illuminate DOMs higher up the DU. The distances of each DOM and the detection times are then considered to synchronize time calibration. Its light reaches as far as other DUs, additionally allowing for inter-DU correlations. At the point of this thesis, a full implementation including a reliable analysis of the inter-DOM time calibration using nanobeacons is not yet in place.

The method utilizing muons, including a full track reconstruction and evaluation of resulting time residuals, is part of the muon-based calibration and will be explained in detail in next chapter's Sec. 4.4.5.

Inter-DU time calibration

From the shore station, the base of every DU can be pinged via a signal through the optical cables to obtain an estimation of this signal travel time. Together with the time, it takes for the signal to reach the DOMs from the base and several assumptions about processing times of the involved devices, an individual offset per DU with about 5 ns precision is calculated [91]. Once

set, the trigger rates deliver a rough estimation about whether this offset is chosen correctly. In this way, a simple test right after deployment is carried out: Some data are taken and are re-triggered offline, assuming different time offsets for one DU. The highest trigger rate in the scan indicates the correct offset. However, the accuracy of this method is limited to a few tens of ns. Instead, a refined approach, using reconstructed track information, has been developed to determine these offsets with sub-nanosecond precision. It is discussed in Sec. 4.2.

3.2 Conventional position calibration methods

The initial positions of the DUs at the seafloor are determined during the deployment sea campaign using an acoustic positioning system independent of the one used later for detector operation. The 3d positions of objects can be triangulated from the boat, which in turn uses GPS to yield an absolute position. With this, the nominal positions according to the planned layout are updated. In the case of ORCA, which is focused on throughout this thesis, these values are accurate to about 40-50 cm [92]. Before acoustic positioning and reliable analysis of its data were available, this was the starting point for any subsequent muon-based calibration, as will be introduced in the next chapter. This includes all data taking periods before the 6 DU constellation (1,2 and 4 DUs from Sep. 2017 to Jan. 2020, with interruptions) and all data presented from ORCA6 in the scope of this work. The fully dynamic position calibration is a relatively new development that has been tested on the ORCA6 data. This allows for cross-checks between the two methods, which will be shown in Chap. 4.4.2. Since these cross-checks helped verify its proper functionality, it is planned to use the dynamic positioning for future processing, as initially envisioned.

Acoustic positioning system

To continuously monitor the swaying of the DUs due to deep-sea currents, a combination of autonomous acoustic emitters in the vicinity of the detector and piezo receivers inside the DOMs are utilized. For most of the ORCA6 data acquisition, three battery-powered emitters were available, each of which is designed to send signals every 10 minutes. They are installed about 100 m away from the DUs and one of them is depicted in Fig. 3.2, left, shortly before deployment. Piezo sensors, such as the one seen as a white spot at the center of the DOM in the same figure on the right, are glued to the glass sphere to pick up on acoustic waves.

Using triangulation, sound propagation models in water and constrains by a mechanical model of the DU, the current position of the DOM can be calculated and interpolated between signals (pings). This information is then available on an event-by-event basis for reconstruction.

Figure 3.2: Left: Acoustic emitter (white) on a tripod right before deployment. Picture: KM3NeT collaboration. Right: Picture of a DOM with a piezo sensor. The device is glued directly to the glass sphere. From [48], p. 8

In order to achieve the desired precision of below 20 cm for the positioning, the positions of emitters and DUs have to be known to greater accuracy than the estimation from the deployment. Hence, an iterative global fit procedure has been developed by the collaboration [93]. It starts with a set of DU and emitter positions, carries out individual fits of the DOM positions for some acoustic data, and then evaluates their overall mean fit quality via a χ^2 per degree of freedom (ndf). DUs and emitters are varied, including their z-positions, and optimal relative positioning is found at the minimal $\chi^2/$ndf. Two examples of scans of the reduced χ^2 after the applied global fit procedure are shown in Fig. 3.3; the xy-position of a DU on the left side and the z-position of a DOM on the right. For the former, a clear minimum around (0/0) can be identified, marking the optimal position of a DU on the seabed and confirming the determined geometry. The same is found for a certain z-position of a DOM on the right, showing best fit results for its currently used height.

From the fact that DOMs are assigned individual heights, it can be seen that not only xyz-positions of DU and emitters are determined, but also the distances between the DOMs on the DU might vary from its nominal design values. Two factors cause such differences of a few centimeters: First, the Dyneema$^©$ ropes are elastic and allow for 1% to 2% stretching when force is applied [94]. Second, a certain pre-stretching, different for each pair of DOMs, can be applied when the DU is loaded onto the *Launcher of Optical Modules* (LOM, Fig. 3.4, left) [95] before deployment. In a process executed by hand, the complete DU is wound around the LOM, which is later unfurled at the seafloor again. This is why in the global fit procedure, a stretching of the entire DU and individual DOM distances are both tested for.

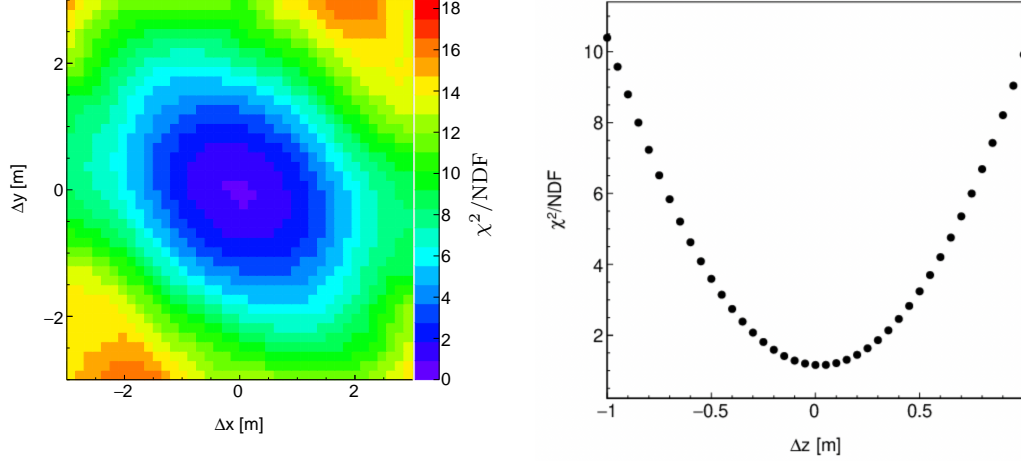Finally, in Fig. 3.4, right, the time residuals between expected and

Figure 3.3: Left: Scan of the *xy*-position of a DU. The color encodes the mean fit quality from fitting acoustic data, displaying a minimum around the input geometry. Right: Fit quality versus different assumptions of the height of one DOM within the DU. Also here, the minimum marks the optimal position of the DOM according to the acoustic positioning.

actual time of arrival of the acoustic signals from all emitters and piezos are combined. The Gaussian shape indicates no significant influences of systematic discrepancies and states a standard deviation of 39 µs. With a sound velocity in water of 1540 m/s, considering temperature, pressure and salinity as encountered in the ORCA detector, this translates to a mean resolution of 6 cm.

Together with the arrival times of acoustic signals at the piezos, information about its orientation is measured by the DOM's compasses and used in the position fit.

In the near future, an additional calibration unit will be deployed 40 m away from the detector that will help reduce degeneracies for both position and time calibration [96]. It is equipped with an additional acoustic emitter, a laser beacon and a hydrophone. The laser will be exciting for inter-DU time calibration. Additionally, it allows for a measurement of the water properties and the use as an energy calibration device is being evaluated.

The final touch to a complete position calibration is the determination of the absolute orientation of the detector. This is particularly important for the ARCA detector, as its pursued physics goals require reliable resolution of the direction of neutrinos in the sky. To this end, the "shadow" from cosmic rays due to the moon and the Sun can be exploited. The region of fewer observed events caused by their shielding can be correlated with their position on the sky.
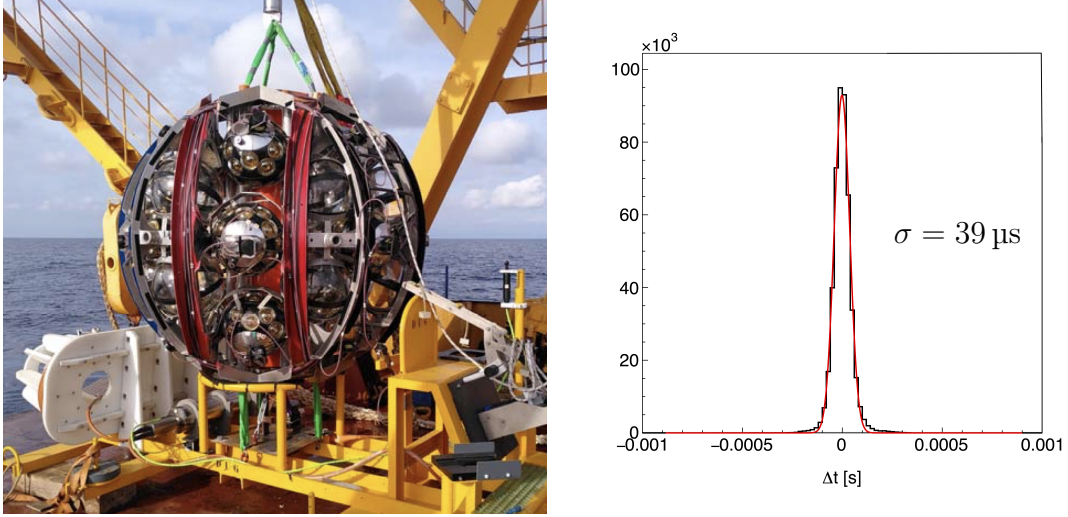
Figure 3.4: Left: Picture of a DU loaded onto the LOM. Below is the anchor of the DU in yellow. From [95]. Right: Time residuals between expected and measured arrival times of acoustic signals at the DOMs (black). A Gaussian fit (red) indicates an accuracy of 39 μs (6 cm).

Development of a muon-based detector calibration

THE subjects of the following chapter cover an introduction to the workings of an atmospheric-muon-based detector calibration scheme, starting with the necessary prerequisites for muon track reconstruction, and then goes on to showcase applications as well as systematic studies in simulated and real data.

The basic principle behind this approach is to find a detector calibration that yields best reconstruction quality. To this end, the hit information is manipulated and the effect on the reconstruction quality monitored. One advantage of this method becomes apparent: Atmospheric muons provide an abundant signal already triggered and recorded by the detector. In this sense, this calibration method can be obtained "for free" from the optical data without the need for any dedicated devices or data acquisition setups. Equally apparent and valuable is that it allows for cross-checking other position and time calibrations with a set of completely orthogonal systematics.

One further, not to be underestimated asset is the constant availability; for any data acquired, a muon-based calibration is attainable. This proved to be essential for the first data taken with the ARCA and ORCA detectors, including the ORCA6 operational period, which was used for several conference contributions and also the deep learning-based studies in this thesis (chapters 6 ff).

4.1 Classical track event reconstruction in KM3NeT

In order to understand the definition of the reconstruction quality for the calibration optimization, the classical track event reconstruction is introduced.

Due to the non-linear nature of the task of fitting a muon track to PMT hit data, an approach with several consecutive steps is employed, of which the principal direction fit adopts a maximum likelihood approach. Internally, this code is called *JGandalf* and the succession of reconstruction scripts around it is referenced to as the *JGandalf chain*. The muon reconstruction is mentioned in the Letter of Intent [48], but has been subject to continuous development and improvements ever since, which is why a few details are updated. The KM3NeT internal notes [97] and [98] contain a helpful overview as well as detailed information, which some of the following parts are inspired by.

Since several parameters need to be fitted simultaneously, the likelihood space is high-dimensional and complex for a minimizer to navigate in. In particular, gradients are close to zero when too far from the solution and no convergence is achieved. Hence, the first reconstruction stage is a linear prefit to a set of selected hits. Including the prefit, each stage identifies a set of causally connected hits to work on. As indicated in Chap. 2.2, the recorded hit information consists of time of the hit, 3d position and 3d direction of the PMT detecting it (deduced from the PMT id and a detector description) as well as the time-over-threshold. The latter is usually omitted as the information about the intensity of the light is already sufficiently encoded in the density of hits on close-by PMTs. The hit selection is more strict than the trigger algorithms and thus can result in discarding an event and not reconstructing it. This will play a role when manipulating the detector calibration later. When diverging further from the optimal description, more events are rejected by these selection criteria.

The coordinate system to solve the problem in is defined such that the direction of the muon track points along the z-axis (see Fig. 4.1, left). The PMT for which the hit is evaluated is located at a distance of closest approach to the track R. For direct light, the angle θ_0 under which the light can be detected is the Cherenkov angle. The PMT is then further characterized by its orientation relative to the track (zenith and azimuth angle, $\theta_\wp$ and $\phi_\wp$).

In the prefit, the problem is linearized by assuming only direct Cherenkov light emitted by a muon. In practice, the track direction is fixed and only the x_0 and y_0 of the track as well as the t_0 are fitted, while z_0 is 0 by definition. In the introduced coordinate system, this point corresponds to the origin and the time the muon crosses the plane of the origin and the PMT (time of the origin). The entire sky is scanned this way, probing directions with
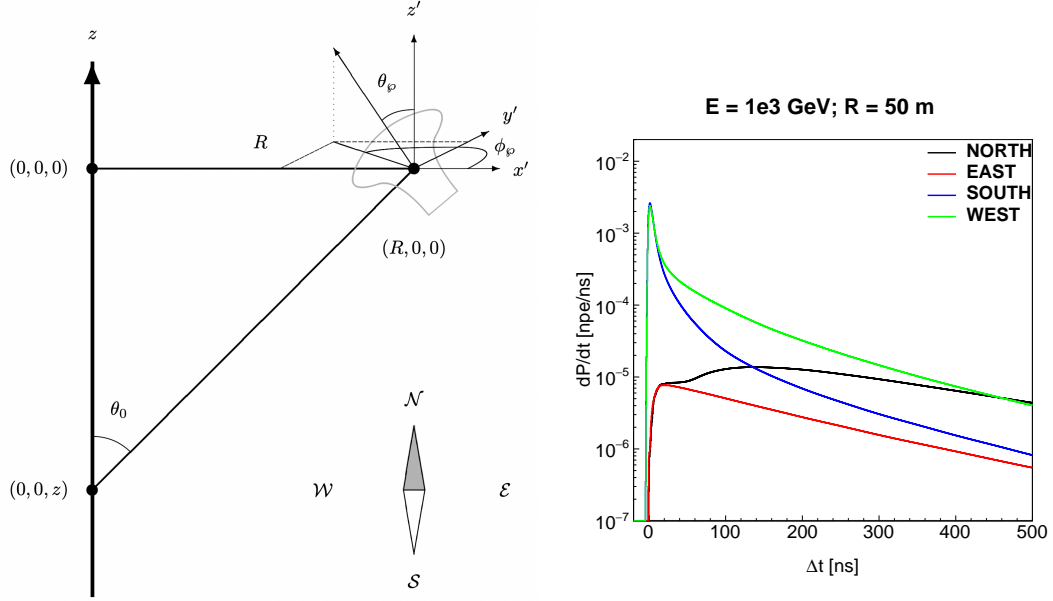
Figure 4.1: Left: Geometry of light being produced by a muon at $(0, 0, z)$ and detected at $(R, 0, 0)$ in the xz-plane. The coordinate system is rotated and translated such that the muon's point of closest approach denotes the origin. The orientation of the PMT with respect to the track is given by its angles $\theta_\wp$ and $\phi_\wp$. Light from the muon that reaches the PMT is emitted under θ_0 along the track. The compass marks specific PMT orientations in the xz-plane ($\phi_\wp = 0$); west, for example, refers to $\theta_\wp = 90°$. Right: Expected photon yield per time relative to the time for the shortest optical path. Different orientations of the detecting PMT (colors) are distinguished following the compass on the left. The absolute numbers refer to a 1 TeV muon at 50 m distance to the PMT.

a mean distance of 5° (for ORCA). The χ^2 from photons' predicted and observed arrival time is then minimized. To stabilize convergence hits passing the initial selection but producing significant outliers ($> 3\,\sigma$) in the χ^2 are excluded. From the mean χ^2 per number of degree of freedom (number of hits minus the three fit parameters), a quality is defined, which is used to select a set of the best directions for the next reconstruction stage.

In an intermediate step, the directions are also fitted in addition to x, y and t. Essentially using Powell's method [99], in which one-dimensional scans in the parameter space are evaluated iteratively without having to calculate gradients, a solution closer to the optimum is searched. This stage would already have difficulties converging if the starting point is too far off the optimum, thus the need for the prefit. However, it is more robust to outliers than a full maximum likelihood approach. In the same way as before, defining a quantity similar to a χ^2 is used to identify a set of best-fitting solutions, which is then passed on for further improvement to JGandalf.

At that stage, the direction is finally fixed. A likelihood between a model

predicting the light yield δP for a certain PMT plus time and the data is maximized. For every selected hit, the probability of it fitting to the currently tested track hypothesis is evaluated. Hence, the term for the likelihood to be maximized is the following product:

$$\mathcal{L} = \prod_{i}^{\text{hits}} \left. \frac{\delta P}{\delta t} \right|_{R_i, \Delta t, \theta_\wp, \phi_\wp} . \tag{4.1}$$

Here, for each hit, the function describing the light yield per time is evaluated at the distance of closest approach R, the difference between expected and measured time of arrival considering the Cherenkov hypothesis Δt and the orientation of the PMT ($\theta_\wp$ and $\phi_\wp$).

It becomes clear that the function P must contain all information about light production, propagation and detection in KM3NeT DOMs. To this end, semi-analytical probability density functions (PDFs) are used that contain all of the above.

- Analytic models describe the amount of direct light that is produced from muons (Cherenkov) and via their energy loss (bremsstrahlung and pair production) as well as single scattered light from these sources. To include all contributions from scattered light, it is integrated over all possible scattering points leading to a detection.
- In addition to scattering, absorption for water with the properties as encountered in the detector is accounted for, attenuating the expected signal.
- The effect of light dispersion, which smears out the expected arrival times, is included.
- For the probability of detecting a photon, angular acceptance of the PMT and its quantum efficiency are taken into account. Furthermore, the time information of the expected hit is smeared with the transit time spread.
- Lastly, the expected optical background is considered.

An example for the light profiles of a muon track is shown in Fig. 4.1, right. There, an energy of 1 TeV and a distance to the PMT of 50 m is assumed. The colors encode different orientations of the PMT and the directions relative to the track can be understood in combination with the compass on the left side of the figure. Since the direct light from the muon hits south and west under a similar incident angle, its main contribution for Δt close to zero is the same in both light curves. For later times, more light is still detected in western direction, as more photons from energy losses further downstream find their way onto this PMT. Contrary to Cherenkov light, the direction of light from energy loss processes is only peaked towards the Cherenkov angle but not restricted to it. For east and north, no direct light can reach the PMT. Only scattered light can be detected in this case, with

north seeing more energy loss photons at later times from the muon as it moves further away.

While the parameters describing propagation and detection are generally fixed, the ones constituting the muon track itself, namely its direction, energy and vertex, are the free parameters for the likelihood fit. Gradients in the likelihood landscape are calculated analytically by interpolating [100] and the negative of the logarithm of their value is then minimized using the Levenberg-Marquardt algorithm [101]. When an optimum is found for all previously input starting points, the likelihood value states the quality achieved and thus allows for identifying the overall best estimation for the direction from the reconstruction chain; the result with the highest quality. In the following, the quality is a positive number, i.e., inverting the negative sign from the log-likelihood for the sole reason to have larger numbers correspond to better quality. However, this does not mean that the quantity is bound to positive values, as the minimized likelihood can have positive values, generally denoting poor agreement.

Not only does such a quality allow for sorting different direction hypotheses of the same event, but it also enables comparisons between events as to how well they can be reconstructed. Hence, a large quality flags an event as one with many signal hits fitting to the track hypothesis over being purely background.

After fixing the direction, energy, vertex position along the track, and its length are further improved. In fact, the one-dimensional likelihood scan of the *z*-position of the interaction along the axis is executed twice in the current setup of the JGandalf chain; once before the energy fit to further restrict the input information to it, and again once afterward to finally fix starting and end position of the reconstructed track. In it, the geometric constraints about direct Cherenkov light are used to project back onto the most probable first and last emission point.

The energy fit itself consists again of a proper one parameter likelihood fit of $\ln(E_\nu)$ using PDFs and a minimizer. Here, all PMTs within a certain road width around the track are considered and probabilities for hit and no-hit are evaluated.

4.2 Muon track quality method for detector calibration

The minimizer in a likelihood fit is naturally designed to change free parameters of the model such that the overall agreement to the data is maximized. As described in the previous section, when fitting a muon track in KM3NeT, this is achieved by varying the inputs to the PDFs that produce light profiles from the track properties and the detector conditions. But as written out in Eq. 4.1, the likelihood is, in the same way, dependent on
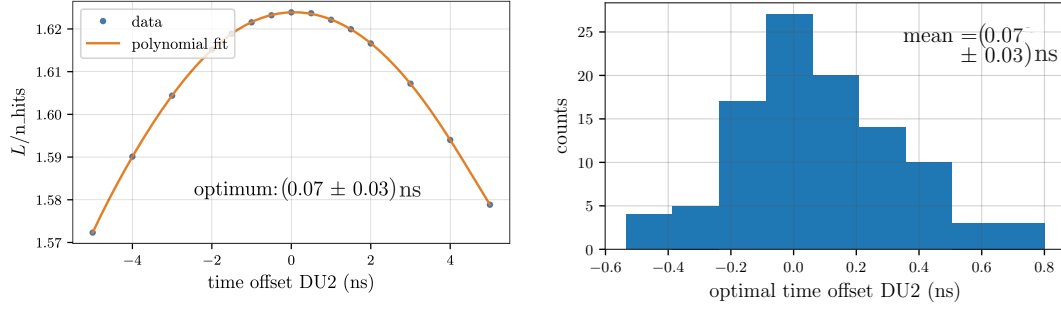
Figure 4.2: Left: Scan of the reconstruction quality (likelihood over the number of hits used in the fit) for different modifications to the time offset of a DU (blue) with fit (orange). Right: Distribution of fitted optimal time offsets when dividing the whole data set into smaller subsets. The standard error is an estimation of the uncertainty of the overall optimal value in the left side's plot.

the hit information. Taking this one step further, it implies that the same input to the PDF (same direction etc.) will yield a different likelihood when being evaluated with different hit information. Hit information in this case is synonymous to calibration, as the individual position and time offsets to each PMT determine $R, \Delta t, \theta_\wp$ and $\phi_\wp$. Consequently, there exists an optimal calibration that achieves the highest likelihood. In this case, the hits follow the predictions from the light curves closest. Any deviation from this calibration, for example, a change in the time offset of a PMT, will result in a lower likelihood, as the observed hit time differs from the maximum of the predicted time. Exploiting this dependence, it is possible to search for the optimal calibration by reconstructing the same data for different manipulations to the calibration and monitoring the quality. This principle is the basis for the muon-based detector calibration, first introduced by Maarten de Jong [102] as it was used similarly in KM3NeT's predecessor experiment ANTARES. This concept has been further refined in the scope of this thesis, and software handling these optimizations has been developed.

4.2.1 Implementation details

The optimization is put into practice by scanning a parameter of the calibration and determining the position of the maximum of the quality. One such example is displayed in Fig. 4.2 for the time offset of a DU. When studying offsets for entire DUs, the calibration of all PMTs on that DU is changed in the same way. In principle, it is also possible to only vary single DOMs and scan for optimal times or positions. However, in practice, this is rather inefficient as for each scan of every DOM, a certain amount of data has to be processed, which is resource-intensive, and strong interplay between neighboring DOMs is expected. Instead, this method is designed to

determine the following properties that mainly describe relations between entire DUs:

- inter-DU time offset: relative timing of one DU with respect to the rest of the detector
- x, y and z offset: changes in the 3d position of a DU relative to other DUs
- rotation: orientation in the xy-plane of a DU within the detector
- stretching: a linear height scaling of the DOM's z-positions on a DU.

As can be seen from the label of the plot in the left of Fig. 4.2, the quality is the likelihood divided by the number of hits used in the JGandalf fit. In other words, this is the mean quality per hit, rather than the absolute sum of all hit contributions. This figure is widely adopted as a quality cut in data analysis inside the experiment. For the use in the context of this muon calibration, slightly better accuracy and less systematic shifts are observed for the normalized likelihood compared to absolute likelihood. Such a reduced quantity is less prone to be dominated by a few elongated muon tracks that produce particularly many hits in the detector and thus high absolute likelihoods.

The data points distributed along the x-axis in Fig. 4.2 mark different assumed detector calibrations for which the time offset of DU1 was modified. The same data are then reconstructed for all these offsets and the average quality is extracted. In order to determine the maximum of this scan, the points are more densely distributed around 0 and fitted for interpolation with a higher dimensional polynomial (8 deg) to read off the position of the maximum reliably. In the case of determining the orientation, a cosine function is fitted, as the data precisely follow this model.

A few moderate cuts are applied to the data before considering them for the analysis. Namely:

- Even before reconstruction, events are filtered to contain at least one triggered hit on the DU currently considered in the scan.
- Events must reach at least reconstruction stage JGandalf in which the likelihood is set. Some events are rejected by the hit cluster selection at the beginning of each stage.
- The events must exhibit a minimum number of DOMs with triggered hits. This serves as the main quality cut without actually using the quality itself, as this is the quantity that is evaluated later. The number depends on the size of the detector and was set to 11 for the ORCA6 configuration.
- Finally, all data sets from each data point in the scan must contain the exact same events. If an event does not survive the aforementioned cuts for at least one tested calibration, it is discarded from the entire scan. This ensures a fair comparison of the reconstruction qualities. Otherwise, larger average qualities could be achieved merely by badly reconstructed events not passing the cuts for certain data points.
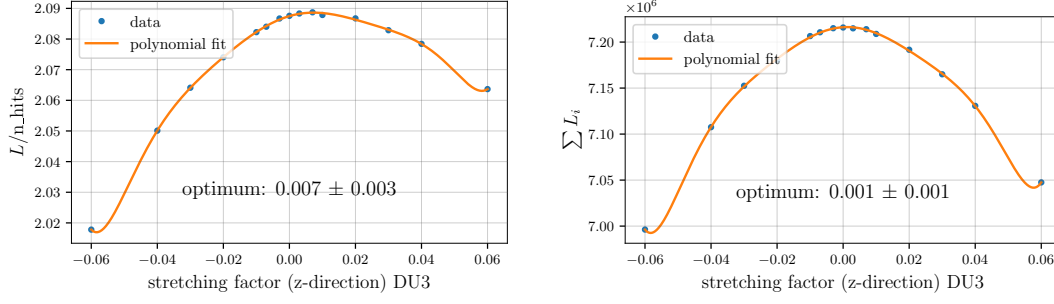
Figure 4.3: Scans to determine the optimal stretching factor f while considering the nominal positions. Left: Using the same analysis as for the other quantities. Right: Adjusting the analysis to improve the capability of recovering the origin.

A statistical uncertainty of the optimal value is estimated by dividing the complete data set used in the scan into subsets, which each still contain enough data to statistically evaluate the individual maximum. A histogram of this is shown on the right side of Fig. 4.2. The standard error of the mean, e, from the distribution of optima of each set is then taken as the uncertainty of the fitted value of the left plot;

$$e = \frac{s}{\sqrt{n}}$$

with the standard deviation of the distribution to the mean, s, and the number of entries n. Both the mean from the histogram and the optimum found via the fit constitute estimations of the average optimum with similar uncertainties, albeit being determined differently. Note that in the displayed example scan for MC data, the absolute numbers with $(0.07 \pm 0.03)\,\mathrm{ns}$ are minimal, considering the desired precision of $1\,\mathrm{ns}$ for time calibration or the fact that hit times on the PMTs are recorded with a $1\,\mathrm{ns}$ resolution.

The special case of height scaling of a DU

The described procedure yields reliable results for all features (see next section) except for the height scaling of the DU, also referred to as "stretching". Here, larger stretching factors than nominal are found when using the reduced likelihood, as shown in Fig. 4.3, left. The factor f is defined as

$$\text{height}_{\text{after}} = (1 + f) \cdot \text{height}_{\text{before}},$$

which means the optimum in the example found to be 0.007 corresponds to a 0.7% linear stretching of the DU. This example is representative of a significant systematic shift present for all DUs. The effect has been studied in [103] and a procedure to recover the nominal stretching has been identified. The following summary lists the found modifications:

Table 4.1: Average sensitivity to a zero offset, combined from all six DUs of the current ORCA setup.

property	mean	standard deviation
time offset (ns)	0.0	0.1
x offset (cm)	0	4
y offset (cm)	0	4
z offset (cm)	-1	5
height scaling (%)	0.15	0.06
orientation (°)	0.0	0.1

- To study the stretching of one specific DU exclusively, the reconstruction is done isolating the DU. If neighboring DUs are considered, only relative stretching factors between the DUs can be determined.
- The sum of the absolute values of the likelihood from the selected events is used, rather than the reduced quantity.
- An additional cut on $-0.99 < \cos\theta < -0.5$ is introduced as a strong dependence on the zenith direction is observed.
- The number of triggered hits requested per event is lowered to 5.

In the context of particle's directions, the terms $\cos\theta$ and *z-direction* are used synonymously in this thesis. The result of a scan following these adjustments can be seen on the right side of Fig. 4.3. In this case, the nominal height scaling can be successfully recovered within the uncertainty estimation.

4.2.2 Consistency and resolution in simulations

In order to get a feeling for the systematic uncertainties involved in determining unknown characteristics of DUs, scans with and without artificial offsets in simulated data are carried out.

The data for the baseline case of no further modifications stem from the ORCA6 detector, where each DU yields one evaluation in how far a zero offset is found back. The mean and the standard deviation of these six values are listed in Tab. 4.1. As can be concluded from the values, the method is generally able to reconstruct the original calibration reliably when considering the specified precision of 1 ns and 20 cm. In particular, no significant systematic shifts can be observed, except for the height scaling case. There, the DU is still reconstructed as 0.15% larger than its nominal height. This corresponds to a dislocation of the highest DOM of an ORCA DU of about 27 cm.

All quoted standard deviations significantly exceed the desired resolution; the 3d position can be generally determined to 4-5 cm, and timing
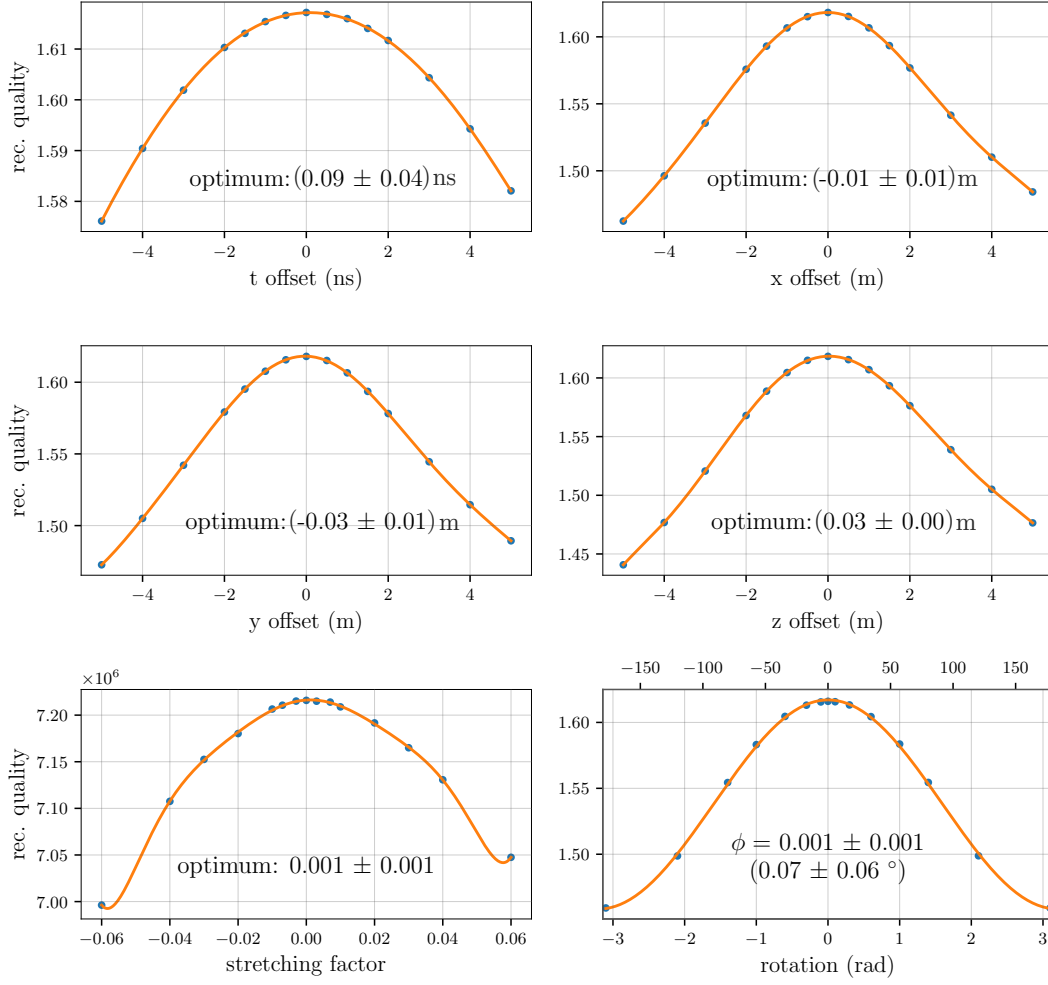
Figure 4.4: Overview of all reconstruction quality scans for DU3 with MC data testing for the accuracy with which the nominal values can be recovered. The quality is defined as the likelihood over the number of hits used for fitting for all plots, except for the stretching, which shows the sum of the absolute values of the likelihood.

information comfortably to sub nanoseconds. Especially the orientation resolution, which was specified with "a few degrees" [48], can be resolved to way below one degree. The high accuracy stems from the fact that a specific model function (cosine) can be used for fitting that utilizes all data points instead of only those around the maximum.

An overview of all determined properties for one DU out of this study is presented in Fig. 4.4. All scanned curves display a smooth behavior due to the high statistics used. This way, the systematic shifts would become visible but prove to be particularly small.

Figure 4.5: Scans for the optimal time offset on two neighboring DUs. DU2, left, has an artificial offset of 1 ns applied, while DU1 does not, but is assigned a fraction of the overall difference between the two DUs by the analysis.
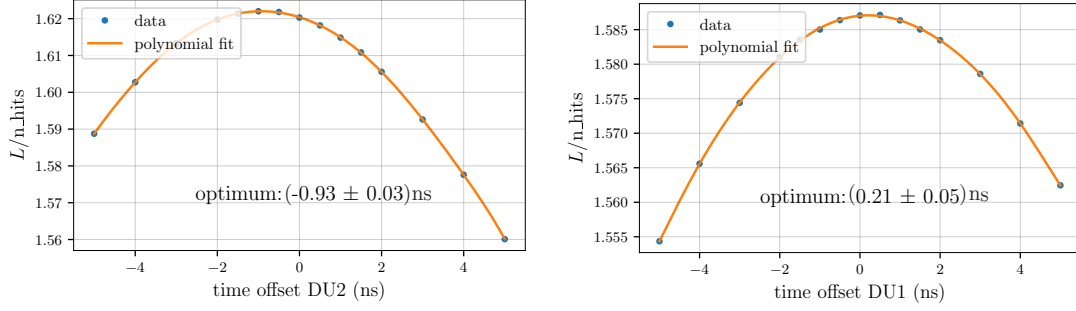
Interplay between quantities and DUs

Naturally, there are a few interactions between different properties of the DUs that should be considered as well as inflections of one DU on another for the same property.

In general, there is a time/space degeneracy. The case of a PMT detecting light earlier than expected due to an incorrect time calibration cannot be distinguished from the PMT being closer to the light source. This automatically causes the offsets for the timing of a DU and its 3d position to be correlated in the scope of this atmospheric muon data-driven approach. Yet, this is only true to a certain extent: Some hit patterns cannot be fixed to better fit the track hypothesis by simply adjusting the time information alone. They require additional translations in space to achieve the best reconstruction quality possible (vice versa for space and time).

In consequence, the influence of a 1 ns offset to a DU is studied. Two example plots from this are shown in Fig. 4.5; DU2 with the modification of one ns and the neighboring DU1. Further values are gathered in Tab. 4.2, including the found optimal x- and z-positions.

Considering the optimal time offsets for each DU, it can be observed that most of the difference is correctly attributed to DU2. All neighboring DUs, however, experience a portion of the shift. This is caused by tracks generating hits on the pair of DU2 and a neighbor. In that case, the time offset is shared between the two DUs without a preference for one or the other. Only events including more DUs can further constrain the system. This way, all close-by DUs get similar inflections, with only the further away DU11 experiencing no change. From this, it becomes clear that in a realistic scenario in which all individual time offsets are unknown, not all determined values should be directly applied in the first iteration. Instead, only a portion considering the number of neighboring DUs n:

$$\text{applied offset} = \frac{n}{n+1} \cdot \text{fitted offset}. \tag{4.2}$$

Table 4.2: Optimal values as found by scans in time and x- and z-position of the DUs. DU2 was manipulated with an additional 1 ns offset. In this and all following tables of this kind, the uncertainties quoted are the statistical ones from determining the optimum. The systematical uncertainties from Tab. 4.1 are not included.

DU	time offset (ns)	x offset (cm)	z offset (cm)
1	0.21 ± 0.05	-5 ± 1	7 ± 1
2	-0.93 ± 0.03	-1 ± 1	-28 ± 1
3	0.43 ± 0.04	-1 ± 1	14 ± 1
9	0.32 ± 0.04	1 ± 1	11 ± 1
10	0.30 ± 0.04	2 ± 1	9 ± 1
11	-0.09 ± 0.05	4 ± 1	-5 ± 1

This corresponds to -0.74 ns for DU2 (from the initial -0.93 ns) and about 0.25 ns for the neighboring DUs in this example, which adds up to the initial 1 ns between them. Still, further iterations would be needed to propagate the differences to DU11, which has to be assigned back the same relative timing with respect to all other DUs except DU2. This principle of partially applying found offsets and then examining their influence forms the basis of the full procedure that is described in Sec. 4.4.1, in which all parameters from every DU are free and need to be optimized.

Furthermore, this example shows how offsets in time can appear as spacial offsets, as alluded to before. No significant adjustments to the x-position can be observed, as changing this dimension does not improve arrival times. However, for the predominantly downgoing muons, a shift of the module's z-position has this effect, which is why a very similar pattern is observed as in the time offsets. A shift of -28 cm translates to about -1.2 ns considering the muon's speed of light in water.

A similar exercise is done introducing a 1 m offset to the z-position on DU2. The values in Tab. 4.3 show the influence analogously to the time offset. The dominant part of the manipulation (97 cm) is attributed to DU2, with neighboring DUs partaking (~ 50 cm). However, the simultaneously found corresponding time offset is smaller with 2.4 ns, meaning 54 cm. This illustrates how changing a time offset cannot fully correct for every z shift due to the directionality of the individual DOMs.

Moreover, the response to the introduction of changing the height scaling of DU2 is studied and reported in Tab. 4.4. The aforementioned adjusted procedure to determine the optimal stretching does not consider other DUs during the analysis. This way, no interplay can occur when scanning for that property in neighboring DUs. Nevertheless, as this constitutes a change in z-direction, an influence on the time and z-position is expected, as observed in the cases before. This is clearly visible, as both time and z offset for DU2

Table 4.3: Influence of a 1 m shift of DU2 in *z*-direction on the determined optimal time and *z* offset.

DU	time offset (ns)	z offset (cm)
1	0.8 ± 0.1	32 ± 2
2	-2.4 ± 0.1	-97 ± 1
3	0.9 ± 0.1	30 ± 1
9	0.7 ± 0.1	29 ± 1
10	0.6 ± 0.1	22 ± 1
11	0.1 ± 0.1	1 ± 1

Table 4.4: Influence of a different height scaling of DU2 (-2%) on the optimal time and *z* offset.

DU	time offset (ns)	z offset (cm)
1	1.3 ± 0.1	52 ± 1
2	-4.7 ± 0.1	-207 ± 1
3	1.5 ± 0.1	59 ± 2
9	1.2 ± 0.1	43 ± 1
10	0.8 ± 0.1	33 ± 1
11	0.1 ± 0.1	-1 ± 1
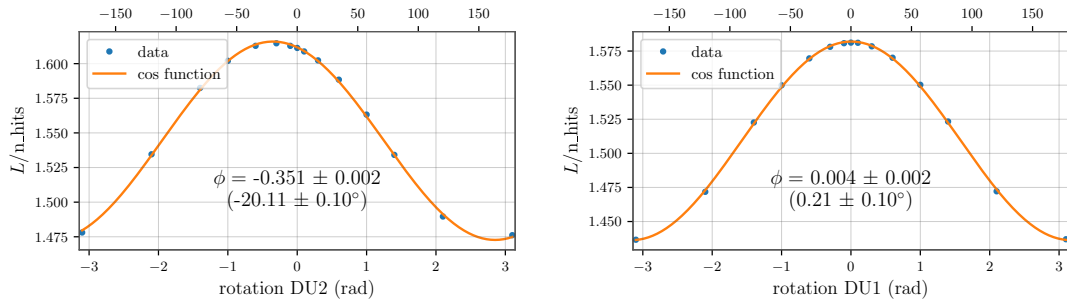
Figure 4.6: Scans of different orientations of a DU for the optimal reconstruction quality. On the left side, DU2 has an artificial offset of $-20°$ applied, while the neighboring DU1 is not affected by such a manipulation. A cosine function is fitted in orange.

show a preference for a shift towards negative values. Shrinking the DU by 2% causes the uppermost DOMs to be around 3.6 m lower than their nominal position with the lowest module remaining in place. The optimal amount to shift all DOMs by equally is found to be 2.07 m. Following the rule of Eq. 4.2, the relative difference between DU2 and the neighboring DU1 is

$$207\,\text{cm} \cdot \frac{4}{5} + 52\,\text{cm} \cdot \frac{4}{5} = 207\,\text{cm},$$

which appears a reasonable attempt to apply an average shift. In the same way as for the artificial *z* offset, the time shift of 4.7 ns, corresponding to 1.06 m, is smaller than the preferred spacial offset.

For geometrical reasons, no strong inflections on other DUs are expected when rotating one single DU with respect to the rest of the detector, as their optimal orientation remains nominal even for a strongly rotated test DU. This is manifested in Fig. 4.6, where a $-20°$ rotation was added to DU2. This value is recovered for the same DU, while its neighboring DU1 only shows a negligible deviation from zero.

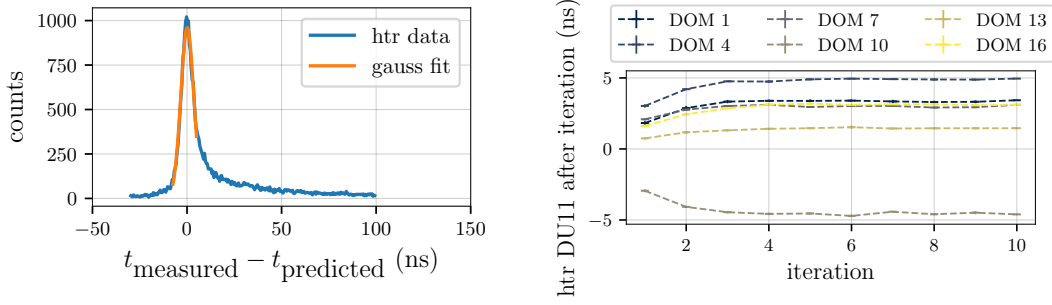Furthermore, rotations do not interfere with the other time and space

Figure 4.7: Left: Hit time residual distribution of a DOM (blue) with fitted main peak (orange) to extract the position of the maximum. Right: Development of fitted hit time residuals of a few selected DOMs over consecutive iterations.

properties because the PMTs' positions hardly change and the expected arrival time of photons is almost not affected.

4.3 Hit time residual time calibration

In a similar fashion to the introduced method for determining DU-specific calibration quantities, the assumptions from the muon reconstruction can be exploited to obtain an inter-DOM time calibration as well. Instead of scanning the reconstruction quality, the hit time residual (htr) distribution is used to evaluate the average timing information of all PMTs in a DOM. The htr are defined as the difference in measured and expected arrival time of photons.

4.3.1 Principle, implementation and expected resolution

For calculating the htr, the expected time has to be defined. The measured time is simply the recorded time of the hit on the PMT. In order to define a predicted arrival time, a muon track has to be fitted to the data first, utilizing the presented reconstruction. Assuming this, the predicted time of light reaching a given PMT is deduced from the shortest travel path of direct Cherenkov light (the first peak in Fig. 4.1, right). An example of a htr distribution from one DOM is shown in Fig. 4.7, left.

The htr data reveal a clear peak around 0 ns, which corresponds to the large majority of direct light that arrives at the expected times, indicating a correct time calibration on this DOM. The width of the peak is influenced by several effects such as scattering, dispersion, transit time spread of the PMT and deviations arising from fitting a non-optimal track hypothesis for this DOM. Scattering also causes the spectrum to be asymmetric, creating more entries for larger measured arrival times. Non-correlated noise hits create the tail ends of the distribution.

The position of the main peak holds information about the average

timing of the DOM; a shift of it implies that this particular DOM registers its hits too late or too early compared to other DOMs that participate in the same events. To determine the maximum, the peak is fitted with a Gaussian between -7.5 and $5\,\text{ns}$. This method was already in place [89] and refined, tested and expanded in the scope of this thesis.

Because a muon only causes hits in a limited part of the detector, the fitting of the htr needs to be applied iteratively, each time re-doing the reconstruction as well using the updated time calibration. Only this way, differences between distant DOMs can be propagated and evened out. The right side of Fig. 4.7 shows the development of the cumulative htr for some DOMs after several iterations. Offsets as large as several nanoseconds only need about four iterations in ORCA until a stable solution is reached. With the detector's smaller dimensions, the downgoing atm. muons lighten up considerable parts, often many DOMs per DU. For the larger scale ARCA, however, it takes up to 40 iterations to fully converge.

As can further be seen from the plot, the final values are stable to considerably below $1\,\text{ns}$ and do not suffer from any systematic drifts that would let the offsets diverge. The individual error bars representing the statistical uncertainty from the fit are tiny even for a moderate number of processed events such as 40,000 in this case. The smooth distribution of the htr and the time calibration resulting from it are two features that heavily profit from the excellent time resolution achievable when equipping water Cherenkov detectors with small diameter PMTs.

Complementary to this, when monitoring the htr found in each iteration, a steady decrease of the values in the ORCA detector is found in Fig. 4.8, left. While the corrections in the first iteration are as large as $5\,\text{ns}$, they do not exceed $0.4\,\text{ns}$ after the 5th iteration. In addition, it can be noticed that the mean of all htr is zero. This is by design as global offsets to the whole detector are compensated for to avoid introducing accidental drifts of the absolute timing.

The resolution in timing that can be achieved with this method is studied in Fig. 4.8, right. Here, two cases are considered: Taking the original time calibration the MC data were produced with (orange) and applying additional offsets that have to be corrected by the introduced procedure (blue). The former serves as a baseline for the average inaccuracy of the time offsets, even for a correct calibration. The latter denotes a more realistic case with non-zero offsets to correct. The individual offsets were created randomly, drawing from a Gaussian distribution with a standard deviation of $3\,\text{ns}$, centered around zero.

From the depicted histogram, it can be concluded that the difference between initial and determined offset is well within half a nanosecond in both cases. The non-zero initial offsets are reconstructed only slightly worse with a standard deviation of the differences of $0.21\,\text{ns}$, compared to the $0.13\,\text{ns}$ baseline. With this, the values are comfortably smaller than the
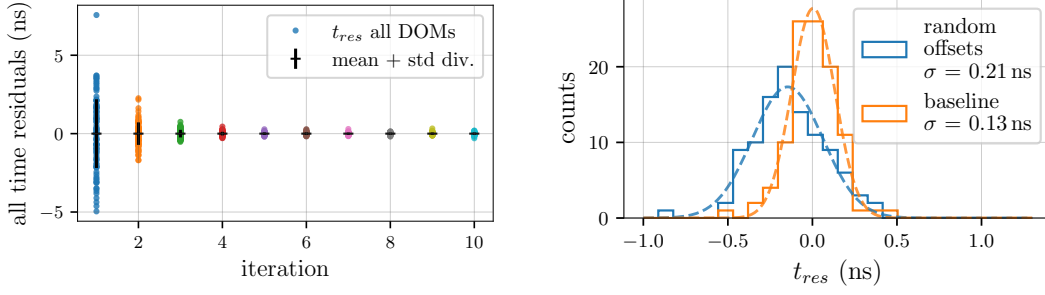
Figure 4.8: Left: Corrections to the time offsets of DOMs per iteration (colors), including the mean and standard deviation (black). Right: Distributions of the difference between initial and reconstructed time offset per DOM t_{res}. Without further offsets applied to the DOMs' time calibration (orange), denoting a baseline resolution. The blue distribution contains the differences of the htr to offsets of random magnitude sampled from a normal distribution with $\sigma = 3\,\mathrm{ns}$. Both data are fitted with a Gaussian (dashed lines).

target resolution of 1 ns. A comment about the shift of the mean is made in the next section.

4.3.2 Systematic effects in htr time calibration

In the following, three special cases of challenging scenarios are discussed and the effect on the htr time calibration is illustrated.

In general, this calibration method aims to focus on inter-DOM relations within a DU and avoid interference from inter-DU timing differences. However, in case there is an additional offset between DUs, there will be an influence on the htr as atm. muon events typically generate hits on several DUs. The response to a scenario in which a 2 ns offset is added to DU4 in the ORCA4 constellation is displayed in Fig. 4.9. There, the fitted htr are plotted in blue next to the input modification in orange for each DU. Below, the difference between input and reconstruction is added.

First of all, a dependence on the floor can be observed in every case. The origin of this feature remains unclear, but could be related to the asymmetry of the distribution of PMTs in the DOM (no upward-facing), making it specific to the downward-going muons, or some geometric effect in the light predictions from the PDFs. In any case, the effect is small with an amplitude of less than 0.5 ns. This representation compliments the resolution plot of Fig. 4.8, right, clarifying where the width and shift of that distribution originate from. The same pattern for the floor dependence persists for the case of random offset.

The influence of the shifted DU4 is clearly visible in the htr, as they differ significantly from zero. However, only a portion of the offset is attributed to that DU with about 1.5 ns, while the missing 0.5 ns are found on each of
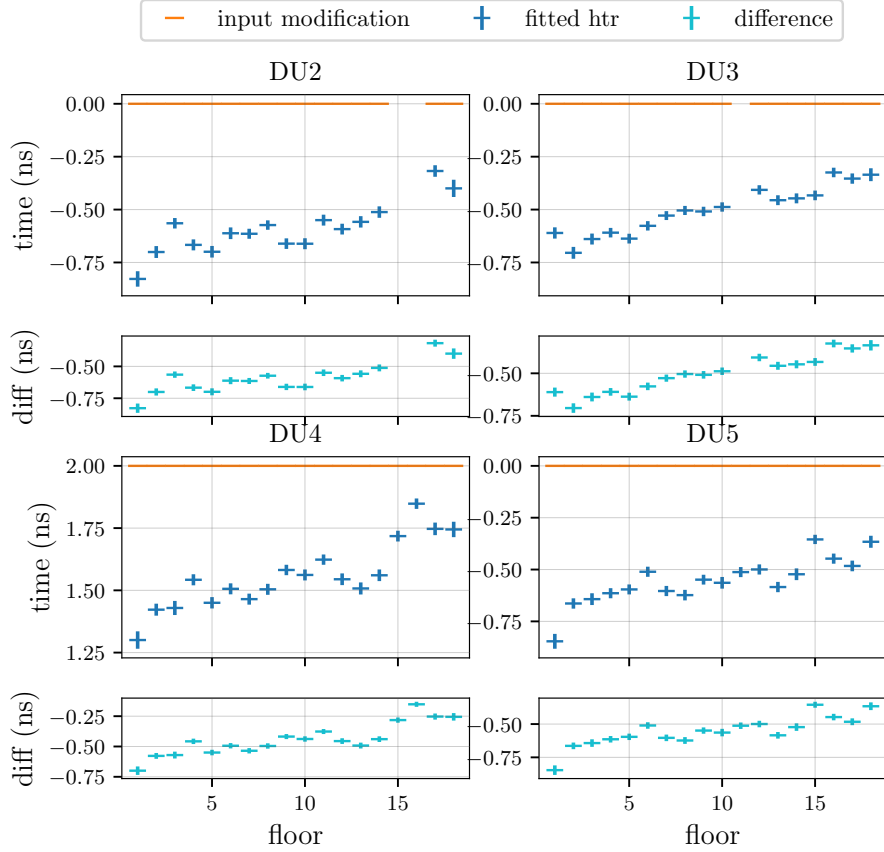
Figure 4.9: Fitted htr (dark blue) as a response to some input offsets (orange) for an ORCA detector consisting of 4 DUs, with the difference between the two data below (cyan). In this case, an inter-DU time offset to DU4 is tested. Non-functioning DOMs are left out.

the other DUs. So, the net offset of 2 ns between DU4 and a neighboring DU is indeed correctly reconstructed. This is similar to the case discussed in Sec. 4.2.2.

This being said, in the case of unknown time offsets for all DUs, the individual offsets cannot be unambiguously identified. Instead, for a real data application, a proper inter-DU time calibration should be ensured first. Another way to handle this would be to correct for possible shifts of the entire DU by centering the htr per DU around zero. This, however, would not allow for testing of the inter-DU time calibration anymore.

Another realistic case that leads to systematic effects is when the information about the height scaling of a DU is incorrect. In Fig. 4.10 the scenario of one stretched and one shrunken DU is analyzed. For DU2, the height scaling of the DOMs is compressed linearly by 2% in the detector description with which the data are reconstructed. Correspondingly, the original, nominal geometry appears *stretched* with respect to the one assumed in the analysis.
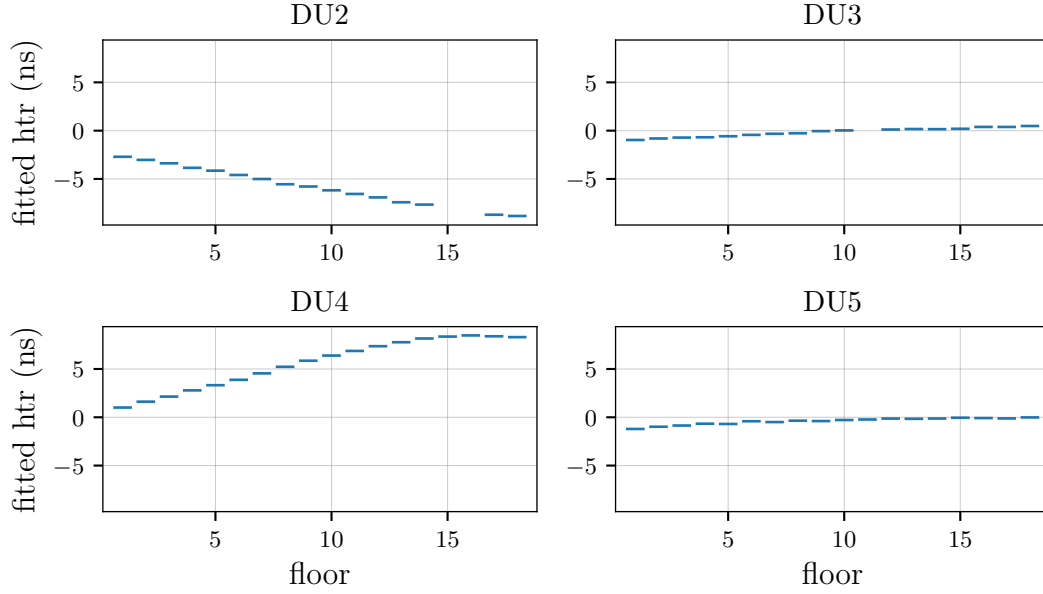
Figure 4.10: Fitted hit time residuals for the case of a reconstruction assuming a smaller DU2 and larger DU4.

Conversely, DU4 shows the response of a *shrunken* DU, as the calibration during reconstruction is 2% larger.

The characteristic patterns can directly be extracted from the plot: for stretched DUs, the htr show a clear dependence on the floor and generally go from more positive values at the bottom of the DU to more negative values towards the top part. Shrunken DUs exhibit the opposite sign of slope for the floor dependence.

Furthermore, the htr of DU2 are all negative. Both effects can be understood by the fact that the actual hit times are taken with a larger DU, which means for downgoing atmospheric muons that they get detected earlier than what is assumed during reconstruction. The time residuals from all photons are shifted towards negative values to counteract that. This is more pronounced for larger displacements on the top. DU4 exhibits exclusively positive shifts, as the modules detect light too late, according to the Cherenkov hypothesis. Due to the 2% enlarged DU in reconstruction, the DOMs are assumed closer to the source, while for the original data acquisition they were not, so their arrival time is shifted to later times.

Additionally, the latter case displays a flattening out for the uppermost DOMs. Here, the limit of how far adjusting the timing can correct for erroneous positions is reached. The upper part for DU4 reaches higher than the rest of the detector in the geometry used in reconstruction and thus makes it more challenging to fulfill the track hypothesis under any circumstances.

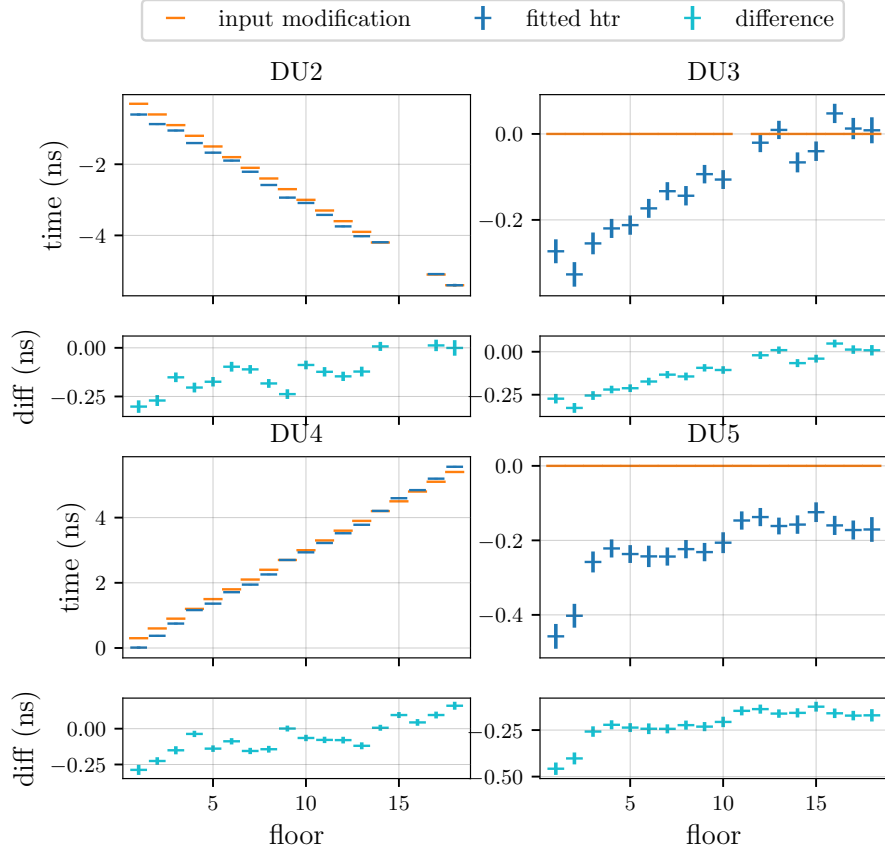The third case covers an increasing time shift for the DOMs along the DUs.

Figure 4.11: Fitted htr (dark blue) for the case of applying a linearly decreasing/increasing time shift to DU2/DU4 (orange). The difference between both is shown in the bottom (cyan).

As depicted in Fig. 4.11, DU2 is shifted to negative times, increasing with height. The DOMs on DU4 are shifted to later times in the same way. The fitted htr can reliably detect such modifications without sacrificing precision on the other DUs. For the entire detector, the differences between input and output of the reconstruction are smaller than 0.5 ns.

The two examples of applying either a stretching/shrinking in space or time illustrate the degeneracy between these two, which cannot be fully solved only using muons. It underlines the importance of a stable position as well as inter-DU time calibration. Only in that case, the resulting htr will yield correct statements about the inter-DOM time calibration. Turning this around, the htr can be utilized to indicate systematic miscalibration whenever a clear floor dependence, for example, is observed. Because of this, it is an essential tool for real data analysis, as is discussed in Chap. 4.4.5.

Attempts to improve the accuracy and eliminate the remaining floor dependency for the baseline case were conducted in the scope of a Bachelor's thesis [103], with the main result that none of the tested selection cuts or adjustments is able to suppress the effect significantly.

4.4 Application to real data

Combining all the knowledge and experience gathered from the studies in simulations, a complete calibration based on atmospheric muon data is outlined in this section. This includes the motivation and description of the procedure for inter-DU time and position calibration as well as the inter-DOM time calibration. Using the results, comparisons to dedicated calibration methods can be performed.

4.4.1 Procedure of a complete muon-based calibration

Following the preceding section, it is clear that a stable inter-DU calibration has to be established before the inter-DOM calibration is evaluated.

The typical starting point consists of the following sources, each providing pieces of information to an initial detector calibration (compare also Chap. 3):

- The xy-positions on the seafloor are determined from the boat during deployment. They usually differ slightly from the planned nominal positions.
- The z-position of the anchor of the DU is determined from the xy-position and a map of seafloor altitudes.
- The inter-DU time offsets are established by sending optical signals from shore and measuring them at the bases of the DUs.
- The inter-DOM time calibration is provided by measurements with lasers in a darkroom.
- The inter-DOM distances on a DU are design values and assumed as nominal. In reality, different distances can be created during mounting of the DU onto the LOM, as mentioned. This can affect the *in situ* observed height scaling of the DU.

Starting from this, carrying out the steps of the following list proves to yield robust and reproducible results. For each step, it is often needed to iterate several times until the values of all DUs simultaneously reach their optimum.

1. Determine stretching factors, as they do not depend on other DUs.
2. Adjust the inter-DU time offsets preliminarily.
3. Determine the orientation, as minimal inflections from subsequently changed xy-positions are expected.
4. Scan for x- and y-position simultaneously. Apply found offsets partly, as described.
5. Search for potential z offsets between the DUs.
6. Establish the final inter-DU time calibration.
7. Control all parameters again simultaneously to see if there is any need for adjustments in this final configuration.
8. Run the hit time residual-based inter-DOM time calibration iteratively

until a stable state is reached. As mentioned, flaws of the preceding steps can be detected here.

This inherently does require a substantial amount of CPU time, as for every iteration in every step a certain amount of data have to be reconstructed several times for each data point in the scan. For this reason, studies were conducted aiming at reducing the resource consumption without sacrificing too much of the precision in the scope of a Bachelor's thesis [104]. In general, about 100,000 events should survive the selection cuts and make it into the final evaluation to get a sufficiently small statistical uncertainty. The efficiency of the initial events is about 70% after cuts.

For the ORCA6 setup, the overall CPU time to produce a single scan is about 400 h. Needless to say that this should be highly parallelized. Utilizing the computing resources available for the KM3NeT experiment, an update of a property for all six DUs can typically be achieved in about 4 h.

In order to yield results that generalize well, the data should be chosen to include many possible data taking conditions to average over. This mainly concerns the different amplitudes and directions of the deep-sea current that cause the detector to sway in the water. A period of at least one month should be considered and more than 20 different runs should be selected spread out over the considered time. Only this way, mean positions and times can be reliably determined.

Instead of independent scans, a minimizer could be developed to optimize the reconstruction quality by varying the DUs' properties and moving accordingly in the parameter space. Still, this would require dedicated reconstructions testing the influence of each change.

To test the robustness of this procedure, it has been applied twice for the ORCA4 detector, each time starting from different initial conditions. The found orientations differ by about one degree and the average difference in the time offset between the two attempts is about 0.3 ns. Only the average distance in the xy-plane is slightly larger with 13 cm but still below the 20 cm precision. It should be noted that with fewer DUs, especially the xy-positions are more challenging to determine precisely. These properties are less constrained, while it is observed in MC studies that time offsets and orientations can be determined rather accurately, even for a two DU case.

4.4.2 Comparisons to static acoustic positioning

There are two ways the results from this method can be compared to the acoustic positioning. First, as described in Chap. 3.2, a static base geometry is deduced from a global fit that follows steps similar to those listed in the procedure utilizing muons. This contains mean positions for each DOM, also averaging over a certain lifetime. Second, there is the dynamic acoustic

Table 4.5: Height scaling factors as found by the muon optimization and the static acoustic fit.

DU	stretching factor muons (%)	stretching factor acoustics (%)
1	1.7 ± 0.2	1.2
2	1.6 ± 0.2	1.6
3	1.3 ± 0.2	1.3
9	1.6 ± 0.2	1.5
10	1.9 ± 0.2	2.0
11	1.7 ± 0.2	1.9

positioning, which updates this geometry for every event and also includes the information from the compasses.

The comparisons between the static geometries are straightforward: The xy- and z-positions as well as the stretching factors can directly be juxtaposed. Starting in the xy-plane, Fig. 4.12 contains a top view of the detector with each DU centered on the position information from the deployment, the mutual starting point for both analyses. Since both methods yield relative positions, the center of the detector has been aligned for all cases for easier comparability. The uncertainty for the deployment information with 40 cm is a rather rough estimate [92]. The acoustic fit's uncertainty is taken from Fig. 3.4, and the error bars for the muon optimization in this section include statistical contributions from the scan as well as the systematical ones found in Tab. 4.1.

In most cases, a satisfactory agreement between muons and acoustics can be observed. The modifications both methods prefer point into similar directions from the starting point and the average distance between the two proposed xy-positions is 12 cm. Only DUs 1 and 11, which are located more separately on either side of the detector, exhibit a slightly more significant discrepancy of about 26 c and 17 cm, respectively. In MC studies, no anomalous behavior is identified for these DUs as far as the capability of reconstructing their xy-position is concerned, hinting at a possible minor systematic deviation in the acoustic positioning procedure.

Table 4.5 contains the estimations of the height scaling factors from the two methods. No uncertainty could be recovered from the intermediate step of fixing the preferred stretching in the acoustic fit procedure. Here again, a very good agreement is found for most DUs within the boundaries of the techniques' uncertainties. Only on DU1 there appears to be a minor disagreement of 0.5%. A difference of 0.1%, as found for some DUs, would correspond to a height difference of the uppermost DOM of about 18 cm. For DU1, the disagreement translates to about 90 cm.

Comparing the determined z-positions, as listed in Tab. 4.6, no obvious

Table 4.6: Shifts to the z-position of a DU as proposed by muon optimization and static acoustic fit.

DU	z offset muons (cm)	z offset acoustics (cm)
1	-10 ± 7	5 ± 6
2	0 ± 8	20 ± 6
3	0 ± 7	-25 ± 6
9	10 ± 6	0 ± 6
10	-10 ± 8	65 ± 6
11	-10 ± 7	-25 ± 6

correlation between the two methods can be identified. Especially the large 65 cm shift of DU10 from the acoustic positioning is not in agreement with the muon-based approach. In general, the modifications from muons are more minor in absolute value, probably due to the time/space degeneracy; since the time offset was adjusted before, a fraction of an initial z offset is already (over)corrected for by adjusting the timing.
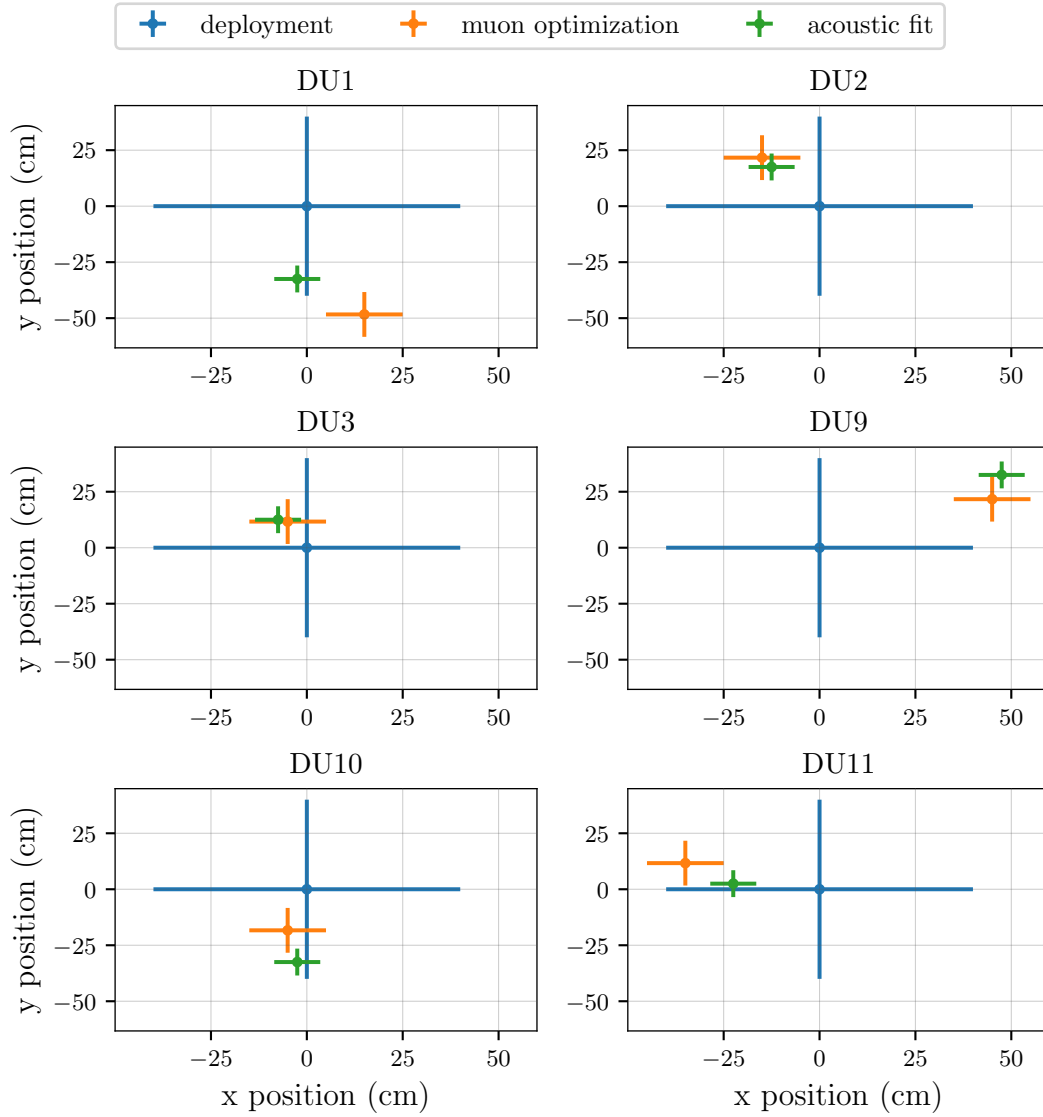
Figure 4.12: Overview of the *xy*-positions from all six DUs in the ORCA detector. Each subplot is centered on the position information measured during deployment (blue), and the reconstructed *xy*-positions from the muon optimization (orange) and the acoustic positioning (green) are shown.
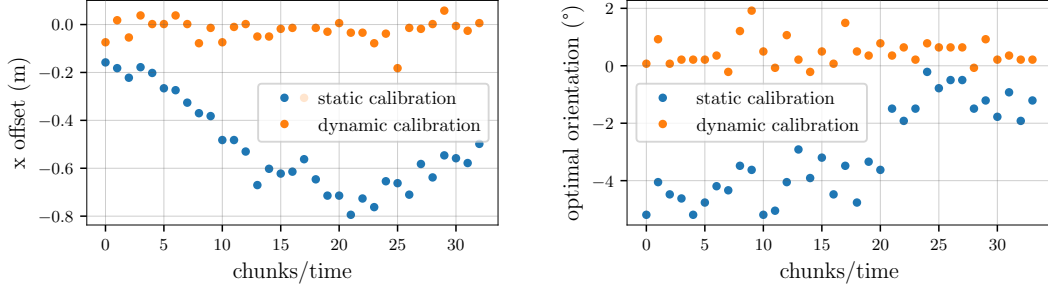
Figure 4.13: Optimal average x offset (left) and optimal orientation (right) for DU9 over one day of data taken under high sea currents. In blue, the analysis was done using the static position information revealing the movement of the DU. In orange, dynamically changing positions are considered, which correct for the time dependent change.

4.4.3 Time-resolved comparison to dynamic positioning

Naturally, there is movement of the detector strings due to the deep sea current. Typically, the DUs oscillate in the water with a frequency of about 17 h, as predicted by the Coriolis force for the detector's latitude. Such a movement can be resolved by measuring the optimal x- or y-position of a DU in a time-dependent manner. For laminar flow, the DOMs of the DU are displaced on average with respect to their nominal position. In a similar way, the DOMs rotate during this movement, as they are constrained by the "ladder-like" form of the Dyneema$^©$ ropes. To make this visible, consecutive data of one day are analyzed with a static geometry and then divided into chunks of 5000 events (like done for the uncertainty estimation). Each chunk corresponds to an independent measurement of the current optimal position or orientation relative to the input. Plotting them successively reveals the movement, as done for the x offset and the orientation of a DU in Fig. 4.13, blue points. Compared to the initial detector calibration, a difference from 20 up to 80 cm for the x-position can be observed for the studied period, chosen because of its particularly high sea current velocities. Needless to say that reconstructing events with the static description in such a scenario reduces performance. For that reason, the dynamic acoustic positioning has been developed by the collaboration, which is able to update the individual DOM positions on an event-by-event basis. Repeating the study of the time dependence like before but now considering this information during the reconstruction of the muon events, the orange points are obtained. They now scatter around zero, predominantly stating an agreement between the updated positions and those preferred by muons of below 10 cm. This continues to be true even for average displacements as large as 80 cm.

A similar behavior is observed for the average optimal orientation of the DU on the right side of Fig. 4.13 in the blue points. Again, considering the updates of the dynamic positioning, the found orientations agree mostly
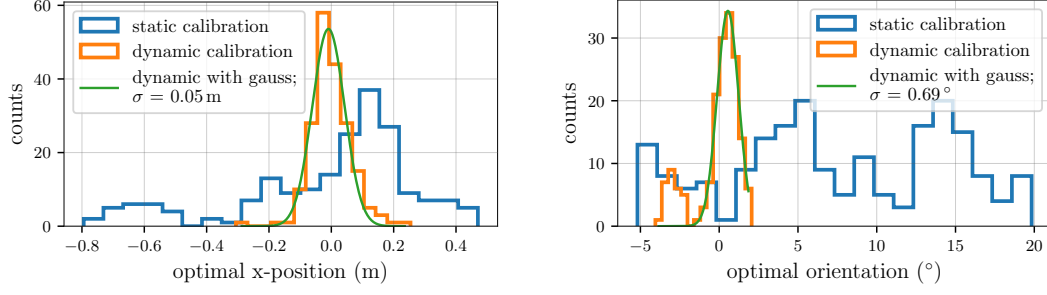
Figure 4.14: Left: Comparison of optimal *x*-positions found by considering a static (blue) and dynamic position calibration (orange). Entries from all chunks of all DUs are combined. The dynamic case is additionally fitted with a Gaussian (green). Right: The same comparison but for the optimal orientation.

within 1° and, in particular, do not exhibit any influence by the underlying movement of the DOMs.

An overall improvement of about 0.7% in reconstruction quality is achieved by using the dynamic calibration over the static one for this period.

In Fig. 4.14, left, the results from all chunks of all six DUs for the *x* offset and orientation are summarized in a histogram. For the *x* offset, contributions from the individual DUs can be made out in the blue curve. They do not peak at zero and reach values from −80 to +40 cm average displacement. In contrast to that, the distribution using the dynamic positioning is centered around zero with a standard deviation of 5 cm, proving an agreement between both methods (and thus precision) far exceeding the requirements.

On the right side of Fig. 4.14, the same exercise for the rotation of the DUs is shown. For the static calibration, there are discrepancies of up to 20° at some points in time. However, using the dynamic calibration yields a peak at zero with a width of 0.69°. Again, this agreement clearly meets the requirements. Interestingly, the contribution of one DU (DU3) diverges from this trend. For it, there appears to be an systematic offset of a few degrees. This is thought to be related to the compasses and their firmware used. An issue with the calibration done onshore of these particular compasses has been identified.

4.4.4 Stretching factor in a changing detector

A unique opportunity to further test the procedure for determining the stretching factor of a DU presented itself by chance. During deployment, the DU is lowered from the boat furled on the LOM. Once arrived at the seafloor and released by a mechanism, it unfurls itself due to buoyancy forces with the LOM floating back to the water surface to be recovered for
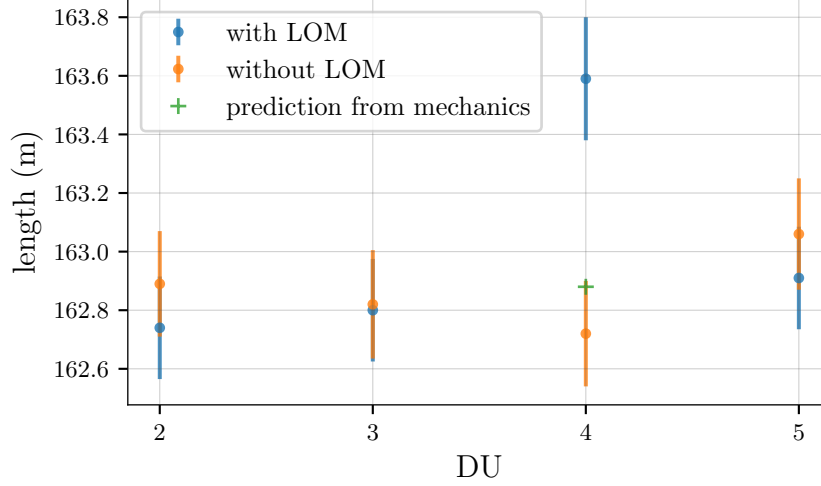
Figure 4.15: Length of the DU, defined as the difference between floor 1 and 18, derived from the optimal height scaling factor. Once determined from a data taking period before (blue) and once after removing the LOM (orange), which added buoyancy forces equal to what is translated to the green cross. From [103] p. 29, modified.

later usage. However, when deploying DU4 in the ORCA4 configuration, the LOM remained at the top of the buoy due to a mechanical failure. This way, it provided an additional buoyancy force, causing the ropes to be further elongated. At a later deep-sea operation, the LOM was successfully removed. The effect of this has been studied in [103] by comparing the optimal stretching factors before and after removal. Considering additional forces of about $2750\,\mathrm{N}$ from the LOM and the stretching of $0.275\,\%$ per $1000\,\mathrm{N}$ measured in the lab [94], the length of the DU should change by about $70\,\mathrm{cm}$. The DU lengths determined with the muon analysis are shown in Fig. 4.15. With about $80\,\mathrm{cm}$ observed difference, good agreement with the predictions from the considerations about the involved mechanical parts is reached. At the same time, no changes for the other DUs are found, as expected.

This example demonstrates that there is indeed an elongation of the DUs in water, which was not probed prior to studies using atm. muons. It also shows that the magnitude of such stretching can be reliably estimated.

4.4.5 Time calibration

To discuss the results of the time calibration provided by muons through maximizing the reconstruction quality for the inter-DU and evaluating the hit time residuals for the inter-DOM time offsets, no complementary *in situ* methods are available at the moment. Still, the time offsets per DU can be checked against the method that measures the signal travel time from

Table 4.7: Changes to the inter-DU time calibration as found by the muons.

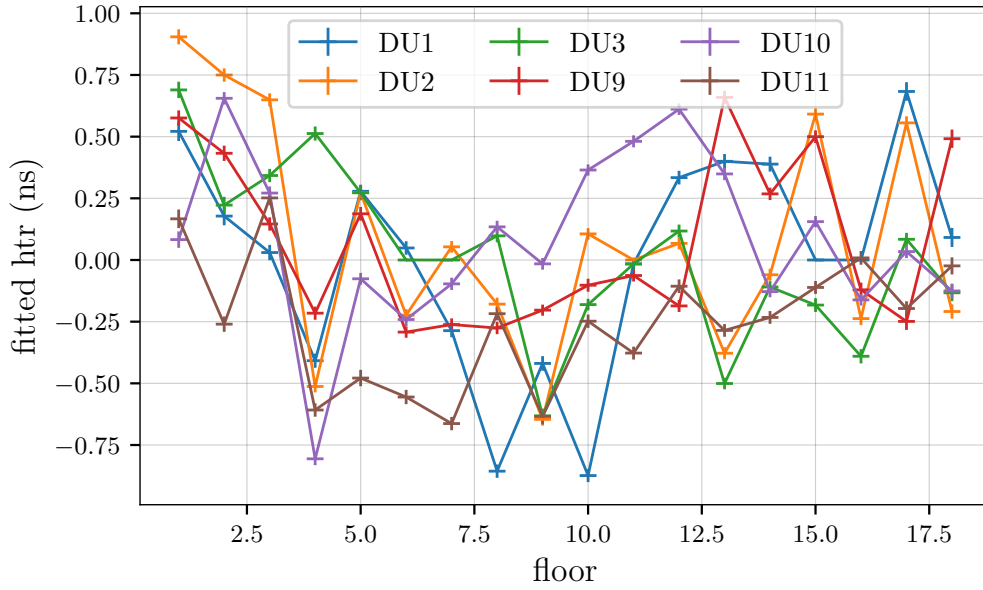| DU	time offset (ns)
1	-0.8 $\pm$ 0.1
2	-0.8 $\pm$ 0.1
3	1.3 $\pm$ 0.1
9	-0.3 $\pm$ 0.1
10	-0.9 $\pm$ 0.1
11	1.3 $\pm$ 0.1

Figure 4.16: Fitted htr for all DUs of ORCA6 versus the floor.

shore to the bases and considers a variety of additional processing times and offsets that together make up the absolute timing of a DU. As Tab. 4.7 suggests, there is decent agreement, as the additional offsets found with muons are smaller than 1.5 ns.

The initial inter-DOM time calibration established in the laboratory is not expected to change significantly during deployment. Thus, more considerable differences, especially those exhibiting a systematic floor dependence or clear outliers, point to insufficient position or inter-DU time calibrations. In Fig. 4.16 the htr for this muon-based calibration are plotted versus the floor for all six ORCA DUs. Firstly, the absolute values are all smaller than 1 ns, basically confirming the darkroom calibration. A slight preference for more positive values at the lower end of the DUs is visible. Compared to the static geometry derived from the acoustic global fit, the most prominent changes to individual z-positions can be found in this region. Running the
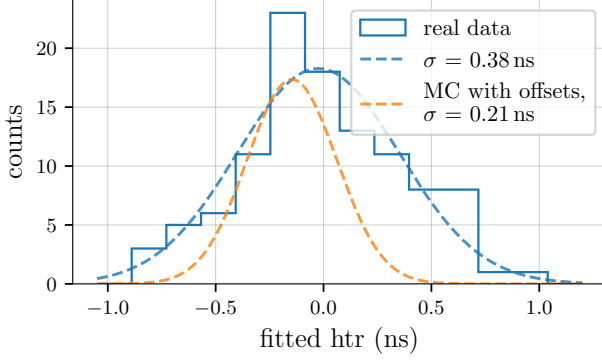
Figure 4.17: Histogram containing the real data htr from all DOMs in ORCA6 (solid blue lines). The distribution is fitted with a Gaussian (dashed blue line) from which the σ is quoted. For a comparison, the expected width of the distribution for offsets in MC (orange, from Fig. 4.8) is included as well.

htr analysis with that geometry as input shows an even flatter distribution without the rise for lower DOMs. Adjusting the distances between DOMs individually in addition to the linear height scaling is not part of this muon procedure due to the large amount of computing involved when scanning properties of single modules.

A histogram with all htr, like in Fig. 4.17 can help evaluate whether the size of the found htr is significant by comparing it to the width such a distribution has for random offsets from Fig. 4.8. From the larger σ of the Gauss fit of 0.38 ns compared to 0.21 ns, the proposed values are significant. Nevertheless, their absolute value is small as far as the required precision of 1 ns is concerned.

4.4.6 Summary for the muon-based calibrations

In this chapter, a method for detector calibration utilizing atmospheric muons has been presented. Systematic studies in simulations have demonstrated that a high precision in time (< 1 ns), position (~ 5 cm) and orientation ($< 1°$) calibration can be achieved, exceeding the experiment's requirements.

For inter-DU correlations, the influences from changes to a central DU on neighboring DUs has been evaluated, which lead to a strategy for a real data application.

In a similar way, a robust method for the determination of the inter-DOM time calibration based on the low-level observable of hit time residuals has been presented. Being able to deal with different kinds of tested offsets, it also achieves a sub nanosecond accuracy.

The application to real data allowed for several interesting comparisons, as different methods typically have different sets of systematics. For the mean x-position of the DUs, for example, an excellent agreement compared to the dynamical acoustic positioning system of 5 cm has been found. Equivalently, the DU orientations agree within less than one degree between the two approaches.

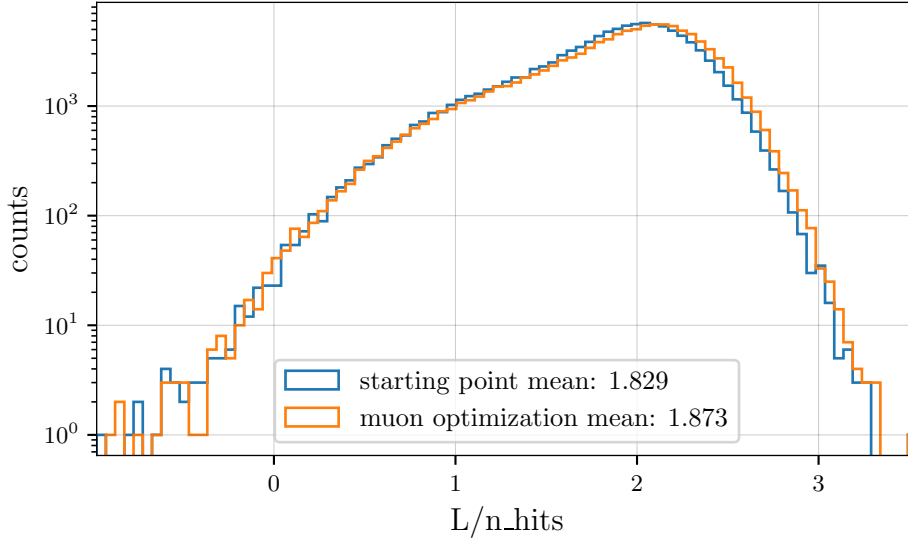For the z-direction of the DUs, the height scaling was precisely

Figure 4.18: Distribution of the reconstruction qualities before (blue) and after (orange) the complete muon optimization. The mean quality could be improved by 2.4%.

determined. The factors for two periods were compared, while in one period there was additional buoyancy force for one DU. The resulting change in length agreed well with predictions based on the properties of the mechanical components involved.

The existing time calibrations (inter-DU and inter-DOM) were generally confirmed and slightly adjusted.

As a final remark on this chapter, the improvement achieved between the starting point (deployment info, laboratory calibrations, nominal values) and the final detector description, including all presented muon adjustments, is shown.

Unlike preceding versions of the ORCA and particularly the ARCA detector in which the muon methods significantly contributed to fixing major issues of miscalibration, no unexpectedly extensive modifications were identified for ORCA6. Still, after confirming and updating the initial calibrations, a noteworthy improvement in reconstruction quality of 2.4% can be observed in Fig. 4.18. Here, especially medium to good quality tracks (likelihood per number of hits of 1-2) were improved to excellent quality tracks (>2), which is the most interesting kind of events for physics analysis.

Looking at the htr in particular, Fig. 4.19 shows the fitted htr for some test data after a single iteration. The distribution now is much narrower with a standard deviation of 0.1 ns, while before, with 0.5 ns, significantly larger values were found. Note that in addition to Chap. 4.4.5, the comparison
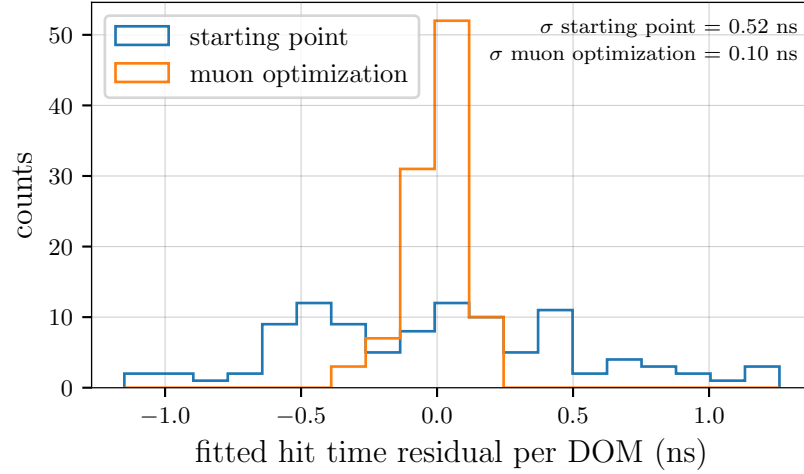
Figure 4.19: Fitted htr as found after a single iteration comparing before (blue) and after (orange) the complete muon optimization. This quoted σ is the statistical standard deviation and not the width from a Gauss fit.

here contains all influences from the altered positions and inter-DU timings on the htr as well.

Chapter

5 Introduction to deep learning and graph neural networks

MACHINE leaning and in particular deep learning have become widely used in the physics community over the past years. They often offer elegant solutions to complex problems, improve the resolution of reconstruction parameters and thus the sensitivity to the physics behind them, and even pave new ways for analyses that are not possible with classical methods. An exhaustive overview for applications in particle physics can be found in [105].

One such application is the use of graph neural networks (GNNs) in KM3NeT [106]. In the scope of this thesis, event classification and reconstruction using GNNs have been studied with the aim of performing a neutrino selection based on its outputs. Finally, some oscillation analysis is conducted on this neutrino set to measure directly the impact of utilizing the deep learning technique. This work builds on the experience gained by Michael Moser who used convolutional neural networks in simulations of the full ORCA detector [107].

In this chapter, a short introduction to deep learning and the implementation of GNNs used in this specific application is given. A more complete description of deep learning can be found in the textbooks [108] and [109]. A review specifically of GNNs used in high-energy particle physics is given in [110].

5.1 Introduction to deep learning

In the recent years, the term "Artificial intelligence" as a buzzword has seen frequent application, not only in computer science but also politics, the private sector and many fields of research. One part of it is machine learning, in which an algorithm is instructed to improve its performance on some task by merely processing data, thus learning from it without

being explicitly programmed to solve the problem [111]. To that end, many different varieties have emerged each specializing on certain tasks, such as decision trees, random forests, linear regression, gradient boosting trees and neural networks.

The latter have come a long way since their invention: As early as 1943, the concept of a computational neural network was first envisioned [112] and later formulated by D.O. Hebb [113]. 1954 this idea was first executed on computational machines [114], before Frank Rosenblatt in 1958 defined the *perceptron*, the first artificial neural network (ANN), trying to model a nervous system to study questions about signal perception and memory storage [115]. Despite the continuous developments in the rapidly growing field of computer science, neural networks would not be widely used until the 2000's, as the computational resources did not allow for complex networks. Also, data to learn from were difficult to compile in the pre-internet era and the performance could not compete with dedicated classical approaches for solving the problems.

The limitations were lifted by implementing the efficient use of graphic processing units (GPUs) in the training, as it allowed for parallelizing the numerous matrix operations (see Sec. 5.1.1). This sparked the *deep learning revolution* in 2011 to 2013 when, for the first time, these models were outperforming conventional methods as well as shallow learning approaches in tasks such as image, speech and object recognition. Here, in particular the introduction of convolutional neural networks (CNNs) (see Sec. 5.2.2) helped to push the envelope.

The term "deep" in this context refers to the number of *layers* (neurons operating in parallel) adding to the complexity of the network and thus its capacity to learn complicated tasks. More specifically, this structure allows for extracting hierarchical features, of which the highest level (last layers) can be directly connected to the physical label. In this way, deep learning is an important, quickly evolving part of machine learning.

Shallow learning, on the contrary, comprises of models with less complexity and the features typically have to be chosen by the user in advance.

In this thesis, supervised learning is utilized, i.e., the data the algorithm learns from are labeled with the true information about their class or parameter. The algorithm can then compare its current prediction to the true label and adjust itself accordingly to improve the accuracy of the predictions. There also exists the possibility to let algorithms train on unlabeled data, called unsupervised learning, in which patterns and clusters, for example, have to be identified.
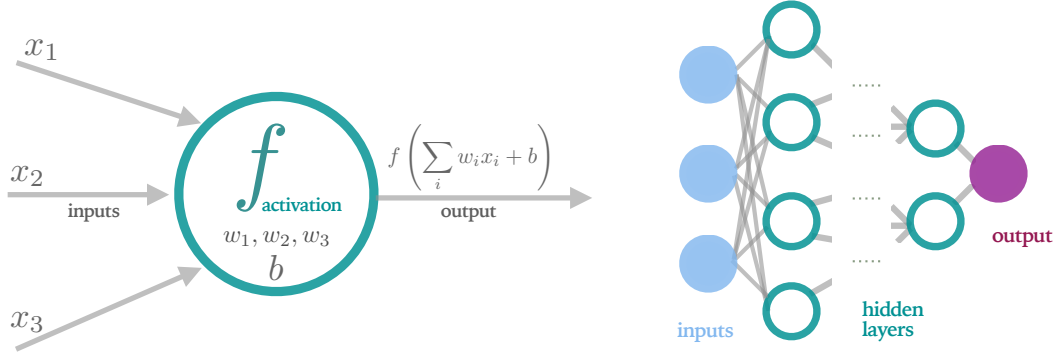
Figure 5.1: Left: Schematic drawing of a neuron (green) receiving inputs from various connections. The neuron's output is created considering the weights of each connection and a constant bias term mapped by a non-linear activation function. Right: Schematic overview of the different types of layers in an ANN. From [116].

5.1.1 Artificial neural networks

An ANN can be thought of as a complex, non-linear function $F(x)$ that maps some input vector x to some output y. The *input layer*, consisting of real numbers, is connected to consecutive layers of so-called *neurons* (see Fig. 5.1). Each neuron of these *hidden layers* sums up the inputs from its connections and maps it via an activation function to create an output of its own. This is then again passed on to neurons of the next layer. Each connection between neurons is characterized by a strength, the *weight w*. In addition to the sum of all incoming connections, which are the inputs multiplied by the weights, wx, a *bias* term b is added to the input of the activation function f. For an individual neuron the output a is then

$$a = f(\sum_i w_i x_i + b). \tag{5.1}$$

In order to enable the predicted variable at the output layer to have a non-linear relation to the inputs (i.e., learn complex features), the activation function must introduce a non-linearity. Functions as simple as the rectified linear unit (ReLU),

$$f_{\text{ReLU}}(x) = \max(0, x), \tag{5.2}$$

are commonly used due to their simplicity and thus efficient computation.

In the case of allowing connections between all neurons of two layers they are called *fully connected*.

By now, there exists a large number of different, more sophisticated *architectures* (type and arrangement of layers) specialized on specific tasks, such as convolutional neural networks (see Sec. 5.2.2), recurrent neural networks [117], or autoencoders [118], with the latter being an example of unsupervised learning.
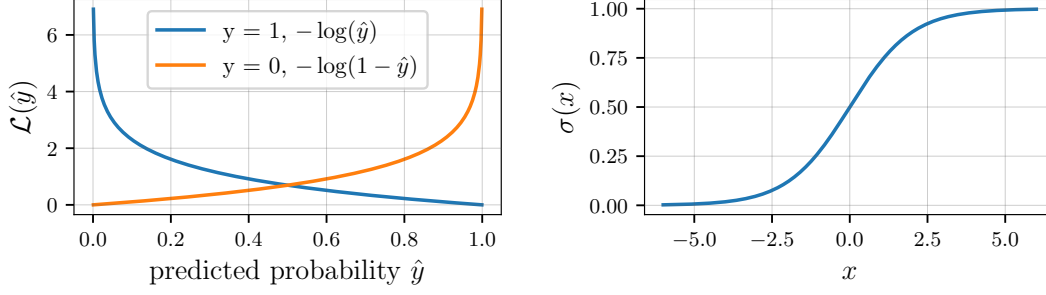
Figure 5.2: Left: Log loss of the two classes ($y = 1$, blue, and $y = 0$, orange) depending on the predicted probability $\hat{y}$. Right: Sigmoid function, also known as logistic function.

5.1.2 Training of neural networks

So far, only the *forward pass* has been considered, i.e., how a response of an ANN to some input is calculated. Before networks are able to output any meaningful answer, the weights and biases of all connections need to be adjusted in the process of *training*. To this end, a *loss* is defined as a quantity related to the difference between currently predicted and true value of the label. In the so-called *back propagation*, the influence of each weight and bias on the loss is evaluated, allowing to modify it accordingly.

Loss functions

Defining the loss function is specific to the problem at hand. In classification tasks, a type of *cross entropy* is often used while regression is commonly tackled utilizing mean squared or mean absolute error [109].

The prefered loss description for a binary classification, i.e., there are only two classes, is the *binary cross entropy* or log loss, which is defined as

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} y_i \ln(\hat{y}_i) + (1 - y_i) \cdot \ln(1 - \hat{y}), \tag{5.3}$$

with y being the label (0 or 1) and $\hat{y}$ is the predicted probability from the network. Since only 0 or 1 are possible values for y, the expression simplifies for each individual training sample to the fist (second) term only for $y = 1$ ($y = 0$), leaving only the respective logarithm. The total loss is then calculated from all N training samples. In Fig. 5.2, left, the loss of the two classes $y = 1$ and $y = 0$ is shown for different predicted probabilities. As can be seen, the penalty is zero for when the true label and predicted probability are the same and exponentially increases when diverging from the true label.

Typically, good performance is achieved by choosing the *sigmoid* activation function at the output neuron, which has the form

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{5.4}$$

and is drawn in Fig. 5.2, right. The sigmoid function maps $\mathbb{R} \to (0, 1)$ yielding the desired, normalized range for $\hat{y}$ because of σ's symmetry with respect to the origin.

For regression purposes, i.e., reconstructing a possibly unbound property, the simple *L1* or *L2* loss can be used, also referred to as mean absolute error

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \text{abs}(y_i - \hat{y}_i) \tag{5.5}$$

and mean squared error

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2}(y_i - \hat{y}_i)^2. \tag{5.6}$$

Here, the penalty the network experiences grows linearly (quadratically) with the deviation from the predicted value $\hat{y}$ to true value y. Yet, expanding from this, more sophisticated losses can be defined, as will be introduced in Sec. 5.2.3.

Backpropagation algorithm

With the definition of the loss function as a quantitative measure for the accuracy of the prediction, the adjustable parameters of the network can be tuned during training. In principle, the influence on the total loss of each individual weight and bias has to be evaluated and the parameter changed accordingly to improve the loss. An efficient way to achieve this is the backpropagation algorithm [119], which is essential to the understanding of how neural networks learn. This section follows [109], which draws a very comprehensible picture.

Assume the output of the jth neuron in layer l of a neural network to be

$$a_j^l = \sigma \left(\sum_k w_{jk}^l a^{l-1} + b_j^l \right), \tag{5.7}$$

where k is the number of neurons in the $(l-1)$th layer and σ can be any activation function. Generalizing this to all entries in j and k the following vector/matrix from is yielded:

$$a^l = \sigma(w^l a^{l-1} + b^l). \tag{5.8}$$

This recursive representation illustrates how the activation of a neuron in a layer is influenced by activations on the previous layer a^{l-1}, the weights between them w^l and its own bias b^l.

To simplify notation even further, $z^l \equiv w^l a^{l-1} + b$ is the weighted input to layer l. So, Eq. 5.8 turns into

$$a^l = \sigma(z^l). \tag{5.9}$$

Now, the influence of the weights and biases of layer l on the loss, $\partial \mathcal{L}/\partial w^l$ and $\partial \mathcal{L}/\partial b^l$, has to be expressed in terms of computable quantities. For this, first, δ^L is the error in the output layer L and can be defined for every component j as

$$\delta_j^L = \frac{\partial \mathcal{L}}{\partial a_j^L} \sigma'(z_j^L). \tag{5.10}$$

Here, the first term gives the change of loss to a change in the activations of the output neurons right at the end of the network. The second part, $\sigma'(z_j^L)$, is the derivative of the activation function evaluated at z_j^L, which is simply a real number. In fact, the derivative $\sigma'(x)$ is fixed and already known when setting the activation function, and z_j^L is already computed during the forward pass.

Rewriting Eq. 5.10 in matrix notation and using the Hadamard product $\odot$, a strictly element-wise vector multiplication, it becomes

$$\delta^L = \nabla_a C \odot \sigma'(z^L). \tag{5.11}$$

Secondly, for preceding layers, the error δ^l depends on δ^{l+1} and is

$$\delta^l = ((w^{l+1})^T \delta^{l+1} \odot \sigma'(z^l) \tag{5.12}$$

using the chain rule.

Interpreting this equation and starting with the error of a layer closer to the output, δ^{l+1}, (which could be L, Eq. 5.10), the error is first propagated backwards through the transpose of weight matrix $(w^{l+1})^T$ (so between layer $l+1$ and l) and then further through the activation function at layer l, thus arriving at the neuron in this layer and yielding its local error.

It can be shown ([109], p. 48) that

$$\frac{\partial \mathcal{L}}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial b_j^l} = \delta_j^l, \tag{5.13}$$

which finally describes the changes to the loss in dependence of the change in the tunable parameters utilizing only known expressions. In particular, the change by the weights of a layer depends solely on the activation of all neurons in the layer before, a_k^{l-1} ("incomming"), and its error δ_j^l

("outgoing"), allowing for the following compact representation omitting indices:

$$\frac{\partial \mathcal{L}}{\partial w} = a_{\text{in}} \delta_{\text{out}}. \tag{5.14}$$

With equations 5.13 it is now possible to calculate the gradient locally ("propagate errors back") for every weight and bias in every neuron and layer and then optimize them in some form of gradient descent. Still, even in this elegant and efficient algorithm, many matrix multiplications and inversions have to be computed, which only can be sufficiently parallelized on GPUs.

Minimizing the loss

In practice, there are a few things to note about the way the gradient descent is realized efficiently.

Firstly, the backpropagation is not executed for each individual training sample all at once, but instead a *minibatch* of the data (e.g. 32 samples) is processed and an average update to the weights and biases is deduced from the batch. This is referred to as *stochastic gradient descent* (SGD). In addition to the faster convergence when using minibatches over the complete set simply because of more frequent updates, there is the practical limit to the amount of data that can be stored in the memory of the GPU.

Secondly, there is a factor with which the proposed updates are scaled, called the *learning rate.* It is similar to the step size in other minimizers. If its value is large, features are learned quickly by the network, especially at the beginning of training. But if it is too large, minima in the loss latent space are overshot, not allowing for further improvements or even no convergence at all. This is why often the learning rate is lowered with the number of *epochs.* An epoch is concluded when the complete training data set was processed once. If the learning rate is too small, the training will be inefficiently slow or easily be restricted to local minima.

Thirdly, there exist extensions to SGD introducing a momentum to the direction of the gradient. The most commonly used and robust one is the Adam optimizer [120], which considers the gradients from earlier updates when estimating the current one. Thus, it is able to speed up learning during earlier phases and slow down the learning rate later to better find smaller minima of the loss function. Even with the additional bookkeeping, it is still able to maintain efficiency, for both speed and memory usage.

An additional common technique that in many use cases speeds up convergence and even increases performance by allowing for better generalizability is to introduce a reduction of the dimensions in latent space, called *pooling.* Average pooling, for example, takes the dimensions of a feature map and returns half of the dimensions initial lengths, averaging over the features ("down sampling").
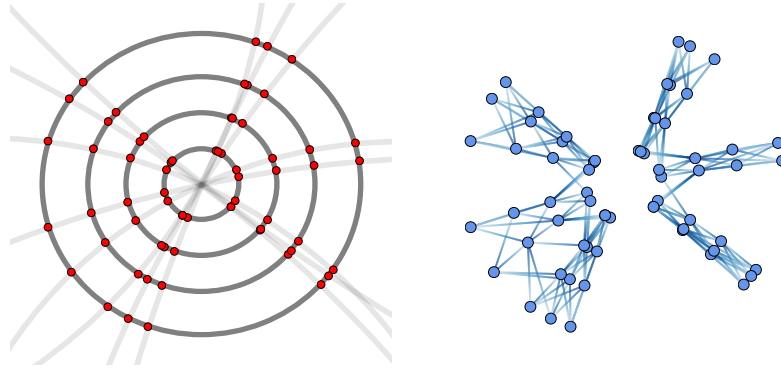
Figure 5.3: Left: Hits on different segments by particles (red points) emerging from a central interaction in a tracking detector. Right: This kind of data can be represented as graphs in which the point cloud data is causally connected. From [125].

Finally, there are countermeasures against vanishing gradients. As seen in Fig. 5.2, right, activation functions such as sigmoid saturate for input values with large $|x|$, meaning the derivative of the activation function at that input σ' in Eq. 5.12 becomes close to zero. In this case, no updates can be propagated to the weights of the neuron, thus no learning is achieved. To avoid this during initialization of the weights at the very beginning, the weights are typically sampled from a Gaussian distribution with a width inversely proportional to the number of connections of the neuron [121] and centered around zero. For the later training, *batch normalization* [122] is commonly used. Here, the weighted input at a neuron (z) is normalized before being mapped by the activation function. The scaling of this normalization is a learnable parameter.

5.2 Graph neural networks in KM3NeT

As with many other fields in machine learning, graph neural networks have been developed a while ago in 2005 [123] but only recently gained interest when they were discovered to solve specific tasks particularly well. Among these are the domains of chemistry, natural language semantics and social networks [124], whose data are point clouds that inherently resemble a graph structure. Also physics data are often point clouds in 3d space holding additional, measured quantities, like a deposited energy, that are causally connected. For example, hits in segments of a tracking detector, as drawn in Fig. 5.3, or in cells of calorimeters can be represented in graphs with connections between the hits.

Since the data of the neutrino telescope KM3NeT also have many features of point clouds, it is only natural to exploit graph neural networks for the
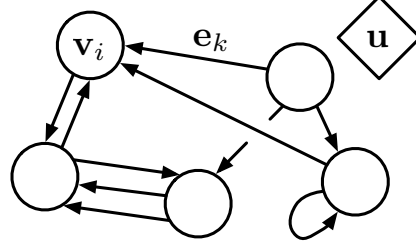
Figure 5.4: A directed graph with a global feature u, consisting of nodes v_i that are interconnected by edges e_k. From [126].

analysis. Consequently, GNNs will be first introduced in general and the specific implementation used in this work is detailed thereafter.

5.2.1 Structure of graph neural networks

The distinctive feature of graph neural networks is the way information is extracted from the input data. It should be noted, however, that there exists a multitude of implementation variants of which only one general one is introduced here. The structure of the input is the aforementioned graph that follows the hierarchy of

- a *global* feature u, which is the entire graph and some physical meaning connected to it,
- the *nodes* v_i, which are the data points that each can have several attributes (time, position, etc) and
- the *edges* e_k, which connect pairs of nodes for their attributes.

An overview of this is drawn in Fig. 5.4. The edges are directed and allow for flow of information between the nodes in certain directions. For smaller absolute sizes of the graphs, each node can be connected to all other nodes. But since the number of edges grows quadratically with the number of nodes, it is often useful to restrict the information flow to k-nearest neighbors. These are typically selected by the geometrical distance between the nodes.

Another characteristic feature of GNNs is the plasticity of the graphs [126]. While the edges are defined at the input layer by the data, they can be different for hidden layers. This is because of how the network is instructed to update the graph through the layer(block)s. The principle is illustrated in Fig. 5.5:

- First, the edges are updated, meaning a e_k' is computed, which can be understood as a "message". Updating in this case means that the parameters of a small kernel network are adjusted to enforce or weaken certain edge features.
- Then, aggregating all incoming edges to a node, its features are updated yielding v_i'. This aggregation can be similar to the convolution in convolutional neural networks.

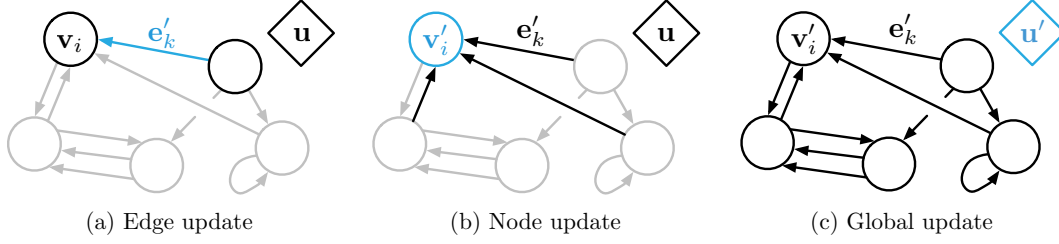(a) Edge update   (b) Node update   (c) Global update

Figure 5.5: Scheme of how information can be extracted from graphs, starting on the small-scale edges, which allow information exchange between nodes (a). All edges are aggregated to update the node features (b) and eventually the entire graph is updated (c). From [126].

- Finally, all nodes are updated and together with the edges constitute the new graph with global attributes u'.

This way, edges in the latent space can hold features learned by the network and thus nodes have dynamic relations to one another. The fact that most of the feature extraction takes place on the small-scale edge and node level is believed to add to the generalizability of GNNs [127].

5.2.2 Comparison to convolutional neural networks

Convolutional neural networks have a slightly longer tradition in computer science [128] and also high-energy physics [129] than GNNs. They are particularly efficient in solving problems involving fixed spatial relations, like images. The main concept here is that filters with a kernel size of a few pixels are slid over the input image and a feature map is extracted as output of a layer block. In the same way as seen for the GNNs where the output in latent space is again a graph, the output feature map can be thought of as a "picture" and then serves as input for the next block. Stacking layer blocks on top of each other allows for extracting higher level features that are eventually connected to some physical properties.

The fundamental difference between CNNs and GNNs of how the input is processed is visualized in Fig. 5.6. For 1, 2 or 3-dimensional convolutions, the relations to fixed neighboring pixels are considered in the aggregation, while for graphs the neighbors and their relations are dynamic.

Furthermore, the need for an image-like input for CNNs often means for physics data that they have to be binned into some finite resolution. For the position of hits in a detector, for example, a reasonable amount of bins per dimension has to be chosen, while the exact position within a bin is not known to the network. For GNNs, however, the precise information is fed to the input as the attributes of the nodes, adding no further smearing.

Another particularity for CNNs is that they require a rectangular, preferably a square, input picture due to the fixed-sized filter. The strategy to mitigate the problem with data that are shaped differently
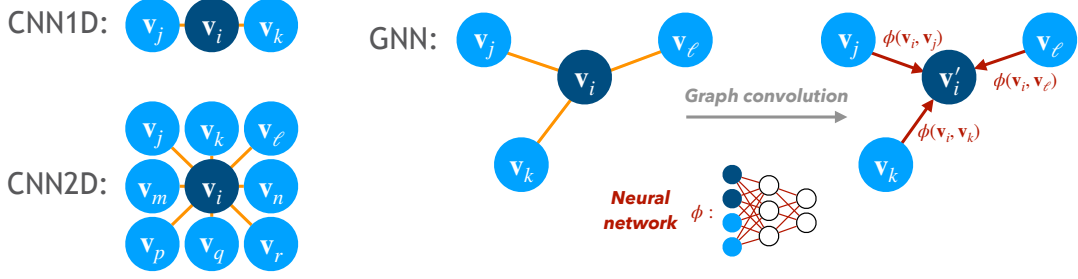
Figure 5.6: Comparison of how CNNs and GNNs process their input data. The relations for image-like data is fixed by definition, while graphs have no inherent spacial orientation. Modified from [110].

is called *zero-padding*, in which empty pixels are added around the initial signal. They do not hold additional information, which does not negatively affect performance but creates additional overhead in the computation. Additionally, for sparse signals in higher dimensions often a large amount of empty bins are considered. Graph neural networks have the advantage that their convolutions inherently only operate on existing nodes.

Moreover, CNNs are currently restricted to 3d convolutions (in implementations based on the popular tensorflow), meaning the filter is at maximum three-dimensional, allowing for 4d data maximally. For many physics applications though, the number of attributes per node is larger than four. For example, an energy deposit of a certain amount, at a certain position in the detector at a certain time is five-dimensional. In order to still utilize CNNs, compromises have to be made sacrificing a spacial dimension, for example, or stacking information of different dimensions. As mentioned before, the attributes of the nodes for GNNs can have any number and only increase the number of components of the edge vector and with it the computing time.

One advantage for CNNs, however, is the aforementioned wide use and with it the larger pool of curated implementations as well as the optimization to many use cases. For the relatively new GNNs, an efficient implementation of the edge convolution, tailored to the needs of the KM3NeT data structure, had first to be written. The details are subject of the next section.

5.2.3 Implementation details

The architecture of the networks used in this thesis is based on [130] and illustrated in Fig. 5.7. As can be seen on the left side, the network consists of three edge convolution blocks, a global pooling layer to reduce the dimensions and eventually fully connected layers that generate the output specific for the task via a final activation function.

Starting at the input, two quantities, the *coordinates* and the *features* have to be defined. The coordinates are only involved in the determination
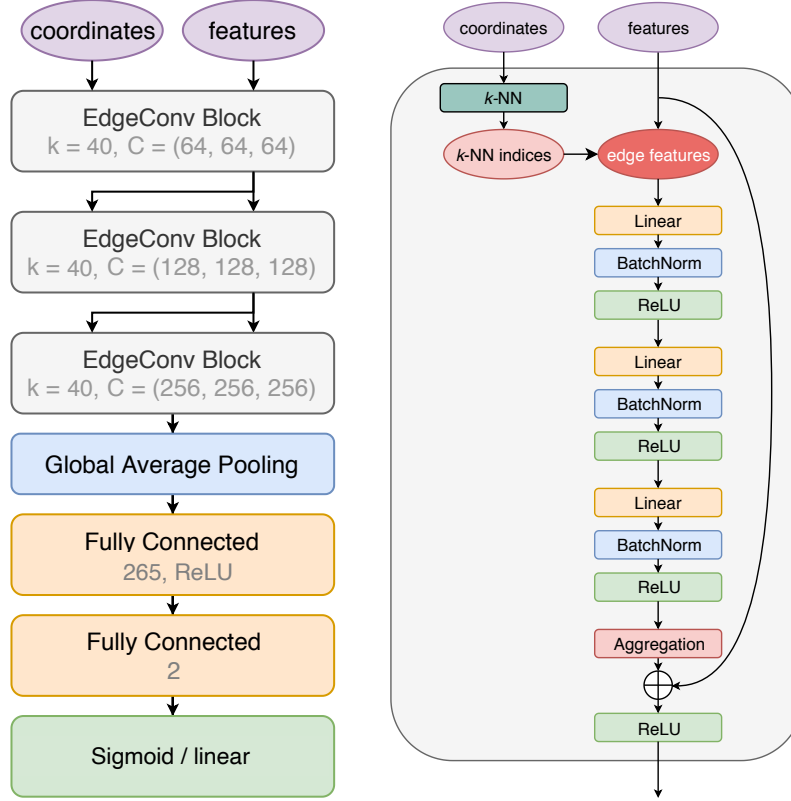
Figure 5.7: Left: Diagram of the used network architecture. The node coordinates and features are passed to three edge convolution blocks after which the extracted features in latent space are connected to some output. In the edge convolution block, k denotes the number of nearest neighbors used and C the number of neurons per layer. Right: Close up of the inner structure of an edge convolution block that updates the graph. Modified from [130].

of the k-nearest neighbors (k-NN). Thus, the hit information used for the coordinates is $\in \{ct, x\text{-}, y\text{-}, z\text{-position}\}$ (Euclidean distance). In these studies, k-NN is set to 40. In the features, the information for each node and edge is defined. The node features are simply all pieces of hit information, including the direction of the PMT, and the edge features are the difference between the neighboring nodes v_i and v_j, i.e., $e_{ij}^k = v_j^k - v_i^k$ where $k \in \{\text{time}, x\text{-}, y\text{-}, z\text{-position}, x\text{-}, y\text{-}, z\text{-direction}\}$.

The edge convolution block is further detailed on the right side of Fig. 5.7. First, the k-NN are calculated that should be considered for the update of the node. From the features, a small kernel network is trained, consisting of three layers with 64 neurons each, accepting the input linearly, applying batch normalization and utilizing the ReLU activation. The changes from all nodes and edges to this kernel network are averaged for one training sample, constituting the convolution. At the end of the block, an updated graph is returned with newly defined nodes and edges. For the output
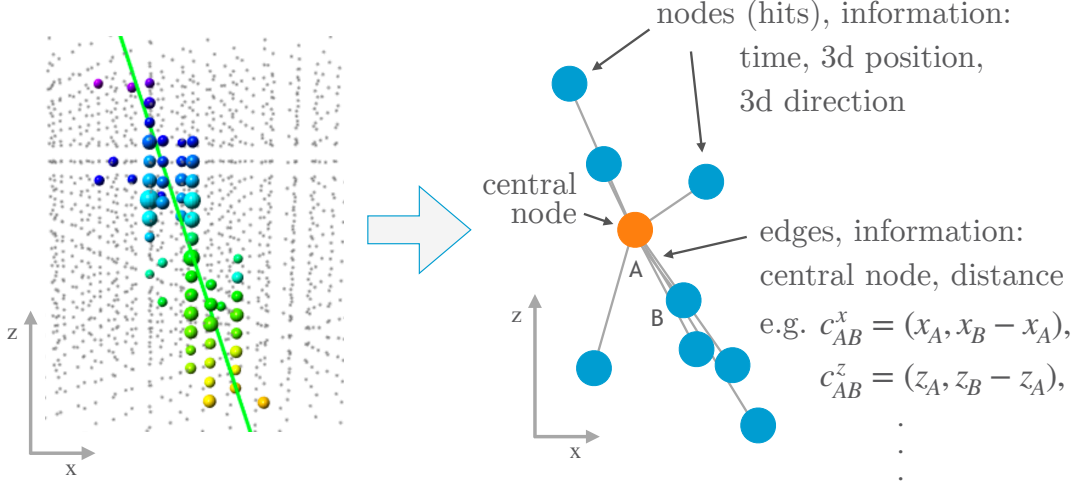
Figure 5.8: Illustration of how KM3NeT data are presented to graph neural networks; left, the original hit distribution as caused by a muon track, and on the right the resulting graph holding the hit information in the nodes (blue, not displaying all nodes). Indicated is the update of a central node (orange) for which information from all edges (connections to other nodes) is aggregated. The event display on the left from [48].

graph, the original features are additionally available via a *shortcut* and can be partly merged with the update.

The second and third edge convolution blocks then receive the graph in latent space from the preceding block and process it the same way. This way, the graph is dynamic throughout the network pass and edges can be redefined repeatedly. The number of neurons per layer increases to 128 and 265 to allow for storing higher level features.

In the context of presenting data form the KM3NeT detectors to such networks, the input node attributes are the time, 3d position and 3d direction of the hits registered on the PMTs, making it seven-dimensional. An illustration shows an update of a central node in Fig. 5.8.

Loss functions

For classifiers, the cross entropy mentioned in Chap. 5.1.2 is also used in this studies to describe the loss of a sample. For the regression tasks, the networks are instructed to not only predict a most probable value but an uncertainty on it as well. This is achieved by defining a loss composed of a distribution with mean value μ and width σ.

A hint in order to find a fitting distribution comes from the cross entropy for categorization in which the loss is chosen to be the negative logarithm of the probabilities the network predicts (see Eq. 5.3) [131]. In the same way, the loss for regression can be defined for a function $f(y, \hat{y}(\mu, \sigma))$ that

describes the expected distribution of the predictions $\hat{y}$ that scatter around μ with σ as

$$\mathcal{L} = -\ln(f(y, \hat{y}(\mu, \sigma))). \tag{5.15}$$

A natural choice for predictions that fluctuate purely statistically around some mean value is a Gaussian distribution

$$\hat{y}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right). \tag{5.16}$$

The resulting loss is then given by

$$\mathcal{L} = \ln(\sqrt{2\pi}) + \ln(\sigma) + \frac{(y-\mu)^2}{2\sigma^2}, \tag{5.17}$$

while for the minimization only the terms including μ and σ are of interest, i.e.

$$\mathcal{L} = \ln(\sigma^2) + \frac{(y-\mu)^2}{\sigma^2} \tag{5.18}$$

when also multiplying by 2. This is the log-normal distribution. For the special case of $\sigma = 1$, the mean squared error from Eq. 5.6 is recovered. With this loss description and two output neurons, one for the μ and one for the σ, the network is instructed to optimize a μ that is closest to the true label y and a σ that denotes the $1\,\sigma$ uncertainty interval of the prediction.

Software used

This work heavily profits from software built by the KM3NeT collaboration and the deep learning community in general.

Preprocessing, i.e., transforming the event hit data into graphs, is achieved with *orcasong*. This and subsequent internal software utilize *km3pipe* for sequential processing of data, including the conversion from ROOT to h5 files.

For handling the training and prediction, *orcanet* is used. It is based on the *Keras* and *tensorflow* [132] libraries.

The implementation for the edge convolution algorithm, *medgeconv*, was written by Stefan Reck [133].

6

Training and performance evaluations of neural networks used

Applying the implementation introduced in the preceding chapter, this part will present for which applications in particular graph neural networks are trained and how. Thereafter, studies of the performance of these networks on some test data are showcased. Often, the terms *track neutrinos* and *shower neutrinos* will be used to refer to the two groups of neutrino interactions that create different signatures in the detector (see Chap. 1.5.2). Namely, track neutrinos have a muon in their outgoing channel, i.e., muon (anti) neutrino CC and tau (anti) neutrino CC with a muon emerging from the tau's decay ($\sim 20\%$). Shower neutrinos are electron (anti) neutrino CC, all NC and tau (anti) neutrino CC interactions emitting no muon in the decay.

6.1 Preprocessing and training

The general approach to provide the possibility of a fully GNN-based analysis is to employ classifier networks for particle identification and regression networks for event reconstruction. More specifically:

- One network learns to distinguish between signal (neutrino induced) and background events (atmospheric muon or random noise), yielding a *neutrino probability*.
- Another network is trained to characterize events as tracks or showers by outputting a *track probability*.
- A further network is instructed to reconstruct the direction of the incident neutrino.
- Finally, one network predicts the energy of the interaction.

In principle, it is possible to use a single network for the classification with several categories. A comparison shows that better performance is achieved by splitting up the tasks and letting the network distinguish between two

binary, opposite categories. Direction and energy reconstruction tasks are split as well due to reasons that are discussed shortly.

As per usual, the raw data from a real or simulated detector need to be preprocessed to serve as input for neural networks. As mentioned before, a custom python package is used to write all hit information of an event into a graph in which the nodes are the PMT hit coordinates in time, space and direction and consequently the edges contain the differences to neighboring nodes.

In KM3NeT, MC data is produced in a *run-by-run* fashion, meaning the data taking conditions are considered in the generation of the simulation. For training neural networks on such data, it is important to select a training set that encompasses a large variety of conditions in order to allow them to generalize well. To this end, dedicated simulations for the training have been produced spanning the same lifetime as the data that will be later used in the analysis.

From this, specific training sets for each application are created providing a balanced set between the categories for the classifiers. For regression, balance is less of a concern as quantity is, which is why more events without restrictions can be used. This is supported by the fact that no apparent correlation between the number of events in a region of direction or energy used during training to the resolution for that region is found. In the case of the energy reconstruction, the training set is limited to showers, lower-energetic muon CC and lower energetic tau interactions ($E_\nu < 50\text{GeV}$). This is motivated by physics, as larger energetic muon tracks are longer than the dimensions of the detector with their length of $4\,\text{m/GeV}$. This makes training the network with higher-energetic events inherently more challenging, as partly contained events with vastly different true energies create similar amounts of photons in the detector volume. Thus, confusion is created and the parameters of the network either do not converge or output a mean energy for all non-contained events to best fit them on average. Showers produce their light in a much smaller volume, meaning they are either mostly contained or not triggered at all when further away from the detector.

The complete overview of the number of events used in each case can be found in Tab. 6.1. The set for the signal/background classifier contains all kinds of neutrino interactions in the signal class, including taus. The track/shower classifier only consists of (anti) muon neutrino charged current interactions for the track and (anti) electron charged current plus any flavor neutral current interactions for the shower class. For the energy reconstruction, atm. muons are excluded, like the the mentioned higher-energetic tracks, while again all kinds of events are used for the direction reconstruction.

For the two classifiers, the true labels for the networks to learn from and calculate the loss with are determined directly from the particle

Table 6.1: Number of events used in training and validating in thousands for each application. The first value indicates the absolute number of events in the training and the second value the number in the validation set, with fractions in brackets.

	signal-background classifier	track-shower classifier
track neutrinos	615 (30.3%) / 158 (30.8%)	627 (50.8%) / 155 (50.6%)
shower neutrinos	394 (19.4%) / 98 (19.1%)	608 (49.2%) / 151 (49.4%)
atm. muons	684 (33.7%) / 173 (33.8%)	0
random noise	337 (16.6%) / 83 (16.3%)	0
total	2,031 / 514	1,235 / 306

	direction reconstruction	energy reconstruction
track neutrinos	1,013 (49.2%) / 260 (49.2%)	601 (48.6%) / 149 (48.4%)
shower neutrinos	538 (26.2%) / 131 (24.9%)	637 (51.4%) / 160 (51.6%)
atm. muons	507 (24.6%) / 137 (25.9%)	0
total	2,059 / 528	1,237 / 309

and interaction type. I.e., an electron CC event is assigned a 1 in the signal/background classifier and a 0 for the track/shower classifier. The binary cross entropy (Eq. 5.3) serves as the loss function.

In the case of the direction reconstruction, the three cartesian directions x, y and z from the MC truth constitute the labels and the loss is derived from the difference between predicted and true value as well as the predicted uncertainty as described in 5.2.3 for all three directions combined. The energy reconstruction operates on the decadic logarithm of the energy, $\log_{10}(E)$, including the uncertainty estimation in the same way as for the direction. For NC interactions, only the visible energy is considered, which is defined according to the bjorken y (Eq. 1.27), because the outgoing neutrino takes away a portion of the energy without producing any light.

The architecture of each of the four neural networks is the same (as illustrated in 5.7) with only the output layer being exchanged.

6.1.1 Training

The networks are trained on a single GPU for as long as either there is no more improvement in the loss for the validation set or *overfitting* sets in. Overfitting is the phenomenon that the loss of the training set is still being minimized while for the validation set no further improvements or even deterioration is observed. In that case, the network commits to features only present in the training set ("learning the training samples by heart"), which decreases generalizability. The training curves for all four networks are drawn in Fig. 6.1 as solid lines. A validation to monitor the current
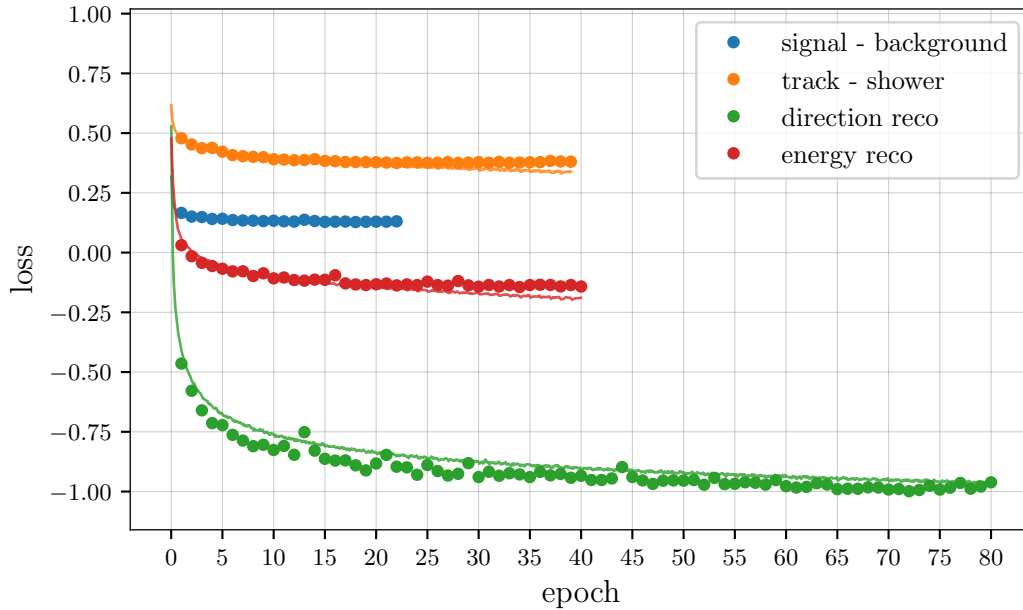
Figure 6.1: Loss over epochs during training (solid lines) and validation (dots, once after every epoch) for the four networks (colors).

performance of the updated weights is launched after each epoch and is added with dots. The absolute values of the loss do not allow for a direct comparison between the displayed cases, as the problem, definition of the loss and the data are different. The final model eventually used for the analysis is the one with the lowest validation loss for each task.

Comparing the training of the two classifiers reveals that both have seen a similar amount of events, because the number of epochs are roughly double the amount for the track/shower classifier, but the number of events per epoch are half of that from the signal/background. Both classifiers exhibit a relatively quick convergence after about 5 and 10 epochs, respectively.

The direction reconstruction seems to be a harder task for this deep learning approach, as it takes more than 70 epochs until no further improvement is observed. This process could be accelerated by choosing a larger learning rate. However, this might lead to a larger loss in the final state (and with it lower reconstruction accuracy), which is why this lengthy training, lasting about 11 days, is preferred. The fact that the training and validation loss are slightly offset for the green curve is due a fluctuation in the composition of the two sets. Here, the validation set appears to contain more events that are generally easier to reconstruct, thus yielding a lower loss. However, the important feature of identifying overfitting can still be extracted from the shape of the validation curve alone.

The smaller training sample for the energy reconstruction converges faster after about 25 epochs.
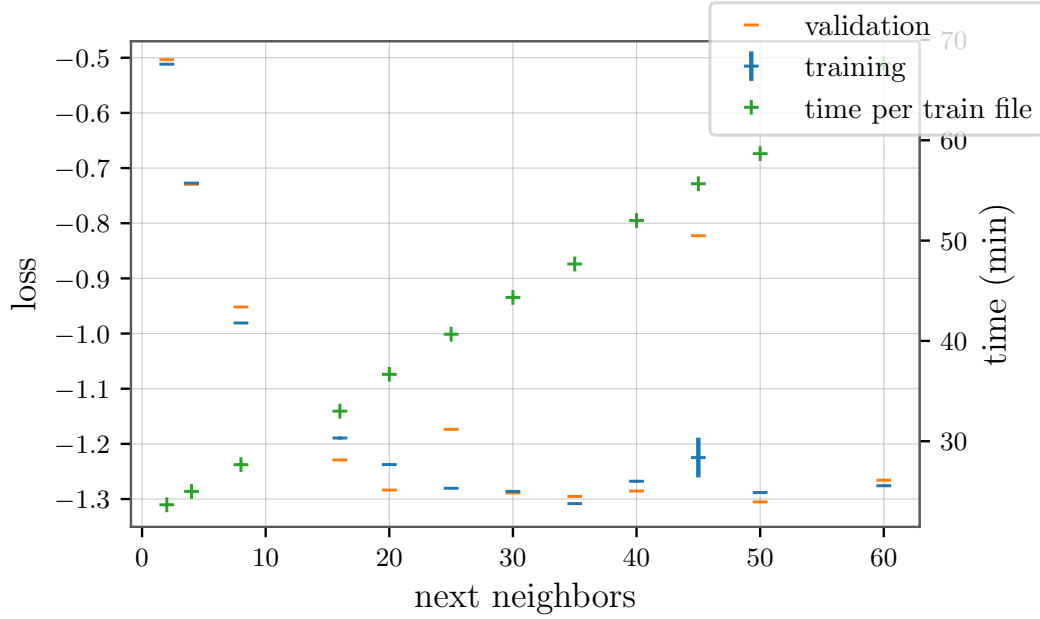
Figure 6.2: Training (blue) and validation (orange) loss and time needed for training on one file (arbitrary quantity) after three epochs versus the number of connections used as input to the edge convolution layers. This example shows the case of the direction reconstruction network.

6.1.2 Hyperparameter optimization

Several scans of hyperparameters are made in order to find close to optimal settings for training the neural networks efficiently and successfully. The parameters tested are learning rate, batch size, number of next neighbors for the edges, optimizer and activation function for the edge convolution layer blocks. Additionally, the architecture of the networks themselves is varied by testing several example cases such as a wider (more neurons per layer) or deeper (more layers) design. But no significant improvement can be found deviating from the ParticleNet implementation.

The scans consist of training and validation for three epochs, then comparing the achieved loss and/or accuracy. One example is the loss versus the number of next neighbors (connections) each hit can trade information with in Fig. 6.2. From the training and validation loss it can be seen that there is substantially less information when restricting the information exchange to fewer than 10 nodes. For large numbers of k-NN, no additional physical information is added any more, as all hits are sufficiently connected throughout the event. This makes sense, as with a mean number of hits of about 90 per average muon neutrino interaction, 40 k-NN means that about half the hits are directly connected and almost all hits are interconnected via only one intermediate node. Because of the increasing processing time (green crosses), 40 next neighbors are chosen to
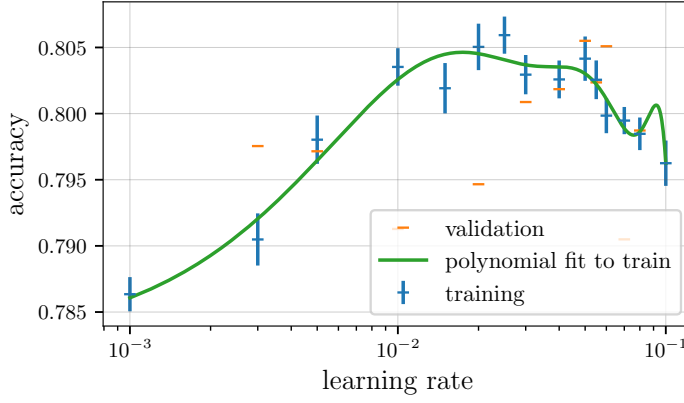
Figure 6.3: Training (blue) and validation (orange) accuracy of the track/shower classifier after three epochs for different learning rates keeping the batch size fixed at 16. A fit (green) to the training accuracy is added to read off the maximum.
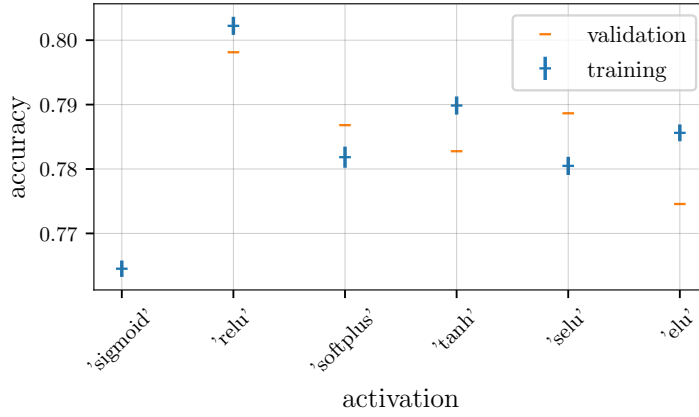

Figure 6.4: Training (blue) and validation (orange) accuracy of the track/shower classifier after three epochs when using different activation functions at the end of the edge convolution layer blocks.

safely ensure best performance. Interestingly, due to its strong relation to physical information transfer, the behavior of this curve is very similar for all four applications.

In Fig. 6.3, the same exercise is done for different learning rates. As can be observed and is indicated by the fitted green line, an optimum around 2×10^{-2} can be determined. It should be noted however, that the effects on the resolution from learning rate and batch size are correlated; both smaller batch sizes and larger learning rates mean faster convergence but are prone to overshoot minima in the high-dimensional loss space. Large batches and small learning rates cause the minimization be constrained to local minima. This is why both parameters are optimized simultaneously, iterating a several times. Also, networks are fully trained for a few selected hyperparameter sets to ensure their long term performance does not differ substantially form the one after three epochs, like displayed here.

In the same way, the activation function used throughout the network at the end of the edge convolution blocks is varied and found to yield best results using ReLU (Fig. 6.4).

Considering these hyperparameter studies, all networks are trained with next neighbors equal to 40, ReLU as activation function, Adam as optimizer
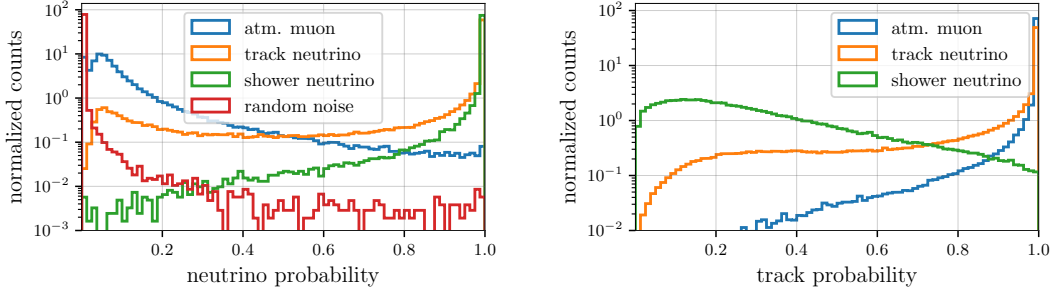
Figure 6.5: Normalized distributions for the different topologies of the neutrino probability (left) and track probability (right).

and the following initial learning rate and batch sizes: signal/background: 0.025 and 32, track/shower: 0.025 and 16, direction reconstruction: 0.01 and 32, energy reconstruction: 0.005 and 32. The quoted initial learning rates are then lowered each epoch to reduce the probability of overshooting the minimum during optimization in the loss space.

6.2 Performance evaluation and comparison to classical methods

The performance on a test set of the fully trained neural networks can now be studied and, where possible, be compared to existing approaches.

One advantage the GNNs hold straightaway over maximum likelihood fits is their run time. When utilizing the GPUs their algorithms are optimized for, the time needed to process a data run is reduced roughly by a factor of 8 compared to the JGandalf reconstruction chain. This does include the preprocessing and the consecutive execution of all four networks on a single GPU. This factor is even larger when comparing to the classical shower reconstruction. The predictions alone for single network take about 15 min for a six hour run containing 150,000 events. Also when relinquishing GPUs due to their high power consumption, cost and sparse availability, all predictions are computed faster on CPUs by a factor 2 to 3 compared to JGandalf.

6.2.1 Signal/background classifier

As previously mentioned, the output of the signal/background classifier is a score between 0 (background) and 1 (signal) for each event with the value indicating the probability with which the network predicts this event to be induced by a neutrino. Fig. 6.5, left, contains the normalized distributions of track- and shower-like signals, and atm. muons and random noise as background class. In this and all following plots, the measured atmospheric
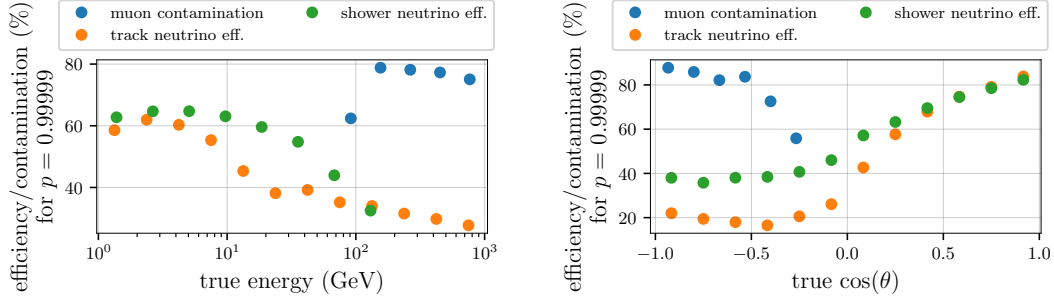
Figure 6.6: Efficiencies for track (orange) and shower (green) neutrinos and muon contamination (blue) considering a cut on the neutrino probability at > 0.99999. Left: Versus the energy of the primary particle. Right: Versus its $\cos\theta$.

neutrino flux [134] and the oscillation probabilities [52] are considered for the weights of each event, allowing for direct transfer to an application in the context of an analysis, as it later be shown (Chap. 7 and 8). Only when specified, like in this first plot, the groups of topologies are normalized to the same integral. As can be seen, the network is generally able to successfully distinguish between the two classes, manifested in the four distributions peaking close to either 0 or 1. Between the two neutrino signatures, there is an even clearer contrast for shower neutrinos, as they look significantly different from the background atm. muons. Naturally, there are muon neutrinos coming from the same direction as atm. muons, which causes the confusion in form of the wide peak at 0.05 for track neutrinos. Even though suppressed by more than 2 orders of magnitude, there is a small peak at 1 for atm. muons, which stems from misidentified muon interactions that appear more ambiguous in their orientation.

The random noise exhibits the strongest separation of all topologies with about 80% of events ending up in the first bin of a neutrino probability of close to 0. Anything above a neutrino score of 0.3 contains less than $10^{-2}\%$ of the random noise events per bin, promising a strong suppression for the cut values of interest.

With $p > 0.99999$, one of the cut values that will also be eventually used to select a pure neutrino set, the resulting efficiency and contamination of the cut can be evaluated for different topologies. In the context of cuts and selection criteria, the term "efficiency" in this thesis will always be defined as the fraction or percentage of surviving events, i.e., number of events after the cut divided by the number before the cut. The "contamination" is the percentage a group of events still occupies of the set after the cut. The dependencies of the neutrino efficiency and muon contamination on energy and $\cos\theta$ for the signal/background classifier are displayed in Fig. 6.6.

Firstly, it should be noted that atmospheric muons only originate from above the detector ($\cos\theta < -0.2$), as they cannot traverse Earth to appear from below, unlike neutrinos. Similarly, they are only observed with energies

above 70 GeV in the vicinity of the detector, i.e., 2500 m below the water surface. The energy spectrum for atmospheric neutrinos creating shower signatures ends at about 200 GeV.

It is this difference in energy and direction the network is able to take advantage of. Even though it does not directly reconstruct these quantities, it assigns lower-energetic neutrinos predominantly a neutrino probability of close to 1, with as much as 65% of the events surviving the strict cut at $p > 0.99999$. With larger energies, the background class comes into the mix reducing the efficiency to around 35% at 100 GeV and allowing a contamination of about 70%. This trend only increases for higher energies. Especially for energies at about 20 GeV, the identification of shower events as neutrinos is 15% more efficient than for track-like neutrinos.

A similar pattern can be observed in the $\cos\theta$ dependence on the right side of Fig. 6.6: For downgoing neutrinos the distinction becomes more difficult dropping down as low as 20% for track neutrinos with a large influx of background events. This suggests that in order to obtain a pure neutrino set, a cut close to a reconstructed $\cos\theta$ of 0 will have to be used with the current size of the detector in conjunction with this cut on the neutrino probability. Recovering some of the downgoing muon neutrino events may be possible in the future, adopting approaches that employ a veto. Thanks to their different signature, it is possible to distinguish shower neutrinos from background even for the downgoing region, where their efficiency is still 40%. For upgoing directions, the neutrino efficiency rises up to above 80% for a true $\cos\theta$ of 1.

Finally, the overall neutrino efficiency and contamination from atm. muons and random noise is drawn in Fig. 6.7. To create this kind of plot, each data point corresponds to a cut value applied on the neutrino probability. Then, the number of neutrinos and background events surviving this cut is evaluated. In the top right corner the cut value of $p = 0.33$ is found, which yields the highest efficiency close to 100% but allows an overwhelmingly large background contamination. The last cut value is $p \geq 1$, which states that without further cuts, the muon contamination is close to 20% with a neutrino efficiency of about 30%. As discussed for Fig. 6.6, selecting only upgoing events will greatly reduce the contamination. The later used cut value at $p > 0.99999$ is indicated in gray and motivated by the fact that in combination with other cuts a decent efficiency is ensured while keeping the contamination to around 1%.

The orange curve of random noise events exhibits generally a lower contamination for the same cut value/efficiency. For this test set, cut values corresponding to about 55% neutrino efficiency already eliminate all random noise events. The lack of statistics in that region of low contamination also causes the groups of points arranged in a downward pointing manner. Here, the same number of background events survive for three or four successive cut values. Only the overall selected set shrinks, as the neutrino efficiency
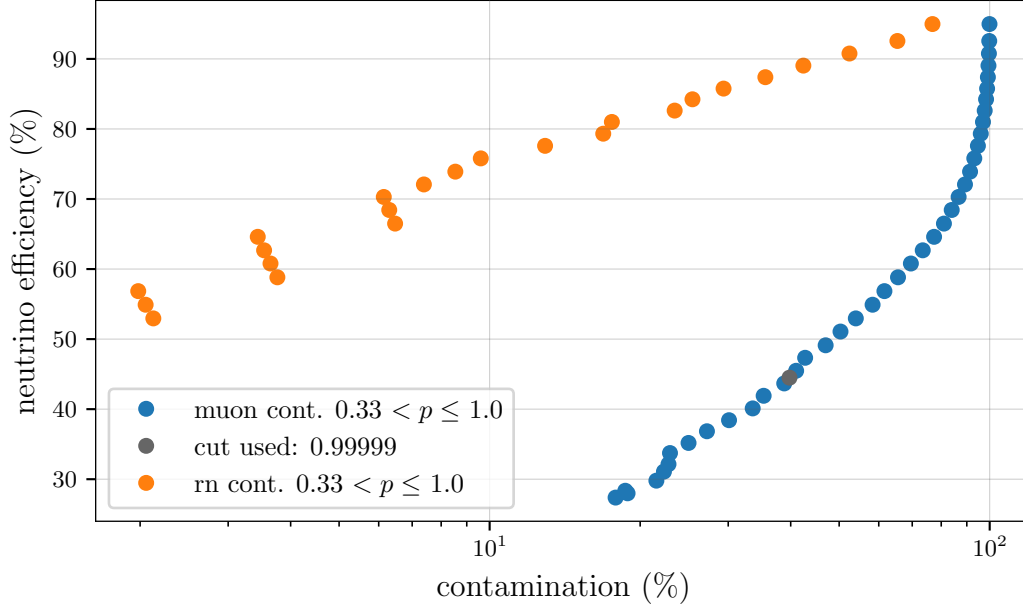
Figure 6.7: Neutrino efficiency of all flavors versus the contamination of atm. muons (blue) and random noise (orange) for different cut parameters.

decreases with more strict cuts. This way, the remaining random noise events make up a larger portion of the selection, causing the contamination to rise.

6.2.2 Track/shower classifier

The track/shower classifier can be used for further particle identification. Its output is shown in Fig. 6.5, right, in form of the normalized distributions of each topology's track probabilities. As is evident from the plot, track-like neutrinos and atm. muons are predominantly assigned a value close to 1, while shower-like neutrino interactions peak less clearly towards 0. In fact, the lowest bin is scarcely populated and the distribution around 0.1 is rather wide. This will make it more challenging to select a pure shower sample with high yield later on.

Form the efficiency versus the energy of the particle in Fig. 6.8, left, it can be seen that when requesting a minimum track probability of 0.8, i.e., selecting the track class over the showers, track-like neutrinos and atm. muons can be efficiently selected for larger energies, especially above 100 GeV with more than 80%. It should be noted that atm. muons were not part of the training set and yet are correctly identified as more track-like than shower-like. Their even higher efficiency is due to the combination of on average higher energies and preferred downgoing direction. It also could be amplified by bundle events in which more than one muon from
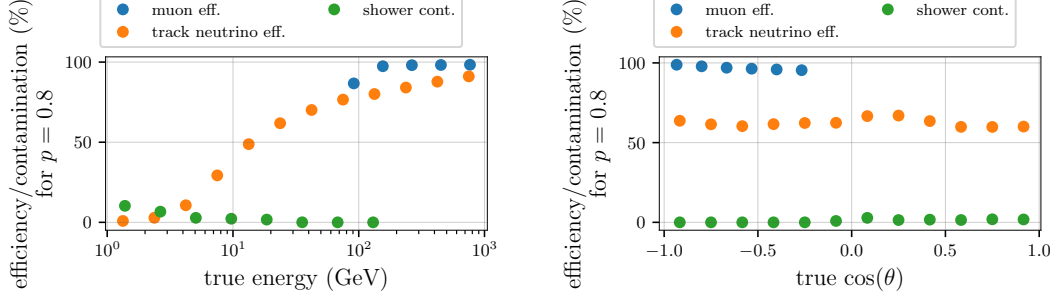
Figure 6.8: Atm. muon (blue) and track neutrino (orange) efficiency and shower contamination (green) considering a cut on the track probability of > 0.8. Left: Versus the energy of the primary particle. Right: Versus its $\cos\theta$.

the primary cosmic ray is observed in the detector and more track-like structures at once are present.

Towards lower energies, the efficiency for track neutrinos drops down to $10\,\%$ at $5\,\mathrm{GeV}$. This can be explained by the physics of lower-energetic muon neutrino interactions: Not only do their track lengths become shorter to about $20\,\mathrm{m}$ at $5\,\mathrm{GeV}$, but also the hadronic shower makes up a larger part of the overall appearance of the event. Put differently, lowest-energetic track neutrino events resemble shower neutrino interactions. Judging by the green points, the suppression of shower events is high for all energies with only a few percent passing this cut.

The strong rejection of the lower-energetic part of the tracks means that for the final selection, this cut should not be chosen too strict as many of the valuable muon neutrinos exhibiting oscillations (see Fig. 2.5, right) are cut away otherwise.

There is no strong dependence on the $\cos\theta$ of the interaction, see Fig. 6.8, right. Atmospheric muons become slightly more ambiguous towards horizontal orientation, as less information is contained in the ORCA6 detector, which extends more in its z-direction than its x and y.

6.2.3 Direction reconstruction

One advantage of this deep learning approach over classical reconstruction algorithms is that a network with its many free parameters has the potential to learn different concepts present in the data, recognize them and act accordingly. This is why there is no need for dedicated track and shower reconstructions. In fact, when trained with events of only one class and then comparing the predictions with those of networks that have also seen other classes, no difference in reconstruction performance is observed.

As manifested in Chap. 2.5.1, it is especially important for the KM3NeT experiment to be able to resolve the vertical direction component, the angle θ, as accurately as possible to achieve its physics goals. Thus, dedicated

classical muon track and shower reconstructions have been developed. The algorithm for the former problem has already been introduced in Chap. 4.1 and is referenced by the name of its main stage for direction fitting, *JGandalf*. The latter problem of fitting a shower hypotheses to an observed hit pattern in the detector is tackled by the reconstruction chain *JShower*, build up similarly to its muon counterpart [135]. In it, the first step is a prefit producing an estimation of the center of the hits. Then, using simple analytical PDFs, the vertex of the interaction is fixed using the same minimizer as in the JGandalf chain. Next, a preliminary energy estimation is done with a likelihood search incorporating more sophisticated analytical PDFs. In the second to last step, a prefit for the direction is executed scanning the sky. In the final stage, the directions and energies displaying best agreement thus far are refined in a final likelihood fit.

Even though each classical reconstruction is designed for its specific use case, they can be run on different data for cross-checks. Naturally, the expectation is to yield better performance when matching the reconstruction to the topology of the event.

Zenith angle resolution for tracks

A way to evaluate the quality of the direction reconstruction is to monitor the difference between the reconstructed and the true angle θ. The absolute value of this difference, $|\Delta\theta|$ is displayed in Fig. 6.9 for the three mentioned reconstruction approaches versus the energy for track-like neutrinos. In this case of interactions creating muons, the median resolution between JGandalf and the GNNs is very similar. For lower energies in particular, the deep learning shows a better resolution by $1 - 2°$ at $10\,\text{GeV}$ with $7°$. This is especially noteworthy, as oscillation research mainly profits from the accuracy in the reconstruction of these low-energetic neutrinos up to $20\,\text{GeV}$. From about $200\,\text{GeV}$ on, the classical method has a slight advantage of $0.4°$ at $1000\,\text{GeV}$. The same can be said about the lower 16% quantile that is representative of only the more well-reconstructed events. The fact that JGandalf performs better for these larger track lengths showcases its strength when dealing with exactly its design case. Here, the assumptions about the energy loss of the muon, the light propagation and eventually detection light yield are accurate and the signal strength is large. For lower energies, as mentioned before when discussing the track/shower classifier (Chap. 6.2.2), more irregularities from the (hadronic) shower part, that are not considered in the PDFs, diverge from the assumptions made by the track hypothesis. This is backed by the studies in [35], chapter 9.5, where a degradation of the direction resolution is reported not only for decreasing energies but also increasing bjorken y.

These observations for lower and higher energies, respectively, allow already for first conclusions about possible improvements for either approach:
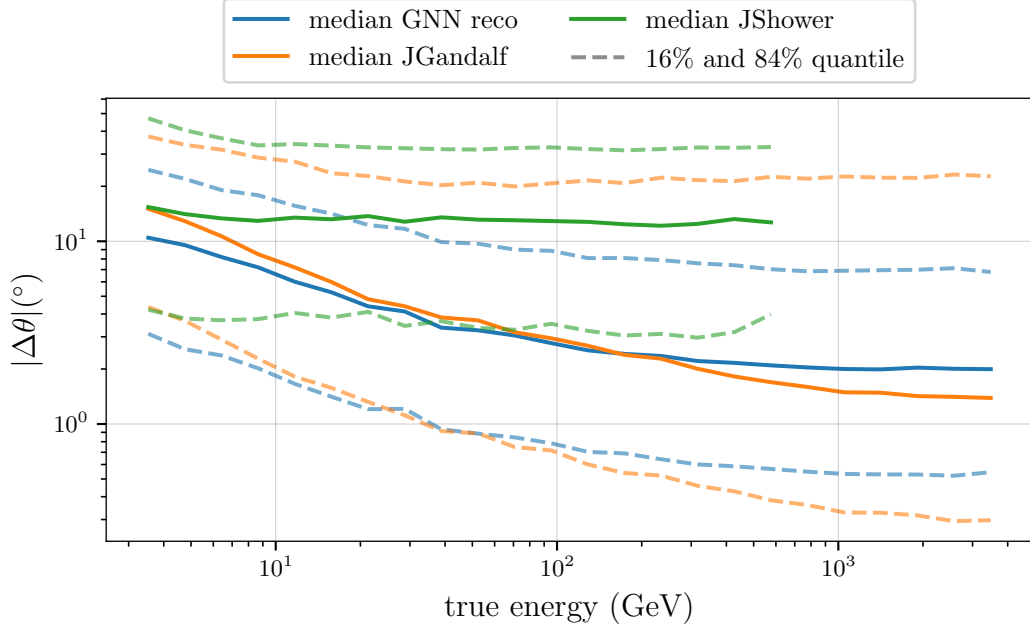
Figure 6.9: Absolute value of the difference between θ_{rec} and θ_{true} versus the energy for track-like neutrino events. Three different reconstruction methods are compared: The GNN based deep learning approach from this work (blue), classical muon track reconstruction JGandalf (orange) and classical shower reconstruction JShower (green). The median (solid lines) and 1σ quantiles (dashed lines) are included to appreciate the distributions per energy bin.

It is worthwhile studying the inclusion of a description of the shower part in the reconstruction of muon tracks. The current PDFs seem to have improvement potential for very low energies, while at least this comparison to the GNNs proves that, in principle, the algorithm itself can produce extremely high resolutions of sub-degree error in θ for selected events of TeV energy in ORCA6. Such an effort to combine PDFs and fit shower and track individually is currently under investigation in the collaboration.

Vice versa, it can be concluded that there is some room for improving the neural networks, as it is indeed possible to extract more information out of the higher-energetic events. The chosen hyperparameters might not yet be optimal and an even larger training sample could be used.

A significant difference in performance is observed in the upper, 84 % quantile: Here, now including more badly-reconstructed events, the deep learning approach clearly produces less severe outliers than the classical counterpart. Below 10 GeV, the discrepancy is about 10° with an overall resolution around 20° for deep learning. In practice, this is usually not a major problem, as this kind of events can be efficiently eliminated by quality cuts. That leads, however, to less selected events. The individual quantile lines differing relatively between the two reconstructions indicates
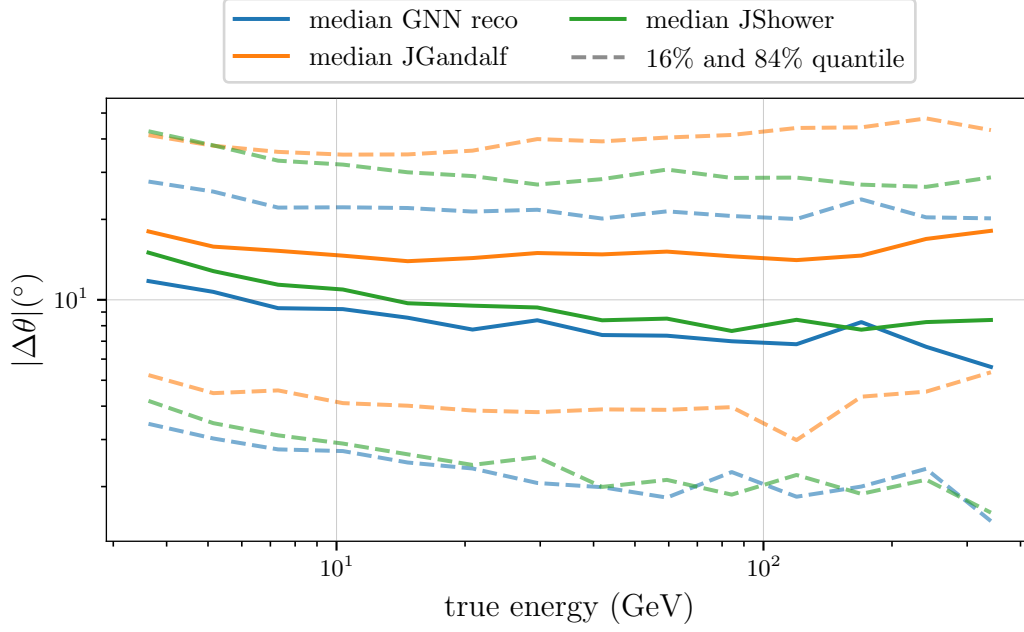
Figure 6.10: Same as Fig. 6.9, but for shower-like neutrinos.

qualitatively different underlying distributions of $|\Delta\theta|$, which are discussed shortly.

Additionally, the reconstruction performance on track-like events for JShower is also added in green to the plot. Its angular resolution is less accurate for most of the energy region, as expected. Only for the lowest energies, it achieves the same resolution as JGandalf. This further supports the statement that the pure muon track assumption at these energies is not correct, as shower PDFs yield similar results. Going to higher energies, there is no visible improvement for JShower, as by design the algorithm limits itself to only hits close by the interaction vertex. It cannot take advantage of the larger lever arm elongated tracks provide. Note, that the green line stops at $500\,\mathrm{GeV}$, simply because higher-energetic muon neutrino productions are not yet processed with JShower at the time of this thesis.

Zenith angle resolution for showers

The same plot displaying the θ resolution is created for shower-like neutrinos in Fig. 6.10. Due to their almost spherical emission of light around the interaction vertex, it is much more challenging to reliably determine the neutrino's incident direction in this case. This causes the absolute values in the plot's y-axis to be larger on average than in the track neutrino case. Comparing the median of the classical shower reconstruction JShower and the GNN-based one, there is a slight advantage for the entire energy range of about $1°$ to $3°$ in θ resolution. Qualitatively, the same behavior is observed
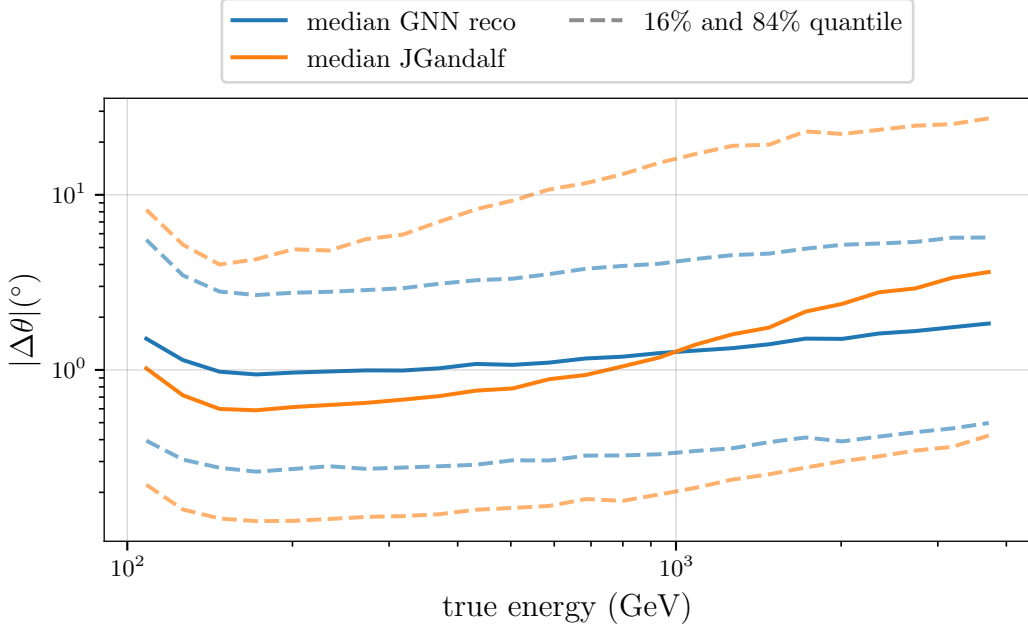
Figure 6.11: Absolute value of the difference between θ_{rec} and θ_{true} versus the energy for atmospheric muon events. GNN-based (blue) and classical muon track reconstruction (orange) are compared, while shower reconstruction was not attempted on these data.

for the 16 % quantile. The non-existent dependence is due to the uniform light propagation, which is similar over a wide energy range.

Similarly to the track case before, outliers occur less often for the deep learning reconstruction, resulting in an upper quantile of about 22° at 10 GeV, versus 32° for JShower.

Again, for a comparison, the JGandalf performance on shower neutrino interactions is included. Not being designed for this application, it is able to estimate the θ angle to about 17° accuracy, consistently to the performance at the lowest-energy track neutrinos. This correspondence further supports the shower-like appearance of low-energy track events.

Zenith angle resolution for atmospheric muons

Thirdly, the resolution of θ is drawn for atmospheric muons for GNN and JGandalf reconstruction in Fig. 6.11. At 100 GeV, the same statements from the discussion about track-like events in that region hold, concerning the performance comparison in median, upper and lower quantile. At 150 GeV, however, the resolution starts to decrease for both cases. This is due to the contribution from higher muon multiplicity events [133] that are more probable for higher energies. Instances of several muons from an incident cosmic ray creating light in the detector are more challenging to reconstruct and lead to misreconstructions more often. The deep learning approach is
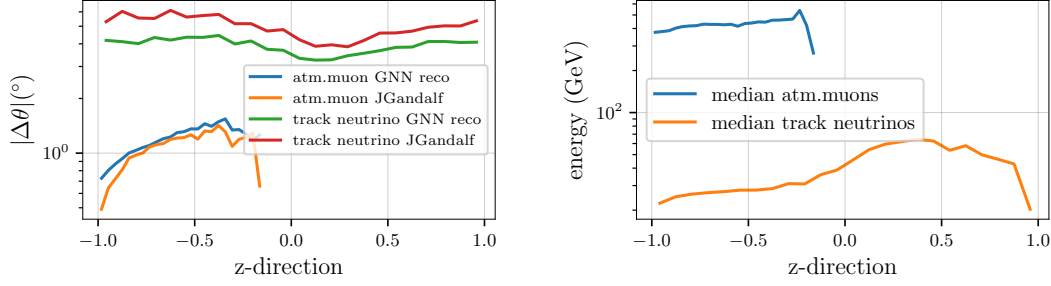
Figure 6.12: Left: Median resolution of θ versus the true z-direction of the event for track neutrinos and atm. muons comparing GNN based and classical reconstruction. Note that for larger z-directions there are very few events in the last bins in the muon case, causing it to have larger statistical fluctuations. Right: Median energy per z-direction bin for atm. muons (blue) and track neutrinos (orange).

superior to the classical one after about $1\,\mathrm{TeV}$. In these cases, JGandalf struggles with the more pronounced appearance of several signal tracks, as it is designed to fit exactly one muon track to the hit data. The neural networks, however, have encountered such events during training and are thus able to reconstruct them like any single muon.

Discussing resolutions to the zenith angle

The achieved resolution can be compared to that of IceCube's low-energy extension DeepCore, who report $10°$ median zenith resolution at $10\,\mathrm{GeV}$ for muon neutrinos with their best performing reconstruction methods [136], very close to this work's $7°$. Similarly, for $40\,\mathrm{GeV}$, $4°$ are found in this case, which are comparable to the $5°$ for DeepCore. Judging from the results in [137], the resolution achieved is fairly close to the intrinsic kinematic limit that arises from the difference between neutrino and muon direction.

Selecting an energy of $200\,\mathrm{GeV}$ and comparing the resolution accomplished by the neural networks for the three discussed topologies clearly exhibits a ranking as to how challenging direction reconstruction is between the classes. Track neutrinos can be resolved to $2.4°$ accuracy, shower neutrinos to $8°$ and atm. muons to $1°$. The difference between tracks and showers is immediately obvious from the signatures illustrated in Fig. 1.13, where the hit pattern from tracks point much more clearly to their origin. To fully understand the higher accuracy for atm. muons, their $\cos\theta$ spectrum has to be taken into account. A more detailed discussion follows, but the fact that they are essentially all close to strictly vertically downgoing, favors their angular resolution potential in the geometry of the ORCA6 detector.

Instead of evaluating the energy dependence of the angular resolution, in Fig. 6.12, left, the dependence on the true z-direction of the event is shown. Z-directions of -1 $(+1)$ correspond to vertical downgoing (upgoing) events

while 0 means horizontal with respect to the detector. Included are track neutrinos and atm. muons. Shower-like events do not exhibit any direction dependence. Starting with tracks from neutrino interactions, the influence of the large number of events with energies lower than $100\,\mathrm{GeV}$ becomes obvious when comparing GNN and JGandalf. Now, for all z-directions equally, the deep learning approach achieves a better resolution, ranging from 3° to 4° versus 4° to 6° in the classical case. Contrary to that, the atmospheric muons have predominantly higher energies and thus JGandalf is more accurate, especially for the strictly downgoing part, where the median resolution of θ is as low as 0.5°, compared to 0.7° for the GNN.

The curve for atm. muons indicates a clear dependence on the z-direction for both reconstructions due to the detector design. The average distance between DOMs on a DU is about $9\,\mathrm{m}$, while the average horizontal distance between DUs is $20\,\mathrm{m}$. Additionally, the current layout of the ORCA detector with its 6 DUs is much more elongated in its z-axis ($180\,\mathrm{m}$) than its $x-$ ($54\,\mathrm{m}$) and $y-$axes ($17\,\mathrm{m}$). Both facts combined lead to an inherently better angular resolution along the detectors z-axis.

Yet, this is not visible for track neutrinos, which should follow the same pattern. Instead, the performance for vertical tracks near z-direction ±1 remains rather flat compared to the rest of the spectrum. The effect causing a deterioration of the precision can be again explained by the geometry of the detector. Smaller distances between photo sensors allow for a higher acceptance of lower-energetic neutrinos. This results in a lower average energy per z-direction bin as displayed on the right side of Fig. 6.12. The orange line drops off to lower energies towards the edges. This is less pronounced for atm. muons.

The asymmetry of the median energy of triggered track-like neutrinos events between up- and downgoing stems from oscillations. Initially, the atmospheric flux and the properties of the detector cause the opposite distribution; more events with lower energies are triggered for upgoing directions (negative z-directions). On the one hand, the additional PMT pointing downwards increases the acceptance from that direction. On the other hand, the light from higher-energetic muons from neutrino interactions does not reach the detector when the vertex is located in the rock below the sea floor. If the interaction takes place in the water above, the moderate absorption length of the sea water still allows for a detection, increasing the mean energy observed for downgoing neutrinos.

Resolution of the azimuth and opening angle

So far, the resolution of θ has been used as the measure for performance evaluation in direction reconstruction. However, there is the other direction
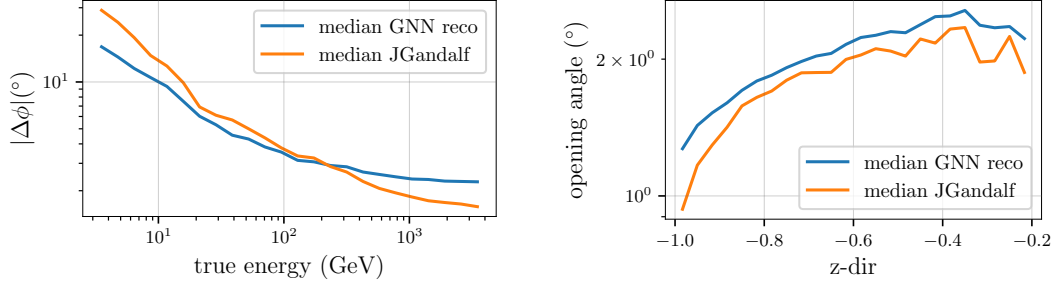
Figure 6.13: Left: Resolution of ϕ versus the energy for track neutrino events comparing GNN (blue) and JGandalf (orange). Right: Opening angle between true and reconstructed direction for atm. muon events versus the z-direction. For deep learning (blue) and classical reconstruction (orange).

component ϕ, which covers the xy-plane and is calculated from the cartesian coordinates x and y following the transformation to spherical coordinates:

$$
\phi = \begin{cases}
\arctan \frac{y}{x} & \text{für } x > 0 \\
\left(\arctan \frac{y}{x}\right) + \pi & \text{für } x < 0, \; y \geq 0 \\
\left(\arctan \frac{y}{x}\right) - \pi & \text{für } x < 0, \; y < 0 \\
+\pi/2 & \text{für } x = 0, \; y > 0 \\
-\pi/2 & \text{für } x = 0, \; y < 0.
\end{cases}
\tag{6.1}
$$

Combining both θ and ϕ to a full direction vector allows for comparing the complete reconstructed direction to the MC truth. The angle between these two directions $\vec{v}_1$ and $\vec{v}_2$ is called the opening angle and is defined as

$$
\text{opening angle} = \arccos \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_2| \cdot |\vec{v}_2|}.
\tag{6.2}
$$

A comparison of the resolutions in ϕ and opening angle generally yield very similar statements and conclusions to those already discussed using θ. As an example, the energy dependence of the resolution of ϕ is drawn in Fig. 6.13, left, for track neutrinos comparing GNN based and classical reconstruction. For example, the energy at which JGandalf's accuracy exceeds the one from deep learning is found to be 200 GeV again (compare 6.9). In general, the potential to resolve the zenith angle is greater in the current constellation ORCA6 than for the azimuth angle (same geometrical reasons as discussed before). For instance, the deep learning's median θ resolution for 10 GeV track neutrinos is 7°, compared to 10° for ϕ.

The right side of Fig. 6.13 is very much comparable to Fig. 6.12, left, showing the same dependence on the z-direction for atm. muon events and the same ranking between the reconstructions. By definition, the opening angle includes both θ and ϕ and is thus limited largely by the less precise azimuth angle. In numbers, the $|\Delta\theta|$ for atm. muons at $\cos\theta = -0.8$ is 1°
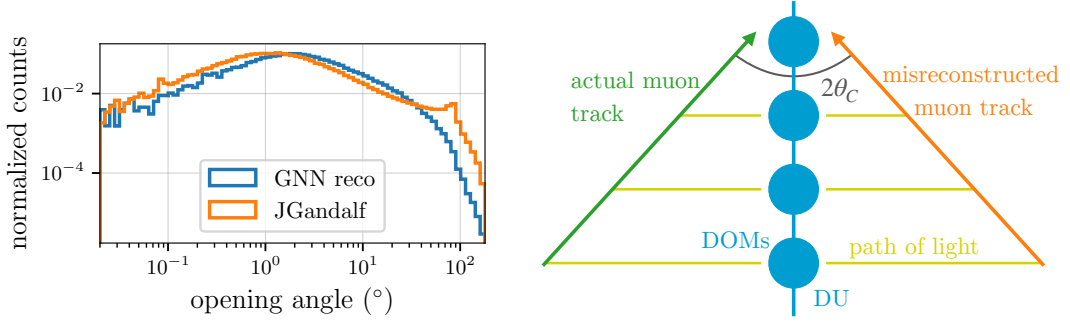
Figure 6.14: Left: Normalized spectrum of the opening angle between reconstructed and true direction for track-like neutrino events. Compared are the deep learning (blue) and the classical approach (orange). Right: Illustration of a case of possible misreconstruction. Events (green track) on the border of the detector (DOMs in blue) can leave hit patterns (light in yellow) that fit well to two hypotheses, two times the Cherenkov angle apart (orange track).

for the GNN reco, $|\Delta\phi|$ for the same point is 2° (not shown here) and the resulting opening angle is 1.7°.

Evaluating a spectrum of the opening angle for track-like interactions and comparing GNN reconstruction to JGandalf, Fig. 6.14, left, reveals the systematically different distributions mentioned earlier. Averaging over all energies now, the shapes of the distributions look similar for the most part, except for larger opening angles above 30°. Overall, the classical approach is shifted towards smaller values with more reconstructions of opening angles of below 1°. For opening angles around 84°, however, there is a second maximum for JGandalf, while the GNN reconstruction falls off monotonously towards 180°, the maximum opening angle possible. The position of the maximum corresponds to two times the cherenkov angle of 42° (see Eq. 1.29) and is caused by misinterpreting the hit pattern registered in the detector. The illustration on the right side of Fig. 6.14 helps to understand this geometrical effect: When projecting kinematics down to two dimensions and following the emitted light from the muon track for a special class of partly contained events, it becomes obvious that there is a certain ambiguity to fit a track model to the data. Both solutions in from of the green and the orange line are comparably valid and either one has to be chosen; the correct one resulting in a small opening angle, or the incorrect one, being $2\theta_C$ off. In principle, this situation is the same *a priori* for classical reconstruction and the neural networks. However, for the GNNs, the erroneous solution is greatly suppressed, as the networks are trained to avoid this case of misinterpretation efficiently by the huge loss (see Eq. 5.18) these events are assigned when they are predicted wrongly by such a large margin. Also, they are not limited to the strict geometrical thinking of having to fit a straight line to the hits.
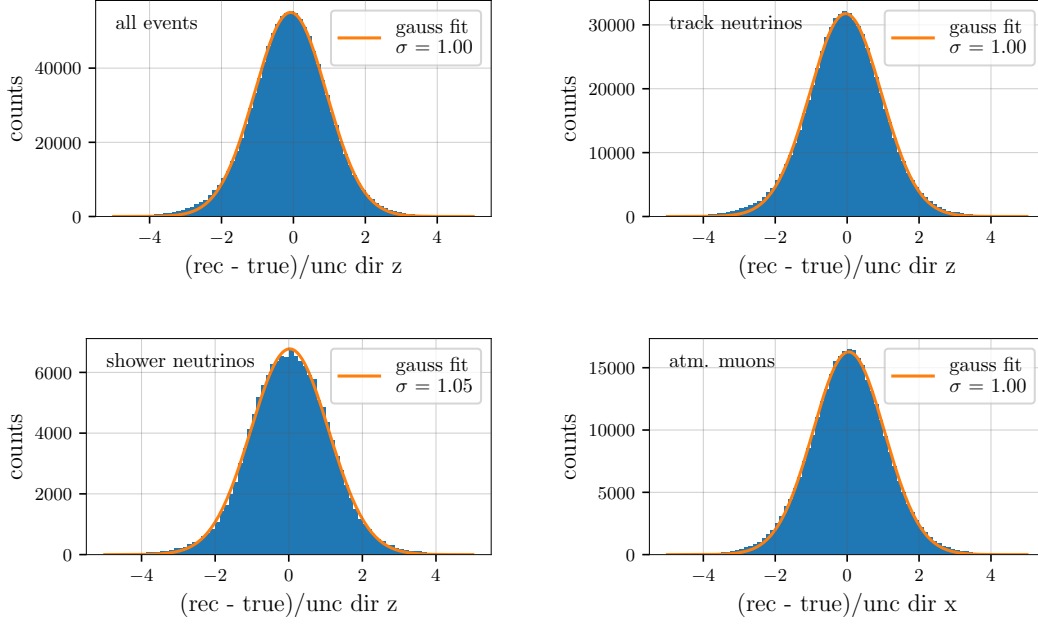
107

Figure 6.15: Pull plots for different data samples and GNN reconstructed quantities (blue). As a test, a Gaussian distribution is fitted to the histogram (orange). From top left to bottom right: z-direction for all topologies mixed, z-direction for only track neutrinos, z-direction for only shower neutrinos and x-direction for only atm. muons.

6.2.4 Uncertainty estimation for direction reconstruction

To start off discussing the direction uncertainty estimation, a statistical test can prove its interpretation as a 1σ interval as it is designed to be in the loss function (Eq. 5.16). For that, the difference between the reconstructed and the true direction component is taken and divided by the uncertainty on an event-by-event basis. This quantity should follow a Gaussian distribution of standard deviation 1. Such *pull plots* are shown in Fig. 6.15 along with a Gaussian fit. The first plot contains all kinds of events, similarly to how they were present during training. It can be concluded that the distribution indeed follows closely a Gaussian distribution with a standard deviation of 1. There are slight deviations from the curve due to an asymmetry that is discussed when studying 2d distributions of reconstructed and true direction components.

The spectra for track neutrinos on the top right and shower neutrinos on the bottom left both display the same agreement with the normal distribution. This allows for the conclusion that during training, the neural network has learned to assign uncertainties for each kind of event individually. As discussed earlier, angular resolution for showers is worse than for tracks from first principles and the uncertainties are thus adjusted for each scenario accordingly. The same is found for the atmospheric muon case and the
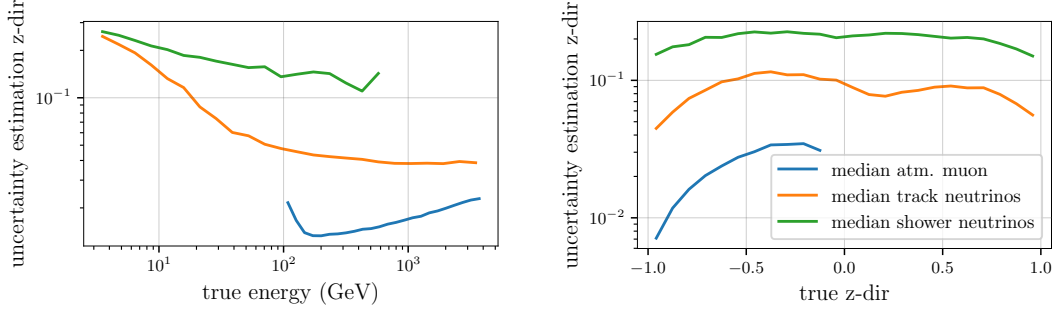
Figure 6.16: The GNN's predicted uncertainty of the z-direction for different topologies (colors, see legend right) versus the energy (left) and the z-direction (right).

other direction components x and y, which is exemplified in the last plot by the x-direction of atm. muons.

Next, the dependence of the uncertainty, in particular the one for the z-direction, on energy and the direction itself are discussed. The two plots in Fig. 6.16 display the dependencies for atm. muons, track-like and shower-like neutrino interactions. First of all, the ordering of the three topologies regarding their accessibility of directional information, as discussed before, is reflected correctly by the network's uncertainty estimation: Shower events are assigned larger values than tracks and low-energetic tracks behave similar to showers. Atmospheric muons are assigned an even smaller uncertainty on average, due to their strictly downgoing nature, combined with the detectors better resolution in vertical direction (see also the right side of this figure).

For both neutrino topologies, a steady decrease in the absolute value is visible as more and higher quality hit information becomes available with increasing energy. Above 1 TeV, it flattens out for single tracks as events are less contained within the active volume of the detector.

The dependence for the atm. muons shows the same feature as was discussed in the resolution of θ; a decrease in the uncertainty estimation until 200 GeV and an increase due to multi-muon bundles thereafter.

On the right side of Fig. 6.16 the dependence of the uncertainty on the z-direction of the event discloses medians falling off to the edges of the spectrum for all topologies. The lower uncertainty for vertical directions is the result of the current dimensions of the ORCA6 detector.

When comparing up- to downgoing, a slightly smaller uncertainty can be noticed for upgoing, stemming from the energy distribution shown in Fig. 6.12, right.

Direction uncertainty as selection criterion

In practice, the estimated uncertainty is supposed to be a quality cut to ensure that a selected set consists of events with reliable direction
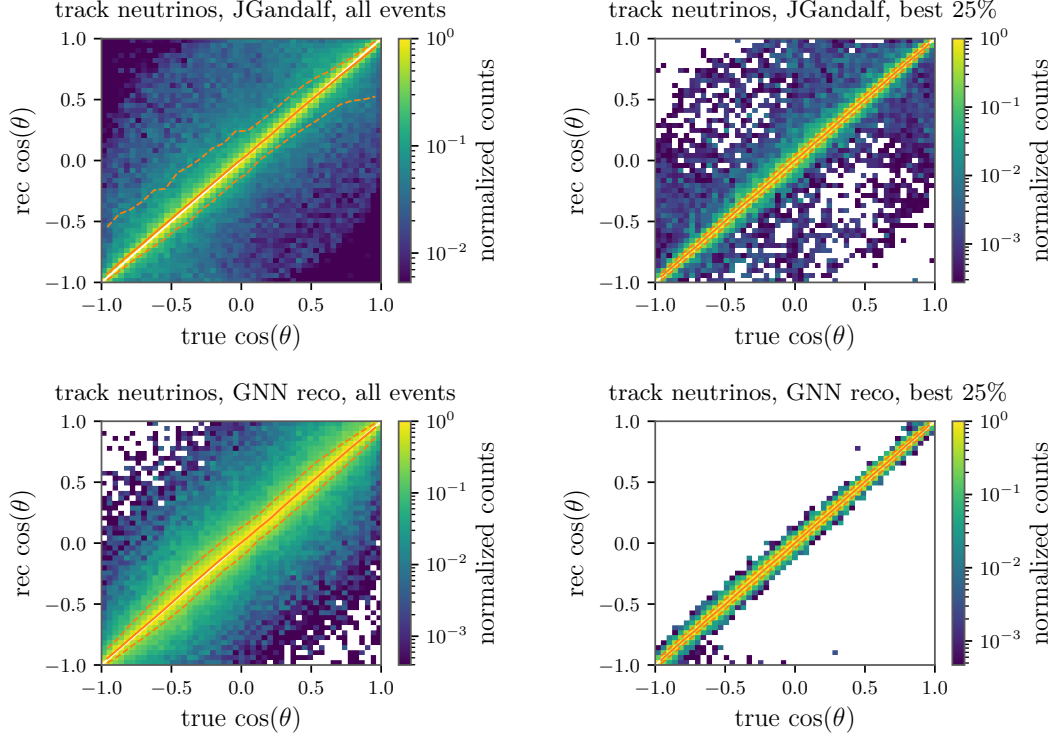
Figure 6.17: Reconstructed versus true $\cos(\theta)$ for track neutrinos. Each plot's column in x is normalized to have the same integral. A white line is added as diagonal to indicate perfect reconstruction. In orange, median (solid) and 1σ interval (dashed) for the 2d data are added. Left column of plots: all events. Right column of plots: best 25% selected by quality parameter. Top row: JGandalf reconstruction. Bottom row: GNN reconstruction.

reconstruction for further analysis. To test for this, the reconstructed and true direction component are plotted against each other and subsequently, a cut requesting the best 25% reconstructed events according to the uncertainty is applied. The result is the overview in Fig. 6.17 for track neutrinos. As mentioned before, the case of the *z*-direction is not only the physically most interesting one but also covers all features present in the other direction components. The plots on the left side, that contain all events, hold additional information concerning the reconstruction of $\cos\theta$ compared to the type of plots used before in Fig. 6.12, which showed the absolute value of $\Delta\theta$. To help guide the eye, a diagonal is added in white, indicating perfect agreement between the reconstructed and true value. The data are normalized to have the same integral in each column of true $\cos\theta$ to better appreciate the distribution of reconstructed values for each individual column. Furthermore, a median per x bin is overlayed with a solid orange line, surrounded by dashed orange lines indicating 1σ intervals.

Starting the comparison with all track neutrino events on the left side, the already discussed difference in the distributions (like Fig. 6.14) becomes clearly visible in this representation as well, with the dashed lines for the
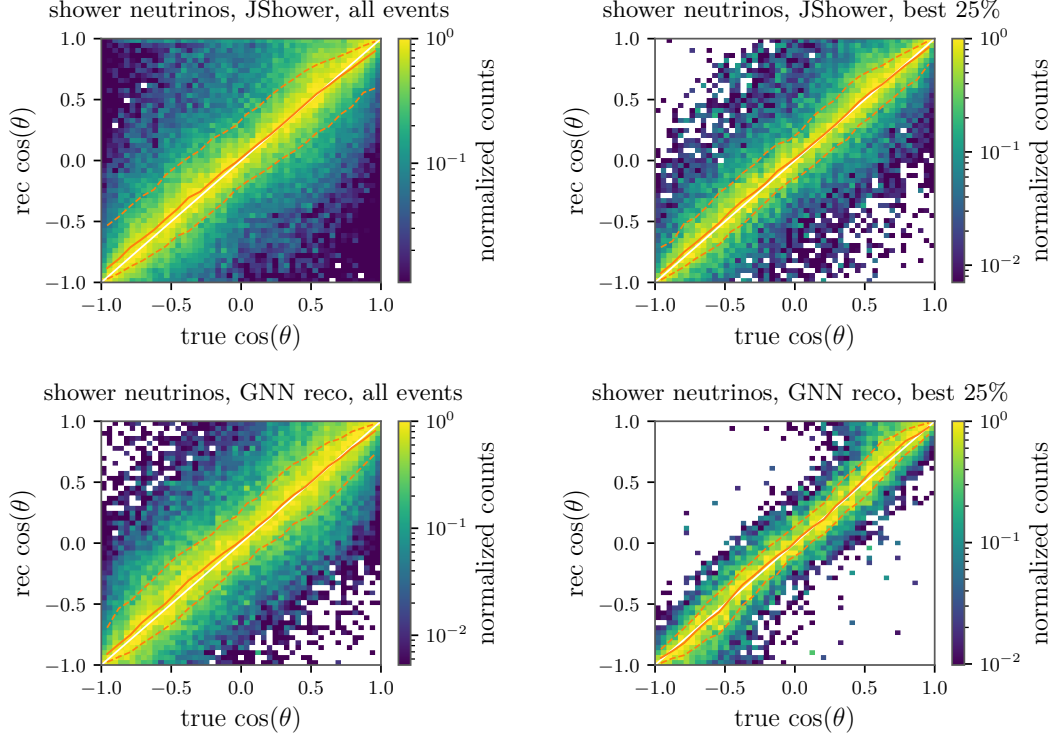
Figure 6.18: Same structure as in Fig. 6.17 but for shower neutrinos and the classical shower reconstruction JShower in the top row.

quantiles indicating a lower spread for GNN (bottom row) than for JGandalf (top row). In general, both reconstructions achieve a median very close to the diagonal, asserting excellent accuracy and no systematic effects for any particular direction.

The selection of the best 25% events for JGandalf is done using the output parameter that is supposed to give a 1σ error interval on the total direction fitted by the algorithm (internally called $beta_0$). For the GNNs, the z-direction uncertainty is used. From the plots on the right side it can be seen that such a selection is very efficient in both cases. The dashed lines for the quantiles are in close proximity to the diagonal implying a very low spread. The cut is even cleaner for the GNN reconstruction, as almost no larger misreconstructions are kept. This supports the statement that the uncertainty is well fit as a quality cut to ensure accurate and precise direction reconstruction.

As expected for shower events, the spread marked by the dashed lines for the two plots on the left of Fig. 6.18 is inherently larger. This also includes a higher probability of the median deviating from the diagonal.

To select events for JShower, the likelihood of its main fit is used. Again, applying such cuts clearly improves reconstruction performance in both cases. The distributions of each x bin for JShower are, however, wider around the diagonal than in the case of the GNNs, for which larger misreconstructions are efficiently eliminated. For the deep learning reconstruction, there are
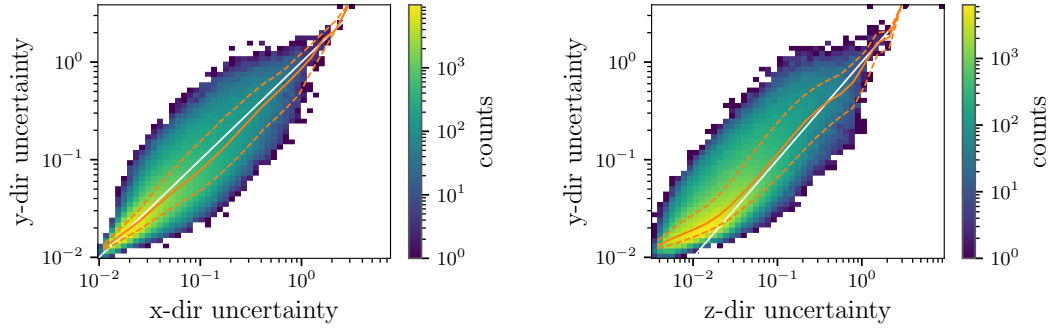
Figure 6.19: Uncertainty estimation for y- versus x-direction (left), and y- versus z-direction (right) for atm. muon events. These histograms are not normalized.

more events kept at reconstructed values of ± 1 because of the better resolution of vertical events, already hinted in Fig. 6.16, right. They are, however, not necessarily all well reconstructed as can be seen by their distance to the diagonal.

To conclude the discussion of the predicted uncertainties, correlations between different direction components are evaluated. If the uncertainty is physical, such correlations are expected, as challenging events, for example, should yield larger uncertainties for all direction components similarly. Figure 6.19 contains correlations between x and y on the left, and y and z on the right side for atm. muon events. Due to the similar dimensions of the detector in x and y, the uncertainties are strongly correlated in a linear fashion. The lower spread for small uncertainties is expected, since events well reconstructed in one azimuth direction should be similarly unambiguous in the other. For higher uncertainties, the spread can be larger, in case these events traverse the detector form one particular side.

As established earlier, the uncertainty in z-direction is smaller due to the detector's geometry. This is reflected in the correlation, as seen in Fig. 6.19, right, especially for the part of low z-direction uncertainties. With this in mind, also the lowest values in the y versus x plot display a slightly better resolution in x-direction, as the orange median line is above the diagonal. This again is supported by the detector's extension, which is larger in x than in y dimension (see also Fig. 2.2).

6.2.5 Energy reconstruction

The final of the four neural networks reconstructs the energy of events recorded with the KM3NeT detector. Along with its predicted value it provides again an uncertainty. To improve convergence of the network's parameters, the $\log_{10}$ of the energy is fed as the true label instead of its natural value. This means that the uncertainty is a factor instead of an

absolute number. An example helps to illustrate this: The predicted value and its uncertainty, as generated from the output layer of the network shall be 1 and 0.11. The predicted energy is then $10^1 = 10\,\text{GeV}$ and the estimated uncertainty should be interpreted as $10^{(1\pm0.11)}$. Then, the boundaries of the error interval are asymmetrical; $10^{1.11} \approx 12.88\,\text{GeV}$ above and $10^{0.89} \approx 7.76\,\text{GeV}$ below. It can be seen, that the uncertainty is the relative factor of $10^{0.11} \approx 1.288$, meaning 28.8%, in both directions. Only in combination with the predicted value, absolute numbers for the interval can be given: $10^{+2.88}_{-2.24}\,\text{GeV}$. In any case, in the context of energy reconstruction a relative error is preferred.

Due to the challenging nature of energy reconstruction in a partly built ORCA detector, more systematic features are expected when evaluating the reconstructed versus true values than for the direction. Especially for partly contained tracks, an underestimation of the particle's energy is expected. This is why it is more instructive this time to first examine the 2d representations and afterward study a comparison of the resolution from the different reconstructions in dependence on the energy. As far as the definition of the true energy is concerned, no change is done to the track neutrinos, even though some light emitted along the track does not even reach the detector. For neutral current interactions though, the true energy is calculated from the bjorken y (visible energy), as only the hadronic part of the interaction is visible.

Similar to Fig. 6.17, track neutrinos processed by deep learning and classical reconstruction are shown, correlating the reconstructed to the true energy, once for all events (left) and once using a quality cut (right) in Fig. 6.20. The first row's results consists of the JGandalf reconstruction stage for energy fitting, also referred to as *JEnergy*, and the second row features the graph neural networks. The 2d plots are again normalized per column for better visibility and diagonal and median lines are added. Note, that the median and quantile lines stem from the un-normalized data.

The selection of the 25% best events is done for the classical reconstruction on the likelihood from JEnergy. For low energies up to $10\,\text{GeV}$ a decent reconstruction is observed as most entries are close to the diagonal or at least generally assigned small energies. From about $20\,\text{GeV}$ on, two bands can be observed with different reconstructed energies for the same true energy. Firstly, there is the naturally occurring underestimation due to the fact that tracks are only partly contained. Secondly, there is a population of nonphysical overestimation, whose origin is unclear and currently under investigation in the collaboration. The presence of this additional bump in higher reconstructed energies poses the problem of how to treat these events in an actual data analysis. Applying a quality cut on the JEnergy likelihood, many of the events from the second bump are discarded, leaving mainly a band of similar reconstructed energies of $10 - 20\,\text{GeV}$ for all true energies.
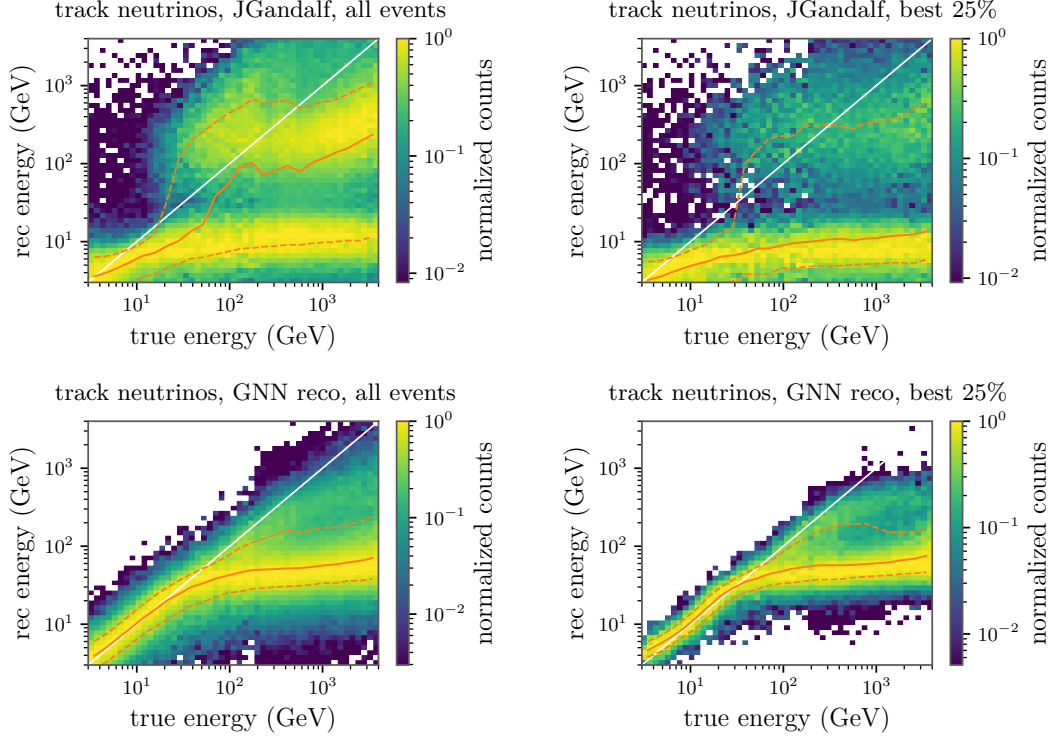
Figure 6.20: Reconstructed versus true energy for track neutrinos. Otherwise same layout as Fig. 6.17.

The second row shows the energy reconstruction by the GNNs for the same events. For energies up to 20 GeV, there is very good agreement between median and diagonal indicating an accurate energy reconstruction with a low spread. Up to about 30 GeV, still many events are reconstructed close to their true energy. For higher energies, the physical limit of the detector is reached for more and more events and underestimation sets in, leaving all higher-energetic neutrinos with the same energy deposited in the detector and being reconstructed with about 40 GeV to 60 GeV. The quality cut on the energy uncertainty factor is able to reduce the spread significantly, but does not identify events of larger true energies than 100 GeV to have an incomplete energy estimation. Still, a slight band (light green) can be observed that continues to follow close to the diagonal even for these higher energies. Overall, no unexpected features appear.

Determining the energy for shower events is simpler than for tracks because of the limited extension the particle cascades have that create these signatures. In the first row of Fig. 6.21, the shower energy as reconstructed by JShower is plotted against the true (visible for the NC contributions) energy. In general, the expected correlation between reconstructed and true energy can be asserted with moderate spread. Especially after applying the quality cut on the right, however, a different slope of the median (and quantiles) is observed. This suggests an inaccurate energy correction, which is applied to account for the hadronic shower component, which is not
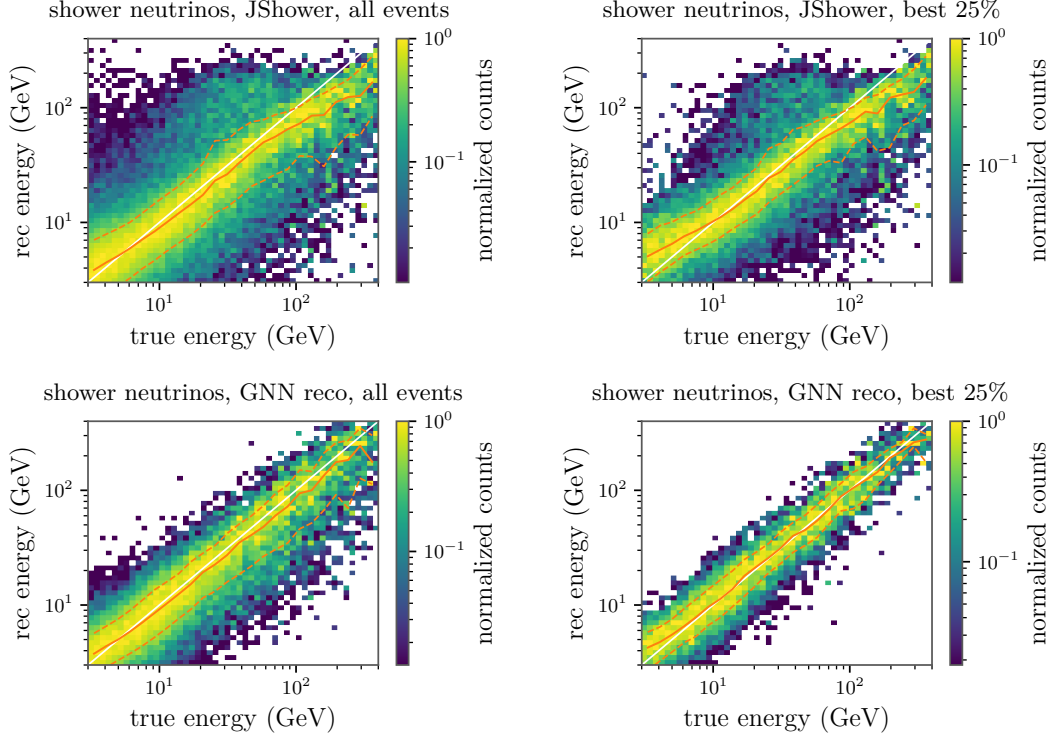
Figure 6.21: Reconstructed versus true energy for neutrino interactions creating shower signatures. Otherwise same layout as Fig. 6.18.

considered in the PDFs of the fit. The parameters of the correction are tuned in simulations and energy dependent. Studies are currently ongoing to improve the parameters, which would change the slope of the median line in this representation. Additionally, a region between $10\,\mathrm{GeV}$ and $100\,\mathrm{GeV}$ exhibits less correlation between true and reconstructed value for a few events (round structure overlayed with the diagonal).

Finally, in the second row the GNN reconstruction results on shower events can be seen. Similarly to JShower, the general trend follows the diagonal closely. Towards higher energies, a slight underestimation due to partially contained events can be observed, as the median lies predominantly below the diagonal. The spread for all events is lower than for JShower (a more quantitative discussion follows in the next part). A cut on the uncertainty factor successfully reduces outliers and brings the median closer to the diagonal, successfully eliminating most of the underestimation, contrary to the track neutrino case. For very low energies of below $5\,\mathrm{GeV}$, the reconstructed energies are larger than the true energies.

Energy resolution for tracks

Defining the energy resolution quantitatively by the relative error $|E_{\mathrm{rec}} - E_{\mathrm{true}}|/E_{\mathrm{true}}$, the three different reconstructions can be compared directly when monitoring the dependence on the true (visible) energy of the neutrino.
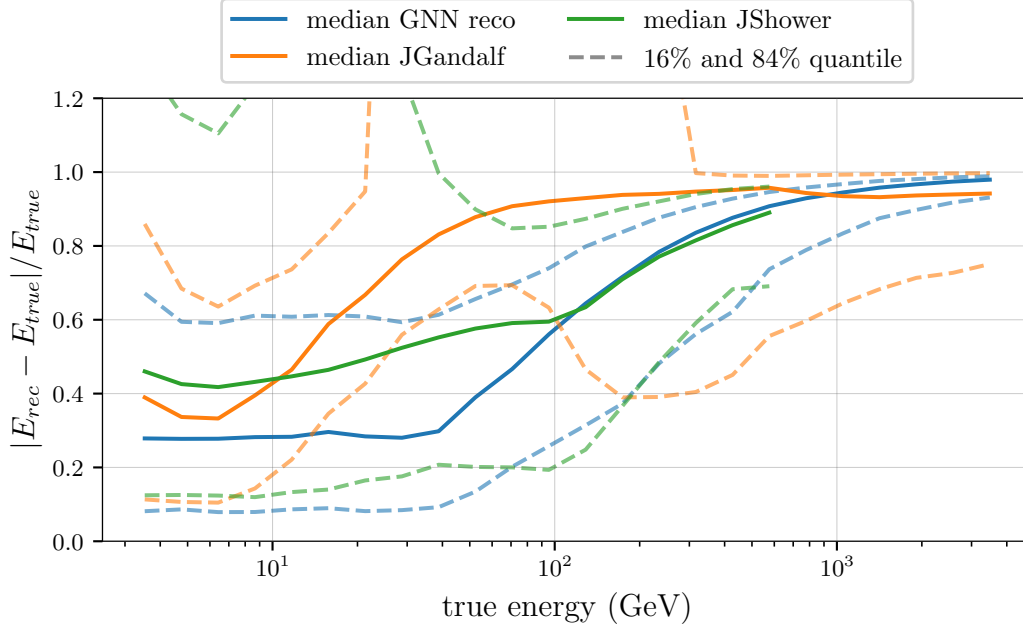
Figure 6.22: Relative error of energy reconstruction as measure for its resolution versus true energy of the neutrino for track-like interactions. Median (solid lines) and 1σ intervals (dashed) are shown for three different reconstruction algorithms (colors).

This is done for track neutrinos in Fig. 6.22. Obviously, relative errors close to zero is the best than can be achieved. In the case of underestimation, the value will in the extreme case converge to 1, as

$$|E_{\text{rec}} - E_{\text{true}}|/E_{\text{true}} \to 1 \ \text{ for } E_{\text{rec}} \ll E_{\text{true}}. \tag{6.3}$$

In the case of overestimation, however, the relative error can explode as

$$|E_{\text{rec}} - E_{\text{true}}|/E_{\text{true}} \to \infty \ \text{ for } E_{\text{rec}} \gg E_{\text{true}}. \tag{6.4}$$

Starting with the median at low energies, the GNN reconstruction achieves an energy resolution of about 30%, even up to 40 GeV. This, with the first 6 DUs in ORCA is already comparable to IceCube's DeepCore, which can determine the energy of muon neutrinos below 10 GeV to 20% accuracy [136]. The longer absorption length in ice allows for more efficient light detection, while scattering processes play a less important role in energy reconstruction. This way, DeepCore is able to achieve resolutions at the intrinsic limits [137]. As muon track lengths get larger, the energy resolution deceases rapidly after 50 GeV and reaches 60% at 100 GeV. After that, the underestimation is so large that the median runs towards the asymptotic value of 1.

Even though the JGandalf reconstruction chain is designed for this application, it only achieves comparable results for energies below 10 GeV

with 30-40% as far as the median is concerned. Underestimation sets in earlier and reaches the plateau close to 1 at 70 GeV, basically losing all sensitivity to the actually deposited energy. An improvement to the procedure is currently being evaluated, promising to increase the performance for the complete energy range.

JShower, however, even though not designed for track reconstruction, is able to better resolve medium energies from 10 GeV on, compared to its classical counterpart. Consistently, the JShower curve exhibits the same behavior above 100 GeV than deep learning, indicating that both operate at the limit of the information available. Only for energies below 10 GeV, the region these track neutrinos should appear the most like showers, the resolution from JShower with constantly over 40% is less accurate than it is for the other reconstructions. This stems from the aforementioned energy correction.

The 16% of the best reconstructed events display a very similar behavior for the lowest energies, as they are all closely together achieving around 10% relative error. This again indicates that this the maximally attainable resolution in this region for stricter cuts. For the energies from 10 GeV to 100 GeV, GNN reco and JShower are closely together with a slight advantage for the deep learning approach, which keeps the 10% resolution up to 50 GeV. Above 70 GeV, JShower is capable of even better resolving track neutrino energies in some selected cases. The curve for JGandalf rises the same way its median rises to 60% relative error at 40 GeV. Above that, centered around 200 GeV, the second maximum is reached in which some events are included that coincidentally have the fitting true energy. In the first plot of Fig. 6.20, fist row, this appears when the upper band intersects with the diagonal.

A similar statement as for the resolution of θ about the 84% quantile, containing also badly reconstructed events, can be made. In deep learning, there are generally less severe outliers, especially when compared to JShower. Considering the comment about possible values for overestimation (Eq. 6.4) and the two-dimensional plots in Fig. 6.20, it becomes clear that particularly the 84% quantile for JGandalf reaches very high relative errors of about 4 for medium true energies. For better visibility of the y axis from 0 to 1, this plot is cut at 1.2. These high values are not of interest as their origin is most likely not physical.

Energy resolution for showers

The same energy dependence of the three reconstruction methods is computed for shower-like neutrino events in Fig. 6.23. Similar to what can be observed for the lowest-energetic tracks, the GNN reconstruction's median energy resolution starts slightly below 30% for energies smaller than 10 GeV. Even towards the highest energies possible with atmospheric
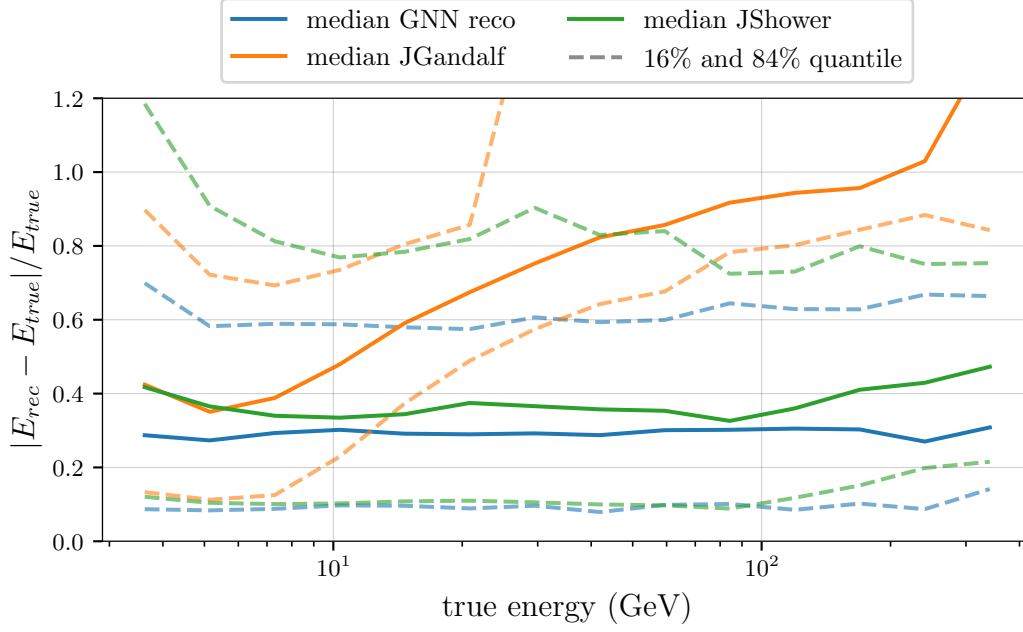
Figure 6.23: Same as Fig. 6.22 but for shower events.

neutrinos creating shower signatures, the resolution does hardly exceed 30%. With this, it maintains the edge for basically the entire energy range over the classical approaches. JShower appears to be lacking some median accuracy in the low-energy description, compared to the deep learning, as well as for energies larger than 130 GeV. Both effects can be attributed to the energy correction. JGandalf only delivers accurate predictions for the region below 10 GeV with relative errors around 40%.

Consistently with the track neutrino plot, 10% is the maximum resolution the best events can achieve for each reconstruction, including JGandalf for the lowest energies. The same differences at the lower and upper end of the energy range between JShower and GNN reconstruction can be identified as well for the 16% quantile.

The larger spread in form of more frequent outliers, which was already visible in Fig. 6.21, first row, is also manifested in the higher 84% quantiles for JShower of 75% at 10 GeV versus the GNN's 60%. This difference is less pronounced towards higher energies.

Influence of containment

When discussing Fig. 6.20, it was briefly mentioned that the cut on the uncertainty of the GNNs for track-like neutrino events (second row) is not able to identify partly contained tracks as having a potentially inaccurate energy reconstruction. This behavior is studied further in Fig. 6.24. First, on the left, the relative error for the energy reconstruction is plotted versus the distance of the interaction vertex to the center of the detector in the
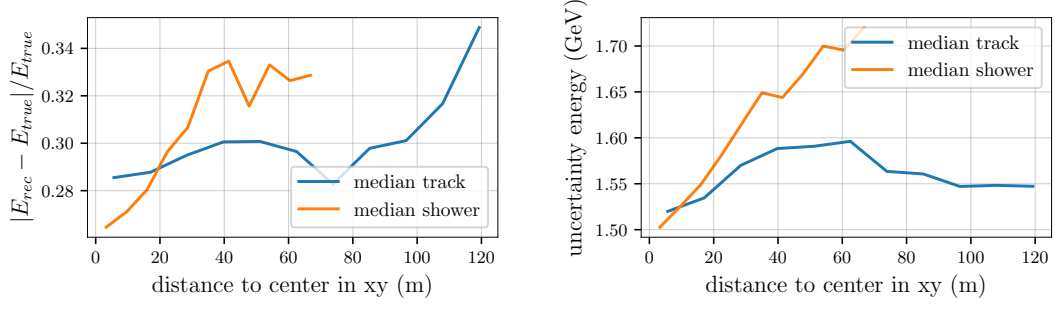
Figure 6.24: Left: Median relative error in energy reconstruction versus the distance of the interaction vertex to the center in the *xy*-plane for tracks (blue) and showers (orange). Right: The uncertainty estimated for the same distance and topologies. For tracks, only events with energies up to $50\,\mathrm{GeV}$ are considered in this plot.

xy-plane. For tracks, the resolution first decreases for larger distances, as they partly leave the detector at $40\,\mathrm{m}$ distance to the center. For distances of around $75\,\mathrm{m}$, a dip is observed, which most likely corresponds to tracks that face the detector from the side and allow for an efficient detection of the Cherenkov light cone. Towards larger distances, the resolution decreases significantly, as the effect of fewer containment dominates. Note that for this plot, only track energies up to $50\,\mathrm{GeV}$ are considered, which causes the overall median to be around 30%. The showers exhibit a monotonously rising relative error, as here, shifting the vertex outside the detector directly causes the spherical, near isotropic light profile to be less contained.

Second, on the right side, the corresponding uncertainties are shown that correctly account for the decrease in resolution for showers, the decrease for tracks until about $50\,\mathrm{m}$, the dip around $75\,\mathrm{m}$, but not for the eventual increase of the relative error for larger distances. Instead, the given uncertainties remain on a low level of 55%. This means the uncertainty cannot be reliably used as a containment cut.

Events with similar amounts of hits

Since energy deposited in the detector is closely related to the cumulative light yield registered by the PMTs, a simple, yet inaccurate way to do energy reconstruction would be hit counting. To prove, that this is not only what the neural network is doing, events with similar number of hits are selected and the median resolution curves are drawn along with those from figures 6.22 and 6.23 in Fig. 6.25.

In fact, the shapes of the medians for both tracks and showers are very similar between all and selected events. This suggests that the number of hits is not the only source of information the network utilizes. Instead, it is able to reconstruct a wide range of different energies for events that have similar amount of detected light.
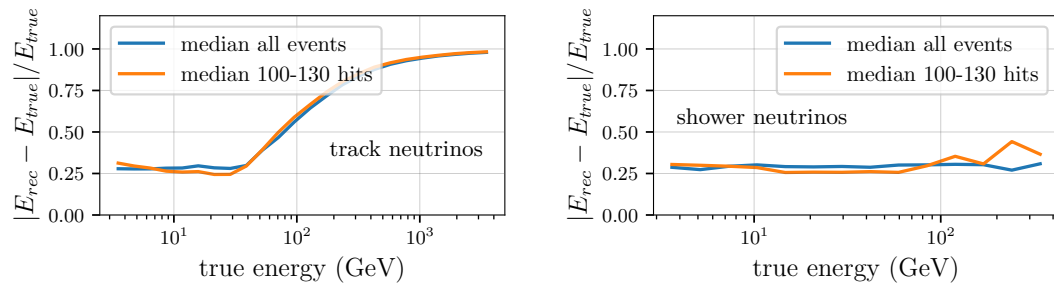
Figure 6.25: Energy resolution versus the true energy of the neutrino. Included are median curves for all (blue) and selected (orange) events with 100 to 130 hits. Left: Track neutrinos. Right: Shower neutrinos.

7

Neutrino selection and oscillation analysis on simulated data

THIS chapter will showcase how the previously discussed outputs of the networks can be used to select a pure neutrino set, which can then be applied to some oscillation analysis. The classifiers are used for particle identification, while the reconstruction networks deliver direction and energy for the analysis. To further enhance the sensitivity of the selection, a combination of cuts based on the classifiers, reconstruction parameters and low-level event observables such as the number of (triggered) hits are employed.

These studies are based entirely on MC simulations to explore the possibilities of the selection without considering any potential discrepancies between real data and MC. The application to real data is then subject to the next chapter.

In general, it is preferable to aim for different selections that concentrate on tracks and showers, respectively, each time focusing on reducing the contributions from the other class. This way, the influence of oscillations can be evaluated separately and the results combined afterward. For a mixed set, the effects from oscillations on showers (ν_e) turning into tracks (ν_μ), and vice versa, cancel each other out, reducing sensitivity. From these different selection strategies, the pure track selection is presented in more detail in the following. After, results can be compared to other selections, including one based on the outputs of classical reconstructions.

7.1 Selection of pure neutrino sets

First, the most important cuts to arrive at a pure track sample are discussed. Thereafter, additional minor cuts are presented that improve the selection. Other samples are then derived from the established set of cuts.
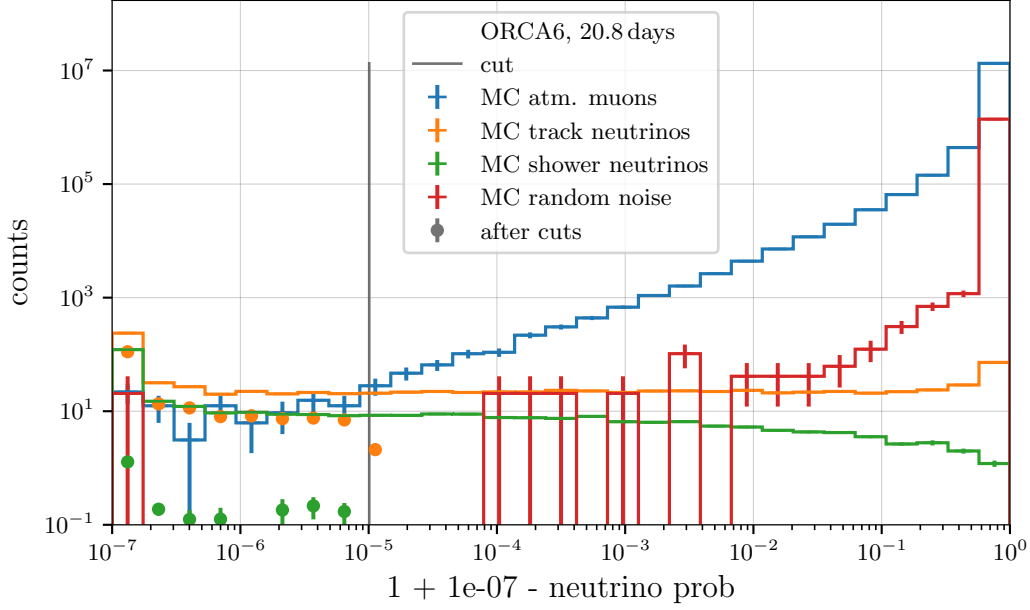
Figure 7.1: Spectrum of the neutrino probability for different topologies (colors) before (solid lines) and after all cuts (points). The gray line represents the cut value on this parameter.

7.1.1 Major cuts for a track set

Starting with the pure track neutrino set, the first cut aimed at reducing the background by several orders of magnitude requests a high neutrino probability. For illustration purposes, the spectrum of $1 -$ neutrino probability is displayed in Fig. 7.1, shifted by some $\epsilon = 10^{-7}$ to avoid 0 in the log space for neutrino probabilities of 1. The plot contains the counts for before cuts as solid lines and after all cuts as points. This way, the influence of this cut can be appreciated for both quantities. Like before, different topologies are resolved, while track neutrinos in orange are dominated by muon (anti) neutrinos. Electron (anti) neutrino and neutral current interactions, responsible for the shower class, behave almost identically. Also, the difference between neutrinos and their antiparticles is not resolved further, which is why in the following for simplicity, they are only referred to as muon or electron neutrinos, omitting the "anti".

Following the discussion in Chap. 6.2.1, it can be seen that this classifier is very efficient in rejecting random noise events. With the chosen cut at $p > 0.99999$, no events from this period of 21 days survive. The number of atm. muon events is still high compared to the neutrino topologies but can be reduced by other cuts. Track- and shower-like neutrinos get both selected in a similar way.

The gray line, which indicates the cut value, is exact. Thus, a part of a bin in this representation may still be selected while the rest is not. This
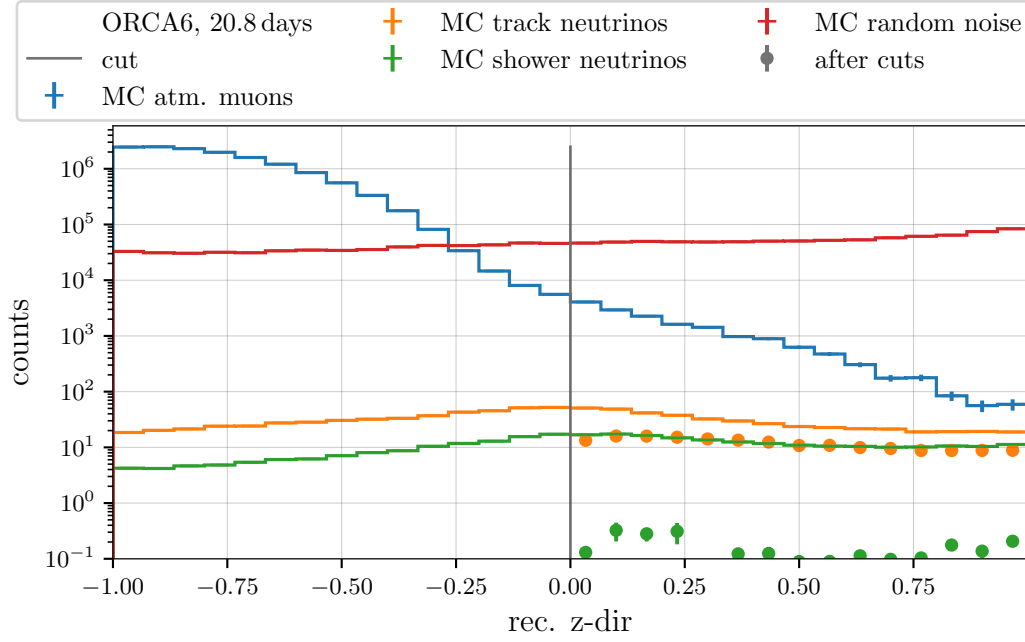
Figure 7.2: Spectrum of the reconstructed z-direction ($\cos\theta$). Same structure as Fig. 7.1.

can cause points from after cuts to seemingly lie beyond the cut value, as they display the bin count of the partially contained bin in its center.

The second important cut to greatly reduce contributions from atm. muons is on the reconstructed $\cos\theta$ of the event. With the dominant downgoing direction of muons, only a tiny fraction of them are falsely reconstructed with a $\cos\theta > 0$. Many of these events are bundles with few hits that create hit patterns that appear much like an upgoing track. The yield of track neutrinos from this cut alone is reduced by a factor of two because of the rapidly increasing contamination below $\cos\theta = 0$ and the resulting low signal to background ratio as shown in Fig. 6.6. The random noise events do not exhibit any preferred direction, as expected.

Finally, the third important cut is to reduce the contributions from shower events, which is achieved by utilizing the track/shower classifier. The cut is chosen to be $p > 0.8$, as two competing objectives have to be balanced. On the one hand, the purity of tracks increases with higher cut values, as is visible in the spectrum of the track probability in Fig. 7.3. On the other hand, the mean energy of the tracks increases with the track probability in the region of interest of higher probabilities (Fig. 7.4), causing a higher cut value to miss out on the critical low-energetic interactions.

This correlation across classifier and regression network further illustrates the shower-like appearance of tracks with low-energies. Interestingly, random noise events peak at the track rather than at the shower side of the spectrum, even though it is less likely that these events display some apparent elongated features.
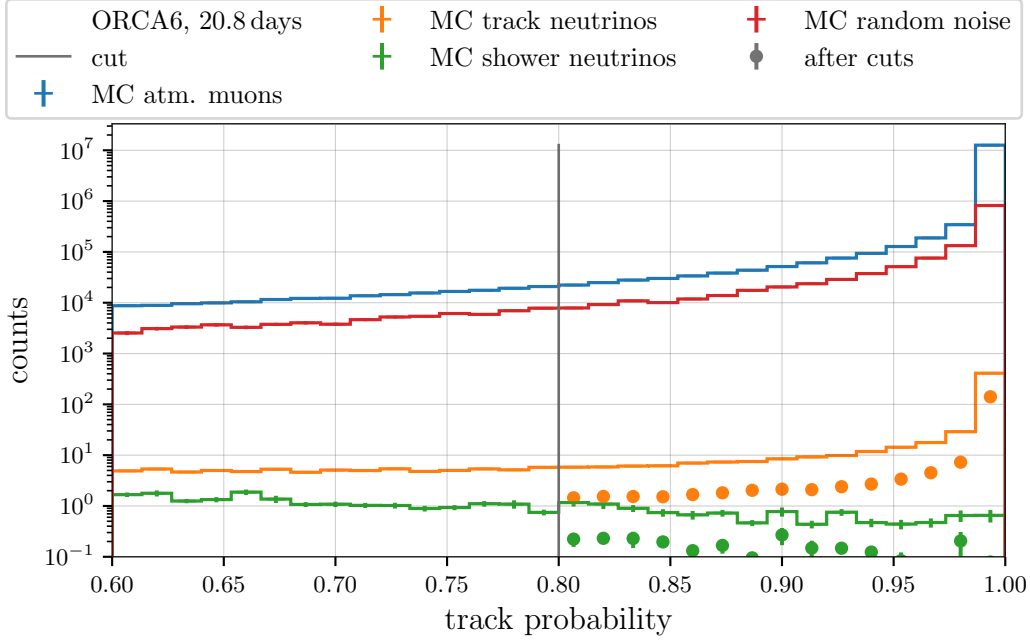
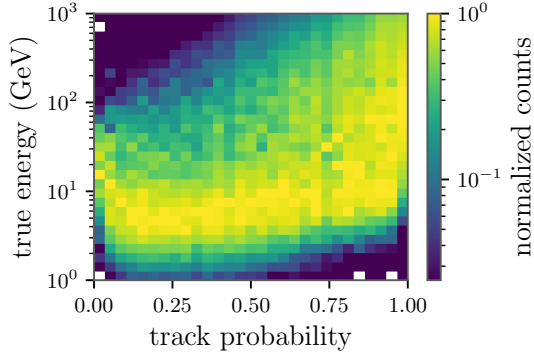Figure 7.3: Track probability spectrum. Same structure as Fig. 7.1.

Figure 7.4: Distribution of the true energy per track probability for track neutrinos. This 2d histogram is normalized to every x bin.

7.1.2 Minor cuts for a track set

In addition to the three major cuts introduced, a few cuts with minor improvements to the reduction of contamination are listed in the following.

- The minimum number of hits and triggered hits per event is requested to be 50 and 8, respectively. Both function to reject ambiguous events of only a few signal hits, which helps to reduce atm. muon contamination. The distribution of the number of hits in Fig. 7.5 shows that in this lower region, only muons are expected. Furthermore, the shower contribution in the lowest bins is rather high.

- A minimum uncertainty to the direction reconstruction components is set to 0.012 for x- and y- and 0.004 for the z-direction. Only atmospheric muons are able to reach such low direction uncertainties. Conversely, for larger uncertainties, the contribution from shower
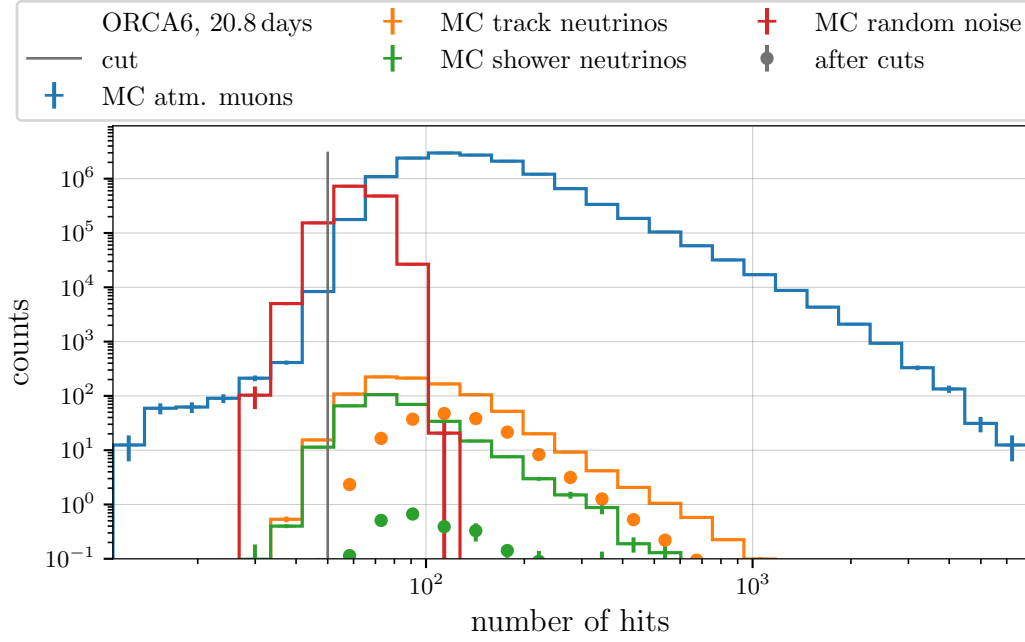
Figure 7.5: Spectrum of the number of (snapshot) hits per event. Same structure as Fig. 7.1.

events is almost as significant as that from tracks. To diminish the overall shower percentage, cuts at 0.4 for x- and y- as well as 0.3 for the z-direction are employed. The spectrum of the z-direction uncertainty is shown in Fig. 7.6 with the cuts added as gray lines.

- Lastly, the uncertainty of the energy reconstruction can be utilized similarly. Since showers have generally smaller uncertainties here, the cut at a factor of 2.6 solely serves to reduce atm. muons.

7.1.3 Shower and all flavor set

Following the same approach of reducing the corresponding background to a minimum, two further selections are made; one aiming at a pure shower sample and one that does not distinguish the topologies but is directed at a higher overall neutrino yield. In analyses, sets of tracks, showers and possibly a mixed class are typically evaluated independently. Afterward, their respective results can be combined to maximize the significance of the fit to oscillation parameters. The sets are kept separate for this section, however, and their individual properties are evaluated. Also, these sets are not exclusive, as the all flavor set encompasses most of the other two sets.

In particular, the cuts for the shower set have been adjusted in the following way:

- The main change is the cut on the track probability at a maximum of 0.05. A zoom of the spectrum around that point is shown in Fig. 7.7.
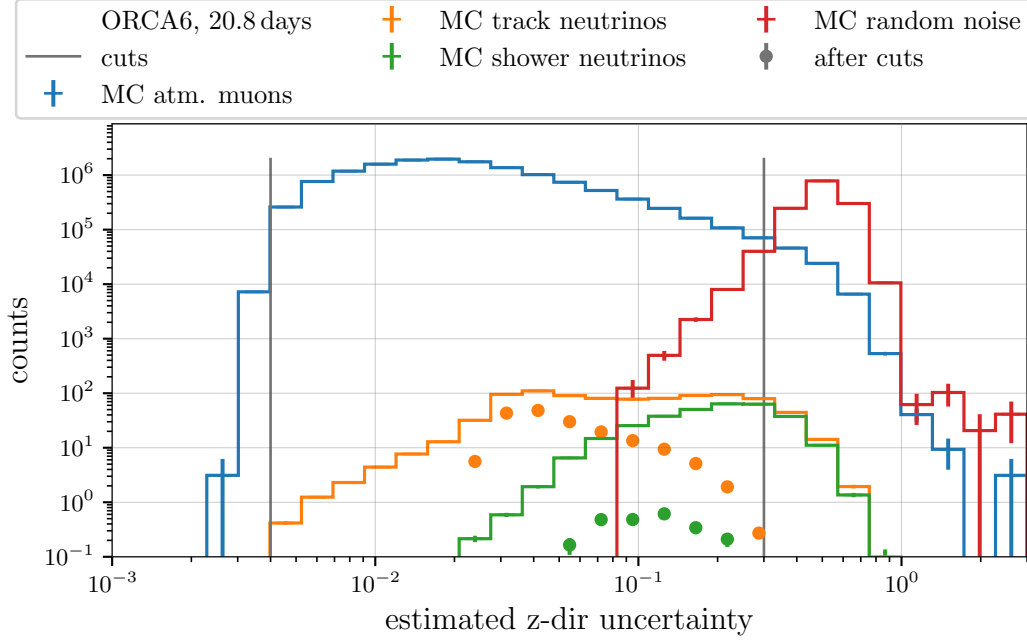
Figure 7.6: Distribution of the estimated uncertainty for the *z*-direction. Same structure as Fig. 7.1.

It is difficult to identify an optimum because the track class continues to be present for the complete range.

- The neutrino probability remains the same at 0.99999, but the limitation on the *z*-direction was lifted, also allowing for downgoing solutions.
- The hit-based cuts have been slightly relaxed to 5 and 40 for the triggered and snapshot hits, respectively, to allow for more lower-energetic shower events. The track score cut efficiently suppresses atm. muon events in this region.
- Cuts on the uncertainty in direction and energy have essentially been lifted, only discriminating against extreme outliers sometimes observed in atm. muon events. Also, the lower bounds remain to exclude extreme small cases.

For the third set of mixed flavors, the following strategy was applied:

- The neutrino probability was required to be even closer to 1 with $0.999999 \, (1 - 10^{-6})$ to ensure the contamination stays low while relaxing some of the other cuts against atm. muons and not using the track/shower classifier.
- The minimum $\cos\theta$ at 0 (horizontal) and the hit-based cuts are adopted from the track case.
- The uncertainties from the reconstruction networks are relaxed from the track selection to be 0.6 for *x*-, 0.7 for *y*- and 0.4 for the *z*-direction and 2.75 for the energy uncertainty factor. The lower bounds are kept throughout.
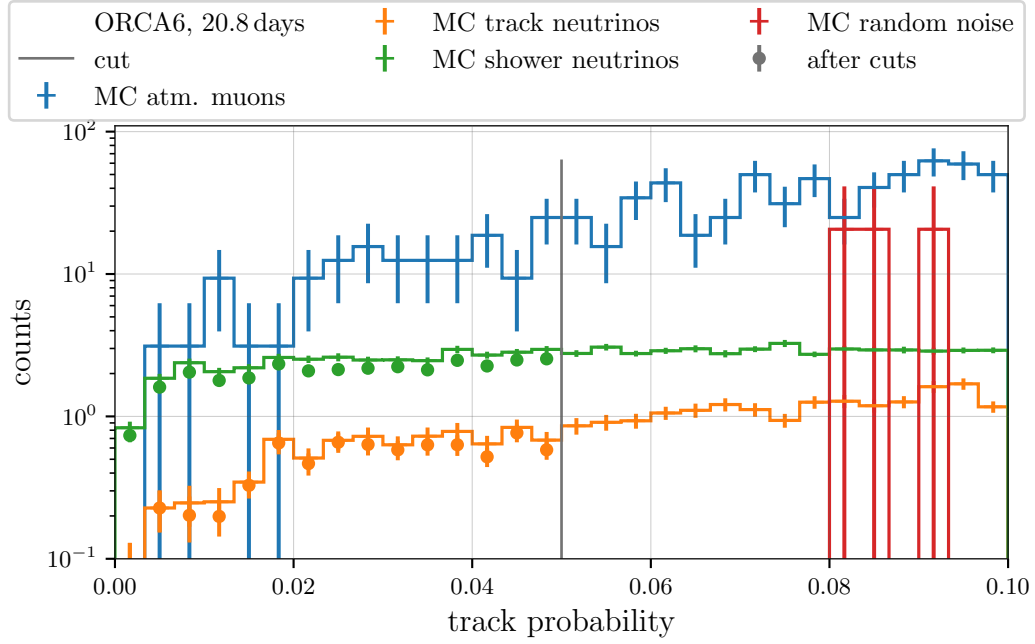
Figure 7.7: Zoom to the lower values of the track probability, illustrating the cut at 0.05. Same structure as Fig. 7.1.

7.1.4 Comparison of selected sets

Scaling all three selections to the available data recorded with ORCA6 from February 2020 to March 2021, which spans 357.2 days of effective lifetime, overall efficiencies and contamination can be determined. These and some additional numbers characterizing the selections are gathered in Tab. 7.1. From these numbers it can be seen that in the pure track selection, an efficiency of 18.4% can be reached while both the contamination by atm. muons and the contributions from shower events are kept to around 1.5%. With such a low contamination of a few percent, an oscillation analysis is not impaired by atm. muons and does not suffer from a washed out oscillation signal due to the shower channel. Random noise contributions are no problem for any of the selections. The overall number of selected events reaches almost 3,000 while 600 of them, one fifth, have reconstructed energies of smaller than 20 GeV, roughly indicating the most interesting region for oscillation studies.

The largest impact in this selection on the energy spectrum comes from the cut on the track probability, while the signal/background classifier keeps most of these events, as was discussed before and is illustrated again in Fig. 7.8. There, the neutrino probability cut rejects even more higher-energetic events, as expected from the discussion of Fig. 6.6. The track/shower classifier cuts away about 80% of the events at 10 GeV, increasing towards lower energies.

127

Table 7.1: Results from three different selections on 357.2 days of ORCA6 data. Overall contamination here means the contributions from atm. muons, random noise and the respective other neutrino topology combined.

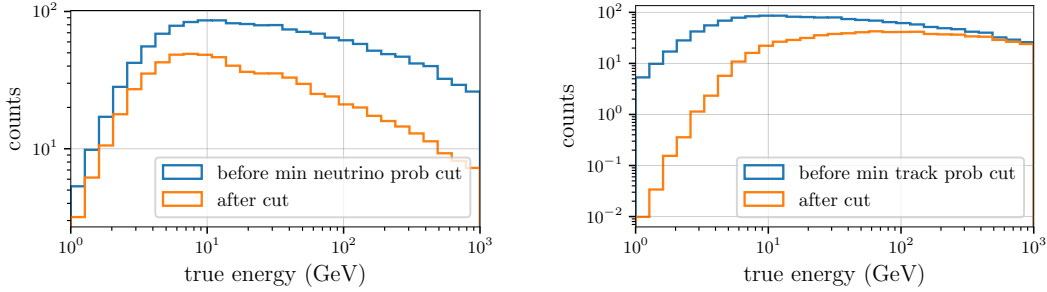| quantity \ selection	track	shower	all flavors
track efficiency (%)	18.4	0.5	23.6
shower efficiency (%)	0.7	7.4	32.6
atm. muon cont. (%)	1.6	3.3	1.1
random noise cont. (%)	0	0	0
overall contamination (%)	2.9	21.0	1.1
selected events	2,965	582	5,642
selected events $E_\nu < 20\,\text{GeV}$	600	419	2994

Figure 7.8: Influence of cuts on the true energy spectrum of all neutrino flavors (blue before, orange after cut). Left: Cut on the neutrino probability at > 0.99999. Right: Cut on the track probability at > 0.8.

Of the 2,965 selected tracks, only 17 events (0.6%) are tau neutrinos with a muon in their decay channel.

A pure shower selection in ORCA6 is challenging in general. In this attempt, the track neutrinos were greatly reduced to 0.5% of their initial count and also the atm. muon background is suppressed to 3.3% contamination in the final set. However, the remaining track neutrinos still account for 17.7% selected events. The fraction of the low-energy events is large with 419 out of 582 (72%), but all remaining muon neutrinos in this region will obscure the oscillation signal from electron neutrinos, as transition probabilities are high. To be more precisely about the counts in the energy and $\cos\theta$ region of interest following Fig. 2.5 ($E_\nu < 10, \cos\theta < -0.4$), there are about 80 significant events for oscillation research.

The composition of selected events features 20.9% NC, 50.9% electron CC interactions and 7.2% tau neutrinos that decay into hadronic or electromagnetic showers. NC interactions do not yield any information about oscillations, so that their contribution essentially can be neglected. In
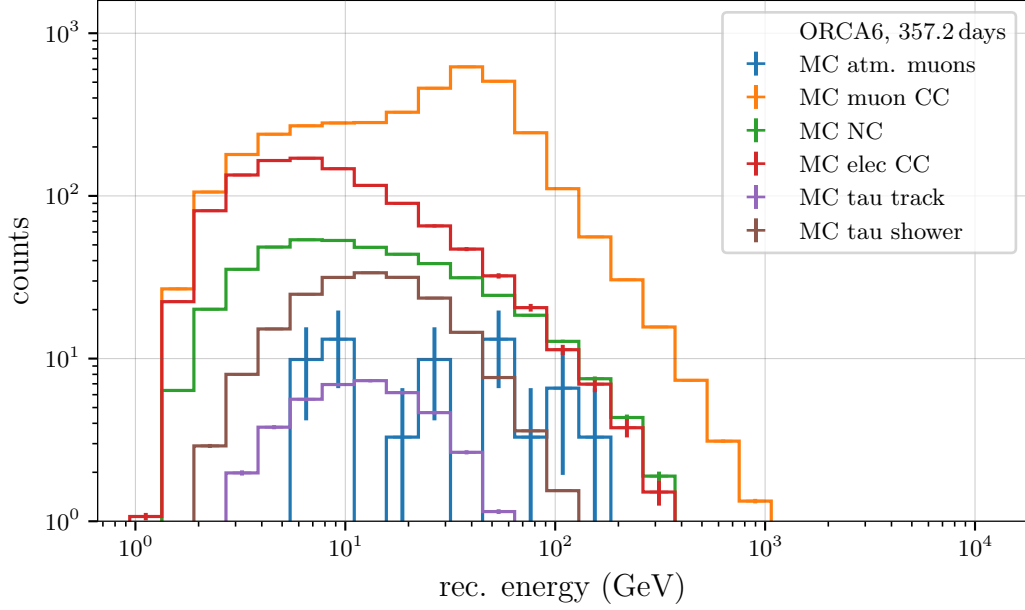
Figure 7.9: Reconstructed energies of the selected neutrinos resolved for all signatures.

ORCA6, these hadronic showers are inseparable from their electromagnetic counterpart.

The third selection accomplishes a high efficiency for both tracks and showers with 23.6% and 32.6%, respectively. The contamination from background events is contained to around 1%. This set excels at high statistics with more than 5,600 events and almost 3,000 with energies smaller than 20 GeV. The complete energy spectrum resolved for each particle type is depicted in Fig. 7.9.

The dominating contribution comes again from muon neutrinos with 66.8%, followed by electron CC interactions with 19.8%, while 242 tau neutrinos (4.3%) are expected from this year of data taking. This hints at the future prospect of tau appearance studies [138].

Next, it is only natural to evaluate the sets' suitability for oscillation research quantitatively. In principle, different strategies regarding splitting the data and combining results are possible. From the sets introduced in this section, two distinct possibilities come to mind: The first one is to combine the pure track and shower sample in an analysis. The second is to divide the all flavors selection at some track probability value and combine the results from individual analyses of these. This way, the first option is based mainly on the track channel with a pure selection and high-quality events, while the second option includes a more mixed selection of events but more extensive statistics.
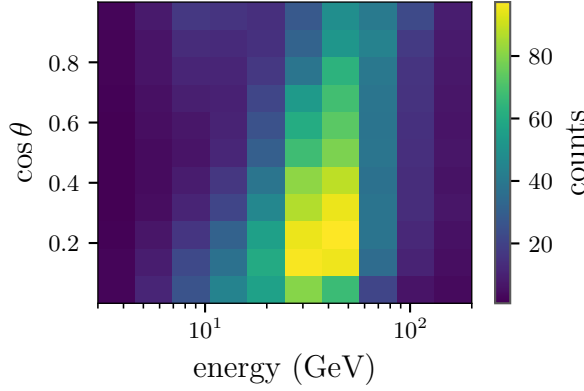
Figure 7.10: Chosen binning in $\cos\theta$ and energy for comparing data to a model. This example contains the pure track selection and the model assuming NuFIT values.

7.2 Sensitivity to oscillation parameters of selected sets

In order to add a quantitative measure about the suitability of the data selected by the outputs of the graph neural networks for oscillation research, the sensitivity to the parameters θ_{23} and Δm^2_{31} is evaluated in a purely statistical way. This means in particular, that no systematic uncertainties are considered, as they will be studied in detail in dedicated future analyses. Nevertheless, for proving the points aimed at in the scope of this thesis, including comparisons to the standard selection based on the outputs of classical reconstructions, such fast comparisons yield already sufficient evidence about the quality of the selected data.

The technique to quantify the sensitivity to θ_{23} and Δm^2_{31} is to compare the expected and observed event counts in bins of energy and $\cos\theta$ and then to vary the oscillation parameters, which change the expected counts like in Fig. 2.5. This comparison is achieved by a χ^2 between the model and the data that is summed up over all bins to state the agreement for a set of oscillation parameters. This is repeated for different θ_{23} and Δm^2_{31}, and from the resulting χ^2 map, the confidence levels can be determined.

In detail, this is realized in the following way. First, the selected data must be binned in a 2d histogram of energy and $\cos\theta$. The bin edges are chosen such that almost all events are considered without creating too many empty bins. Also, it needs to be possible to still resolve the impact of oscillations by events changing bins for different assumed parameters. If the bin size is too large, this effect is no longer visible. For the track selection, for example, this results in 10 logarithmic bins from $3\,\text{GeV}$ to $200\,\text{GeV}$ in energy and 11 bins in $\cos\theta$, ranging from 0 to 1 as can be seen in Fig. 7.10.

Such a histogram is then created for the data and the model (expectation from MC) by weighting the events with the oscillation probability according to the currently tested parameters. For the studies in this chapter, the "data" are represented by an Asimov data set of the same MC events as in the model,
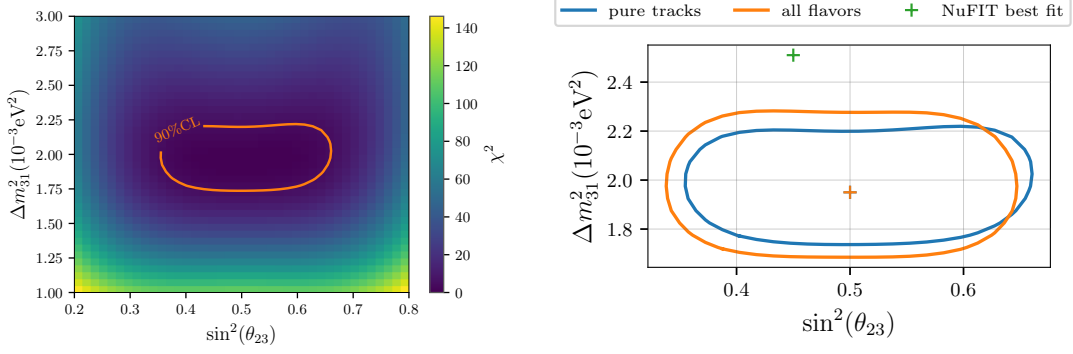
Figure 7.11: Left: χ^2 map in Δm_{31}^2 and $\sin^2\theta_{23}$ for the pure track selection. The dark blue areas indicate good agreement between data and model. The contour line for the 90% CL (orange) is determined from the $\Delta\chi^2$ relative to the minimum. Right: Only the extracted contour lines comparing the pure track (blue) and all flavors selection (orange). The orange and blue crosses (overlapping) mark the minimum, which was chosen to be (0.5/1.95). As a reference, the NuFIT best fit value is added (green) [52].

weighted with specific "true" oscillation parameters. The term "Asimov" data set refers to a set that represents the same distributions real data is expected to have, but delivers mean expectation values to otherwise discrete observables so that analyses do not suffer from statistical fluctuations. These are chosen to be $\sin^2\theta_{23} = 0.5$ and $\Delta m_{31}^2 = 1.95 \times 10^{-3}\text{eV}^2$, following the previously published results based on the classical selection [139]. This is, however, only an arbitrary choice to be able to simulate the expected sensitivity for such a specific parameter set and to use it to state a significance with which other sets, like the NuFIT best fit (see Tab. 1.1), can be excluded.

The χ^2 that measures the difference between the data (O, observed) and the model (E, expected) histogram counts is defined as [140]

$$\chi^2 = 2\sum_i O_i \ln(O_i/E_i), \tag{7.1}$$

also known as the G-test, which is similar to Pearson's widely used χ^2 test [141] but more robust for smaller sample sizes. The sum of the χ^2 over all bins yields a total value, which then quantifies the agreement between the two histograms. This comparison is repeated for different combinations of θ_{23} and Δm_{31}^2, which allows to draw a χ^2 map in the parameter space such as done in Fig. 7.11, left.

The minimal χ^2 signals the best fitting parameter set. Next to the minimum, a confidence level (CL) can be deduced from the difference in χ^2 value to the optimum. The 90% CL, for example, for a distribution with two free parameters (Δm_{31}^2 and $\sin^2\theta_{23}$) can be sampled from the cumulative χ^2 distribution with the corresponding number of degrees of freedom and hence is found to be 4.61 [142]. Meaning a contour for this level can be drawn over bins that exhibit a $\Delta\chi^2$ of 4.61, as is added in orange in Fig. 7.11,

left. Similarly, the difference of any point in the parameter space can be assigned a significance with which it differs from the minimum. This is done translating the $\Delta\chi^2$ to a p-value/confidence level (like it is illustrated in [142], figure 40.1) and then expressing it in terms of standard deviations (σ). In the context of this study, the $\Delta\chi^2$, and thus the expected significance of exclusion, to the NuFIT value is an interesting point to probe. While the chosen combination of the test data's true value at (0.5/1.95) and the difference to NuFIT's best fit allows for the desired quantitative evaluation, it should be kept in mind that this is only one possible application in which the neutrino sets can be tested for their suitability to oscillation studied.

With this procedure of producing a contour from a neutrino selection, the precision for the two oscillation parameters given by the size of the enclosed area, can be directly compared. To this end, contours of the pure track and the all flavor selection are contrasted in Fig. 7.11, right. The shower selection alone does not yield significant sensitivity due to the small sample size. A $\Delta\chi^2$ of larger than 4 is hardly realized anywhere in the studied parameter space. For most combinations of test parameter sets, the sensitivity achieved by the pure track sample is higher than for all flavors. The higher statistics cannot make up for the fact that oscillation signatures from ν_e and ν_μ cancel each other out for some of the regions in $\cos\theta$ versus energy. The significance of the difference to the NuFIT values, for example, is determined to 4.0σ and 3.0σ, respectively. Yet, for larger $\sin^2\theta_{23}$, the additional shower component adds some information to the analysis.

In order to improve the results of the all flavors selection further, the set can be split according to the track probability. As cut, the 0.8 used in the pure track cuts is applied. The resulting sets then feature one track-dominated part with 537 events exhibiting a reconstructed energy below $20\,\mathrm{GeV}$ and a shower percentage of 1.3%, and a mixed part with 54.2% shower contributions (34.3% electron neutrinos) and 2456 events below $20\,\mathrm{GeV}$. With this, the former set is very similar to the initial pure track sample and the latter encompasses the initial pure shower selection with a lot of additional events from the middle region. It is also possible to further divide such sets, even utilizing other quantities than the track probability, which helps to constrain systematics. One possibility is to use different bins in the uncertainties estimated for the direction components or the energy, with the highest quality class of low uncertainties and others with higher values.

When dividing selected sets in such an orthogonal way, i.e., having no overlap in events, the resulting χ^2 maps can be combined (added) to boost the sensitivity. This is realized in Fig. 7.12 for the track and shower parts of the all flavor selection. The newly added green and red contours denote the sensitivities achieved by only analyzing the separate parts from the all flavors selection. The track part appears to yield similar results to the former pure track selection, as expected from its characteristics. The shower
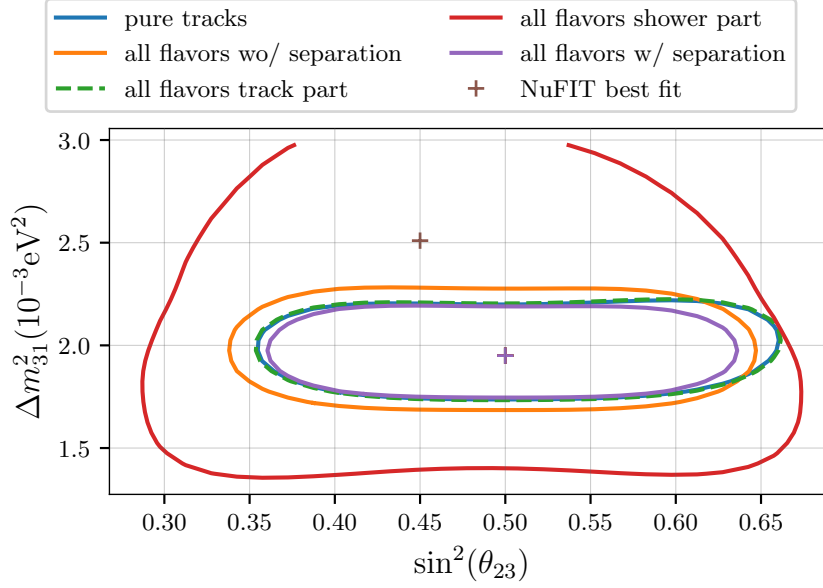
Figure 7.12: 90% confidence intervals for different selections and reconstructions in Δm_{31}^2 and $\sin^2 \theta_{23}$. For comparison, the selections from Fig. 7.11, right, are included (blue and orange). Green and red mark the two samples the all flavors set is split into. The purple contour combines the sensitivities from the two split parts. As a reference, the NuFIT best fit value is added [52]. (brown)

part does not exhibit significant sensitivity to Δm_{31}^2, but it is similar to the track selections when restricting the upper boundary of θ_{23}. Consequently, the combination of both curves is shown in purple and agrees mostly with the track selections in the Δm_{31}^2 dimension, while the sensitivity to θ_{23} can be improved by the additional shower component.

Evaluating the $\Delta \chi^2$ between true and NuFIT parameters, the track part of all flavors reaches 3.9σ, comparable to the 4.0σ for the pure track selection. The shower/mixed part alone only achieves 0.9σ, but together, the split set from all flavors can exclude such a point with 4.2σ.

To put these results into perspective, the same evaluation can be done for the selection based on classical reconstruction outputs, which will be carried out at the end of the next section.

7.3 Comparison to classical selection

A selection based on the outputs of the JGandalf reconstruction has been established for the use to produce physics studies from the ORCA6 data spanning one year [139] [143] [144]. It was developed by several people over the course of some iterations and consists of hand-crafted cuts on reconstruction and hit features. Since there is no distinction between track and shower events in the current version, the comparisons is made to the all flavor selection from the proceeding section.

Table 7.2: Numbers characterizing the all flavors GNN and classical selection for 357.2 days of ORCA6 data.

quantity \ selection	all flavors GNN	classical
track efficiency (%)	23.6	9.1
shower efficiency (%)	32.6	7.8
atm. muon cont. (%)	1.1	2.0
selected events	5,642	1,898
selected events $E_\nu < 20\,\text{GeV}$	2,994	1,002

Some general quantities are compared in Tab. 7.2. From these, it becomes clear that overall higher efficiencies can be achieved by the deep learning approach, while maintaining an even lower contamination by atmospheric muons. The classical selection naturally discriminates more against showers as it is based only on track reconstruction. This way, the showers only make up 22% versus the 31% for the GNN selection. The yield of track and shower neutrinos is thus 2.6 and 4.2 times higher for deep learning. The fraction of low-energetic events with 53% is the same for both selections. However, the energy reconstruction of JGandalf exhibited some underestimation already around the 10 GeV region (Fig. 6.20, first row), which causes some events with higher true energies to be wrongly reconstructed with low energies.

One way to study the intricate features distinguishing the two selection approaches is to look at distributions of true quantities of the events. One such example is the true $\cos\theta$ from the simulations, displayed in Fig. 7.13. To allow for better comparison, only muon neutrino CC interactions and atmospheric muons are shown. From this, the susceptibility of the selection to actually downgoing particles can be evaluated. For both cases, the overwhelming majority of events is correctly identified as upgoing (both selections require a reconstructed z-direction of >0). However, a fraction of events survives all cuts, which decays with the difference to the horizontal direction. Consequently, all selected atm. muons are naturally misreconstructed. This fraction is slightly higher for the GNN selection with 6.8% versus the 3.5% for the classical selection. Here, the stricter quality cuts imposed on the latter selection, initially aimed at reducing the muon contamination, cause this focus on higher-quality events. The inclusion of the shower-like interactions causes the fraction of downgoing solutions to increase due to the lower resolution of the direction for those events.

For the upgoing region, slightly different shapes of the distributions can be observed. In general, in both cases the trend from the original flux as shown in 1.4 is visible, where the rate peaks at horizontal directions and falls off towards vertical. This trend is modified by the quality cuts that discriminate against lower-energetic neutrinos that traverse the detector
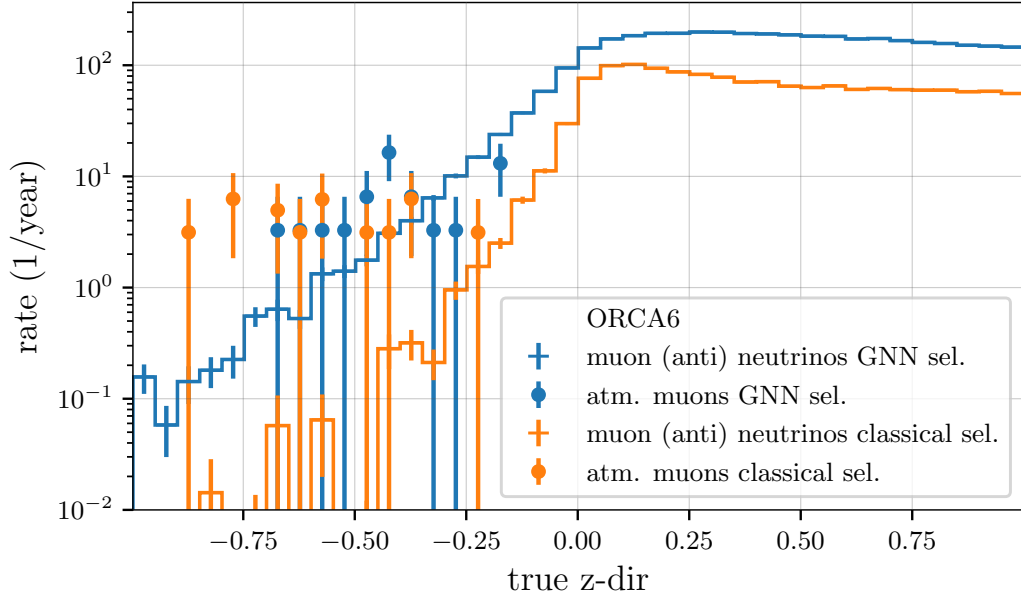
Figure 7.13: Spectrum of the true *z*-direction (−1 is downgoing, +1 is upgoing) for muon neutrinos (crosses) and atm. muons (dots) as selected by either GNN (blue) or classical cuts (orange).

from the side, causing them to be potentially badly reconstructed. Their lack is manifested in the drop of the selection rate already for slightly higher values than $\cos\theta = 0$. This effect is different between the samples, as the GNN selection is flatter and exhibits its maximum around 0.25, while the classical selection peaks at 0.1.

Furthermore, the spectrum of the true energy, plotted in Fig. 7.14, indicates that especially for energies below $50\,\mathrm{GeV}$, the efficiency from the GNN selection is higher to all neutrino flavors combined. This is partly due to the larger fraction of showers that are considered in the GNN selection. Towards higher energies, the selected rates become comparable. Similar conclusions can be drawn from the distributions of the true vertex position, where the selection rates are the same for large distances but higher for GNN selected events for when the interaction happened in or close by the detector. In these regions, also lower-energetic particles can cause a trigger to fire, while events further away need to be brighter and thus are easier to reconstruct. This trend also causes more events with a low number of hits to end up in the GNN selection, while larger numbers of hits are similarly frequent.

Despite the overlap in the spectra, comparing the selections on an event-by-event basis reveals that only 58.8% of the classical is contained in the GNN selection and 22.6% of the GNN in the classical selection. Due to technical reasons, these numbers refer to the real data selection presented later. The numbers should be, however, similar in simulations.
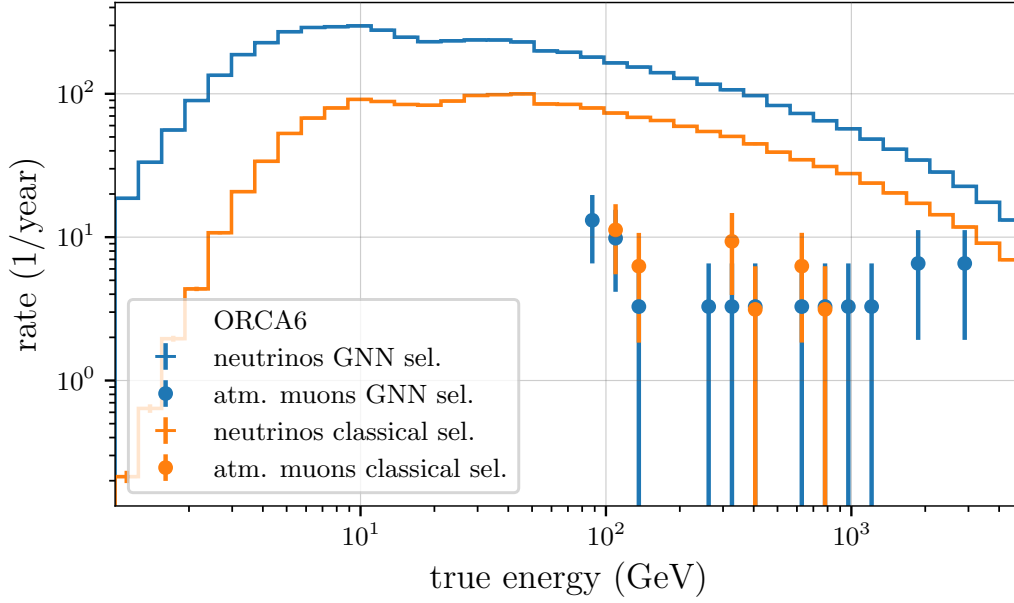
Figure 7.14: Spectrum of the true energy of atmospheric neutrinos (crosses) and muons (dots) as selected by either GNN (blue) or classical cuts (orange).

One interesting variable to monitor is the quality parameter from the classical reconstruction (likelihood over number of hits used in the fit). This is shown for the selected events in both cases in Fig. 7.15, left. The orange distribution is capped to the minimum of 2, which was the cut value chosen to help with the suppression of ambiguous atm. muon and random noise events. Events that are not processed by the main direction fit of the classical reconstruction JGandalf are assigned a 0 in this spectrum to also include them in the discussion.

Starting at these lowest values, the tiny fraction of 20 events (0.3%) pass the complete GNN selection but did not pass the hit selections during the classical reconstruction chain. Similarly, the lower classical qualities are also strongly suppressed, increasing nearly exponentially towards medium qualities of 2.5. Overall, about 1000 neutrino events are gained in the deep learning approach when considering the cut at two on JGandalf's likelihood. Half of the selected atm. muons are also in that region. The two distributions are very similar in shape for the rest of the spectrum, indicating no particular differences in selection efficiencies other than the overall larger yield for the neural networks.

The most important figure of merit for the quality of an event selection is its sensitivity to the physics the experiment aims at. Hence, the comparison to the classical selection encompasses the 90% CL in θ_{23} and Δm^2_{31} once of the all flavors set without any further splitting and once with the aforementioned separation into a track and a shower sample. The result is depicted in Fig. 7.15, right. There is currently no obvious way to split the
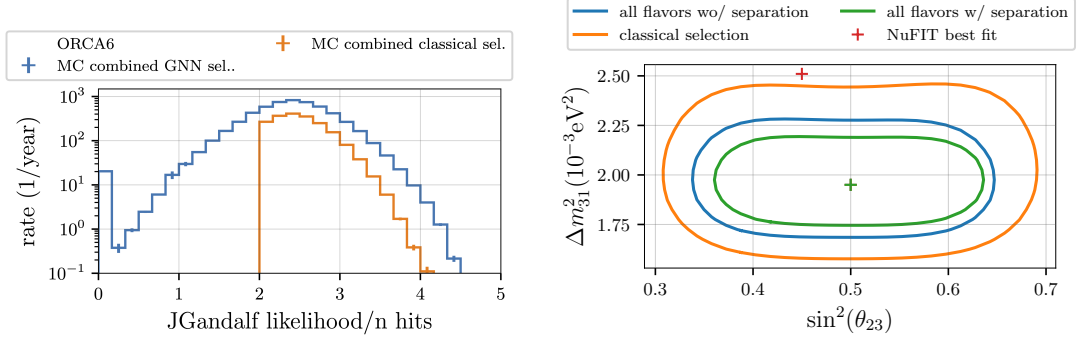
Figure 7.15: Left: Rates of the quality measure from the classical direction reconstruction in GNN (blue) and classical selection (orange). In the latter case, the cut value is 2. Events selected by the deep learning selection that were not reconstructed with the JGandalf stage are plotted in the first bin at 0. Right: Comparison of the 90% CL contours in Δm_{31}^2 and $\sin^2 \theta_{23}$ from the deep learning approach, with (green) and without (blue) splitting the set into track and shower part, and the classical selection (orange). As a reference, the NuFIT best fit value is added (red) [52].

set for the classical selection to improve sensitivity. Because of this, it is directly comparable to the all flavors set without separation. The sensitivity of the split set is added to better appreciate the overall improvement possible when taking advantage of the complete deep learning approach.

From the size of the contours, it becomes directly apparent that utilizing the GNN selection, a significantly higher sensitivity to oscillation parameters is reached for the same initial data set. The improvement is similar for the entire region of the phase space, only marginally smaller for lower than it is for higher θ_{23}. The equally distributed gain in sensitivity is expected from the similar compositions of the selections of electron and muon neutrinos.

Considering the significance with which a value such as the NuFIT best fit can be excluded from a (0.5/1.95) truth, the differences can be expressed in single numbers. For the classic selection, 1.9σ are determined from the $\Delta\chi^2$, while 3.0σ and 4.2σ are yielded from the GNN selection without and with splitting the set, respectively. It should be remembered that these studies do not consider systematic effects, which typically slightly reduce the significance quoted.

From the discussion so far, two main factors contribute to the enhancement in sensitivity; more accurate reconstruction of direction and, in particular, energy by the neural networks and the higher statistics of events exhibiting oscillation, enabled by utilizing the classifier outputs. In order to get a feeling for the importance of the two effects, a contour is generated based on the GNN event selection but taking the reconstruction variables from JGandalf. This is carried out in Fig. 7.16 on the complete all flavors set that then can be compared to the cases from before, once
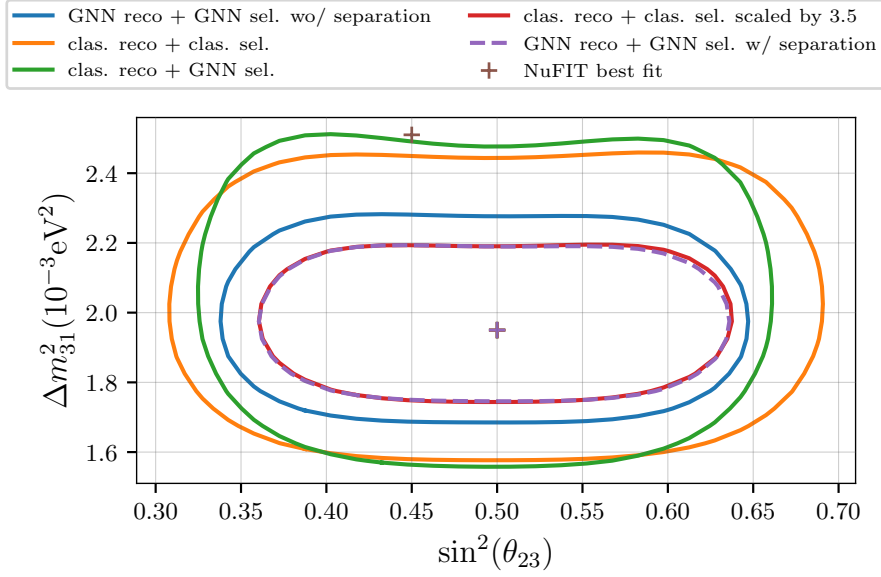
Figure 7.16: 90% confidence intervals for different combinations of deep learning and classical selection and reconstruction algorithms in Δm_{31}^2 and $\sin^2 \theta_{23}$. For the red case, the selected events have been weighted with an additional factor of 3.5. As a reference, the NuFIT best fit value is added (brown) [52].

keeping the reconstruction method fixed and once the event selection. In addition, the statistics of the classical selection plus classical reconstruction are scaled up by a factor to match roughly the sensitivity achieved by the GNN reconstruction and selection.

Starting with the two cases maintaining the same selection, the influence of the reconstruction method can be evaluated. In the plot, this corresponds to the blue and green lines. The higher precision from the GNN reconstructions is manifested clearly in the improvement in sensitivity, as for most regions in the parameter space, a significant improvement is observed. This effect is most substantial in the mass splitting and less evident for the oscillation angle. The specific test point of the NuFIT parameters, starting from the same true parameter set as before, can be excluded with 1.8σ for the green contour, while 3.0σ are achieved in the blue case.

The same reconstruction method is used in orange and green, but the selections differ. The most crucial difference is then simply the number of events relevant to the analysis. The effect, however, does not seem to be very significant. Only in θ_{23} the sensitivity can be enhanced, while in Δm_{31}^2, the resulting contour for GNN selection is even larger than the classical one. These mixed results can be explained by the fact that the additional results suggested by the neural networks are poorly reconstructed by the classical approach and thus do not add abundant information to the χ^2 map. Especially the shower events (31%) pose a challenge to the track

reconstruction. Consequently, the significance to the NuFIT set is even higher for the purely classical analysis with 1.9σ.

To further quantify the overall improvement that can be achieved with the full deep learning approach, which includes the possibility to split the dataset into a track and shower part, the statistics of the classical case are scaled up by a factor of 3.5 to reach the same sensitivity. This is demonstrated with the red line in Fig. 7.16. In other words, either the data taking lifetime would have to be 3.5 times longer when relinquishing the GNNs or the active volume would have to be more extended by the same amount. This assumes that the uncertainties on the results are statistically dominated. That is not entirely the case, as there are also considerable systematic influences to the overall uncertainty for the current ORCA6 data set [145] [146], like currently ongoing studies indicate [147]. Yet, there are contributions to the systematics that can be constrained by the data itself, resulting in an approximately linear scaling of the χ^2 values with statistics, i.e., lifetime or volume.

Selection and analysis applied to real data

T<small>HE</small> natural next step is to apply the trained neural networks to real data in the same way as done for MC and to evaluate their behavior. Any discrepancies between expected and observed data can have two different origins; the networks may not generalize well to real data, as they are confused by specific, not necessarily physical features that are different between data and simulations. Alternatively, parts of the actual underlying physics are not simulated sufficiently. The latter occurs naturally, as software within a physics experiment in its earlier stages has to be developed, expanded, improved and fine-tuned to model real data reliably. Such discrepancies, including some for neutrino simulations, are already indicated by preceding internal studies. However, these deep learning selections allow to probe the MC with higher statistics than previously possible and thus to identify features that were not apparent before.

Even though the focus of this chapter lies on the verification of the neutrino simulation chain utilized, the resulting selected data can be used for some first oscillation research. Thus, the first section will demonstrate how the selection strategy is adjusted for real data. Several cross-checks on the potential influence of data taking periods and conditions on the selection are presented thereafter and finally, the preliminary oscillation analysis from one year of ORCA6 data will be performed.

8.1 Neutrino selection in real data

For a model that has been developed and optimized only using simulated data, it is crucial to verify its application to real data. To achieve that, thorough comparisons between data and MC are carried out, evaluating the outputs of the GNNs before and after cuts. If a quantity does exhibit significant discrepancies, it cannot be reliably used as a cut variable, as the effect on data and MC is different. From the studies summarized in this section, it becomes clear that an adjusted strategy has to be adopted for
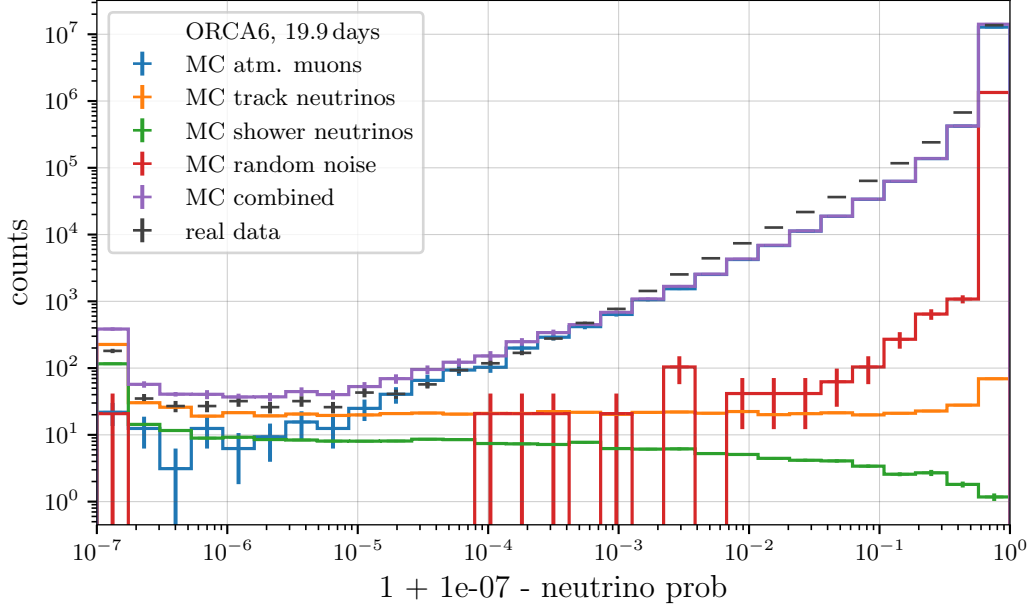
Figure 8.1: Distribution of the neutrino probability for different contributions in MC (colors) and real data (black). For better visibility, the quantity is reversed and plotted as a log scale on the x-axis. With the shift of $+10^{-7}$, also the scores of 1 can be displayed.

real data. The motivation for this, the execution details, and the resulting spectra are discussed in the following.

8.1.1 Selection strategy for real data

As presented in Chap. 7.1, the natural choice for identifying a pure neutrino set is by using the signal/background and the track/shower classifier for further distinction. In Fig. 8.1 and 8.2 these quantities are shown before any cuts for a few days of lifetime. For the most part, this only allows for comparisons to the atm. muon and random noise predictions, as the rate is dominated by background.

The spectrum of the neutrino probabilities exhibits a discrepancy for high scores (on the left side in the chosen representation), where the combined predictions from the MC are up to three times larger than the observed real data. This constitutes a different shape of the spectrum, which peaks less distinctively for real data at $p = 1$ than it does in MC. Instead, more entries spread out over the entire region of medium values from 0 to 0.999.

A similar pattern is observed in the track score (Fig. 8.2). There, the shapes are also different between the combined MC and real data, as the latter is assigned overall larger track probabilities. The discrepancy is as large as one order of magnitude for the most shower-like bins around zero.

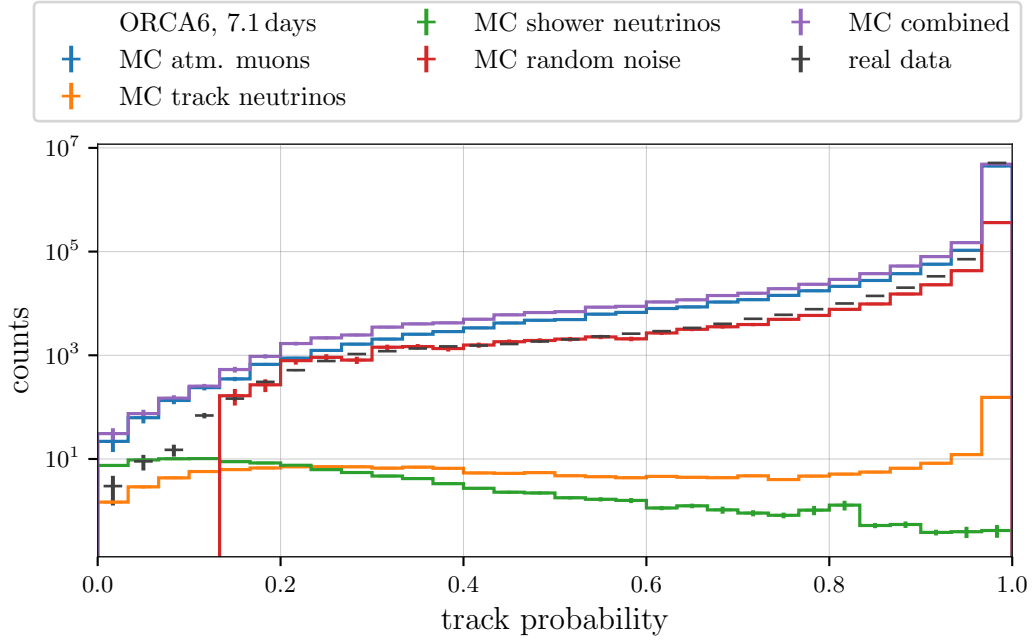Since these discrepancies also persist for any neutrino dominated

Figure 8.2: Distribution of the track probability for different contributions in MC (colors) and real data (black).

selections, i.e., "after cuts", as will be revisited in Sec. 8.1.2, the cut on the neutrino probability is chosen to be not as strict to lower its impact and contribution to the overall data/MC disagreement. The track probability is disregarded altogether. Instead, most of the discrimination power is shifted to the reconstruction networks, as they display reasonable agreement. In Fig. 8.3, for example, the estimated uncertainty of the z-component of the direction is shown, which is well modeled by atm. muons, random noise and also after cuts by neutrinos. In addition to what is observed in simulated events, some real data events are assigned huge values of over 3 and minimal values of below 2×10^{-3} (the spectrum is cut at 1×10^{-3}). The large uncertainties stem from events caused by another trigger on afterpulses following a bright event. Afterpulses are more likely to occur when a PMT is exposed to intense light such as that from high-energy atmospheric muons passing close by a DU. In that incident, residual gas inside the PMT can be ionized and travel to the photocathode, releasing photoelectrons, which then, up to several microseconds after the initial light detection, create additional signals [148], p. 440-445. In rare cases, afterpulses persist to trigger a new event. These events are not simulated and therefore appear particularly unphysical to the neural networks, confusing them to output extreme values like vast directional uncertainties. The smallest uncertainties have a similar origin. Here, the afterpulses contribute significantly to the light registered within the same event. For the selection, these regions are already excluded by the existing cuts on the directional uncertainties.

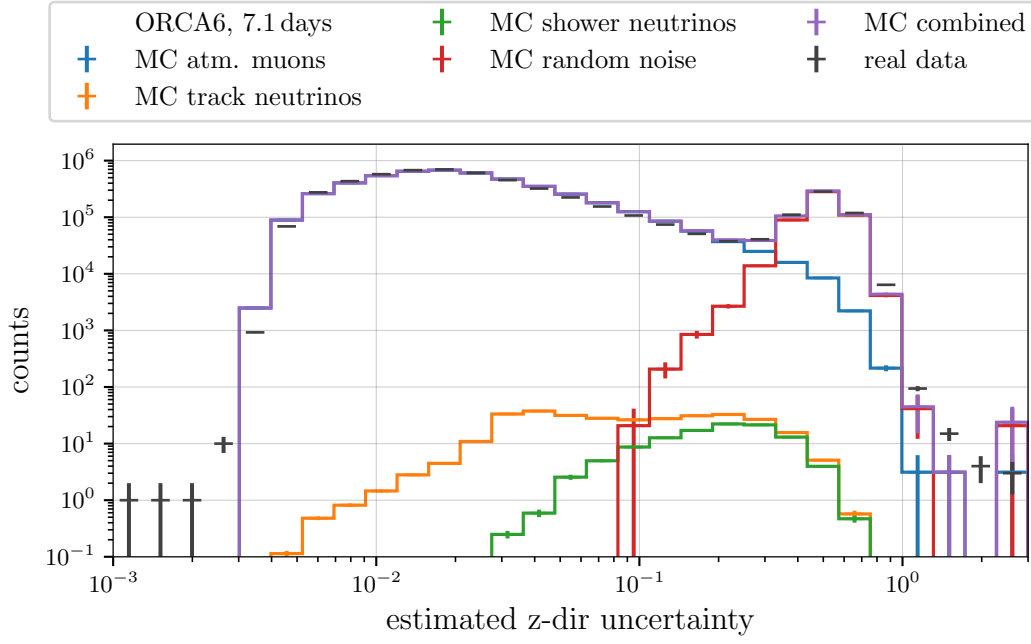Simply cutting on a fixed maximum uncertainty for all events equally

Figure 8.3: Distribution of the z-direction uncertainty for different contributions in MC (colors) and real data (black)

will remove a significant part of the neutrino spectrum when aiming to reduce the influx of poorly reconstructed atm. muons. This is especially true for lower-energetic interactions, which exhibit the strongest oscillations. To this end, the correlation to the reconstructed energy is utilized and a 2d cut is adjusted to select neutrinos and suppress background efficiently.

The focus of this approach lies on the identification of a particularly pure set of track-like neutrinos, meaning also the contributions from shower neutrinos are reduced. The motivation and the efficiency of all cuts are discussed in the following.

Cut on upgoing

In this strategy, there are two main cuts aimed at background reduction. Namely, requesting upgoing events and setting a maximum uncertainty dependent on the reconstructed energy. The first one is chosen to be at $\cos\theta = 0.13$, rather than 0.0, to exclude a region of relatively large data/MC discrepancies in addition to further atm. muon suppression. In Fig. 8.4, the cut is relaxed to explore the region towards horizontal directions. There, the upper part of the plot features the MC predictions in red and its individual contributions from the different topologies in other colors. Random noise is omitted as no entries are left after cuts. The black crosses mark the observed event counts for 357.2 days of ORCA6. In the lower part, the
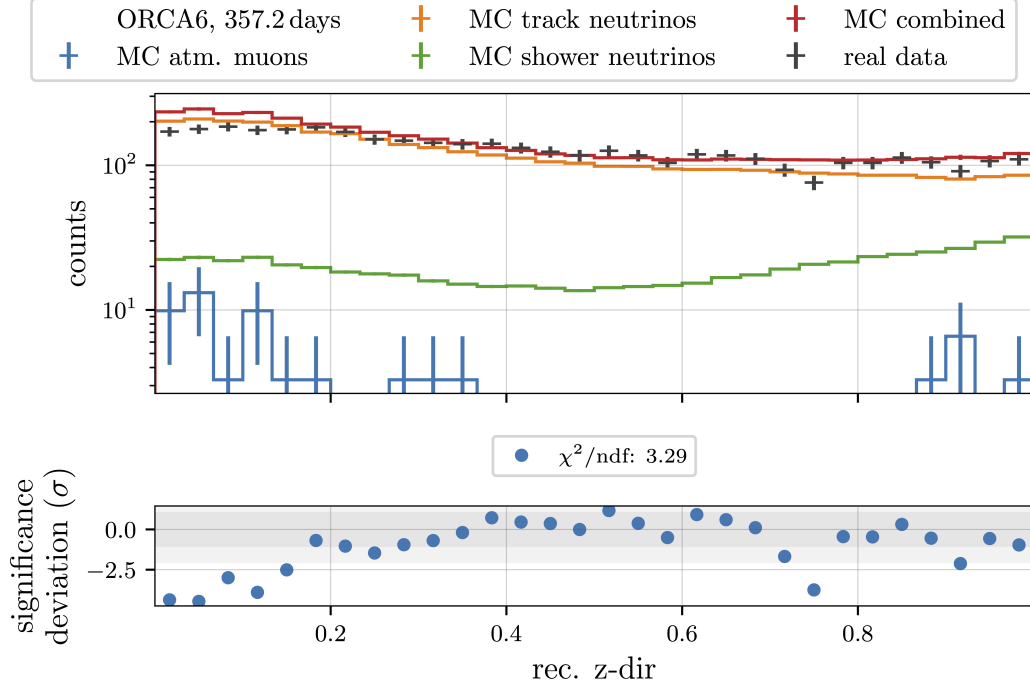
Figure 8.4: Spectrum of the reconstructed $\cos\theta$ for data (black) and MC (red), including different components (other colors). This is after all cuts, but relaxing this cut value to 0.0 rather than 0.13. The lower part shows the significance of the deviation (see Eq. 8.1) and a mean χ^2 per bin. The shaded areas denote 1σ and 2σ, respectively.

significance of the deviation between MC combined (E) and real data (O) is given, defined as

$$\frac{O - E}{\sqrt{O^2_{\text{unc}} + E^2_{\text{unc}}}}. \tag{8.1}$$

Also, the sum of the bin-wise χ^2 comparison (Eq. 7.1) divided by the number of bins is displayed.

A more significant discrepancy is found for the close-to-horizontal events, predicting more events than actually observed considering this selection. Moreover, the influx of misreconstructed atm. muon events is higher towards $\cos\theta = 0$, pushing the eventually chosen cut to 0.13.

Energy dependent cut on direction uncertainty

The second significant group of cuts is based on the estimated directional uncertainty and the reconstructed energy. A correlation between these two quantities is expected from first principles, as with increasing brightness of an event the visible energy and the potential for a accurate direction
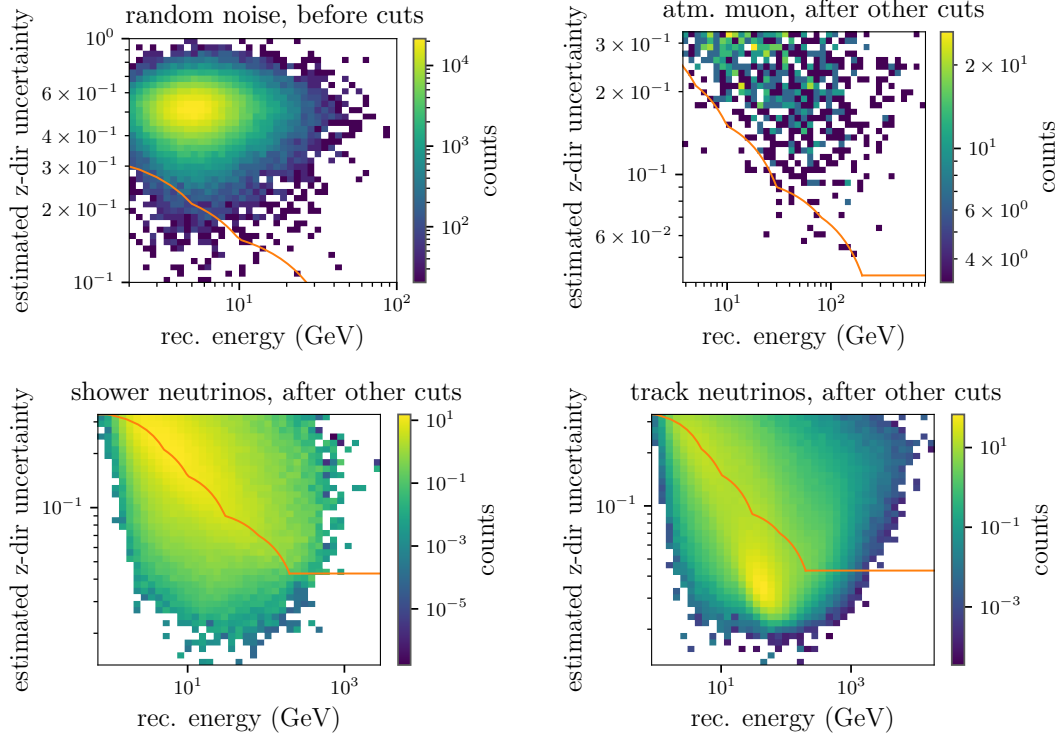
Figure 8.5: Estimated *z*-direction uncertainty versus the reconstructed energy of an event for different topologies. The energy dependent cut applied on the uncertainty is drawn as the orange curve. All events above the line are rejected.

reconstruction also increase. The correlation becomes apparent from Fig. 8.5, which contains the 2d histograms for the four topologies.

For random noise, no such correlation is anticipated and the size of the events is minimal. Thus, the reconstructed energies are predominantly below 10 GeV and the uncertainty is large (>0.4), compared to the particle-induced triggered events. The proposed cut indicated by the orange line (requesting a maximal uncertainty of 0.2 at 5 GeV, for example) is on its own able to exclude the vast majority of random coincidence events. Since the sum of the events passing the other cuts is small, this distribution is shown before applying any other cut.

The criterion applied to atm. muons is an efficient, orthogonal way to suppress those muons that pass all other cuts. As can be seen on the top right of Fig. 8.5, the remaining muons exhibit high uncertainties on the *z*-direction considering their energy. The largest cluster is found around 10-20 GeV and 0.3 uncertainty, which is distinctly excluded by the orange line. Without this cut, the contamination from muons alone would amount to 29%.

The shower neutrinos on the bottom left are also considered background for this selection. As mentioned before, shower neutrinos of a few up to tens of GeV behave similarly to track neutrinos of that energies. Consequently, the two topologies are difficult to distinguish in reconstruction parameters.

Because of that, a trade-off between track neutrino efficiency (for low energies) and shower contamination has to be made. With the chosen uncertainties per energy, the majority of the showers are successfully excluded, while the higher-quality track events are kept (see bottom right of that figure). This means that the main discrimination power for the neutrino topologies in the absence of the dedicated classifiers comes from this 2d cut.

The same strategy is followed for the x and y component of the direction. The uncertainty of the energy does not display a sufficient data/MC agreement to allow for the use as a cut parameter.

Cut on neutrino probability

To keep the influence on atm. muons and neutrinos of the cut on the neutrino probability low, a value of > 0.999 is chosen. The effect on the neutrino efficiency is small, having 3367 expected events with and 3435 without the cut passing the selection ($\sim 2\%$) (see also Fig. 8.1). The muon contamination only rises from 1.0% to 1.4% if the cut is omitted. Nevertheless, the neutrino probability is still able to discriminate efficiently against random noise events. As was discussed in Fig. 8.5 top left, a certain fraction of these events still survives the 2d cut but are completely removed by the classifier. Applying this cut, the contamination due to random coincidences is reduced from 88% to 0%. It should be noted that the number of simulated random noise events is not very high, allowing for some statistical uncertainty on the quoted value. The uncertainty of the zero counts is about four events (considering Poisson error and the scaling factor for these events), meaning the overall maximum contamination from random noise is most likely below 1%.

Cuts on hit quantities and summary

For completeness, two very lose cuts are applied on the number of hits and the number of triggered hits per event. In the same way as for the MC based selections, 50 and 8, respectively, are chosen to suppress background events with minimal impact on the signal neutrinos.

The results from this chapter's selection strategy can now be compared to the selections developed in MC data in Chap. 7.1, being mainly based on the classifiers. Table 8.1 gives an overview of the numbers for the MC events. Concluding from these numbers, the hand-crafted cuts in the data track selection land somewhere between the selection not distinguishing the topologies and the one mainly aiming at a pure track set. The efficiency for tracks is slightly higher in the data case, but significantly more shower events are included. Namely, the shower contributions are 1.3% for the MC pure track and 14.6% for the data pure track selection. At the same

Table 8.1: Characterizing quantities of the selections purely optimized on MC data ("MC") and the selection aimed at maintaining data/MC agreement ("data"), introduced in this chapter.

quantity \ selection	MC all flavors	MC track	data track
track efficiency (%)	23.6	18.4	18.9
shower efficiency (%)	32.6	0.7	9.9
atm. muon cont. (%)	1.1	1.6	1.0
random noise cont. (%)	0	0	0
overall contamination (%)	1.1	2.9	15.6
selected events	5,642	2,965	3,400
selected events $E_\nu < 20\,\mathrm{GeV}$	2994	600	1397

time, the contamination from background events is kept to a minimum in all cases. The main difference between the selections on the right side is the number of lower-energetic events, which is more than double the amount for the data selection. This can be understood from the less strict suppression of shower events that also allows for more track-like neutrinos with lower energies to be kept.

The comparisons of the resulting sensitivities are subject to Sec. 8.2. Next to the fact that the additionally introduced selection allows for the transfer and subsequent analysis on real data, this approach is an interesting test as to whether a slightly larger mixture but significantly higher yield of low-energy tracks can lead to higher sensitivity. Furthermore, the different angles the selections have (once mainly utilizing the classifiers, once mainly the reconstruction networks) showcase the versatility and proficiency for the application to neutrino selection and additionally allow for cross-checks.

8.1.2 Correcting for larger data/MC discrepancies

Before the selected events from real data and MC are compared in spectra, two measures are discussed that are developed to fix some of the currently present systematical discrepancies between simulation and reality that the studies with the graph neural networks revealed. Namely, the difference in the number of hits per event and a shift in the reconstructed energy. Most probably, both symptoms have a common origin. Several candidates could explain the observations that are currently under investigation. After detailing the features of the discrepancies and their countermeasures, the potential causes are briefly discussed.
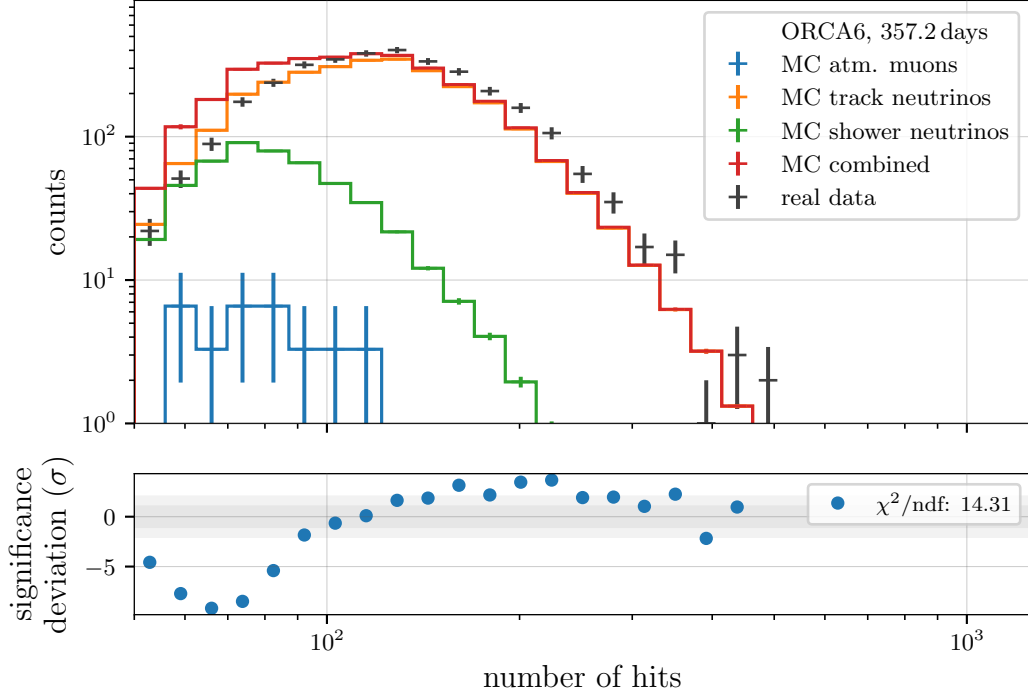
Figure 8.6: Distributions of the number of hits per event for data and MC. Same structure as Fig. 8.4.

Discrepancy in the number of hits

The initial spectrum of the number of hits after all cuts is shown in Fig. 8.6. It can be observed that the spectrum of the combined MC and the real data seem to be shifted, with generally more hits in the actual events than in the simulations. The significance decreases only with lower statistics towards 0 for the upper and lower end of the spectrum.

Such an observation can be caused, in principle, in two different ways: Either the number of hits is mismodeled in simulations, or this particular selection is more sensitive to low-count events in MC than in real data and simultaneously less efficient for higher hit counts in MC than in data. The latter is disfavored by the fact that also other selections exhibit the same feature, although with less statistical significance. Namely, the introduced classical selection as well as another one based on the same original input information from the classical reconstruction, but relying on Boosting Decision Trees [149], a type of shallow learning, for identifying the neutrino set. Also, the shift persists when probing different cut combinations of the GNN outputs, i.e., creating differently composed neutrino-dominated sets.

If this is indeed a characteristic of the current version of the simulated data, a re-weighting based on this observable can be used to still perform analyses, as the data/MC agreement is restored or at least significantly improved. The procedure for the re-weighting can be followed in Fig. 8.7
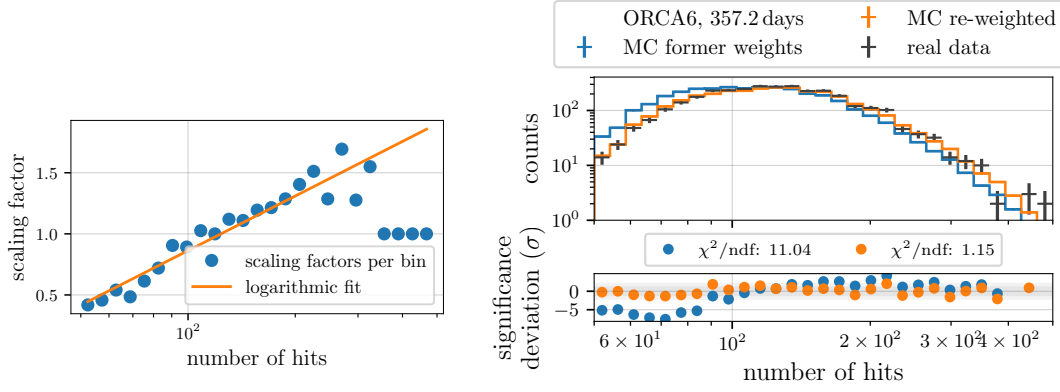
Figure 8.7: Left: Scaling factor (blue) for differently sized events, defined as the ratio of data and MC counts from Fig. 8.6. A logarithmic fit is added (orange). Right: Spectra before (blue) and after (orange) applying the correction extracted from the fit on the left side to the weights of the MC events. The lower part of the plot probes the agreement with real data for the different weights like in Fig. 8.4.

and consists of first extracting the ratio of counts observed in the binned histogram of Fig. 8.6. As discussed before, this discrepancy follows closely a logarithmic dependence (visible as linear shift in the spectrum using log scale) and is thus fitted with a function that allows mapping the former MC weights w to the updated ones, namely

$$w_{\text{re-weight}} = w \cdot 1.484 \log_{10}(\text{n}_{\text{hits}}) - 2.105, \qquad (8.2)$$

based on the two constants extracted from Fig. 8.7, left, and the number of hits of the event. Bins with low counts (6 in either data or MC) are excluded from the evaluation. The right side of Fig. 8.7 shows the number of hits spectrum (short: *hits spectrum*) considering the former and updated weights, which, by design, now exhibit better modeling of the real data.

An indication of whether this adjustment of weights reliably counteracts the underlying mismodeling comes from the fact that most of the quantities probed in data/MC comparisons do improve. One example is shown in Fig. 8.8, where the comparison for the uncertainty of the *y*-direction displays higher agreement with a reduced χ^2 of 2.1 instead of the 4.8 from before. For two different regions, the data are better modeled by MC now: low uncertainties around 0.03 are generated by higher-energetic tracks that can be reconstructed particularly well. Their contributions have been increased by the re-weighting, as they are characterized by high numbers of hits (see the modification in Fig. 8.7, right). The weight of events with larger uncertainties between 0.1 and 0.4 (lower number of hits) has been decreased, which now fits the observed distribution better. In this region, the contribution form shower events is large (see Fig. 8.6).

In summary, the agreement for the direction reconstruction networks is improved by the re-weighting, including the discrepancy for the *z*-direction
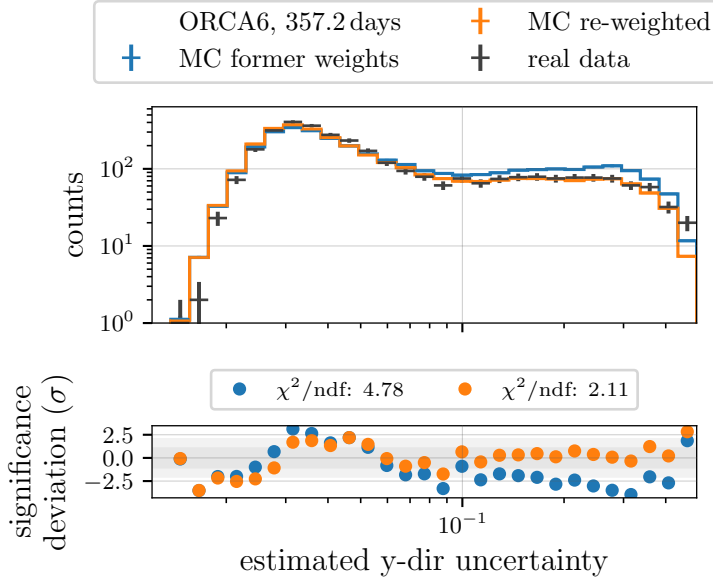
Figure 8.8: Re-weighting effect on the data/MC agreement for the estimated y-direction uncertainty. Same details as Fig. 8.7, right.

of horizontal events. The reduced χ^2 is lowered from 1.4 to 1.0. For the other directions, similar effects are observed; their χ^2 values are lowered by a factor of 2.

The energy reconstruction network is discussed separately in the next section. As expected from changes to the number of hits, the agreement for the number of triggered hits is also improved from 5.7 to 2.3 χ^2/ndf. The most notable change takes place at low numbers in the spectrum.

The signal/background classifier does, however, not show an overall improvement (Fig. 8.9). Most notably, the MC counts in the highest bin are slightly lower, getting closer to the observed quantity. Nonetheless, the effect is counteracted by the remaining bins so that quantitatively the agreement remains poor regarding the χ^2/ndf. The improvement in track probability is significant, with the reduced χ^2 improving from 44 to 29. The spectrum is discussed again in the following section for data/MC comparison, but the overall disagreement remains significant.

Another interesting change can be observed for the energy spectrum from the JEnergy stage of JGandalf, still considering the GNN selection. The distribution shown in Fig. 8.10 displays an significantly improved data/MC agreement with the χ^2/ndf being reduced by more than a factor of 3. The same discussion about lower-energetic showers and higher-energetic tracks and the change of their respective weights applies here as well.

Lastly, the potential influence of this re-weighting on oscillation analysis is evaluated. After all, the probing of different sets of oscillation parameters relies solely on generating different weights modified by the oscillation probabilities for the MC data. Any additional changes to the weights might interfere with those studies. To that end, two cases are studied in which the MC is re-scaled based on the hits spectrum, once assuming NuFIT values
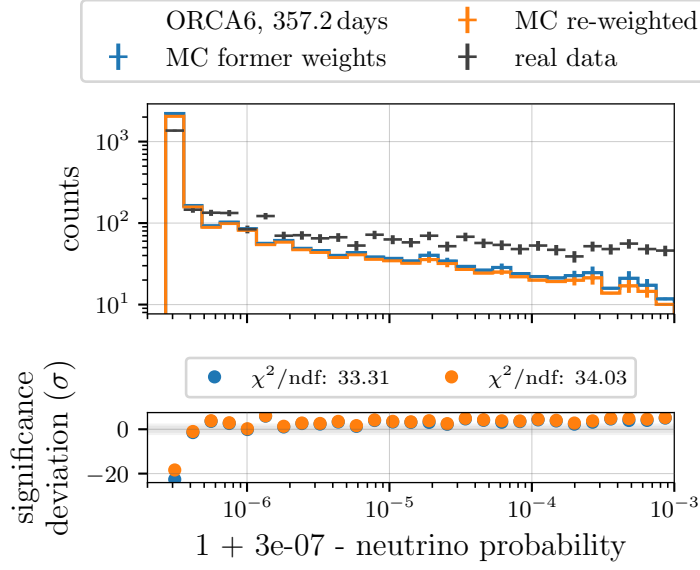
Figure 8.9: Re-weighting effect on the data/MC agreement for the neutrino probability. Same details as Fig. 8.7, right.
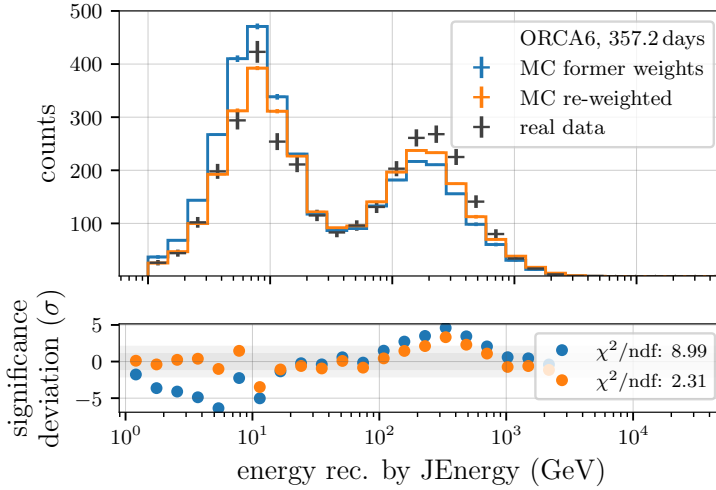

Figure 8.10: Energy of the GNN selected events as reconstructed by JEnergy. With (orange) and without (blue) re-weighting based on the number of hits. Same details as Fig. 8.7, right.

(as before) and once considering the previously mentioned "best fit" from the first ORCA analysis, namely $\sin^2\theta_{23} = 0.5$ and $\Delta m_{31}^2 = 1.95 \times 10^{-3}\text{eV}^2$. The strength of extracting the re-weighting from a low-level observable, such as the number of hits, is that the oscillations have minimal influence on this distribution. Thus, the impact of weighting events to a specific model should not compromise the analysis. This is reflected in the fact that the same statement from a high-level observable, such as the energy, which is the single quantity most sensitive to oscillations, is obtained for scaling to different models, as can be seen in Fig. 8.11. In it, the spectrum of reconstructed energies from real data is compared to the two different oscillation scenarios, NuFIT and best fit, each time starting from a different scaling inferred from the hits spectrum. A shift of the energy, that will be discussed in the next section, is already applied for better comparability. Also in this observable, the cases of not applying the scaling (blue and orange) are
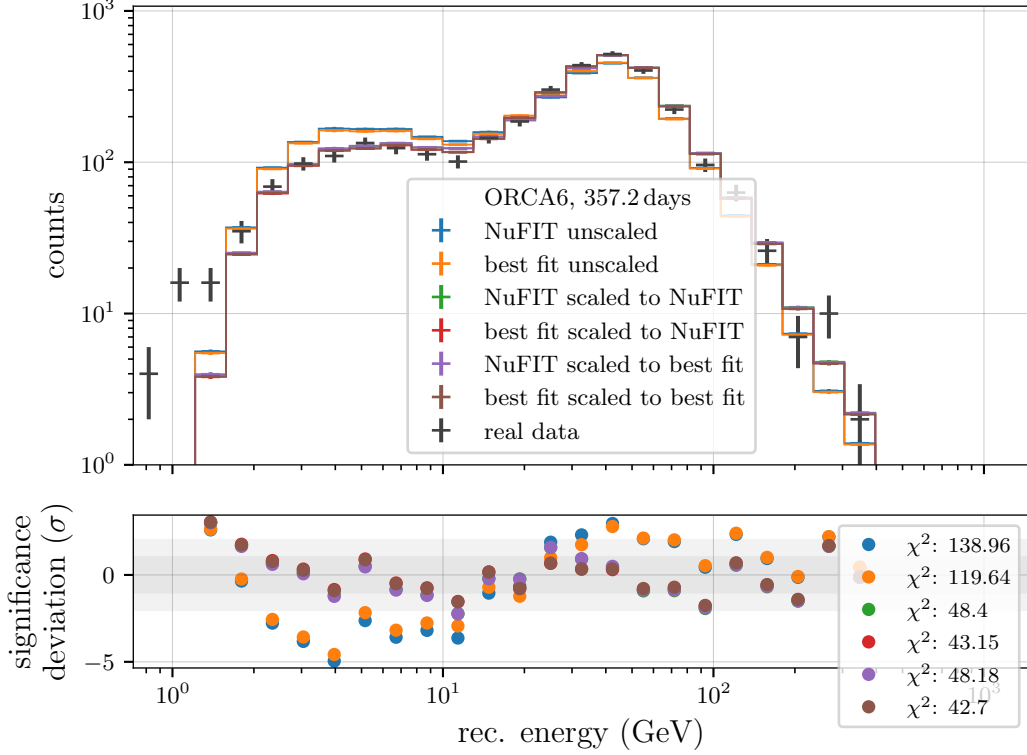
Figure 8.11: Spectrum of the GNN reconstructed energy for real data (black) and MC (colors) assuming different oscillation scenarios and re-weightings based on the hits spectrum. The lower part evaluates the fit of each scenario to the real data (like in Fig. 8.4) and the quoted χ^2 is here the sum from all bins.

strongly disfavored by about a factor of three larger disagreement between expected and observed data. However, scaling to the number of hits based on the NuFIT parameters (green and red) yields very similar χ^2 values as for considering the best fit values in the hits spectrum (purple and brown). Most importantly, the difference between the two oscillation parameter sets stays nearly the same, with a $\Delta\chi^2$ of 5.25 and 5.48, respectively.

Energy shift in real data

Evaluating closer the energy spectrum for re-weighting the events as illustrated, reveals a better fit between real data and MC than before, Fig. 8.12, but the two distributions appear shifted with respect to each other. In particular, the energies reconstructed in real data are higher on average, while the shape of the distribution does not change. Most probably, a common origin in the simulations causes both the number of hits to be smaller and the energy estimation to be lower. Since the re-weighting of events does not fully solve the discrepancies in the energy spectrum but still improves the data/MC agreement, there are most likely two effects involved. On the one hand, a direct impact from simulations on the energy
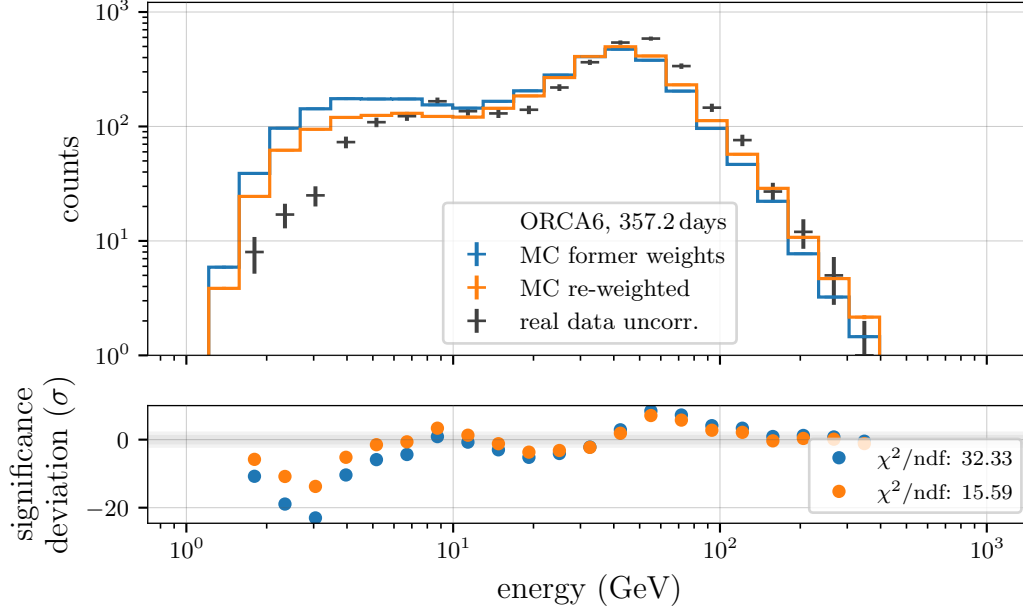
Figure 8.12: Reconstructed energies for combined MC with (orange) and without (blue) re-weighting based on the hits spectrum and real data (black). Evaluation of the similarity in the bottom part like Fig. 8.4.

reconstruction manifested in the remaining shift. And on the other hand, an indirect component via the number of hits enhancing the direct effect, as the number of detected photons naturally influences the predicted energy. The different options that could lead to such discrepancies are currently under investigation and are reviewed at the end of this discussion.

Since determining the cause and fixing and fine tuning the simulations is a delicate, prolonged process, a temporary solution is to shift the energies of the real data spectrum to match the MC distribution. A linear mapping in the log space allowing for a scaling a and shift b is a natural choice, as the quantity the GNNs receive as input is the decadic logarithm of the energy. The corrected energy E_{corr} is then

$$E_{\mathrm{corr}} = 10^{a \log_{10}(E)+b} \tag{8.3}$$

starting from the initially predicted energy E. By minimizing the χ^2 between data and MC, the parameters are determined to be $a = 1.17$ and $b = -0.365$. This correction is visualized in Fig. 8.13, where comparing with the diagonal of no modification (orange) reveals that these numbers correspond to a shift to smaller corrected energies up to around $100\,\mathrm{GeV}$ and to larger energies thereafter.

This modification to the neutrino selection is applied in Fig. 8.14, which contains the corrected and uncorrected real data counts along with the MC predictions. The agreement is improved by a factor of 10 (χ^2/ndf from 18.56 to 1.83) with only minor inconsistencies for unrealistically low energies
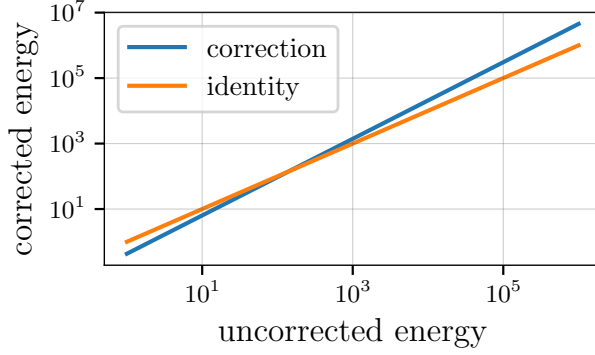
Figure 8.13: Mapping of the uncorrected to the corrected energy following the formula derived from minimizing the data/MC agreement (blue). As a reference, the identity for no change in energy is given (orange).
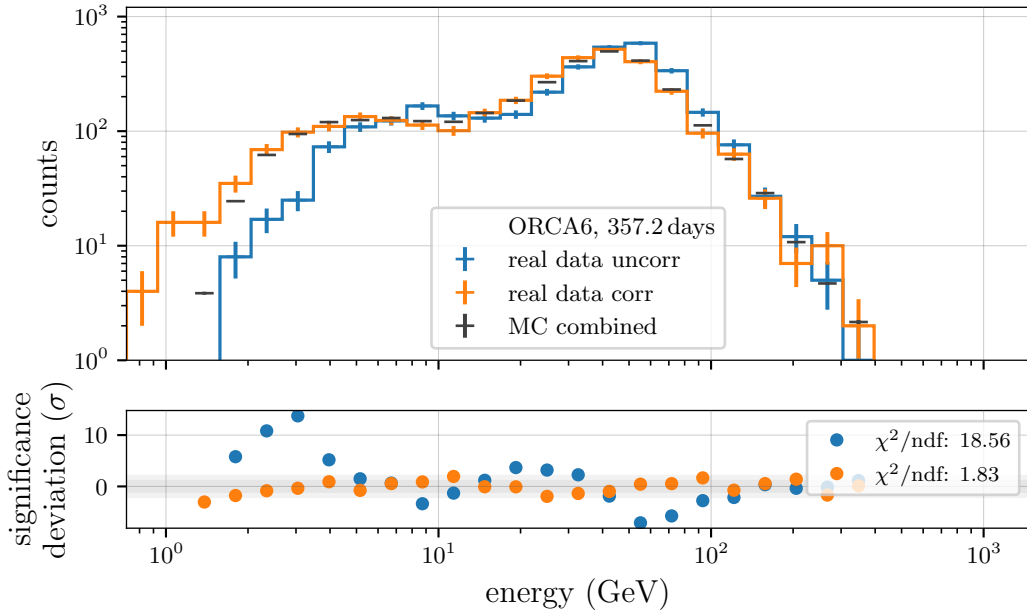

Figure 8.14: Spectrum of the reconstructed energy showing the influence of the energy shift (blue to orange), comparing both to the distribution expected from MC (black crosses). Evaluation of the similarity in the bottom part like Fig. 8.4.

around 1 GeV remaining, where a few real data events end up, indicating the simple nature of the correction may not be applicable for that region.

Typically, such modifications stemming from suspected inadequacy in the simulations are applied to the MC predictions rather than the real data. In this case of training a neural network on the simulation, however, all the physical information and connections between detector response and the true energy are learned from the MC. It is here where the predictions make correct estimations for the energy, as was shown, for example, in figures 6.20 and 6.21. Because the events are distributed around the diagonal without larger indications for a systematic shift, the energy reconstruction for MC data has to be kept fixed and the real data spectrum has to be adjusted to it, rather than the other way around. This assumption requires the initial
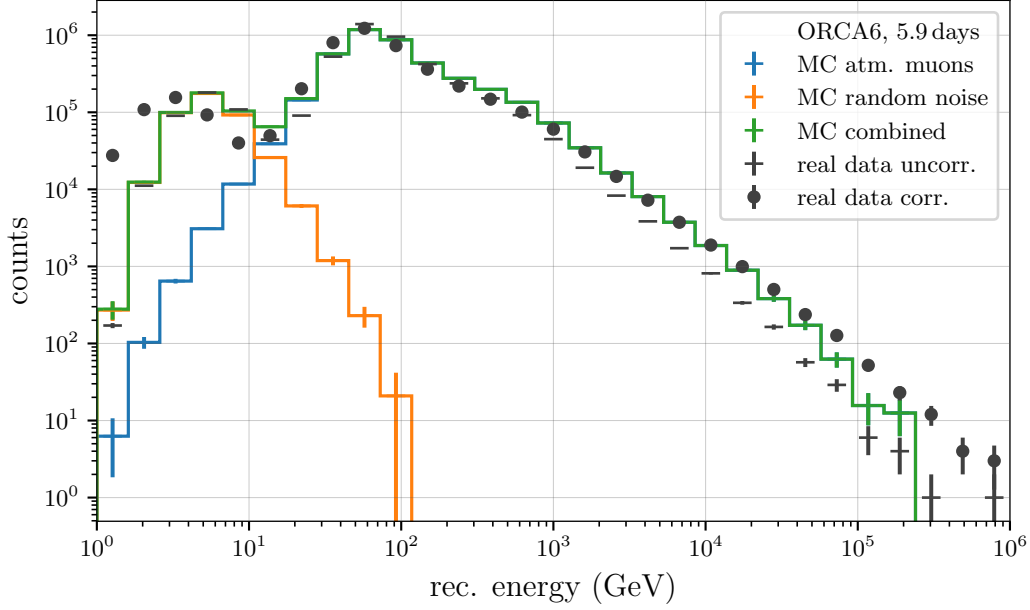
Figure 8.15: Reconstructed energy before cuts, which is dominated by random noise (orange) and atm. muon events (blue). Included are the MC combined prediction (green) and the uncorrected (black crosses) and corrected real data (black dots).

spectra to have very similar spectra to begin with (neutrino flux), which is given, as the distributions align well in Fig. 8.14.

An interesting test of the established energy shift would be to apply it to atm. muon data as well. This is straightforward to do, as the quantities before cuts are background-dominated. However, it should be noted that the generation and simulation software in this case is completely different from the neutrino simulation chain. Yet, an improvement of the data/MC agreement also for muons would point to a common origin. The comparison between with and without the shift is drawn in Fig. 8.15. From 30 GeV reconstructed energy onward, the atmospheric muons constitute the signal and exhibit a better description by the MC when considering the modified energy in real data. The distinction is pronounced for energies larger than 1000 GeV. Between 10 GeV and 100 GeV, where the energies are corrected towards smaller values, the preference is more ambiguous, as effects from the random noise contribution interfere.

This consideration is in line with the observation that the number of hits for atm. muons is overestimated in simulations, leading to larger reconstructed energies (Fig. 8.16). Even though the modification to the energy is determined only based on the neutrino data, which contain only a few entries of above 100 GeV, the behavior for the highest energy muons is still reasonably well modeled by the correction.

The factors specific to atm. muon simulations that affect the number of hits and energy distribution are the multiplicity (number of muons from one
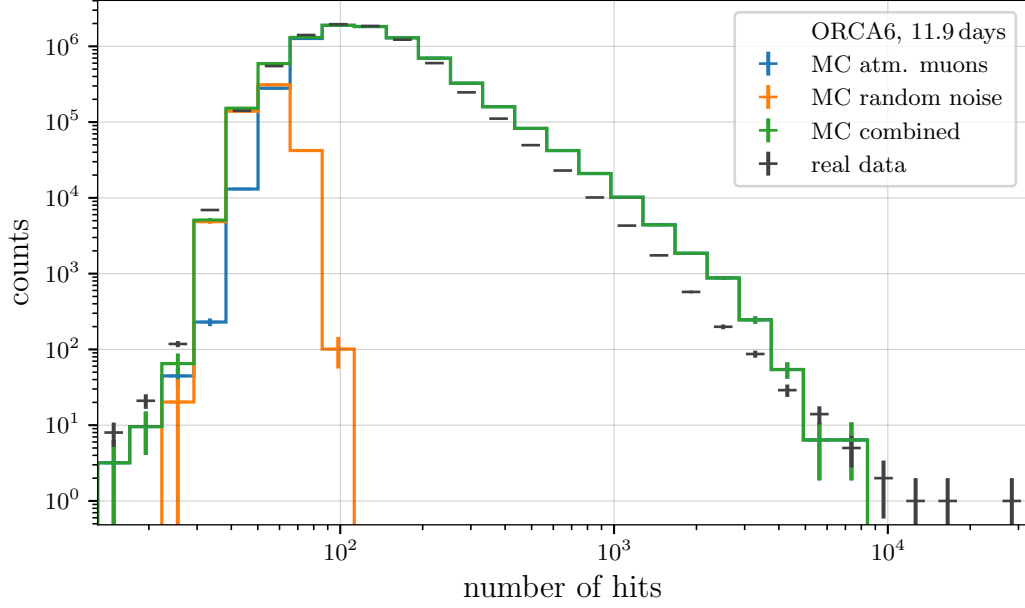
Figure 8.16: Number of hits per event before cuts in the background dominated regime. The combined MC predictions (green) can be compared to the real data (black crosses).

cosmic ray interaction) and the flux. Both are currently under investigation, especially considering the flux is parameterized and can be tuned to data.

Completing this discussion is a consideration about the random noise events that describe well the uncorrected data (Fig. 8.15). In other words, the response of the neural networks is the same to both kinds of randomly generated data. This can be understood by the simplicity of the simulation in that case; hits are drawn randomly from distributions that are extracted from the real data acquisition (see Chap. 2.4). No additional physics need to be simulated. This further supports the statement that there is an underlying mismodeling in the simulation of particle-induced events.

Possible explanations for data/MC discrepancies

First of all, it should be stated that fixing the discussed data/MC discrepancies is currently one of the topics with the highest priority within the KM3NeT collaboration. Also, as initially mentioned, it is only natural to discover such inconsistencies when probing neutrino simulations with higher statistics for the first time; something the application of the GNNs developed in this thesis enabled. Now, several hypotheses can be tested by producing updated MC data and repeating the tests and comparisons.

One candidate are the water properties assumed in the simulations. There were several measurements conducted at the detector sites, and the ANTARES collaboration published their studies inferring parameters from

the light transmission [150]. Since the absorption length is dependent on the wavelength of the light, temperature, density (depth), and salinity of the water, the best estimate for the currently used value has an uncertainty of about 10% [151]. The scattering length may be similarly affected in magnitude but ultimately has less impact due to the low probability of light scattering over the relevant distances. Furthermore, the ANTARES paper reported a time variability of 15% over a three-year period. Some studies in the scope of this PhD work were conducted on the low-level observables, such as the incident angle of the photon on the PMT and hit time residual distributions for light exhibiting different travel lengths. However, no clear evidence of a preference for certain water properties could be identified.

Another group of causes are characterized by uncertainties at the generation level. This includes the neutrino flux, as indicated in Chap. 1.2, as well as atm. muons. After all, both stem from cosmic ray interactions (Eq. 1.1) where large uncertainties in the hadronic processes are involved. This results in uncertainties for energy and direction dependencies, the overall flux normalization and composition. Below 10 GeV, especially the electron (anti) neutrinos are affected [25]. Similarly, the final states for hadronic processes, as present for the shower component when the neutrino scatters off the nucleon, are challenging to describe in simulations and thus strongly model-dependent [152].

Other candidates are the simulations at the particle and light propagation level. In KM3NeT, different software is used for different particles and energy ranges. Various cross-checks have, so far, not exposed any major differences that would indicate the observed mismodeling.

At the end of the detection chain is the DOM with its PMTs. Any features in its properties not or wrongly accounted for in the simulation of its response function can equally cause data/MC discrepancies. The overall detection efficiency calibration, which is influenced by the gain settings of the PMTs and any covering on the glass spheres from sedimentation or biofouling, could be compromised. Here, several minor effects are known to be only approximations, like the handling of instances of strong sedimentation and the fact that the shadowing provided by the mounting structure is different for the downgoing muons (the efficiency calibration is based on) than for neutrinos.

One outcome of the aforementioned analysis of the low-level hit features was a slightly different angular acceptance of the PMT between MC and real data. Following these findings, simulations were carried out and a new PMT model was established. Moreover, PMT properties like the transit time, transit time spread and quantum efficiency all depend on the position of the incident hit, exhibiting differences of a few ns for the transit time and significantly larger spread at the edges of the PMT [153]. This is currently not considered in the response function.

Yet another assumption with indications for the data/MC agreement,

especially for the lower-level features, is the exact shape of the transit time distribution. For the one assumed in the simulations, different distributions measured in the lab are averaged over, allowing for an effect of a few percent variations. Finally, additional hits are expected from afterpulses, to which the neural networks may be sensitive, as they are not simulated. In addition to those exhibiting long delays (mentioned in Sec. 8.1.1), afterpulses with short delays (several up to tens of nanoseconds, created by electrons elastically scattering on the first dynode [148]) may also contribute to the recorded signals and thus cause higher numbers of hits in real data.

Moreover, a non-optimal calibration would lead to hit information that displays slightly different features than those encountered during training. This could either be an overall difference in calibration quantities, like the positions or timings of the DOMs or be an influence of the moving detector in reality, as these changes are not considered for the simulations. Studies regarding this point are going to be presented in Chap. 9.

Needless to say that it is well possible that the final solution involves a combination of the contributions listed and is not restricted to them.

8.1.3 Data/MC comparisons

Considering the re-weighting, the data/MC agreement of particularly interesting and relevant spectra is discussed in the following.

Starting with the direction reconstruction network, the estimated z-direction is shown in Fig. 8.17 and displays an overall excellent agreement with a χ^2/ndf of 1.05. No particular region of the spectrum exhibits more extreme outliers or systematic discrepancies. Until the cut at 0.13, the combined MC also describes horizontal events well. The largest fraction of shower contributions is found for the upgoing direction, while most of the sparse atm. muon events are reconstructed close to horizontal, as expected from misreconstructed atm. muon tracks.

The direction components describing the azimuth angle display a similar agreement, with a reduced χ^2 of 1.51 for the combined angle. In both zenith and azimuth, the distribution of the significance of the deviation appears normally distributed, indicating no systematic preference for a more or less precise agreement in any specific region of the reconstructed quantity.

As an example for the direction uncertainty distributions, the x-direction component is shown in Fig. 8.18. In general, it displays a good data/MC agreement with $2.03\,\chi^2$/ndf. The structures in the deviation plot below indicate separate regions of slight systematic shifts. For the muon neutrino-dominated uncertainties from 0.04 to 0.07, for example, more events are observed in data with the trend of a decreasing ratio of data to MC events continuing to 0.09. Still, the deviations with the highest significance do not exceed $2.5\,\sigma$. As established in earlier discussions, the shower events
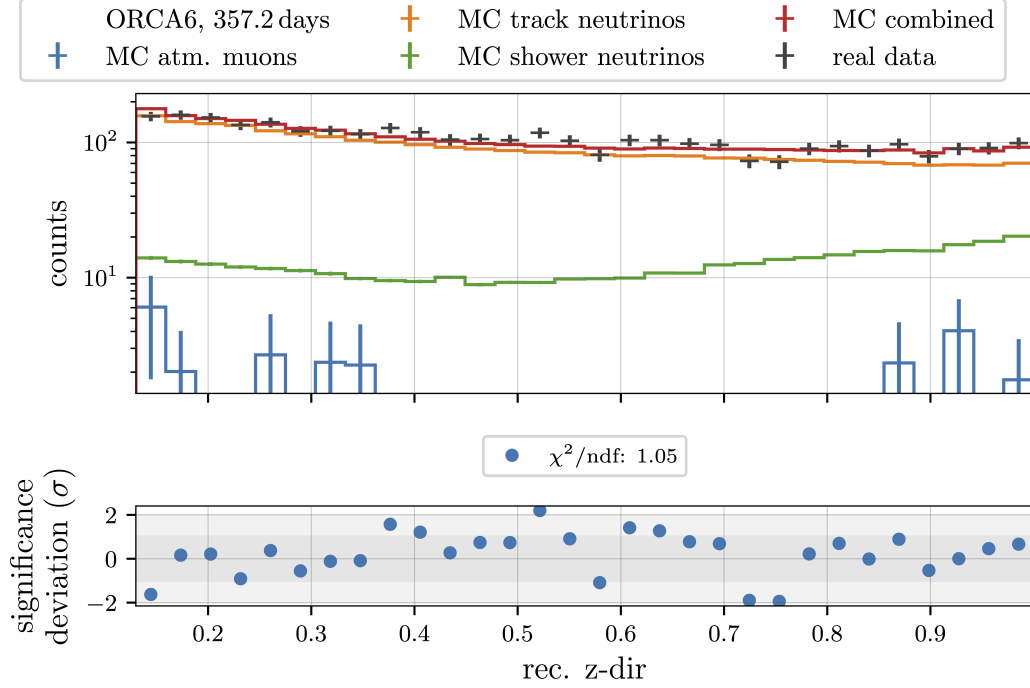
Figure 8.17: Spectrum of the $\cos\theta$ after cuts and corrections for data (black) and MC (colors). Evaluation of the similarity in the bottom part like Fig. 8.4.

naturally exhibit higher uncertainties, which are equally well modeled, while the data/MC ratio rises steadily for the highest entries from 0.23 to 0.5.

Similar observations can be made for the other two direction components, which score a reduced χ^2 of 2.03 in y- and 2.80 in z-direction.

The energy is already shown in Fig. 8.14, with the real data energy in orange and the combined MC in black crosses. The agreement is found to be rather good after correcting the real data's reconstructed values and the re-weighting in MC. The largest discrepancies remain for the lowest energies with about $3\,\sigma$. This may be an artifact of a non-optimal correction extrapolated to that region. The overall χ^2/ndf equals 2.24 when considering MC as the expected and real data as the observed quantity again.

The $\cos\theta$ and the energy are the primary quantities for the oscillation analysis, which is why it is essential to ensure accurate data/MC agreement to allow for the precise testing of different oscillation scenarios. With the quoted χ^2 values, this prerequisite is sufficiently fulfilled.

The uncertainty estimation of the energy is not used in the real data analysis, as it too exhibits a shift between real data and combined MC, visible in Fig. 8.19. The generally higher uncertainty factors correspond to the higher energies initially predicted by the GNN in data. A correction based on the modification by the energy shift is conceivable, but no immediate transfer was successful.

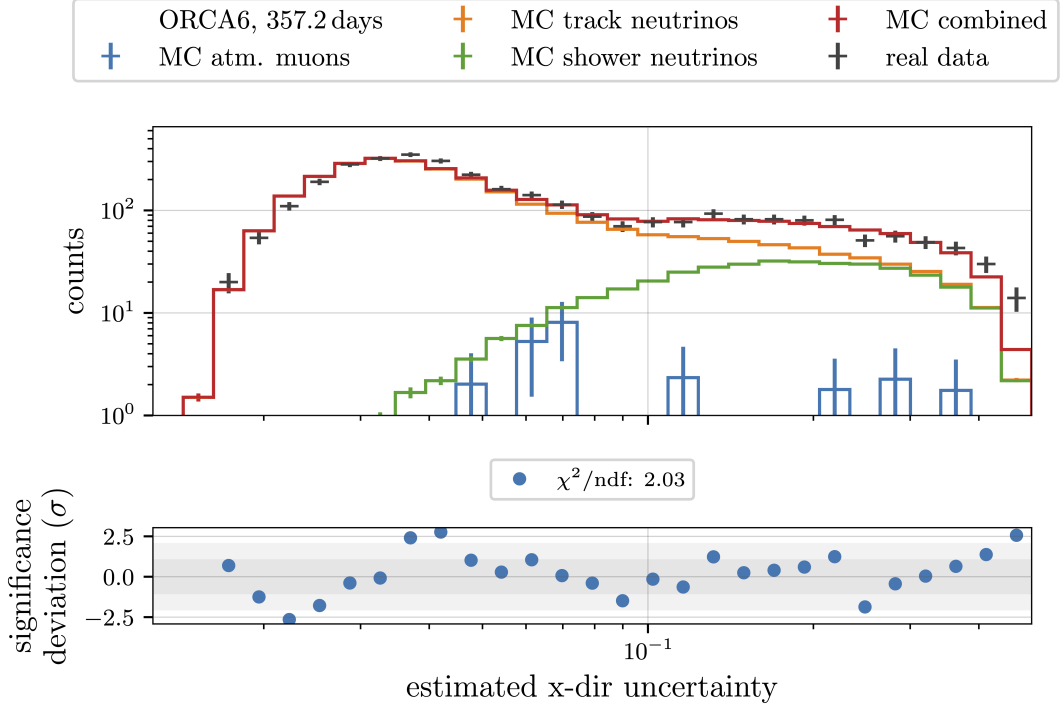The discrepancies in the classifier spectra have been alluded to already.

Figure 8.18: Spectrum of the *x*-direction uncertainty after cuts and corrections for data (black) and MC (colors). Evaluation of the similarity in the bottom part like Fig. 8.4.

The spectrum for the neutrino probability was reviewed in Fig. 8.8, right, pointing out a high statistical significance of the difference in the shape of the distributions between simulated and real data. This is especially accumulated in the bin at $p = 1$. The track probability for the neutrino sample can be viewed in Fig. 8.20. It confirms the statements about the distributions before cuts made earlier, that the spectrum appears to be shifted towards higher track scores in real data, leaving almost no events identified as showers. With this, the track probability is the quantity most sensitive to the differences between simulations and reality. The influence of the calibration as one possible candidate for the origins for these discrepancies will be discussed in Chap. 9 and indeed delivers a possible explanation.

Finally, there is the number of triggered hits per event in Fig. 8.21. It, again, displays a decent data/MC agreement ($\chi^2/\mathrm{ndf} = 2.27$) with a region at lower values up to 11 that shows fewer hits in data, followed by a middle region with predominantly more hits than predicted by the MC up to 60 triggered hits.
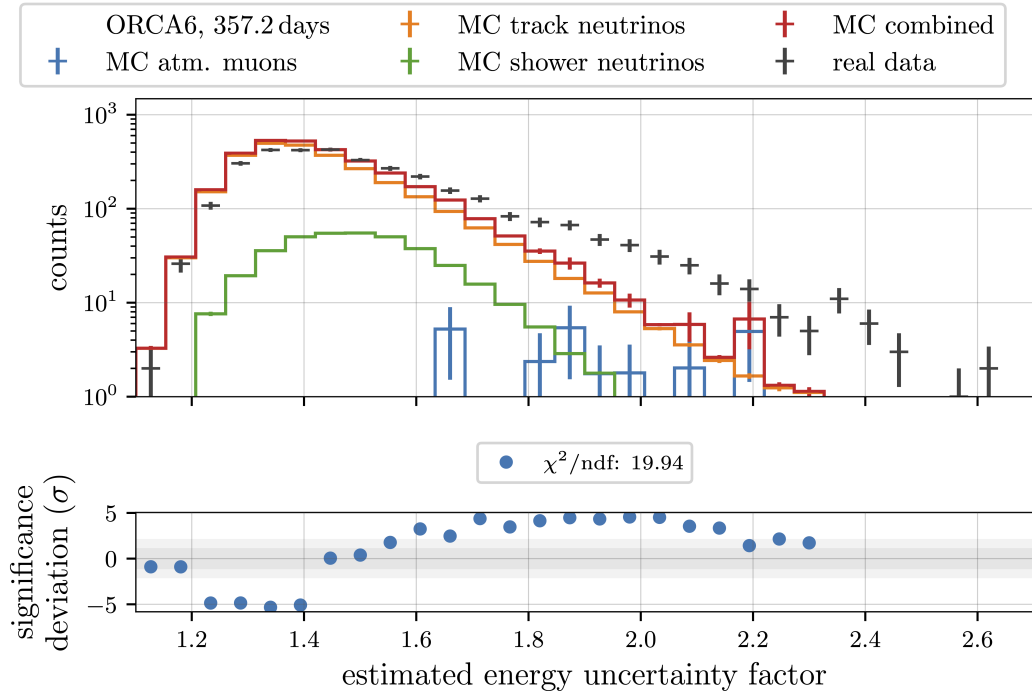
Figure 8.19: Spectrum of the energy uncertainty factor after cuts and corrections for data (black) and MC (colors). Evaluation of the similarity in the bottom part like Fig. 8.4.
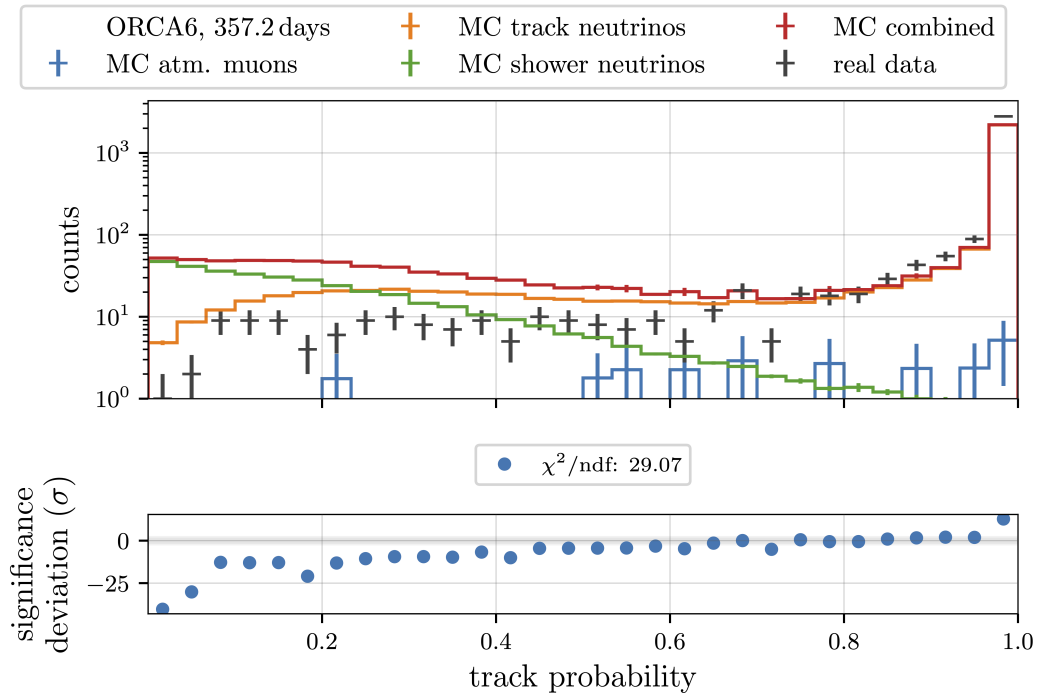

Figure 8.20: Spectrum of the track scores after cuts and corrections for data (black) and MC (colors). Evaluation of the similarity in the bottom part like Fig. 8.4.
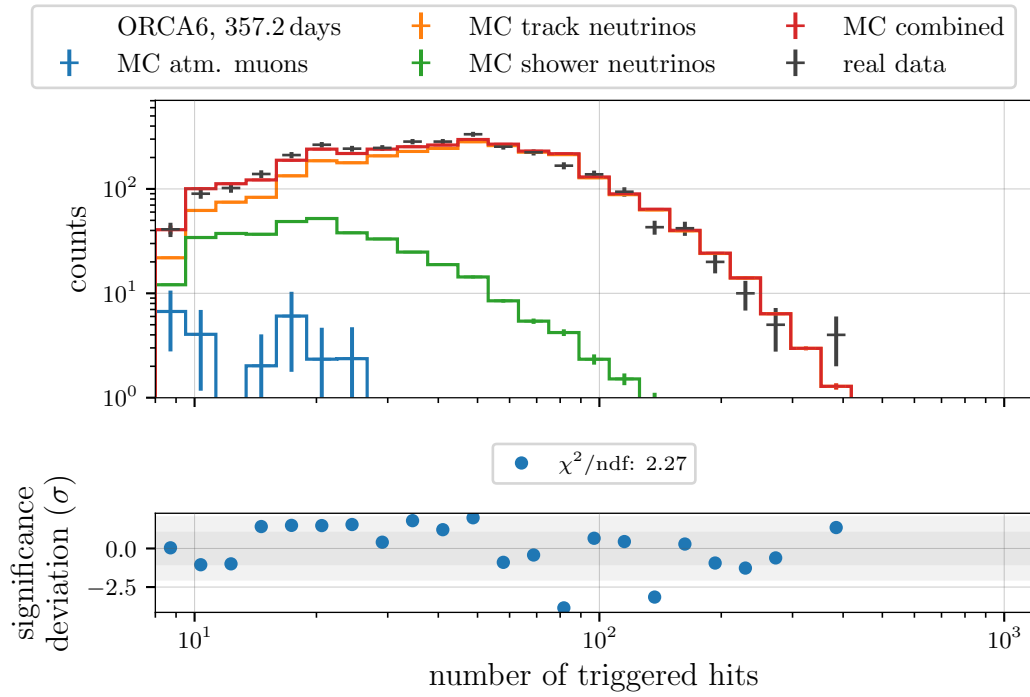
Figure 8.21: Spectrum of the number of triggered hits per events after cuts and corrections for data (black) and MC (colors). Evaluation of the similarity in the bottom part like Fig. 8.4.
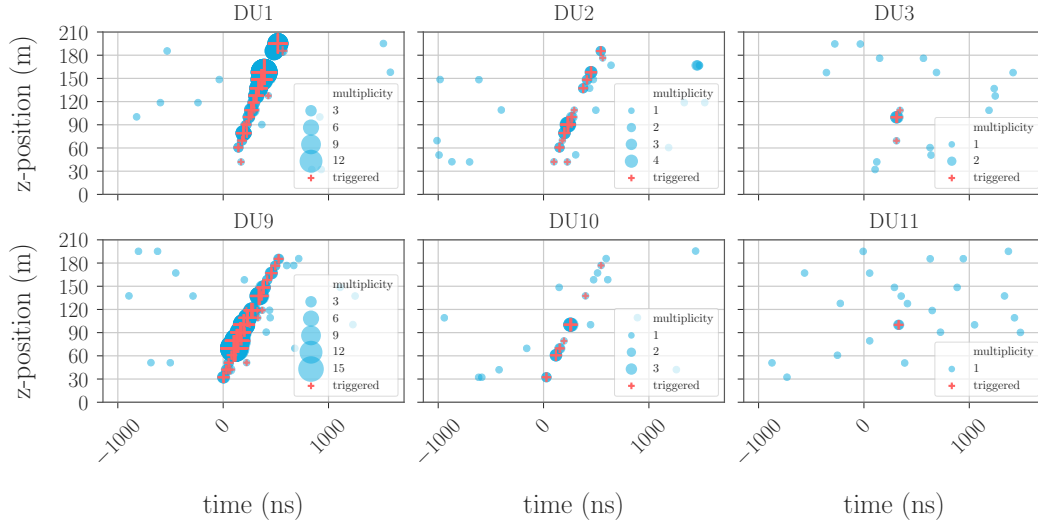
Figure 8.22: Projection of the hit's *z*-position and time for a real data event. Event information and the outputs of the graph neural networks are included in the top part. See the main text for further explanations.

Displaying selected events

After comparing reconstructed values in spectra and examining their compositions, looking at event displays helps confirm that the selected data exhibit the hit features they are expected to have as neutrinos, given the selection cuts. To that end, the available software to produce *z-t* projections, also used in the online monitoring, has been slightly modified. The following plots display the height in the detector and the time of hits per DOM for each DU as a circle. The size of the circle denotes the number of detected photons per optical module ("multiplicity") and a red cross is added if the hits caused a trigger algorithm to fire. The *z*-direction of the detected particle can be roughly read off by considering the *z*-position of earlier and later hits. For an upgoing track, the later hits are expected at higher positions than the first and vice versa for downgoing. Information about the event and the outputs of the GNNs are added to the top. In the following, a few examples are showcased.

First off, Fig. 8.22 displays one of the higher-energetic (106 GeV) neutrino candidates in real data that is reconstructed as almost perfectly upgoing ($\cos\theta = 0.98$). This causes many hits on the DUs closest to the particle's track, which are DU1 and 9. The circles with the largest size denote the point of closest approach to the DU, which happens first for DU9 in the lower part and later in DU1 at the upper part. DU2 and 10 are further away and only receive a portion of the light, while about 40 m away, only a faint, yet triggered signal is seen on DU3 and 11 (see footprint in Fig. 2.2). The classifiers assign high neutrino and track scores.

ORCA6, run id 7397, event id 138, hits (trig.) 87 (16), neutrino prob 1.0, track prob 0.1719, rec. energy $(4.26^{+2.13}_{-1.42})$ GeV, rec. x-dir -0.55±0.29, rec. y-dir 0.76±0.17, rec. z-dir 0.34±0.2 - MC info: ν_e (CC), true z-dir 0.34, true energy: 5.89 GeV
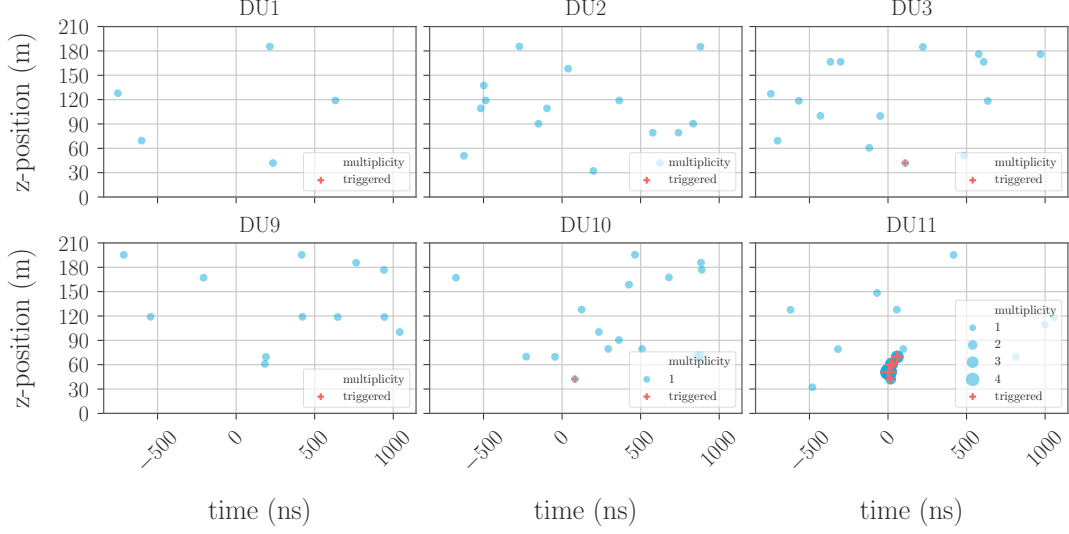

Figure 8.23: *Z-t* projection of hits from a simulated event induced by an electron neutrino. Some true information is also added to the top.

The resulting hit pattern from the simulation of an electron charged current interaction is shown in Fig. 8.23. The energy and $\cos\theta$ reconstruction of this example are rather accurate even though the uncertainty on the direction is considerable for this kind of events. The track/shower classifier correctly predicts a more shower-like topology. The hits from the particle interaction are almost exclusively recorded on DU11, with single photons being triggered on DU3 and 10. This is the expected hit distribution for electrons due to their short mean free path.

Figure 8.24 contains the signature of a simulated muon neutrino of 7.5 GeV true energy. Its *z*-direction and energy are well reconstructed, but the low brightness of the event causes hits mostly on a single DU, with others only detecting a few photons. This topology looks similar to that of the shower before and thus causes the classifier to assign a low track score of 0.22. This is the visualization of the phenomenon referred to several times, that the lower-energetic track neutrinos look very much shower-like.

Evaluating the few events that are described by low track scores in real data, with one example displayed in Fig. 8.25, no obvious pattern can be identified from these hit distributions that distinguishes these real data events from MC shower events. Rather than being outliers in any particular way, they are simply the examples of the lowest values of an otherwise continuous track score, which merely appears to be shifted for real data towards higher values.

Finally, one of the simulated atm. muons passing the selection is depicted in Fig. 8.26. This gives a feeling about how the muon events look like that survive the neutrino probability as well as upgoing cut. They leave behind

ORCA6, run id 7397, event id 48, hits (trig.) 83 (31), neutrino prob 1.0, track prob 0.2181, rec. energy $(7.74^{+4.0}_{-2.64})$ GeV, rec. x-dir 0.62±0.11, rec. y-dir -0.15±0.16, rec. z-dir 0.77±0.1 - MC info: ν_μ (CC), true z-dir 0.83, true energy: 7.5 GeV
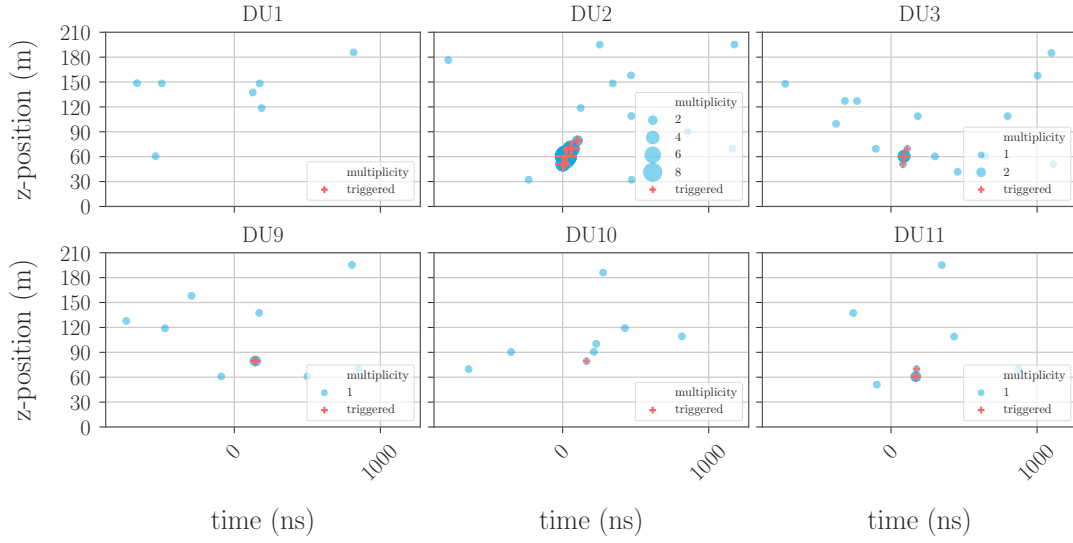
Figure 8.24: Hit pattern in *z-t* projection for a simulated muon neutrino event of lower energies.

ambiguous hit patterns that can coincidentally appear identical to an upgoing muon, even to the human eye.

ORCA6, run id 7581, event id 36039, hits (trig.) 64 (15), neutrino prob 1.0, track prob 0.1518, rec. energy $(2.51^{+1.3}_{-0.86})$ GeV, rec. x-dir 0.04±0.21, rec. y-dir -0.16±0.3, rec. z-dir 0.99±0.09
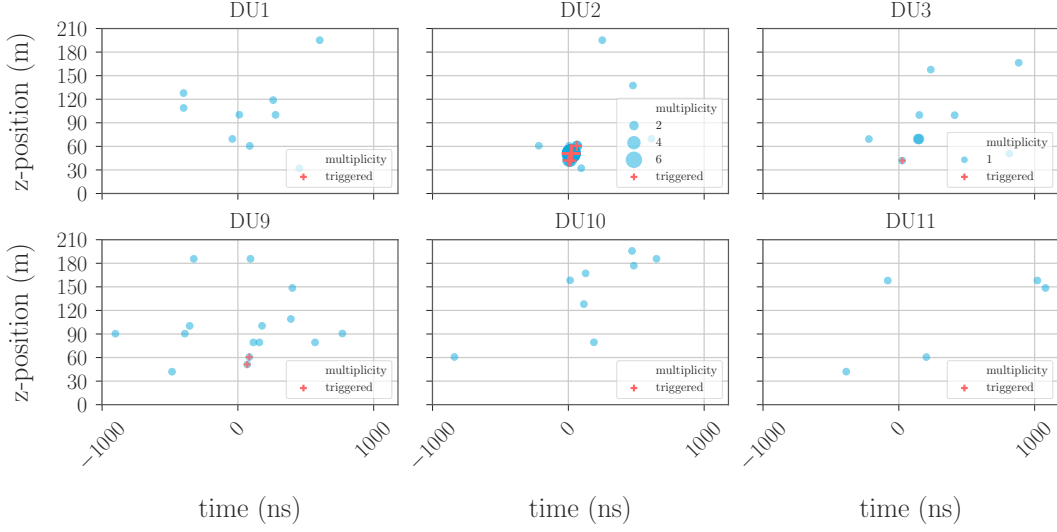
Figure 8.25: *Z-t* projection of the hits in a real data event that was assigned a low track score.

ORCA6, run id 7483, event id 7201, hits (trig.) 79 (19), neutrino prob 0.99934, track prob 0.8754, rec. energy $(18.82^{+12.21}_{-7.41})$ GeV, rec. x-dir -0.34±0.11, rec. y-dir 0.32±0.14, rec. z-dir 0.88±0.1 - MC info: $\bar{\mu}$ (n/a), true z-dir -0.4, true energy: 82.19 GeV
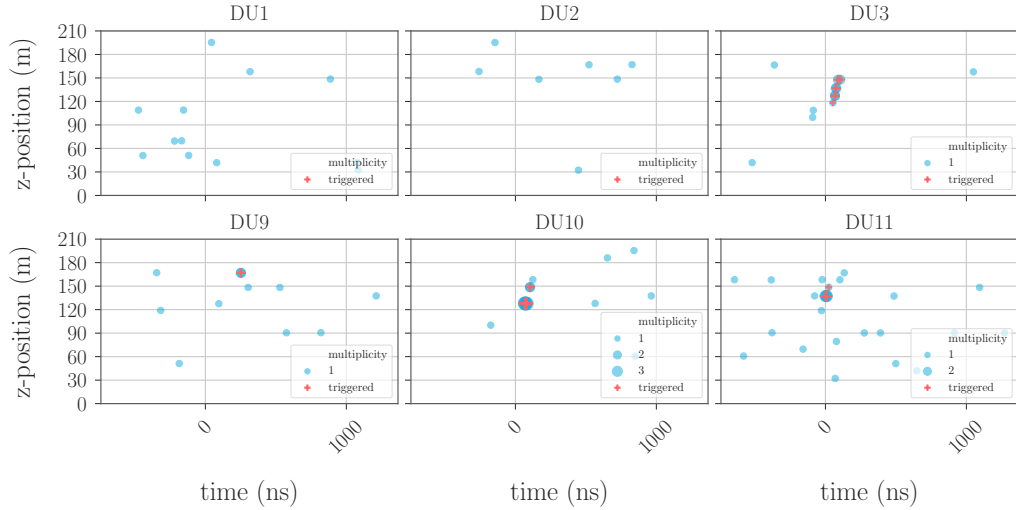

Figure 8.26: Seemingly upgoing hit pattern in the *z-t* projection caused by an initially downgoing atmospheric muon.
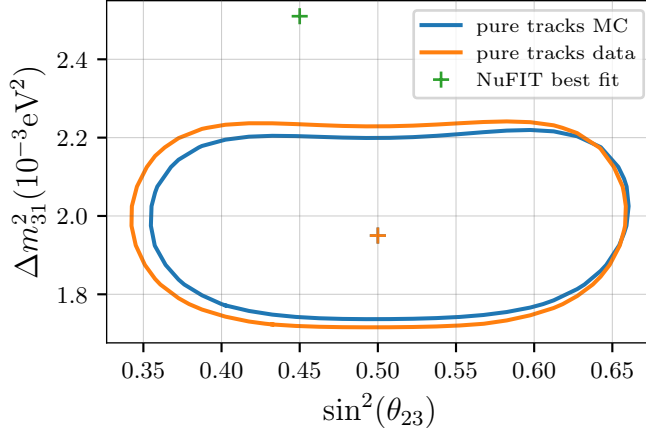
Figure 8.27: 90% CL in large mass splitting and oscillation angle θ_{23} for the analysis based on two different underlying data sets, one designed for MC only (blue) and one which also ensures data/MC agreement (orange). As a reference, the NuFIT best fit value is added (green) [52].

8.2 Oscillation analysis in real data

Utilizing the established selection of real data neutrino candidates, their energy and $\cos\theta$ are employed to extract a set of oscillation parameters that best describes the data. The same procedure as detailed in Chap. 7.2 is applied to yield the limiting contour lines.

8.2.1 Expected sensitivity for adjusted selection

Before revealing the real data results, the evaluation of the impact of the change in selection strategy, described at the beginning of this chapter (8.1.1), is given. To obtain the most accurate prediction for the real data analysis, the contour for the data selection does include the effect of the re-weighting and allows for an additional normalization to the overall real data counts to account for some of the uncertainties involved with the atm. neutrino flux. This is realized for the orange line in Fig. 8.27 which is compared to the pure track selection solely focusing on MC from Chap. 7.1 in blue. As before, the true oscillation scenario is chosen to be $\sin^2\theta_{23} = 0.5$ and $\Delta m^2_{31} = 1.9 \times 10^{-3}\text{eV}^2$, represented by an Asimov data set.

As expected from the properties characterizing both selections (Tab. 8.1), the more prominent shower component for the data case slightly reduces the sensitivity in Δm^2_{31} and lower values for θ_{23}. The additional shower events better constrain the region of larger oscillation angles around 0.65. The impact can be quantified by the significance for the NuFIT value exclusion, which decreases from 4.0σ to 3.5σ.

8.2.2 Measurement of θ_{23} and Δm^2_{31} in real data

Now, the Asimov test data set is exchanged with the real data and the best fitting values are determined to be $\sin^2\theta_{23} = 0.42$ and $\Delta m^2_{31} = 1.9 \times 10^{-3}\text{eV}^2$, as indicated by the blue cross in figure 8.28. The figure also draws the 90%
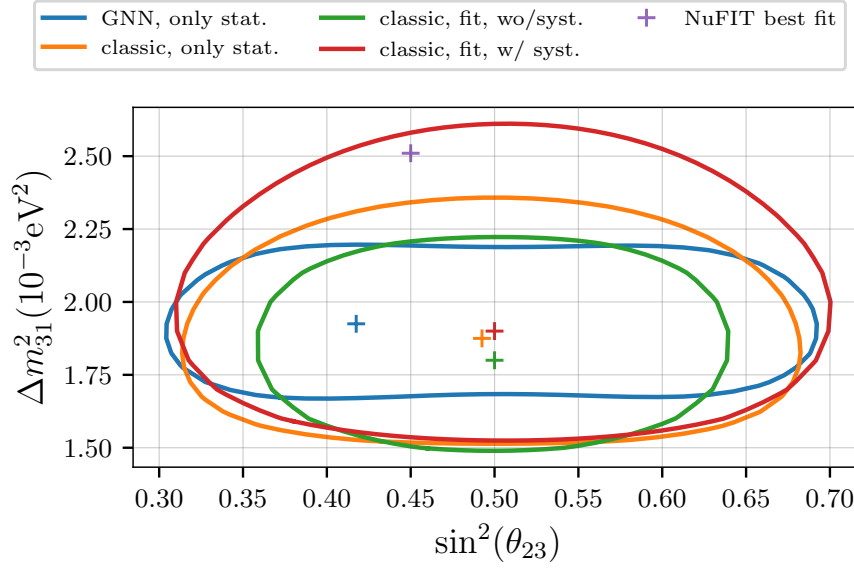
Figure 8.28: Best fits (crosses) and 90% CL (solid lines) in Δm_{31}^2 and $\sin^2\theta_{23}$ for different analyses and selections based on the one-year ORCA6 data set. Blue and orange are evaluations based on the statistical approach for the GNN and classical selection, respectively. The fits for the classical case (green and red) are taken from [139]. As a reference, the NuFIT best fit value is added (purple) [52].

CL contour, which is found to be slightly more extended in θ_{23} than was expected from figure 8.27. Instead of the lower threshold at $\sin^2\theta_{23} = 0.34$ and the upper one at 0.66, the χ^2 map, when compared to the real data, produced 0.31 and 0.69 as 90% ranges. For the mass splitting, the limits at $\sin^2\theta_{23} = 0.5$ are predicted to be 1.71 to 2.23, respectively, and found in real data to be 1.69 to 2.18. Consequently, the actual tension to the world's combined best fit (purple cross) is calculated to 3.8σ, which is in the same region as the MC studies that started from a slightly different true value.

With that the transfer from MC to real data is successful. The cause for the moderately minor discrepancies between the predicted and observed contours are a combination of the remaining differences in the underlying spectra between data and MC (due to the discussed potential differences in simulation and reality) and potential biases introduced by the correction of the energy shift and the re-weighting based on the number of hits.

In addition to the blue curve from the GNN selection plus GNN reconstruction, the contour for the classical selection utilizing the classical track reconstruction is included in orange to discuss the outlook to an analysis encompassing the consideration of a complete set of systematics. First, the classical curve exhibits similar features regarding the transfer from MC to the actual data. The initially expected limits were shown in Fig. 7.15, right, and turn out to be smaller in Δm_{31}^2, including a different shape for small and large oscillation angles, while they are comparable

in $\sin^2\theta_{23}$. In this thesis, all contour lines are produced scanning the χ^2 in the parameter space and then reading off the minimum and the position of the $\Delta\chi^2$ corresponding to the 90% CL. A more sophisticated approach is to use all points of the χ^2 map in the form of a fit, which can better constrain the optimal parameters. For this purpose, the KM3NeT collaboration developed dedicated software that handles the influences of different systematics, such as the normalization, spectral index or flavor ratios. The complete description of the analysis, including the list of all systematics, can be found in [139]. The resulting contour lines denoted with "fit" in figure 8.28 are taken from these studies as a reference. Performed on the classical selection for the same data set, the fit neglecting all uncertainties except for statistical contributions is added in green and proves that the sensitivity can be significantly increased over considering only single points of χ^2 values (orange). Then, to account for all uncertainties, the red curve is generated, which constitutes the final result of the first oscillation analysis. Using these, it is possible to predict the type of improvement that can be expected from the presented GNN pure track selection to the present results already utilizing the current state of data and simulations. If the involved systematics are assumed to be comparable between deep learning and classical approach, the differences between the orange and the red lines can be transferred to the blue and the future contour describing the full fit for the GNN case. Doing this, an ample increase in sensitivity in Δm^2_{31} is expected, while the capability to determine the oscillation angle suffers from the still mixed set of electron and muon neutrinos and the discovered data/MC discrepancies. This is particularly interesting, as the preference for a significantly smaller Δm^2_{31} of 1.9 is only reinforced by this work's studies. Additional contributions to the uncertainty from the energy shift and re-weighting may need to be considered if the effects causing the data/MC discrepancies remain. Efforts to integrate the GNNs into the existing frameworks are currently ongoing.

These discussions underline once more the need for an excellent data/MC agreement, as not only is the analysis compromised in the form of the underlying spectra, but also the possibility utilize the track/shower classifier will be instrumental for realizing the full potential the deep learning-based selection offers, allowing for the subsequent splitting of data sets.

To showcase the type of contribution KM3NeT/ORCA can make to the world's combined efforts in oscillation research, Fig. 8.29 contains the 90% CL of the leading experiments contributing their data to produce the NuFIT best value. Added are the two results discussed from the preceding figure; the full systematics fit for the classical analysis and the preview the statistical approach gives for the GNN selection.

Both demonstrate that after only one year of data taking with a partly built detector, ORCA is on the map to contribute to the determination of oscillation parameters. Due to the different techniques, sizes and lifetimes
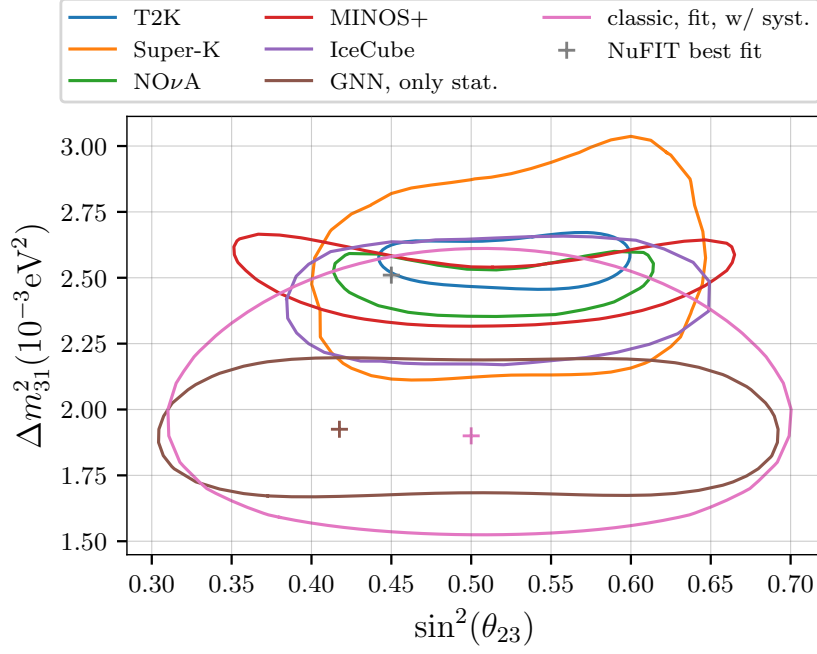
Figure 8.29: 90% contour lines for different experiments constraining the parameter space of large mass splitting and oscillation angle θ_{23}, including the results from the ORCA6 data set [139]. More information about the world's data can be found in [53].

of the involved experiments, it is often not possible to directly compare the contributions. Furthermore, the systematic uncertainties are vastly different, and the size of the contour in the parameter space does not linearly shrink with statistics. However, it should be noted that most of the established experiments run for several years, like Super-Kamiokande for 26, T2K for 10 or NOνA for 8 years. With the ongoing improvements in simulations and analysis, the steady enlargement of the detector and the continuous stable data taking, KM3NeT/ORCA is well on its way to establish its position in the oscillation research community.

8.2.3 Oscillation signature

One instructive way to visualize the impact oscillations have on the data collected is to compare predictions from different models, including the no oscillation hypothesis, to the observed data by combining energy and $\cos\theta$ to some L/E, where

$$L = 2\cos\theta \cdot r_{\text{Earth}}, \quad \text{with } r_{\text{Earth}} = 6371\,\text{km},$$

the mean radius of the Earth [154]. This is not entirely accurate, as the Earth is not completely round, neutrinos are created in the atmosphere and not at the surface, and the path through the atmosphere is different for

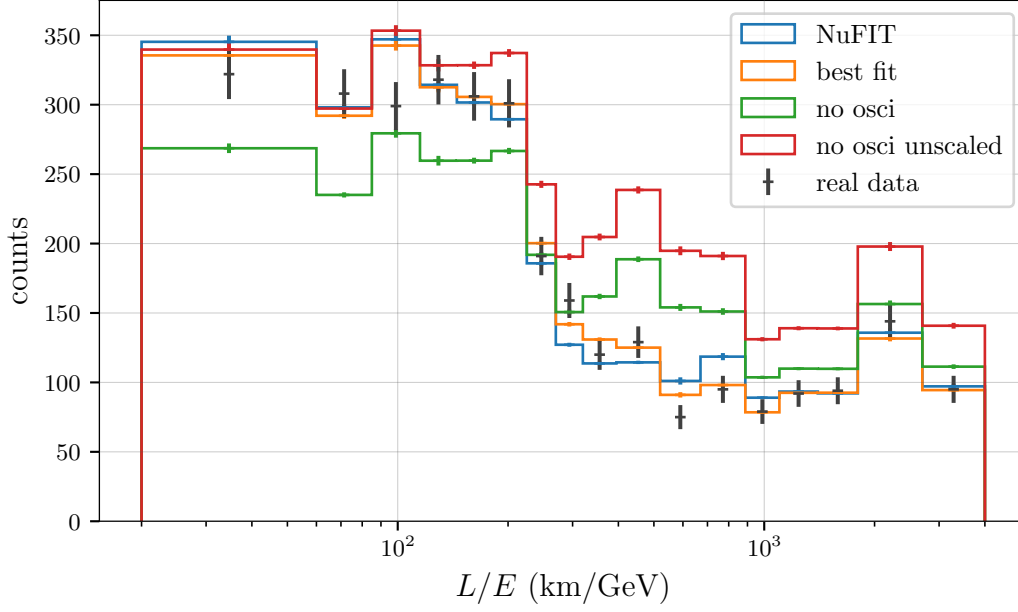Figure 8.30: Histogram of the L/E for different oscillation scenarios (colors) compared to real data (black). The first three cases are normalized to yield the same integral as the real data.

different angles. Nevertheless, since the height of the atmosphere with about 10 km is small compared to the radius, and absolute values for the following representation are not as important, this approximation is sufficient. The L describes the length of the path the neutrino had to oscillate (see Eq. 1.15), and a few values were mentioned when discussing the ranges different kinds of experiments are sensitive to (Chap. 1.3.4). For atm. neutrinos, ~ 500 km/GeV were quoted to exhibit the first oscillation minimum.

To clearly depict this, the data are first binned with handpicked ranges like shown in Fig. 8.30 to have a denser coverage from 300 km/GeV on, while maintaining statistically significant counts per bin. Included are the predictions from the NuFIT parameters, the best fit found with the GNN and the no oscillation scenario, as well as the real data counts. For the first three entries, the normalization is left free, i.e., the spectra are normalized to have the same integral between real data and MC. In the red histogram, the flux is kept fixed.

The largest differences already become apparent for the higher L/E and consequently, a total χ^2 can be calculated, quantifying the agreement between a model and the data. The values are displayed in Fig. 8.31 and each bin's counts are divided by the unscaled non-oscillation hypothesis. This causes the red line to stay at 1 by design. The green line thus represents the average best fit of no oscillations to the data, preferring 80% fewer events overall than the unscaled scenario.

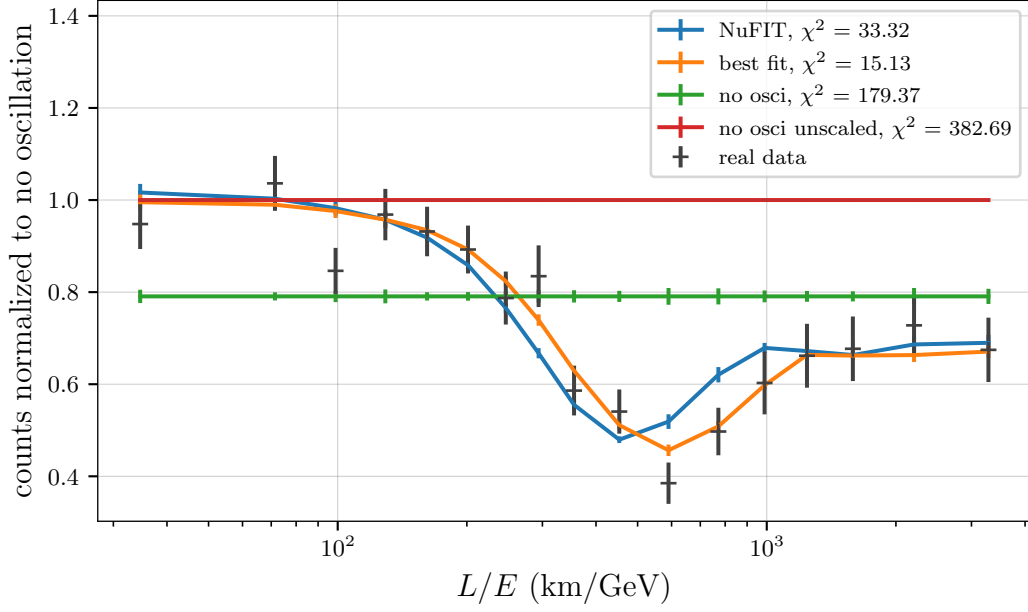For small values of L/E, all models are similar, as there is simply no

Figure 8.31: Ratio from the bins of the preceding figure with respect to the unscaled non-oscillation case. The resulting χ^2 compared to the real data (black) is given in the legend.

distance over which oscillations can take place. Only after $100\,\mathrm{km/GeV}$ the oscillation signal can be distinguished, rising in significance until around $500\,\mathrm{km/GeV}$. After the first, most distinctive dip, further oscillating maxima and minima follow but are not yet well resolved, due to the low number of statistics for events with particularly low energies and close to vertically upgoing directions.

The quantity L/E not only visualizes the influence of oscillations but also allows for quantitative statements about the tested models and their fit to the data. In this case, the best fit parameter set is preferred by a $\Delta\chi^2$ of 18.27, which translates to 4.2σ assuming one degree of freedom. Most of the discrimination power in favor of the best fit comes from the region of $600\,\mathrm{km/GeV}$ to $1000\,\mathrm{km/GeV}$. This is similar to the statement derived from the contour plot (3.8σ, Fig. 8.28), which is expected, as this is merely another way of conceptualizing the same data.

The no oscillation scenario can be excluded with a $\Delta\chi^2$ of 164, which is difficult to translate to a significance because of the limit of floating-point precision. An approximation assigns this purely statistical evaluation more than 12σ. With this, neutrino oscillations are evidently measured in the ORCA6 data set utilizing the tools presented.

As initially illustrated in the introduction in Fig. 2.5 and now demonstrated with the discussions on the analysis, the oscillation signature is most apparent when correlating energy and $\cos\theta$. However, other single quantities reveal the influence of oscillating flavors as well, even though they are not

as powerful. One example is the direction uncertainty. The most significant effect is visible in the y-direction, which yields a 2.4σ tension between the determined best fit and NuFIT. The major contribution comes from medium uncertainties and thus muon neutrinos of lower to medium energies. This sensitivity stems from the close correlation to the energy, which was discussed for Fig. 8.5. No oscillations are also here clearly disfavored by 73.8 $\Delta\chi^2$.

In principle, the classifier outputs, especially the track probability, would be an exciting quantity to study, as tracks changing into showers and vice versa are expected to be observed. However, due to their large data/MC discrepancies to begin with, no meaningful conclusions can be drawn from these.

As a cross-check, specific distributions should not change under the oscillation hypothesis. One example is the azimuth angle, which indeed only allows differentiating between NuFIT and this best fit with 0.7σ.

8.3 Influences of data taking conditions

The detector exhibits different data taking conditions, mainly governed by the bioluminescence activity and the sea current. Additionally, different high-voltage settings were run during the ORCA6 lifetime resulting in adjusted time calibrations. Influences from those on the selection, data/MC agreement and the oscillation analysis are evaluated in this section.

This constitutes a test for the generalizability of the GNNs, as no additional effects, other than those explicable by the conditions, are expected to be observed.

8.3.1 Influence of the mean PMT rate

An instructive way to evaluate some of the influences of particular data taking conditions is to divide the lifetime into periods based on the quantity to study and then to assess the data/MC agreement. This is executed in Fig. 8.32, where the dependence of the selected events in real data and MC on the mean PMT rate per run is plotted. Additionally, the same dependency for the classical selection is shown. This version of the classical selection has an additional cut on the reconstructed energy at 100 GeV to remove a region of data/MC discrepancies. This causes the rates to be lower than previously mentioned in the discussions about MC only. Hence, both selections are now adjusted to do real data studies. The bins of the x-axis are chosen to contain the same number of MC events, which leads to the unequal spacing but conveys an idea about the underlying distribution of mean PMT rates; runs exhibiting around 8,000 Hz are much more abundant than 13,000 Hz. The y-axis has been normalized to an event rate per year for better readability and the lower part of the plot contains
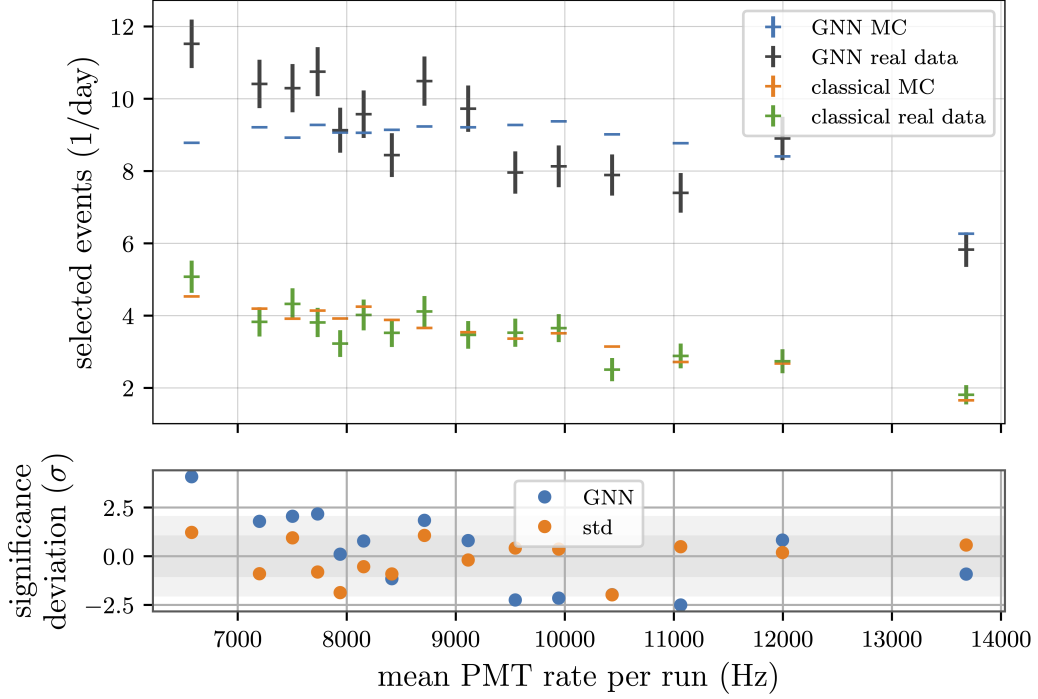
Figure 8.32: Selected event rate for GNN (blue and black) and classical selection (orange and green), comparing data and MC, versus the mean PMT rate observed in a run. Evaluation of the similarity in the bottom part like Fig. 8.4.

the same quantitative comparison as used before (Eq. 8.1) to spot significant differences between observation and prediction.

With higher overall PMT rates, the probability of the high-rate veto being triggered for single PMTs is higher and thus the number of selected events will decrease. When fewer PMTs are actively contributing to the data taking, the detection efficiency deteriorates. Such changes are caused by bioluminescence, as the organisms become more active depending on the sea current and the seasonally changing nutrition levels. The rate of triggered events, on the other hand, does go up, as more background noise is generated. The decrease in selected real data rate is clearly visible, while the MC predicts a flatter plateau.

The most significant disagreement is found for rates below 7,000 Hz with more than 4σ. A significant overestimation in MC is apparent in the middle region between 9,500 Hz and 11,000 Hz. Interestingly, the agreement is excellent again for the highest rates, which are potentially the most difficult to correctly model. Regions with more significant disagreement might point at specific features not well reproduced in the simulation. For the mean PM rate, for example, the capability to select events for low background rates appears to be higher in reality than it is possible in MC.

The same decreasing trend is observed for the classical selection, though data and MC display a closer agreement, partly due to the lower statistics
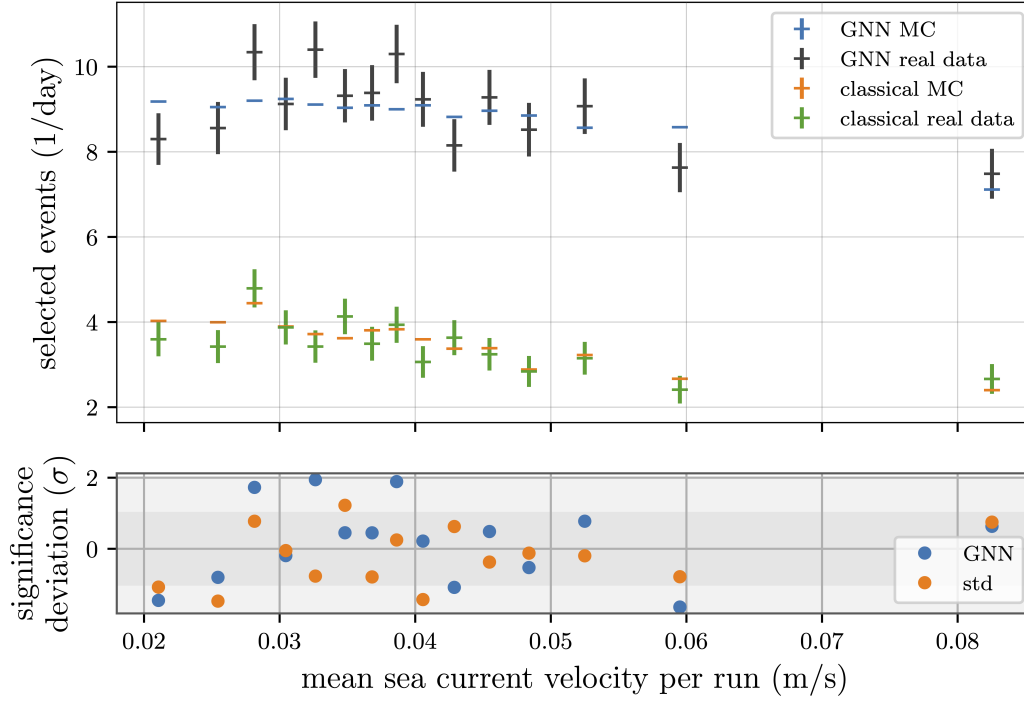
Figure 8.33: Same structure of displaying the selected event rate as in Fig. 8.32, this time versus the mean sea current velocity.

resulting in lower significance. Most notably, the classical MC rates reflect the decrease in selected events with the PMT rate more clearly.

8.3.2 Influence of the sea current

The sea current as measured nearby the ORCA detector can be evaluated in the same way as the mean PMT rate by binning the data into periods of different sea current velocities, realized in Fig. 8.33. In addition to affecting the bioluminescence activity, larger sea currents also cause swaying of the detector's DUs as they get dragged by the forces. On average, the static position and orientation calibration used in this ORCA6 data set should provide a less accurate description for increasing currents.

Also in this evaluation, a slight trend of fewer selected events with increasing sea current is observable. For velocities as high as 8 cm/s the number of events has decreased to 7.5 per day, while more than 10 events per day are selected for slower currents. The significance of the deviation between data and MC rate does not exceed 2σ for any bin. Similar conclusions can be drawn from the classical selection.

Especially the accurate modeling for the highest sea currents suggests that the systematic effect introduced by the missing dynamical position calibration on the selection efficiency is minor.

Table 8.2: Characteristic properties of the complete selection and the lifetime subdivided into periods of different high voltage settings and thus intra-DOM time calibrations. Detailed explanations about the meaning of the quantities are given in the text.

	complete lifetime	1st period	2nd period	3rd period
lifetime (days)	357.24	69.40	205.76	82.07
mean HRV fraction (%)	10.8	5.5	14.5	7.6
mean PMT rate (Hz)	9155	8449	9977	7938
mean sea current (cm/s)	4.1	4.3	3.9	4.1
rate real data (1/year)	3311 ± 58	3597 ± 137	3142 ± 74	3495 ± 124
rate MC (1/year)	3235 ± 7	3298 ± 21	3198 ± 11	3268 ± 13
ratio data/MC (%)	102.3 ± 1.8	109.1 ± 4.2	98.2 ± 2.3	107.0 ± 3.8
muon cont. (%)	0.7 ± 0.2	1.1 ± 0.6	0.7 ± 0.3	0.5 ± 0.4
reduced χ^2 energy spectrum	2.24	1.22	1.89	2.02
$\Delta\chi^2$ NuFIT to no osci in L/E (one year)	149	173	177	57

8.3.3 Characteristics of different data taking periods

The data taking over the 357.24 days of ORCA6 lifetime was operated with three different high voltage settings. As explained in Chap. 3.1, each PMT's high voltage is adjusted from time to time to ensure equal gains and thus detection capabilities. With it, the intra-DOM time calibration is adjusted, while the inter-DOM and inter-DU time offsets have also been monitored. Very slight modifications where found with the updated PMT timings. These different periods of high voltage settings/calibrations can be used to divide the complete lifetime and evaluate whether they behave systematically different. The characteristics studied are normalized to a one year lifetime where possible to ensure comparability. All values are summarized in Tab. 8.2, starting with the individual lifetimes. The first and third period are rather short, yielding a generally higher statistical uncertainty on their statements.

Three further quantities describing the data acquisition conditions are included to put the physics results into perspective. The high-rate veto fraction reduces, as discussed, the effective detection capability and is found to be particularly low for the first and third period, with the second doubling the amount. The PMT rates are also highest for the middle period. The mean value of the sea current, on the other hand, stays relatively constant over the entire year.

The rate of selected events in real data follows the HRV fraction closely, as roughly 14% more events are found in period one than in period two.

Even though the noise level is extracted from real data for the run-by-run simulation, this effect is not as pronounced in the MC rate with only 3% more events in the period with a low HRV fraction than for the second one, which showed a higher fraction. This is strongly reflected in the ratio of data and MC rates, which is close to a 10% underestimation in MC for the first period and even an overestimation for the middle period, though the statistical uncertainties in the latter case still allow for perfect agreement. These correlations are similar to the conclusions drawn when discussing the dependence on the mean PMT rate.

The low statistics, paired with the considerable uncertainty of selected muon events, do not allow for any meaningful conclusions, as all periods are compatible with the average muon contamination of 0.7%. These, and the other values concerning the numbers of events, are slightly different from the ones quoted in Tab. 8.1 because of the now considered re-weighting.

Next, the data/MC agreement can be compared in binned spectra, like the energy. The other quantities allow for similar conclusions. The re-weighting on the hits spectrum is performed in the same way for all periods, considering the values obtained from the complete lifetime. It can be seen that the number for the shortest period is smaller (1.22) and the complete spectrum displays the largest reduced χ^2 (2.24). This can be understood by the following: Regardless of the size of a set, a χ^2/ndf close to 1 always states that the two quantities are drawn from the same underlying distribution. For sets of lower statistics, the counts per bin are lower and consequently, the difference between the two compared quantities is smaller, resulting in a generally lower χ^2. At the same time, statistical fluctuations are relatively large, yielding larger χ^2 values. Sets of higher statistics are thus more sensitive to actual differences between the distributions. This is the case for the complete lifetime, while for period one, for example, the discrepancies are hidden in the statistical fluctuations. Following this argumentation, the third period exhibits a relatively large reduced χ^2 of 2.02 and can be examined in Fig. 8.34. In it, a slight preference for a different energy shift is observable, primarily noticeable for energies from 20 GeV to 50 GeV. Other bins with lower counts are simply not significant enough. This is particularly interesting, as it points towards a possible change over time to the origin that causes the need for the correction in the first place. It can be connected to the efficiency calibration, which is impacted by more or less sedimentation, or the time calibration itself (see discussions Sec. 8.1.2).

Finally, Tab. 8.1 in the last row lists the difference in χ^2 with which the NuFIT best fit can be distinguished from no oscillations in the L/E spectrum. Due to the small statistics, it is challenging to use the 2d histograms needed to create the χ^2 map in the Δm_{31}^2 versus θ_{23} parameter space and to evaluate the difference of points in there. Instead, the L/E spectrum and the comparison between two fixed oscillation scenarios, NuFIT
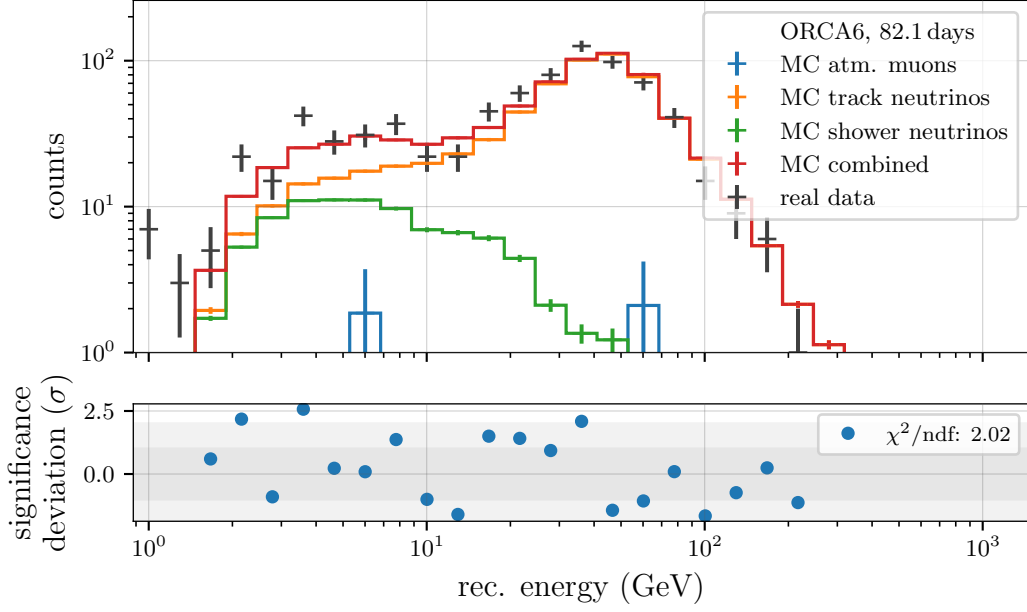
Figure 8.34: Energy spectrum for the third period only, showing contributions from MC (colors) and the real data (black). Evaluation of the similarity in the bottom part like Fig. 8.4.

and no oscillation, can be used to quantify the sensitivity normalized to one year for each period more reliably.

The resulting values indeed attribute the third period the lowest discrimination potential with only a third of the $\Delta\chi^2$ achieved in the first two periods. This is directly connected to the discussed discrepancies for data/MC in the quantities, such as the energy. For periods one and two, the close sensitivities are reassuring, suggesting steady conditions.

8.3.4 Poisson nature of neutrino counts

Concluding the discussion about potential systematic influences on the GNN neutrino selection, the number of selected events per data run can be evaluated. If the number of selected neutrinos per run follows a Poisson distribution, each run is an independent measurement under the same conditions, i.e., with the same probability to contain a selected neutrino. This is probed in Fig. 8.35, where a Poisson distribution is fitted to the occurrences of the different neutrino observations per run. The blue curve with an expected value of 2.0 indeed describes well the measured data, underlining the low influence of any of the mentioned effects on the selection. Especially for low numbers of selected events (0 and 1), no significant deviation is found that would point to an abundance of problematic runs that do not produce the expected number of neutrinos. Instead, the number
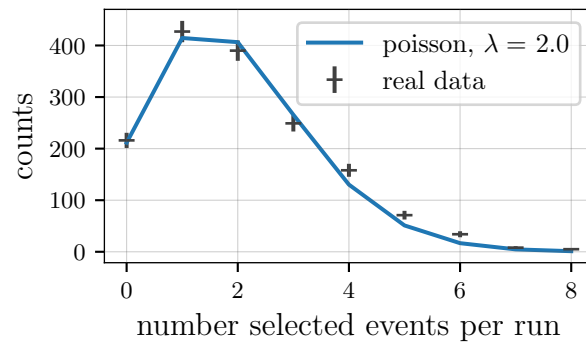
Figure 8.35: Histogram of the number of selected events in a data run (black). The resulting counts are fitted with a Poisson distribution (blue).

of short runs or those rendered challenging by data taking conditions is very low for those considered in the analysis.

9

Influence of calibration on data analysis

T HE final chapter explores the influence of a miscalibration of the detector on the various stages of the data analysis. To accomplish that, the knowledge gained from the studies presented in chapters 3 and 4 is used to define a set of minor and a set of major miscalibration. Uniting the two main projects of this PhD work, the deep learning data processing and analysis is then repeated for the same ORCA6 data, starting from the different calibrations. This way, the influence on the quantities before cuts, the neutrino selection, and the oscillation sensitivity can be evaluated.

9.1 Definition of miscalibration

In Chap. 4.4, the base calibration that is used for all analyses so far was established. Now, modifications to different calibration properties are introduced, which affect the hit information the neural networks receive as input. This can *a priori* influence the networks in two ways: Either the output is different because the physics have changed - different hit patterns in time and space cause different observable directions, for example - or the GNNs are confused by the modified input, outputting extreme or significantly shifted values. The latter is also referred to as "adversarial examples" [155].

The modifications are applied to both MC and real data in order to study the influence on the data/MC agreement. When comparing the outputs from the differently processed MC sets, not only the impact of a potential miscalibration can be assessed, but also the implications from missing dynamic positioning, as also here the positions differ from what the networks are used to during training on the static simulations.

The actual values have been chosen to reflect a case of "small miscalibration", which considers values in the order of the uncertainties and differences that were found when dealing with real data and comparing to other calibration methods, and a "large miscalibration", which features

Table 9.1: Values characterizing the modifications for the small and large miscalibration case. The numbers apply to the six DUs consecutively, i.e., DU1, DU2, DU3, DU9, DU10, DU11.

	small miscalibration	large miscalibration
inter-DU time offset (ns)	2, -2 ,-3 ,2 , 1, -1	20, -15, -25, 20, 20, -30
inter-DU x offset (m)	0.5, 0.0, -1.0, 0.3, 0.1, -0.6	5, 0, -3, -3, 4, 4
inter-DU y offset (m)	0.4, -0.8, 0.5, 0.2, 0.5, -0.4	2, 4, -4, 5, 2, -3
height scaling (%)	0.3, -1.1, -0.4, 0.5, -0.4, 0.9	-0.8, -2.0, 1.3, 1.2, 1.6, -2.1
orientation offset (°)	-2.3, -1.7, 4.0, -2.9, 1.1, 6.3	25.8, -24.6, 28.6, -37.2, 24.1, 16.6
inter-DOM time offsets (σ of Gauss)	1	4

more extreme values to better probe the limits. All modifications to the different quantities are applied simultaneously, meaning for the first case that all potential inconsistencies add up, which describes the most pessimistic scenario.

An overview of all modifications is given in Tab. 9.1. The individual values were randomly chosen, but attention was paid that their effects do not cancel each other out. These modifications are kept constant over time. The inter-DOM time calibration is changed on a DOM-by-DOM level with additional offsets sampled from a Gaussian distribution centered at zero with the standard deviation quoted in Tab. 9.1.

The change of calibration is introduced when creating the input for the networks, which means explicitly that the influence from the trigger stage is not considered. The effect of events not getting triggered should be negligible, considering the settings for ORCA6 are rather loose in the first place to not miss out on any physics events. Applying the trigger assuming the modification will eventually only lead to a reduction in events that instead are now kept and may be challenging to reconstruct.

9.2 Influence on quantities before cuts

The natural first question would be how the raw outputs of the GNNs change compared to the original calibration. This is demonstrated for a few selected reconstruction parameters and the classifiers.

9.2.1 Reconstruction parameters

To start off, the uncertainty for the x-direction is examined in Fig. 9.1. These and the following plots contain spectra before cuts for a few days
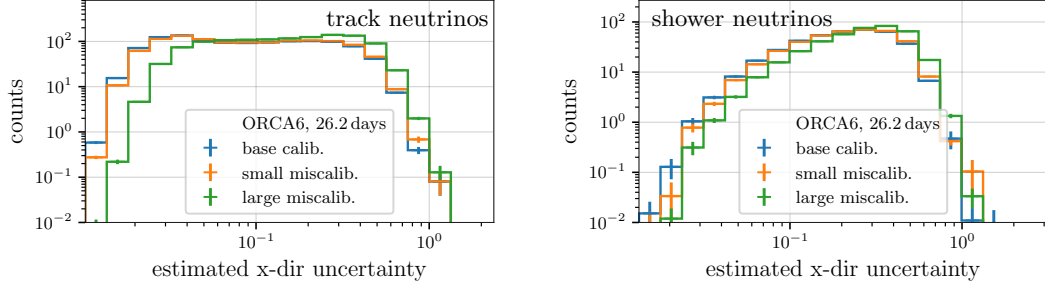
Figure 9.1: Spectrum of the estimated *x*-direction uncertainty for reconstructing with different calibrations (colors). Left: Track neutrinos. Right: Shower neutrinos.
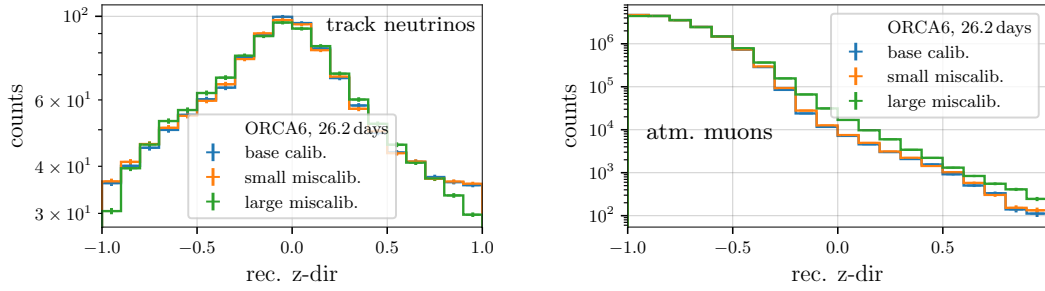

Figure 9.2: Distributions of the reconstructed *z*-direction for processing with different calibrations (colors). Left: Track neutrinos. Right: Atm. muons.

of data taking, weighted to atmospheric flux as always, for only a single topology and compare the three different calibrations used in the processing. The left side shows track neutrinos and the right side the shower counterpart. For both topologies, the base calibration and the small miscalibration arrive at very similar distributions. The larger miscalibration exhibits the desired apparent effect of a substantial impact on the GNN outputs. For the direction uncertainty estimation (and this is true for atm. muons and the other direction components as well), the values of the green curve are generally shifted towards higher uncertainties, visible from 0.1 on in the track case and lack entries below 0.04. The general shape of the distributions remains similar, suggesting that the observed influence is of actual physical nature. The uncertainty increases because hit distributions do not fit that well to the learned signature anymore. Thus, according to this control parameter, the direction cannot be determined as reliably. The same qualitative statements are valid for the shower neutrinos, with the difference appearing slightly smaller.

Next, the influence on the parameters relevant to the physics analysis, energy and $\cos\theta$, are evaluated. Figure 9.2 contains the spectra of reconstructed *z*-direction for tracks (left) and atm. muons (right). The differences for most of the track neutrinos are small, except for the vertical directions. Here, the solution close to ± 1 is less often found for the larger
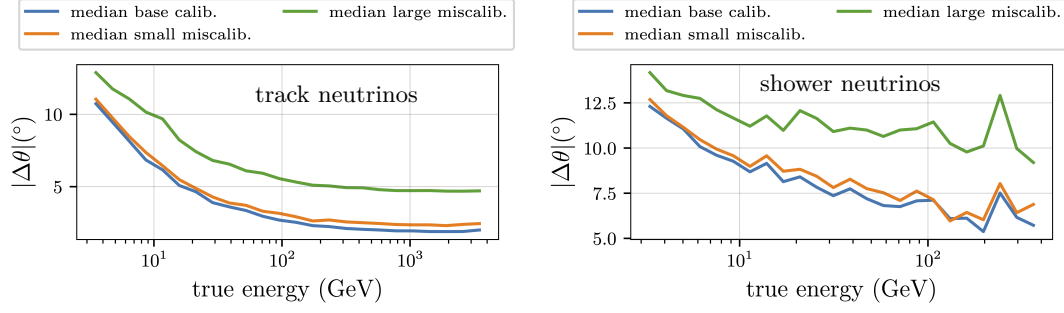
Figure 9.3: Median of the absolute value of the difference between reconstructed and true zenith angle for track-like neutrinos on the left and shower-like data on the right. The differently colored lines assume different calibrations during processing.

miscalibration. Avoiding these systematically indicates that the GNNs are confused by the introduced changes, rather than it is likely to explain this by the different geometry the detector exhibits due to the modification in the 3d space. The same feature is observed for shower neutrinos.

The leading contribution to the differences in the distribution of atm. muons on the right side is the increase of misreconstructions at positive $\cos \theta$. An additional effect for the edges of the spectrum is not visible. This will affect the neutrino selection, as more muons will pass the upgoing cut.

Constituting that even larger modifications do not act as clear adversarial data to the direction reconstruction network still leaves the question of how much the resolution is affected by the physical changes to the detector. This is answered in Fig. 9.3, where for track (left) and shower neutrinos (right) the median of the resolution in the zenith angle reconstruction is compared for the different calibrations.

Over the whole energy region, the difference between optimal and slightly modified calibration is minimal ($<1°$). For energies lower than 10 GeV, it almost vanishes, as the uncertainties from kinematics dominate this region.

In the case of the large miscalibration, though, the degradation is about 3°, giving a $|\Delta\theta|$ of 10° at 10 GeV. Such loss in resolution will significantly impact the oscillation analysis, as differences between oscillation models get washed out over bins in $\cos \theta$.

On showers, the impact is similar. Only the already lower resolution causes the relative effect to be smaller. At 10 GeV a deterioration in the large miscalibration case by also about 2.5° is observed, which increases slightly for higher energies.

Next up is the energy, of which the spectra for tracks and showers are presented in Fig. 9.4. In the same way as for the direction reconstruction network, the shape of the distribution does not change significantly, even for the substantial modifications. For both topologies, higher energies are predicted more often for the large miscalibration. This could be because hit patterns in the 3d space are distorted and thus are interpreted by the
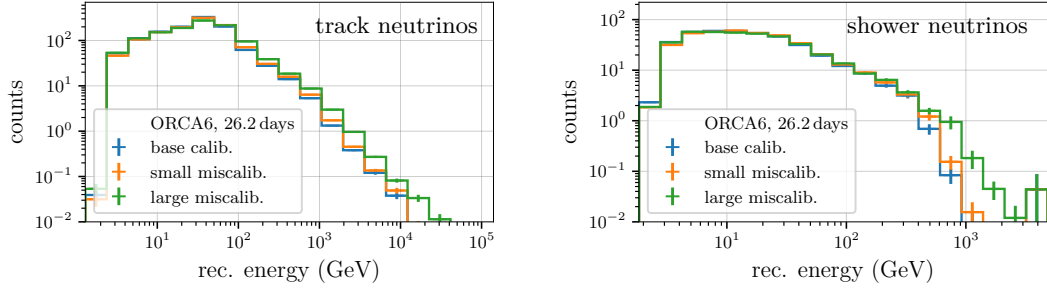
Figure 9.4: Spectrum of the reconstructed energy (left for track neutrinos, right for shower neutrinos) when using different calibrations (colors).
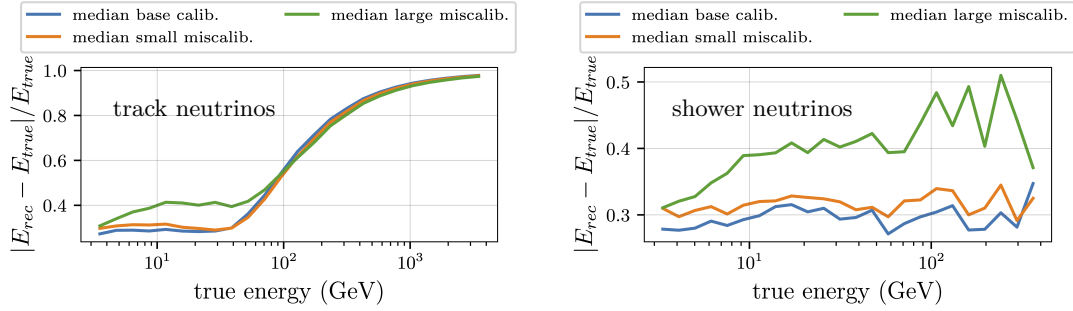

Figure 9.5: Resolution of the energy in form of the relative error. For both neutrino topologies (left and right) and all three calibration scenarios (colors).

network as a more extended event, with potentially missing hits in between due to fluctuations. However, this fraction is tiny and particularly for the tracks concerns the region with considerable underestimation. These events already exhibit no direct connection between the true energy and their detector response.

In addition to the distribution, the resolution of the energy is studied, drawn in Fig. 9.5 for track- and shower-like neutrinos. For the lowest-energetic track neutrinos, a visible difference between base and small miscalibration can be identified. At $10\,\text{GeV}$ it is 3% in relative error. The larger miscalibration rises before the effect of the detector limit sets in to about 40% resolution. In the limit, there is no distinction, as the significant underestimation dominates. For a few GeV, the inter-DOM relations are less significant, resulting in similar performances across all calibrations.

This feature is also observed for showers, with a similar increase in relative error around $10\,\text{GeV}$ for the green line, that is not present for the other cases. The order of magnitude of the differences between base and small miscalibration is the same as for tracks and does not depend on the total energy. The difference to the large miscalibration continues to increase for higher energies, as the events become more extended and thus suffer more from modified inter-DU properties. The number of events for
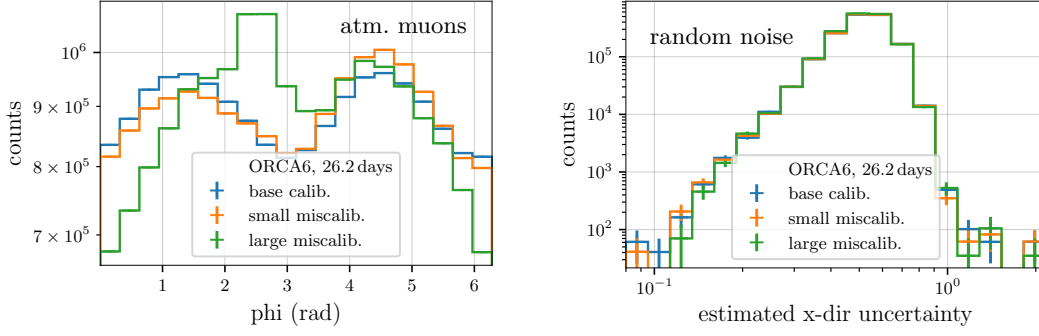
Figure 9.6: Further examples for the influence of the calibration (colors) on reconstructed parameters. Left: The azimuth angle for atm. muons. Right: The x-direction uncertainty for random noise events.

energies higher than $100\,\text{GeV}$ is low, causing larger statistical fluctuations, especially visible in the green curve. Overall, even the extreme case of the significant modifications does not exceed 45% resolution for the most part, which means that the energy estimation is relatively robust against influences of wrong hit times and positions. This is in line with the model of the energy reconstruction being more of a calorimetric measurement than heavily relying on causalities in space and time among the detected photons.

From the discussion of the energy reconstruction network, it can be concluded that also in this case the presented examples are hardly able to substantially confuse the network and cause it to output meaningless predictions. The observed effects are of physical nature, as a resolution deterioration is expected. This will negatively impact the oscillation analysis, but considering a reasonable uncertainty on the actual calibration, like represented by the small miscalibration case, the significance of the impact is limited, compared to other sources of uncertainties, like atmospheric flux or statistics.

Before concluding the discussion about the reconstruction, two additional examples are presented that underline the physical (geometrical) origin of the influence, rather than a technical one, and thus the robustness of the reconstruction graph neural networks. In Fig. 9.6, left, the reconstructed azimuth angle is plotted for the three calibrations. Exemplified by atm. muons, the original distribution is almost symmetric with higher acceptances (and thus counts in this spectrum) for the y-direction (see Fig. 2.2) because of the larger extension of the detector in x-direction with $50\,\text{m}$ than in y-direction with $15\,\text{m}$. When changing the xy-positions of the DUs, this regular pattern is disturbed and deviations from the original shape of the distribution are visible in the large miscalibration case.

Finally, the random noise events are not expected to exhibit any influence from the calibration, as their hits are uncorrelated by definition. This is
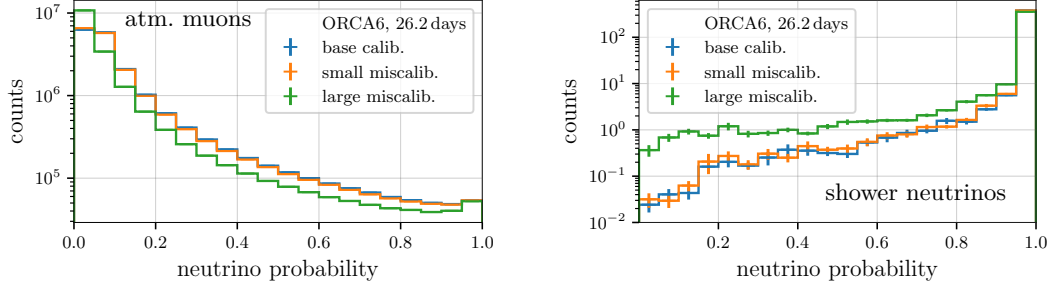
Figure 9.7: Distributions of the neutrino probability for different calibrations (colors). Left: Atm. muons. Right: Shower neutrinos.
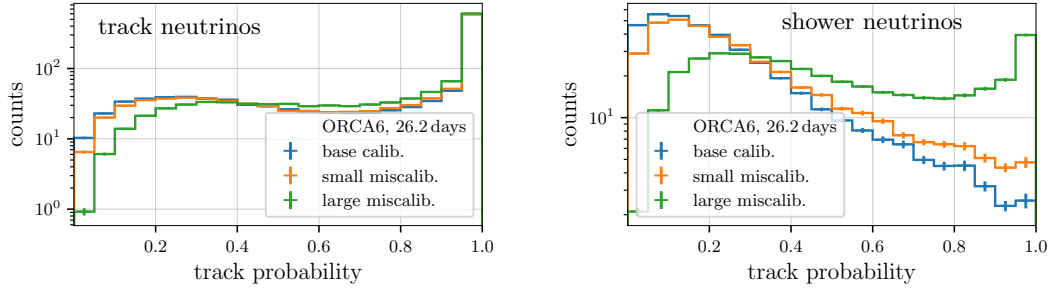
Figure 9.8: Spectrum of the track probability for neutrino topologies as it is influenced by different calibrations (colors).

also observed for all quantities, of which one example is shown in Fig. 9.6, right.

9.2.2 Classifier quantities

Next, the influence of the calibrations on the classifiers is evaluated. In the data/MC comparisons, these quantities displayed the largest discrepancies, which renders these studies even more important to possibly clarify the origin of the differences. Figure 9.7 starts with the neutrino probability for atm. muons and shower neutrinos. Similar to before, the base and small miscalibration do not differ much. However, the large miscalibration introduces a more extreme shape to the distribution of the background class for the signal/background classifier. Now, significantly more events are found in the first bin close to $p = 0$, while the counts from all other bins decrease. For signal events, the distribution is slightly less peaked at its true value of 1, and more entries are found for lower neutrino scores, resulting in a shift towards smaller scores. The shower neutrinos are used as the example, but the same is true for track neutrinos.

Furthermore, the influence on the track score for neutrino events is displayed in Fig. 9.8. For the track neutrinos on the left, the influence of a significant deviation from the original calibration appears to be a slight

shift towards higher track scores. The same effect is much more pronounced for the shower neutrinos on the right side. The shape of the distribution completely reverses in the green case, predicting more track-like events on the pure showers, which manifests itself in a clear peak at $p = 1$. The small miscalibration already indicates the change in shape, which is pushed to the extreme by the large miscalibration.

Apparently, distorting the near-spherical shapes of the shower event topology leads the track/shower network to assign a higher track scores much more often. That makes this network especially sensitive to potential differences between some test data (real data for example) and the data seen during training, as far as the hit distributions are concerned.

In summary, both classifier networks exhibit systematic influences by the changing calibrations. Especially for the track score, shower-like events act as adversarial examples that confuse the network. A robust classifier would still be able to identify shower events, as with the introduced changes, the appearance in the detector should not change dramatically. I.e., hits mainly on a single DU, like shown in Fig. 8.23, remain as such. One reason for the high sensitivity to the miscalibration could be the fact that the GNN has invested a lot of discrimination power to be able to distinguish lower-energetic tracks from showers, as they make up an important fraction of the training set. A short muon track is observed in those track examples on top of the hadronic cascade. This additional information has to be used to identify the tracks. If the hit patterns from particle cascades are modified, they may easily display features resembling an outgoing muon.

Consequently, these influences on the shapes of the spectra will affect the neutrino selection and subsequent analysis, as cuts have different effects for different calibrations. It also suggests an explanation for the discrepancies observed in the track probability spectrum for the real data case (Fig. 8.2), which implies a case of rather significant miscalibration. This, however, is in stark contrast to the conclusions from Chap. 4.4, where an excellent agreement between different calibration methods is found.

Of course, these evaluations only probe two distinct sets of miscalibrations that incorporate different kinds of modifications simultaneously. More systematic studies could be done examining the influence specific changes in time or space have on the spectra or resolutions. One special case of different height scalings is examined, as the uncertainty for those values is more prominent during the muon calibration (Tab. 4.5). The study reveals that neither a stretching nor a shrinking alone can explain the differences in the spectra observed. It instead suggests that the influence from different height scalings is relatively minor.

In the case of the neutrino probability, the atm. muon dominated spectrum (Fig. 8.1) does not show the same differences between data and MC as those observed in Fig. 9.7, left. For real data, all entries except the ones with the lowest score are populated more, indicating a shift towards

Table 9.2: Numbers characterizing the selections using the same cuts as for the real data track selection but different calibrations. For all selections, the re-weighting derived from the base calibration is used.

quantity \ calibration	base calibration	small miscalibration	large miscalibration
track efficiency (%)	18.9	17.4	10.2
shower efficiency (%)	9.9	8.0	5.1
atm. muon cont. (%)	0.7	1.0	7.2
random noise cont. (%)	0	0.6	0
overall contamination (%)	12.5	11.2	17.5
selected events	3,166	2,918	1,733
selected events $E_\nu < 20\,\text{GeV}$	1397	894	575

higher probabilities for muons. On the other hand, the neutrino-dominated regime is reflected by the shape of the larger miscalibration, where fewer events end up close to 1 and are instead distributed over the entire parameter range.

Another indication for a calibration not perfectly describing the real data is found for the direction uncertainty, which appears indeed larger for certain regions in real data (Fig. 8.18), as predicted from Fig. 9.1. However, for muons, the effect is not visible (Fig. 8.3).

9.3 Influence on selected data

The same neutrino selection can now be applied using the data processed with the three different calibrations for the complete ORCA6 lifetime. Not adjusting the cuts for any shifts in the spectra due to miscalibration corresponds to the realistic case of optimizing the selection on the MC data of nominal calibration and then transferring the cuts to some test data (real data or MC with compromised calibration). For that, the track selection for real data, developed in Chap. 8, is utilized. This way, further statements about the data/MC agreement can be included in the discussion.

9.3.1 Effects in MC selections

First, the numbers characterizing the selected set, listed in Tab. 9.2, are evaluated. From the discussion in the preceding section about the influence on the relevant spectra for the selection, a smaller efficiency for both neutrino topologies is expected. This is indeed observed as the track efficiency drops by 1.5% and 8.7% for smaller and larger miscalibration, respectively. The shower efficiency follows suit with decreases of 1.9% and 4.7%, corresponding
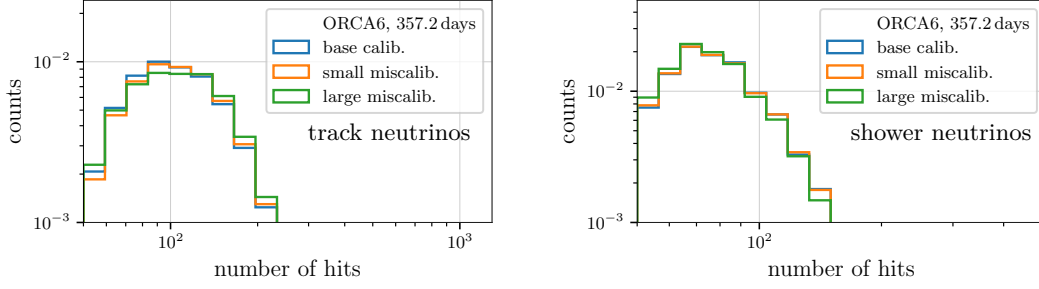
Figure 9.9: Spectra of the number of hits normalized to the same integral for track (left) and shower neutrinos (right) and different calibrations in colors. These are the quantities after cuts.

to similar relative changes. For neutrinos, the drastic modifications of the large miscalibration are visible in which only about 55% of the former selected events remain.

At the same time, the muon contamination increases significantly, while for the small miscalibration, the same atm. muon events pass the selection as for base, only the final neutrino set is smaller in comparison. With the more considerable statistical uncertainties on the random noise contribution, a single event is selected for the small miscalibration and then again discarded for the constellation of shifts to spectra introduced by the larger miscalibration.

For both the neutrino efficiency and the background contamination, the contributions to the changes do not stem from any one specific cut that displays a remarkably different behavior. Instead, the cuts on the energy-dependent uncertainty, neutrino probability and $\cos\theta$ contribute equally to the deterioration of the selection.

In addition to the mentioned decrease of the overall events, the reduction in the lower-energetic region is even more severe. With a fraction of 44% events with energies of smaller than 20 GeV for the base calibration, only 30% and 33% of interesting events for oscillation studies are left in the small and large miscalibration, respectively.

The fractions of selected shower events, which represent additional contamination to the track sample, stay roughly the same.

As far as the spectra of the GNN outputs after cuts are concerned, no further features can be identified other than those discussed before cuts. Except for the hits spectrum, which is the same for all calibrations before cuts, by definition. After cuts, the influence of the selection on that quantity can be evaluated. This is displayed in Fig. 9.9 for both neutrino topologies. The spectra for each calibration are normalized to visualize the relative differences better. Otherwise, the comparison would be dominated by the difference in rate only. Regarding the shape of the distribution for track neutrinos, the same features as those from the comparison with the
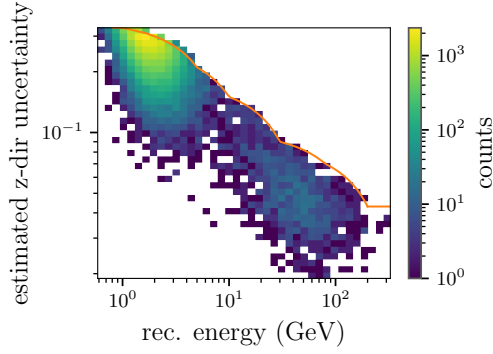
Figure 9.10: 2d representation of the uncertainty for the *z*-direction versus the reconstructed energy of real data. It shows a significant population of selected events for small energies and large uncertainties. The orange line indicates the cut used.

real data, Fig. 8.6, are observed. This suggests a possible impact from the selection, which behaves differently for MC and real data due to non-optimal calibration. However, the rates are comparable between MC and real data and would be predicted to be substantially lower if the calibration is the dominating influence. Also, the effect in the calibration study is less pronounced for the lower boundary of hits close to 50, while in the real data comparison, the difference continues to grow with smaller events.

For showers on the right side of Fig. 9.9, the influence is smaller and inverted. This effect is hardly noticeable for the real data, since the overall shower percentage is low.

In conclusion, the observed selection effect may only account for a minor contribution attributed to miscalibrations. Rather, other systematic differences between reality and simulation (see 8.1.2) appear to dominate the discrepancies. It is also possible that a combination of a miscalibration and other effects particularly magnify the importance of accurate hit information.

Nonetheless, the studied case of modifying the calibration is one approach to probe the networks for generalizability by generating differences between training data and test data and evaluate their behavior.

9.3.2 Effects on data/MC comparison

The processing assuming the modified calibrations is also carried out for real data. This allows for studying whether the influence is different on real data and MC. Under the hypothesis that both are similar, the same level of agreement (and discrepancies) in the data/MC comparison in spectra as observed thus far is expected. With the lower statistics of the selected sets of the miscalibrations, an even smaller reduced χ^2 is possible.

One major difference for the large miscalibration, apparent in all spectra, is an excess of selected events in real data. They can undoubtedly be identified with random noise, as they populate the regions of the output parameters that typically host the random noise contributions. One example is shown in Fig. 9.10, where the uncertainty of the *z*-direction is plotted versus the reconstructed energy. Entries orders of magnitudes larger than
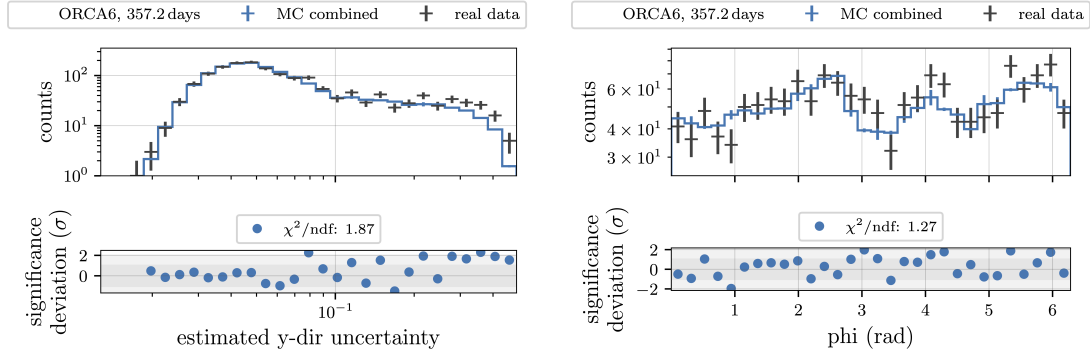
Figure 9.11: Spectra of reconstructed quantities after cuts for the large miscalibration. Displayed are the combined MC prediction (blue) and the real data (black). Left: Uncertainty for the y-direction component. Right: Reconstructed azimuth angle ϕ. Evaluation of the similarity in the bottom part like Fig. 8.4.

expected are found for low energies and high directional uncertainties (compare Fig. 8.5). This effect is not visible for the base and small miscalibration. The main reason for the random noise passing the selection is the signal/background classifier, which appears to be unable to identify these events reliably in real data. As was stated in Tab. 9.2, the selection efficiency for random noise in MC is still zero, also for the large miscalibration. One of the mentioned differences between simulations and reality leads to confusing the signal/background classifier, especially for random noise events.

To be able to continue the comparisons without being dominated by this effect, a cut on the number of triggered hits is applied for all three cases to remove random noise efficiently.

Considering the more extreme case of the large miscalibration, a few representative data/MC comparisons are discussed. First, the outputs from the direction reconstruction network exhibit the same level of good agreement for the neutrino-dominated regime as was observed for the base calibration. Two examples are shown in Fig. 9.11, one for the uncertainly of the y-direction on the left side and the azimuth angle on the right. Both distributions underwent significant changes (Fig. 9.1 and 9.6, left) when modifications were made to the calibration. Yet, the same effect is observed on real data causing the spectra to exhibit the same general shape. A smaller discrepancy is found for uncertainties larger than 0.2, where probably a few noise events survive the cuts for real data. The overall reduced χ^2 is calculated to 1.87, while for base and small miscalibration 1.93 and 1.72 is yielded (assuming also for these cases the added cut on the number of triggered hits). The irregular features for the azimuth angle are well modeled by the simulation with 1.27 χ^2/ndf, while 1.37 and 0.83 are assigned for base and small miscalibration, respectively.

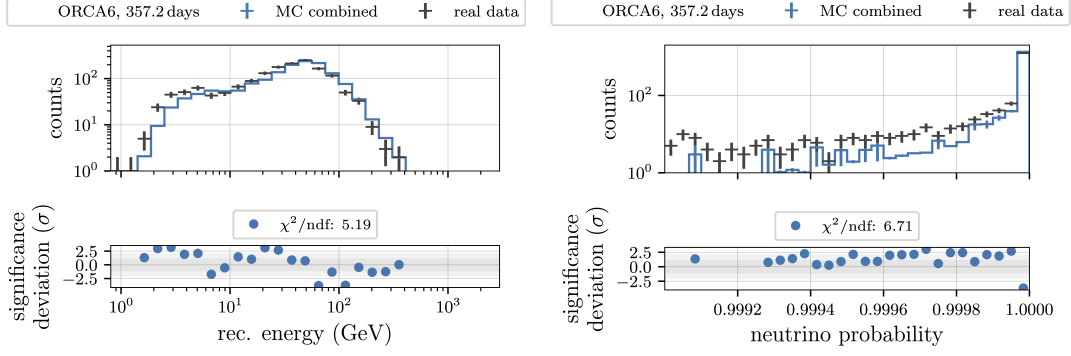These considerations lead to the conclusion that the direction reconstruc-

Figure 9.12: Data/MC comparison for the large miscalibration after cuts. Left: Reconstructed energy. Right: Neutrino probability. Evaluation of the similarity in the bottom part like Fig. 8.4.

tion network is robust against the differences in data and MC and changes in the calibration are reproduced for both in the same way.

Next, the reconstructed energy is plotted for the large miscalibration in Fig. 9.12, left. The simulation appears to be shifted to larger values equally for the complete range when inspecting the two distributions for data and MC. This suggests that this reconstruction is indeed affected differently by changing calibrations in data than it is for MC. A slight shift towards higher energies was observed in Fig. 9.4, which is now found to be weaker for real data. This serves as another indication that the calibration is one of the contributions to the overall data/MC discrepancy, as the need for the presented energy correction in Chap. 8.1.2 could stem from such miscalibration. Consequently, the reduced χ^2 of 5.19 is significant, while base and small miscalibration yield lower χ^2/ndf of 1.82 and 1.95, asserting a fitting energy correction in these cases.

Alongside the energy spectrum, the distribution of the neutrino probability for the large miscalibration is plotted on the right side of Fig. 9.12. With data and MC both changing as discussed for the right side of Fig. 9.7 (neutrinos are less peaked at 1), and considering there are overall lower statistics involved, the disagreement for the largest bin is now less significant. For the base calibration, the significance is 5σ, for the small miscalibration 6σ and for the large miscalibration, 3σ.

Finally, the track probability is shown in Fig. 9.13. In principle, the real data show a similar shift to larger values than is predicted by the MC. With the few real data events for track scores of smaller than 0.8, it is difficult to quantify. However, with the combined MC curve changing drastically in shape (as discussed for Fig. 9.8) and the real data only shifting additionally towards higher values, better agreement is found, as the reduced χ^2 is 9.12, compared to the 20.69 and 15.9 for base and small miscalibration.
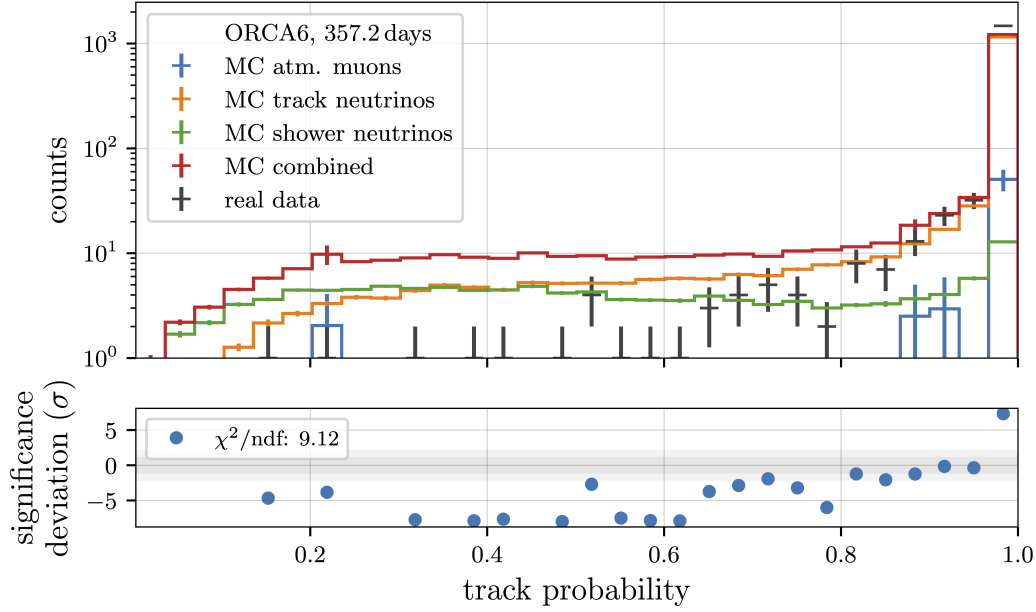
Figure 9.13: Track probability for the large miscalibration. The MC contributions (red) are resolved (colors) and compared to the real data (black). Evaluation of the similarity in the bottom part like Fig. 8.4.

9.4 Influence on oscillation analysis

The concluding remark in this chapter and the thesis as a whole is about quantifying the influence of miscalibration on the eventual physics analysis. To this end, the selections (pure track for real data from Chap. 8.1.1) without the additional cut on the number of triggered hits are utilized again, as the evaluations are done in MC only. The same procedure as in Chap. 7.2 is used, in which the results were obtained with respect to a true Asimov data set at $\theta_{23} = 0.5$ and $\Delta m^2_{31} = 1.95 \times 10^{-3} \text{eV}^2$. The 90% CL contour lines are drawn around that point in Fig. 9.14 for the three tested calibrations. The performance for the base calibration and the smaller miscalibration is very similar. The significance with which the NuFIT parameter set can be excluded is calculated to 3.5σ and 3.4σ, respectively. The more significant modifications in the larger miscalibration cause a severe decrease in oscillation sensitivity. The contour line enlarges similarly for the entire parameter space to almost include the NuFIT values within 90% CL. This translates to a 1.9σ significance for the rejection of that point.

In summary, the effects contributing to the decrease in sensitivity are the lower rate of selected events and the lower resolution of energy and direction. The potential need for additional cuts to restore data/MC agreement, such as the one on the number of triggered hits, would further diminish sensitivity.

Judging by the $\Delta\chi^2$, the difference between not doing any dedicated
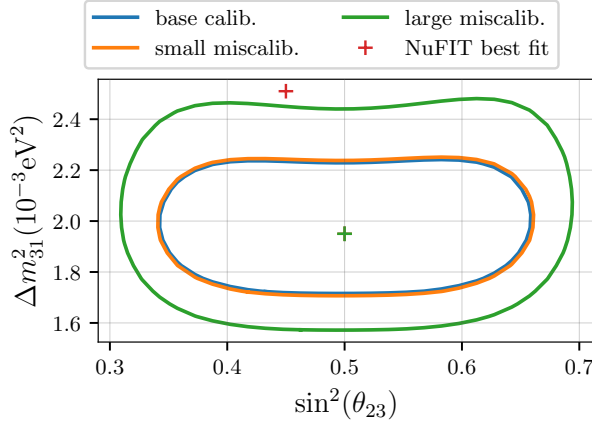
Figure 9.14: Contour lines for the 90% confidence level in Δm_{31}^2 and $\sin^2 \theta_{23}$ surrounding the chosen true value at $(0.5/1.95)$ for different calibrations (colors). As a reference, the NuFIT best fit value is added (red) [52].

calibrations at all (large miscalibration) and the currently used muon calibration developed in Chap. 4.4, the improvement to the oscillation sensitivity corresponds to a factor of 2.7 in effective lifetime or volume.

Summary and outlook

THIS thesis was dedicated to the improvement of the sensitivity of KM3NeT/ORCA to neutrino oscillations, an active field of research for the last decades. Even though properties like the mass, the exact oscillation parameters and the mass ordering are fundamental to describe neutrinos, they still have not been precisely measured. With one year of data acquired with an ORCA detector consisting of six detection units, the first contributions to solve these conundrums are made by the KM3NeT experiment. To help maximizing the physics potential that can be reached with these data, two important topics have been taken on in the scope of this PhD work: Ensuring a accurate detector calibration utilizing atmospheric muon data and improving the reconstruction resolution and neutrino selection efficiency by employing deep learning techniques to extract oscillation parameters.

The concept and the implementation details of a complete, muon-based calibration have been presented, along with consistency studies in a controlled MC environment. Searching for optimal time offset and geometry by evaluating the reconstruction quality revealed excellent resolution of well below 1 ns in time and about 5 cm in space when probing the method's reliability in simulations. Furthermore, height scalings of the DOMs on a DU can be determined with about 0.15% precision and the orientation to below one degree, comfortably meeting requirements.

For a realistic environment with several simultaneous unknowns, interplay in the determination of different quantities was studied before applying the procedure to real data. The recently developed (dynamic) acoustic positioning system allowed for comparisons to a static and a time-dependent geometry. The average distance in the xy-positions was found to be 12 cm and for the height scaling, both approaches predicted a mean stretching of around 1.5%. Excellent consistency was asserted for the dynamic case as well, in which the average difference between acoustics and muons for the x-position of a DU in a period exhibiting high sea currents was determined to be 5 cm. The significantly varying orientations of the DU were also resolved to 0.7°, well exceeding the specified accuracy.

Similarly, the inter-DU and inter-DOM time calibration generally

confirmed the existing values, constraining, for example, the per DOM offsets to ± 1 ns. Also here, consistency studies were conducted to predict the effect of systematic influences.

Starting with the deep learning part, the motivation for the use of graph neural networks was stated and their implementation detailed. Four different networks were employed to allow for a complete deep learning-based analysis, namely a signal/background classifier, a track/shower classifier, a direction and an energy reconstruction, including uncertainties.

Dependencies of the performance on energy and direction have been presented, affirming a powerful separability for tracks and showers (60% track efficiency at $\sim 2\%$ shower contamination) and a high neutrino efficiency for upgoing directions ($\sim 70\%$). Especially the background from random noise events could be strongly suppressed with a neutrino efficiency of 55% at 1% random noise contamination.

Comparing the direction resolution reached with the GNNs to the classical methods, 2° for track-like signatures and about 1° for shower-like events of improvement were achieved at 10 GeV. The uncertainty estimated on an event-by-event basis has proven to work well for the use as a quality cut.

A significant enhancement for the energy resolution of tracks was attained, as their relative error decreased from 60 to 30% at 20 GeV compared to the classical track reconstruction. The improvement over the classical shower reconstruction is determined to be around 6% for the same energy, and in both cases, fewer outliers of poorly reconstructed cases were observed with GNNs.

In addition to the increased performance, the processing times for data are significantly shorter, with a factor 8 compared to the classical track reconstruction.

Taking advantage of the outputs of the neural networks, sets of neutrinos were selected and their suitability for oscillation research was probed. Comparing to the selection based on classical reconstruction parameters, which exhibited a lower neutrino yield of one third of the selection based on the GNNs, the sensitivity to a specific combination of parameter sets is found to be 3.5 times higher.

Furthermore, the successful transfer of the deep learning algorithms to real data is crucial for the eventual use in analyses. Thorough studies evaluating the behavior in real data have been conducted, identifying several discrepancies. Possible causes have been discussed and their effects have been partly corrected to restore agreement between data and MC and allow for a preliminary oscillation analysis. The purely statistical result confirms the previously published preference for a smaller mass splitting $\Delta m_{31}^2 = 1.9 \text{eV}^2$ and favors a $\sin^2 \theta_{23}$ of 0.42. The analysis considering the full set of systematics is currently prepared and is expected to yield a higher significance of the tension to the world's current best fit.

The final chapter elaborates on the influence of miscalibration, quantify-

ing the decrease in resolution and neutrino selection efficiency (46% fewer events surviving for large miscalibration). The considerations were also extended to data/MC comparisons and the evaluation of the sensitivity to oscillations, asserting a drop from $3.5\,\sigma$ significance to $1.7\,\sigma$ for inaccurate calibrations.

Considering the improvements achieved with the GNNs developed in this thesis and the ever-increasing amount of data a growing ORCA detector will take, KM3NeT will soon be able to contribute significantly to the field of oscillation research. Next, it will be interesting to see indications for the octant of $\sin^2 \theta_{23}$ and tau appearance studies will be possible utilizing the particle identification methods.

In order to pursue these physics results, further improvements to the presented tools will be advantageous. The hyperparameters of the neural networks could be optimized more systematically, for which packages exist that accomplish this in an automated way. This will directly increase resolution and selection efficiency. Moreover, different adjustments to the training data set could be explored; an event selection filtering for only contained events may improve in particular the energy resolution.

Different approaches to the training that focus on ensuring better generalizability may be interesting, as is, for example, demonstrated in [156] where data drawn from different distributions is trained to yield comparable results. This could help with data/MC discrepancies as observed for the energy.

Inherently different selections (purely classical, purely deep learning) will allow for valuable cross-checks in analyses. For oscillations for example, the sets could be combined to reach an even higher sensitivity. Likewise, the set of overlapping events between both selections should exhibit a higher quality.

Glossary

- **ARCA**: Short for astroparticle research in the abyss; the Italian detector site.
- **atm. muon**: Short for atmospheric muon; a muon created in cosmic ray interactions in the upper atmosphere.
- **bjorken y**: Inelasticity of a neutrino interaction, defined as the ratio of energy going into the hadronic part and the neutrino energy.
- **CC**: Short for charged current; in a charged current interaction, weak hypercharge is exchanged.
- **contamination**: Fraction or percentage of an undesired class in the set of selected events.
- **cut**: A value for a quantity (like the energy) that is used to select events by.
- **DU**: Short for detection unit; the string the photo sensors are lined up on in the KM3NeT detectors.
- **DOM**: Short for digital optical module; the photo sensors housing several PMTs.
- **efficiency**: Percentage or fraction of the number of surviving and initial events for a cut.
- **floor**: Height of the DOM on a DU, referred to in integer numbers.
- **downgoing**: The direction coming from above the detector. Corresponding to a zenith angle of $\cos\theta < 0$.
- **GNN**: Short for graph neural network.
- **hit**: An incident of a detection of light of a PMT, characterized by the time the amplitude rises over a certain threshold, the time it stays above it and the PMT identifier.
- **hits spectrum**: Short for the number of hits spectrum; an histogram containing the number of snapshot hits per events.
- **HRV**: Short for high-rate veto; a veto excluding specific PMTs from the online data acquisition due to high rates, typically caused by bioluminescence.
- **htr**: Short for hit time residual; the difference between expected and observed arrival time of photons on a PMT.

- **IO**: Short for inverted ordering as one possibility for the neutrino mass ordering.
- **JGandalf**: Multi-stage classical reconstruction algorithm for tracks.
- **JShower**: Multi-stage classical reconstruction algorithm for showers.
- **LOM**: Short for Launcher of Optical Modules, the device detection units are deployed with.
- **MC**: Short for Monte Carlo simulations.
- **NC**: Short for neutral current; in a neutral current interaction, no weak hypercharge is exchanged.
- **ndf**: Short for number of degrees of freedom.
- **NO**: Short for normal ordering as one possibility for the neutrino mass ordering.
- **NuFIT**: Combined fit to world data in oscillation research, see `http://www.nu-fit.org`.
- **ORCA**: Short for oscillation research in the abyss; the French detector site.
- **ORCA6**: Refers to a 6-DU ORCA detector.
- **PMT**: Short for photomultiplier, the device for photon detection.
- **shower neutrinos**: Neutrino interactions in which either hadronic or electromagnetic showers are created. Namely caused by electron charged current, neutral current and tau charged current interactions without muon in the decay channel. Including the corresponding antineutrinos.
- **track neutrinos**: Neutrino interactions in which a muon track is created. Namely muon (anti) neutrino charged current and tau interactions with a muon in the decay. Including the corresponding antineutrinos.
- **triggered hit**: A hit that contributed to the triggering of an event, as opposed to snapshot hits, which are the remaining hits around the trigger time.
- **TT**: Short for transit time; the time it takes from the interaction of the photon in the photocathode of the PMT to form the electric signal.
- **TTS**: Short for the transit time spread.
- **upgoing**: The direction coming from below the detector. Corresponding to a zenith angle of $\cos\theta > 0$.
- **visible energy**: Energy of the hadronic shower for neutral current neutrino interactions. It is the same as the neutrino energy for charged current events.
- **z-direction**: The z-component of the direction of an event/particle. Used synonymously to $\cos\theta$.

Bibliography

[1] W. Pauli, *On the earlier and more recent history of the neutrino (1957)*, Neu- trino Physics, Cambridge Univ. Press.

[2] D. J. Griffiths, Introduction to elementary particles; 2nd rev. version, Physics textbook, Wiley, New York, NY, 2008. URL https://cds.cern.ch/record/111880

[3] C. L. Cowan, et al., *Detection of the free neutrino: a confirmation*, Science 124 (3212) (1956) 103–104. URL http://science.sciencemag.org/content/124/3212/103

[4] P. Zyla, others (Particle Data Group), *2020 review of particle physics*, to be published in Prog. Theor. Exp. Phys. 2020 083C01.

[5] M. K. Gaillard, P. D. Grannis, F. J. Sciulli, *The standard model of particle physics*, Reviews of Modern Physics 71 (2) (1999) S96–S111. URL http://dx.doi.org/10.1103/RevModPhys.71.S96

[6] S. Abachi, B. Abbott, M. Abolins, B. S. Acharya, I. Adam, D. L. Adams, M. Adams, S. Ahn, H. Aihara, J. Alitti, et al., *Observation of the top quark*, Physical Review Letters 74 (14) (1995) 2632–2637. URL http://dx.doi.org/10.1103/PhysRevLett.74.2632

[7] G. Aad, et al. (ATLAS Collaboration), *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B716 (2012) 1–29, Preprint 1207.7214.

[8] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. 13 (1964) 508–509. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.508

[9] G. Danby, J. R. Gaillard, K. A. Goulianos, L. M. Lederman, N. N. Mistry, M. S. Schwartz, J. Steinberger, *Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos*, Physical Review Letters 9 (1962) 36–44.

[10] K. Kodama, et al., *Observation of tau neutrino interactions*, Physics Letters B 504 (3) (2001) 218 – 224. URL http://www.sciencedirect.com/science/article/pii/S0370269301003070

[11] C. Spiering, *Towards high-energy neutrino astronomy*, The European Physical Journal H 37 (3) (2012) 515–565. URL https://doi.org/10.1140epjh/e2012-30014-2

[12] L. Evans, P. Bryant, *LHC machine*, Journal of Instrumentation 3 (08) (2008) S08001–S08001. URL https://doi.org/10.1088/1748-0221/3/08/s08001

[13] K. Murase, *Active galactic nuclei as high-energy neutrino sources*, Neutrino Astronomy (2017) 15–31. URL http://dx.doi.org/10.1142/9789814759410_0002

[14] M. G. Baring, *Diffusive shock acceleration: The Fermi mechanism*, in: Very high-energy phenomena in the universe. Proceedings, 32nd Rencontres de Moriond, Les Arcs, France, January 18-25, 1997, 1997, pp. 97–106, Preprint astro-ph/9711177. URL https://inspirehep.net/record/452485/files/C97-01-18_97-106.pdf

[15] U. Katz, C. Spiering, *High-energy neutrino astrophysics: Status and perspectives*, Progress in Particle and Nuclear Physics 67 (3) (2012) 651–704. URL https://doi.org/10.1016/j.ppnp.2011.12.001

[16] Aartsen, et al, *Observation of high-energy astrophysical neutrinos in three years of IceCube data*, Physical Review Letters 113 (10). URL http://dx.doi.org/10.1103/PhysRevLett.113.101101

[17] Aartsen, et al., *Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A*, Science 361 (6398), Preprint https://science.sciencemag.org/content/361/6398/eaat1378.full.pdf. URL https://science.sciencemag.org/content/361/6398/eaat1378

[18] R. Stein, S. v. Velzen, M. Kowalski, A. Franckowiak, S. Gezari, J. C. A. Miller-Jones, S. Frederick, I. Sfaradi, M. F. Bietenholz, A. Horesh, et al., *A tidal disruption event coincident with a high-energy neutrino*, Nature Astronomy 5 (5) (2021) 510–518. URL http://dx.doi.org/10.1038/s41550-020-01295-8

[19] B. P. Abbott, et. at, *Observation of gravitational waves from a binary black hole merger*, Physical Review Letters 116 (6). URL http://dx.doi.org/10.1103/PhysRevLett.116.061102

[20] B. P. Abbott, et al. (LIGO Scientific Collaboration and Virgo Collaboration Collaboration), *GW170817: Observation of gravitational waves from a binary neutron star inspiral*, Phys. Rev. Lett. 119 (2017) 161101. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.161101

[21] T. Kajita, *Atmospheric neutrinos and discovery of neutrino oscillations*, Proceedings of the Japan Academy. Series B, Physical and biological sciences 86 (2010) 303–21.

[22] M. Honda, M. S. Athar, T. Kajita, K. Kasahara, S. Midorikawa, *Reduction of the uncertainty in the atmospheric neutrino flux prediction below 1 GeV using accurately measured atmospheric muon flux*, Physical Review D 100 (12). URL http://dx.doi.org/10.1103/PhysRevD.100.123022

[23] J. P. Yáñez, A. Kouchner, *Measurement of atmospheric neutrino*

oscillations with very large volume neutrino telescopes, Advances in High Energy Physics 2015 (2015) 271968.

[24] M. Honda, T. Kajita, K. Kasahara, S. Midorikawa, *Improvement of low energy atmospheric neutrino flux calculation using the JAM nuclear interaction model*, Physical Review D 83 (12). URL http://dx.doi.org/10.1103/PhysRevD.83.123001

[25] G. D. Barr, S. Robbins, T. K. Gaisser, T. Stanev, *Uncertainties in atmospheric neutrino fluxes*, Physical Review D 74 (9). URL http://dx.doi.org/10.1103/PhysRevD.74.094009

[26] R. Patterson, *The NOvA experiment: Status and outlook*, Nuclear Physics B - Proceedings Supplements s 235–236 (2012) 151–157.

[27] K. Abe, N. Abgrall, H. Aihara, Y. Ajima, J. Albert, D. Allan, P.-A. Amaudruz, C. Andreopoulos, B. Andrieu, M. Anerella, et al., *The T2K experiment*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 659 (1) (2011) 106–135. URL http://dx.doi.org/10.1016/j.nima.2011.06.067

[28] A. Akindinov, et al., *Letter of interest for a neutrino beam from protvino to KM3NeT/ORCA*, The European Physical Journal C 79 (9). URL http://dx.doi.org/10.1140/epjc/s10052-019-7259-5

[29] H.-T. Janka, *Neutrino emission from supernovae*, Handbook of Supernovae (2017) 1575–1604. URL http://dx.doi.org/10.1007/978-3-319-21846-5_4

[30] S. Aiello, A. Albert, S. A. Garre, Z. Aly, A. Ambrosone, F. Ameli, M. Andre, G. Androulakis, M. Anghinolfi, M. Anguita, et al., *The KM3NeT potential for the next core-collapse supernova observation with neutrinos*, The European Physical Journal C 81 (5). URL http://dx.doi.org/10.1140/epjc/s10052-021-09187-5

[31] X. Qian, J.-C. Peng, *Physics with reactor neutrinos*, Reports on Progress in Physics 82 (3) (2019) 036201. URL http://dx.doi.org/10.1088/1361-6633/aae881

[32] B. Verberck, A. Taroni, *Nuclear fusion*, Nature Physics 12 (5) (2016) 383–383. URL https://doi.org/10.1038/nphys3771

[33] M. Maltoni, A. Yu. Smirnov, *Solar neutrinos and neutrino physics*, The European Physical Journal A 52 (4). URL http://dx.doi.org/10.1140/epja/i2016-16087-0

[34] C. Giunti, K. C. Wook, Fundamentals of Neutrino Physics and Astrophysics, Oxford Univ., Oxford, 2007.

[35] J. Hofestädt, *Measuring the neutrino mass hierarchy with the future KM3NeT/ORCA detector*, Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (2017). URL https://ecap.nat.fau.de/wp-content/uploads/2017/04/Dissertation_JHofestaedt.pdf

[36] J. P. Yáñez Garza, *Measurement of neutrino oscillations in atmospheric neutrinos with the IceCube DeepCore detector*, Ph.D. thesis,

Humboldt-Universität zu Berlin (2014).

[37] P. D. Group, P. A. Zyla, et al., *Review of Particle Physics*, Progress of Theoretical and Experimental Physics 2020 (8), Preprint https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf, 083C01. URL https://doi.org/10.1093/ptep/ptaa104

[38] Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino, Y. Itow, T. Kajita, J. Kameda, S. Kasuga, et al., *Evidence for oscillation of atmospheric neutrinos*, Physical Review Letters 81 (8) (1998) 1562–1567. URL http://dx.doi.org/10.1103/PhysRevLett.81.1562

[39] A. Bellerive, J. Klein, A. McDonald, A. Noble, A. Poon, *The Sudbury Neutrino Observatory*, Nuclear Physics B 908 (2016) 30–51. URL http://dx.doi.org/10.1016/j.nuclphysb.2016.04.035

[40] J. Wolf, et al., *The KATRIN neutrino mass experiment*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 623 (1) (2010) 442–444. URL https://doi.org/10.1016/j.nima.2010.03.030

[41] L. Wolfenstein, *Neutrino oscillations in matter*, Phys. Rev. D 17 (1978) 2369–2374. URL https://link.aps.org/doi/10.1103/PhysRevD.17.2369

[42] M. Gonzalez-Garcia, M. Maltoni, *Phenomenology with massive neutrinos*, Physics Reports 460 (1-3) (2008) 1–129. URL http://dx.doi.org/10.1016/j.physrep.2007.12.004

[43] T. Ohlsson, H. Snellman, *Three flavor neutrino oscillations in matter*, Journal of Mathematical Physics 41 (5) (2000) 2768–2788. URL http://dx.doi.org/10.1063/1.533270

[44] S. P. Mikheev, A. I. Smirnov, *Resonant amplification of ν oscillations in matter and solar-neutrino spectroscopy.*, Nuovo Cimento C Geophysics Space Physics C 9 (1986) 17–26.

[45] Q. R. Ahmad, R. C. Allen, T. C. Andersen, J. D.Anglin, J. C. Barton, E. W. Beier, M. Bercovitch, J. Bigu, S. D. Biller, R. A. Black, et al., *Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory*, Physical Review Letters 89 (1). URL http://dx.doi.org/10.1103/PhysRevLett.89.011301

[46] J. Bernabéu, S. Palomares-Ruiz, A. Pérez, S. Petcov, *The earth mantle-core effect in charge-asymmetries for atmospheric neutrino oscillations*, Physics Letters B 531 (1-2) (2002) 90–98. URL http://dx.doi.org/10.1016/S0370-2693(02)01358-8

[47] M. FREUND, T. OHLSSON, *Matter enhanced neutrino oscillations with a realistic earth density profile*, Modern Physics Letters A 15 (13) (2000) 867–874. URL http://dx.doi.org/10.1142/S0217732300000864

[48] S. Adrián-Martínez, M. Ageron, F. Aharonian, S. Aiello, A. Albert, F. Ameli, E. Anassontzis, M. Andre, G. Androulakis, M. Anghinolfi, et al., *Letter of intent for KM3NeT 2.0*, Journal of Physics G: Nuclear and Particle Physics 43 (8) (2016) 084001. URL http://dx.doi.org/10.1088/0954-3899/43/8/084001

[49] W. Winter, *Neutrino mass hierarchy: Theory and phenomenology*, AIP Conference Proceedings 1666 (1) (2015) 120001, Preprint https://aip.scitation.org/doi/pdf/10.1063/1.4915577. URL https://aip.scitation.org/doi/abs/10.1063/1.4915577

[50] V. Barger, P. Huber, D. Marfatia, *Supernova neutrinos can tell us the neutrino mass hierarchy independently of flux models*, Physics Letters B 617 (3-4) (2005) 167–173. URL http://dx.doi.org/10.1016/j.physletb.2005.05.017

[51] D. Franco, C. Jollet, A. Kouchner, V. Kulikovskiy, A. Meregaglia, S. Perasso, T. Pradier, A. Tonazzo, V. Van Elewyck, *Mass hierarchy discrimination with atmospheric neutrinos in large volume ice/water Cherenkov detectors*, Journal of High Energy Physics 2013 (4). URL http://dx.doi.org/10.1007/JHEP04(2013)008

[52] *nu-fit.org*, http://www.nu-fit.org, accessed: 2021-11-25.

[53] I. Esteban, M. Gonzalez-Garcia, M. Maltoni, T. Schwetz, A. Zhou, *The fate of hints: updated global analysis of three-flavor neutrino oscillations*, Journal of High Energy Physics 2020 (9). URL http://dx.doi.org/10.1007/JHEP09(2020)178

[54] M. Gonzalez-Garcia, M. Maltoni, T. Schwetz, *Global analyses of neutrino oscillation experiments*, Nuclear Physics B 908 (2016) 199–217. URL http://dx.doi.org/10.1016/j.nuclphysb.2016.02.033

[55] Dunne, Patrick, *Planetary talk at neutrino 2020: Latest Neutrino Oscillation Results from T2K*, URL: https://doi.org/10.5281/zenodo.3959558 (July 2020).

[56] M. A. Acero, et al. (NOvA Collaboration), *An Improved Measurement of Neutrino Oscillation Parameters by the NOvA Experiment*, preprint Preprint 2108.08219.

[57] Y. Nakajima, *SuperKamiokande*, Talk given at XXIX International Conference on Neutrino Physics and Astrophysics (July 2020).

[58] S. Aiello, et. al, *Determining the neutrino mass ordering and oscillation parameters with KM3NeT/ORCA*, The European Physical Journal C 82 (1). URL http://dx.doi.org/10.1140/epjc/s10052-021-09893-0

[59] A. Diaz, C. Argüelles, G. Collin, J. Conrad, M. Shaevitz, *Where are we with light sterile neutrinos?*, Physics Reports 884 (2020) 1–59. URL http://dx.doi.org/10.1016/j.physrep.2020.08.005

[60] S. Aiello, et al., *Sensitivity to light sterile neutrino mixing parameters with KM3NeT/ORCA*, Journal of High Energy Physics 2021 (10). URL https://doi.org/10.1007/jhep10(2021)180

[61] J. A. Formaggio, G. P. Zeller, *From eV to EeV: Neutrino cross sections across energy scales*, Reviews of Modern Physics 84 (3) (2012) 1307–1341. URL http://dx.doi.org/10.1103/RevModPhys.84.1307

[62] B. Cleveland, T. Daily, R. Davis, J. Distel, K. Lande, C. Lee, P. Wildenhain, a. Ullman, *Measurement of the solar electron neutrino flux with the Homestake chlorine detector*, The Astrophysical Journal 496 (2009) 505.

[63] J. Abdurashitov, T. Bowles, B. Cleveland, S. Elliott, V. Gavrin, S. Girin, V. Gorbachev, P. Gurkina, T. Ibragimova, A. Kalikhov, N. Khairnasov, T. Knodel, I. Mirmov, J. Nico, A. Shikhin, W. Teasdale, E. Veretenkin, V. Vermul, J. Wilkerson, V. Yants, G. Zatsepin, *Measurement of the solar neutrino capture rate in Sage*, Nuclear Physics B - Proceedings Supplements 118 (2003) 39–46, proceedings of the XXth International Conference on Neutrino Physics and Astrophysics. URL https://www.sciencedirect.com/science/article/pii/S0920563203013021

[64] W. Hampel, et al. (GALLEX Collaboration), *GALLEX solar neutrino observations: Results for GALLEX IV*, Phys. Lett. B 447 (1999) 127–133.

[65] M. Altmann, M. Balata, P. Belli, E. Bellotti, R. Bernabei, E. Burkert, C. Cattadori, G. Cerichelli, M. Chiarini, M. Cribier, et al., *GNO solar neutrino observations: results for GNO I*, Physics Letters B 490 (1-2) (2000) 16–26. URL http://dx.doi.org/10.1016/S0370-2693(00)00915-1

[66] P. A. Cherenkov, *Visible emission of clean liquids by action of γ radiation*, Doklady Akademii Nauk SSSR 2 (1934) 451+. URL http://ufn.ru/en/articles/2007/4/g/

[67] T. Gluesenkamp, *On the detection of subrelativistic magnetic monopoles with the IceCube Neutrino Observatory*, Master's thesis, Rheinisch-Westfälischen Technischen Hochschule Aachen (10 2010).

[68] K. Eguchi, S. Enomoto, K. Furuno, J. Goldman, H. Hanada, H. Ikeda, K. Ikeda, K. Inoue, K. Ishihara, W. Itoh, et al., *First results from KamLAND: Evidence for reactor antineutrino disappearance*, Physical Review Letters 90 (2). URL http://dx.doi.org/10.1103/PhysRevLett.90.021802

[69] G. Alimonti, C. Arpesella, H. Back, M. Balata, D. Bartolomei, A. de Bellefon, G. Bellini, J. Benziger, A. Bevilacqua, D. Bondi, et al., *The Borexino detector at the Laboratori Nazionali del Gran Sasso*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 600 (3) (2009) 568–593. URL http://dx.doi.org/10.1016/j.nima.2008.11.076

[70] J. Tiffenberg, *UHE neutrino searches with the Pierre Auger*

Observatory, http://users.ictp.it/~smr2246/monday/tiffenberg-NUSKY.pdf, 2011.

[71] A. Margiotta, *The ANTARES detector*, International Journal of Modern Physics A 20 (2009) 6965–6967.

[72] M. G. Aartsen, et al. (IceCube Collaboration), *The IceCube Neutrino Observatory: Instrumentation and Online Systems*, JINST 12 (03) (2017) P03012, Preprint 1612.05093.

[73] B. Herold, *Study of k40-induced rates for a KM3NeT design option with multi-PMT optical modules*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 626-627 (2011) S234–S236. URL https://doi.org/10.1016/j.nima.2010.04.137

[74] S. H. Haddock, M. A. Moline, J. F. Case, *Bioluminescence in the sea*, Annual Review of Marine Science 2 (1) (2010) 443–493, Preprint https://doi.org/10.1146/annurev-marine-120308-081028, pMID: 21141672. URL https://doi.org/10.1146/annurev-marine-120308-081028

[75] M. Ageron, S. Aiello, et al., *Dependence of atmospheric muon flux on seawater depth measured with the first KM3NeT detection units*, The European Physical Journal C 80 (2). URL https://doi.org/10.1140/epjc/s10052-020-7629-z

[76] G. Carminati, M. Bazzotti, A. Margiotta, M. Spurio, *Atmospheric MUons from PArametric formulas: a fast GEnerator for neutrino telescopes (MUPAGE)*, Computer Physics Communications 179 (12) (2008) 915–923. URL http://dx.doi.org/10.1016/j.cpc.2008.07.014

[77] S. Aiello, A. Albert, S. A. Garre, Z. Aly, et al., *gSeaGen: The KM3NeT GENIE-based code for neutrino telescopes*, Computer Physics Communications 256 (2020) 107477. URL http://www.sciencedirect.com/science/article/pii/S0010465520302241

[78] C. Andreopoulos, A. Bell, D. Bhattacharya, F. Cavanna, J. Dobson, S. Dytman, H. Gallagher, P. Guzowski, R. Hatcher, P. Kehayias, et al., *The GENIE neutrino Monte Carlo generator*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 614 (1) (2010) 87–104. URL http://dx.doi.org/10.1016/j.nima.2009.12.009

[79] S. Agostinelli, et al., *Geant4 - a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (3) (2003) 250–303. URL https://www.sciencedirect.com/science/article/pii/S0168900203013688

[80] J. Evans, *The MINOS experiment: Results and prospects*, Advances in High Energy Physics 2013 (2013) 1–18.

[81] S. Fukuda, et al., *The Super-Kamiokande detector*, Nuclear In-

struments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 501 (2) (2003) 418–462. URL https://www.sciencedirect.com/science/article/pii/S016890020300425X

[82] T. Kajita, E. Kearns, M. Shiozawa, *Establishing atmospheric neutrino oscillations with Super-Kamiokande*, Nuclear Physics B 908.

[83] Y. Ashie, et al. (Super-Kamiokande Collaboration), *Measurement of atmospheric neutrino oscillation parameters by Super-Kamiokande I*, Phys. Rev. D 71 (2005) 112005. URL https://link.aps.org/doi/10.1103/PhysRevD.71.112005

[84] R. Abbasi, Y. Abdou, T. Abu-Zayyad, M. Ackermann, J. Adams, J. Aguilar, M. Ahlers, M. Allen, D. Altmann, K. Andeen, et al., *The design and performance of IceCube DeepCore*, Astroparticle Physics 35 (10) (2012) 615–624. URL http://dx.doi.org/10.1016/j.astropartphys.2012.01.004

[85] M. G. Aartsen, M. Ackermann, J. Adams, J. A. Aguilar, M. Ahlers, M. Ahrens, I. Al Samarai, D. Altmann, K. Andeen, T. Anderson, et al., *Measurement of atmospheric neutrino oscillations at 6–56 GeV with IceCube DeepCore*, Physical Review Letters 120 (7). URL http://dx.doi.org/10.1103/PhysRevLett.120.071801

[86] T. Ehrhardt, *IceCube/DeepCore results and PINGU*, in: Proceedings of The 19th International Workshop on Neutrinos from Accelerators NUFACT2017 — PoS(NuFact2017), Sissa Medialab, 2017, pp. 100–106. URL https://doi.org/10.22323/1.295.0060

[87] A. Ishihara (IceCube Collaboration), *The IceCube Upgrade - Design and Science Goals*, PoS ICRC2019 (2021) 1031, Preprint 1908.09441.

[88] *A table of frequently used radioisotopes*, https://cds.cern.ch/record/1309915/files/978-3-642-02586-0_BookBackMatter.pdf, accessed: 2021-10-15.

[89] K. Melis (on behalf of the KM3NeT collaboration), *In-Situ Calibration of KM3NeT*, 35th ICRC, Busan, South Korea (2017) PoS(ICRC2017)1059. URL https://pos.sissa.it/301/1059/pdf

[90] D. Real, D. Calvo, *Nanobeacon: A time calibration device for KM3NeT*, EPJ Web of Conferences 207 (2019) 07002. URL https://doi.org/10.1051/epjconf/201920707002

[91] A. Sánchez Losa, personal communication.

[92] V. Bertin, personal communication.

[93] M. de Jong, personal communication.

[94] E. Berbee, personal communication.

[95] S. Aiello, et al., *Deep-sea deployment of the KM3NeT neutrino telescope detection units by self-unrolling*, arXiv e-prints (2020) arXiv:2007.16090 Preprint 2007.16090.

[96] R. Le Breton, M. Billault, C. Boutonnet, C. Champion, S. Colonges, A. Cosquer, A. Creusot, S. Henry, A. Ilioni, P. Keller, et al., *The*

calibration units of KM3NeT, Journal of Instrumentation 16 (09) (2021) C09004. URL http://dx.doi.org/10.1088/1748-0221/16/09/C09004

[97] B. O. Fearraigh, *Track reconstruction in KM3NeT*, internal note (2021). URL https://common.pages.km3net.de/jpp/TrackReconstruction.PDF

[98] M. de Jong, *The probability density function of the arrival time of light*, internal note (2021). URL http://pi1139.physik.uni-erlangen.de/documentation/Jpp/latest/JPDF.PDF

[99] M. J. D. Powell, *An efficient method for finding the minimum of a function of several variables without calculating derivatives*, Comput. J. 7 (1964) 155–162.

[100] M. de Jong, *Multi-dimensional interpolations in C++*, arXiv e-prints (2019) arXiv:1907.02597 Preprint 1907.02597.

[101] D. W. Marquardt, *An algorithm for least-squares estimation of nonlinear parameters*, Journal of the Society for Industrial and Applied Mathematics 11 (2) (1963) 431–441. URL http://www.jstor.org/stable/2098941

[102] M. de Jong, *Detector calibration (from a jpp perspective)*, internal presentation, KM3NeT collaboration meeting Caserta (Oct 2018). URL https://indico.cern.ch/event/748460/contributions/3095697/attachments/1736979/2811838/Detector-calibration.pdf

[103] T. Stellhorn, *Studien zur Optimierung der Detektorkalibration mit atmosphärischen Myonen bei KM3NeT/ORCA unter Verwendung einzelner vertikaler Detektoreinheiten*, Bachelor's thesis, Uni Münster (2020).

[104] W. Achtermann, *Studien zur Myon Rekonstruktion bei KM3NeT ORCA im Hinblick auf die Optimierung der Detektorkalibrierung mit atmosphärischen Myonen*, Bachelor's thesis, Uni Münster (2020).

[105] *A living review of machine learning for particle physics*, https://github.com/iml-wg/HEPML-LivingReview, accessed: 2021-12-02.

[106] S. Reck, D. Guderian, G. Vermariën, A. Domi, *Graph neural networks for reconstruction and classification in KM3NeT*, Journal of Instrumentation 16 (10) (2021) C10011. URL http://dx.doi.org/10.1088/1748-0221/16/10/C10011

[107] S. Aiello, A. Albert, S. A. Garre, Z. Aly, F. Ameli, M. Andre, G. Androulakis, M. Anghinolfi, M. Anguita, G. Anton, et al., *Event reconstruction for KM3NeT/ORCA using convolutional neural networks*, Journal of Instrumentation 15 (10) (2020) P10005–P10005. URL http://dx.doi.org/10.1088/1748-0221/15/10/P10005

[108] I. J. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016, http://www.deeplearningbook.org.

[109] M. A. Nielsen, Neural Networks and Deep Learning, Determination

Press, 2015.

[110] J. Duarte, J.-R. Vlimant, *Graph Neural Networks for Particle Tracking and Reconstruction*, arXiv e-prints Preprint 2012.01249.

[111] A. Burkov, The Hundred-Page Machine Learning Book, Andriy Burkov, 2019. URL https://books.google.de/books?id=0jbxwQEACAAJ

[112] S. C. Kleene, Representation of Events in Nerve Nets and Finite Automata, Princeton University Press, 2016. URL https://doi.org/10.1515/9781400882618-002

[113] D. O. Hebb, The organization of behavior: A neuropsychological theory, Wiley, New York, 1949.

[114] B. Farley, W. Clark, *Simulation of self-organizing systems by digital computer*, Transactions of the IRE Professional Group on Information Theory 4 (4) (1954) 76–84.

[115] F. Rosenblatt, *The perceptron: a probabilistic model for information storage and organization in the brain.*, Psychological review 65 6 (1958) 386–408.

[116] F. Psihas, M. Groh, C. Tunnell, K. Warburton, *A review on machine learning for neutrino experiments*, International Journal of Modern Physics A 35 (33) (2020) 2043005. URL http://dx.doi.org/10.1142/S0217751X20430058

[117] A. Sherstinsky, *Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network*, Physica D: Nonlinear Phenomena 404 (2020) 132306. URL http://dx.doi.org/10.1016/j.physd.2019.132306

[118] D. Bank, N. Koenigstein, R. Giryes, *Autoencoders*, arXiv e-prints (2020) arXiv:2003.05991 Preprint 2003.05991.

[119] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Learning representations by back-propagating errors*, Nature 323 (1986) 533–536.

[120] D. P. Kingma, J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv e-prints (2014) arXiv:1412.6980 Preprint 1412.6980.

[121] K. He, X. Zhang, S. Ren, J. Sun, *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*, arXiv e-prints (2015) arXiv:1502.01852 Preprint 1502.01852.

[122] S. Ioffe, C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, arXiv e-prints (2015) arXiv:1502.03167 Preprint 1502.03167.

[123] M. Gori, G. Monfardini, F. Scarselli, *A new model for learning in graph domains*, Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005. 2 (2005) 729–734 vol. 2.

[124] Y. Li, D. Tarlow, M. Brockschmidt, R. Zemel, *Gated Graph Sequence Neural Networks*, arXiv e-prints (2015) arXiv:1511.05493 Preprint 1511.05493.

[125] J. Shlomi, P. Battaglia, J.-R. Vlimant, *Graph neural networks*

in particle physics, Machine Learning: Science and Technology 2 (2) (2021) 021001. URL http://dx.doi.org/10.1088/2632-2153/abbf9a

[126] P. W. Battaglia, et al., *Relational inductive biases, deep learning, and graph networks*, arXiv e-prints (2018) arXiv:1806.01261 Preprint 1806.01261.

[127] G. Yehudai, E. Fetaya, E. Meirom, G. Chechik, H. Maron, *From Local Structures to Size Generalization in Graph Neural Networks*, arXiv e-prints (2020) arXiv:2010.08853 Preprint 2010.08853.

[128] Z. Li, W. Yang, S. Peng, F. Liu, *A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects*, arXiv e-prints (2020) arXiv:2004.02806 Preprint 2004.02806.

[129] K. Albertsson, et al., *Machine Learning in High Energy Physics Community White Paper*, J. Phys. Conf. Ser. 1085 (2) (2018) 022008, Preprint 1807.02876.

[130] H. Qu, L. Gouskos, *Jet tagging via particle clouds*, Physical Review D 101 (5). URL http://dx.doi.org/10.1103/PhysRevD.101.056019

[131] C. M. Bishop, *Mixture density networks*, technical Report (1994). URL https://publications.aston.ac.uk/id/eprint/373/

[132] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, et. al, *TensorFlow: Large-scale machine learning on heterogeneous systems*, software available from tensorflow.org (2015). URL https://www.tensorflow.org/

[133] S. Reck, T. Eberl, U. Katz, *Muon bundle reconstruction with KM3NeT/ORCA using graph convolutional networks*, PoS ICRC2021 (2021) 1048.

[134] M. Honda, M. S. Athar, T. Kajita, K. Kasahara, S. Midorikawa, *Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model*, Phys. Rev. D 92 (2015) 023004. URL https://link.aps.org/doi/10.1103/PhysRevD.92.023004

[135] A. Domi, *Shower Reconstruction and Sterile Neutrino Analysis with KM3NeT/ORCA and ANTARES*, Ph.D. thesis, Aix-Marseille U., Genoa U. (2019).

[136] M. Aartsen, et. al, *Neutrino oscillation studies with IceCube-DeepCore*, Nuclear Physics B 908 (2016) 161–177. URL https://www.sciencedirect.com/science/article/pii/S0550321316300141

[137] S. Adrián-Martínez, et. al, *Intrinsic limits on resolutions in muon and electron neutrino charged-current events in the KM3NeT/ORCA detector*, Journal of High Energy Physics 2017.

[138] M. Moser, *Sensitivity studies on tau neutrino appearance with KM3NeT/ORCA using deep learning techniques*, doctoral thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) (2020).

[139] L. Nauta, et al. (KM3NeT Collaboration), *First neutrino oscillation measurement in KM3NeT/ORCA*, PoS ICRC2021 (2021) 1123.

[140] J. Hoey, *The Two-Way Likelihood Ratio (G) Test and Comparison to Two-Way Chi Squared Test*, arXiv e-prints (2012) arXiv:1206.4881 Preprint 1206.4881.

[141] K. Pearson, *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50 (302) (1900) 157–175, Preprint https://doi.org/10.1080/14786440009463897. URL https://doi.org/10.1080/14786440009463897

[142] P. D. Group, et. al, *Review of Particle Physics, Chapter Statistics*, Progress of Theoretical and Experimental Physics 2020 (8), Preprint https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf, 083C01. URL https://pdg.lbl.gov/2020/reviews/rpp2020-rev-statistics.pdf

[143] D. Stavropoulos, V. Pestel, Z. Aly, E. Tzamariudaki, C. Markou, *Atmospheric neutrinos from the first KM3NeT/ORCA data and prospects for measuring the atmospheric neutrino flux*, PoS ICRC2021 (2021) 1125.

[144] J. Manczak, N. R. Khan-Chowdhury, J. J. Hernandez-Rey, S. Navas Concha, *Neutrino non-standard interactions with the KM3NeT/ORCA detector*, PoS ICRC2021 (2021) 1165.

[145] J. Coelho, personal communication.

[146] P. Sinervo, *Definition and treatment of systematic uncertainties in high energy physics and astrophysics*, Statistical Problems in Particle Physics, Astrophysics, and Cosmology.

[147] Z. Aly, personal communication.

[148] A. G. Wright, The Photomultiplier Handbook, Oxford University Press, 2017.

[149] H. Drucker, C. Cortes, *Boosting decision trees*, in: NIPS 1995, Vol. 8, 1995, pp. 479–485.

[150] J. Aguilar, A. Albert, P. Amram, M. Anghinolfi, G. Anton, S. Anvar, F. Ardellier-Desages, E. Aslanides, J.-J. Aubert, R. Azoulay, et al., *Transmission of light in deep sea water at the site of the ANTARES neutrino telescope*, Astroparticle Physics 23 (1) (2005) 131–155. URL http://dx.doi.org/10.1016/j.astropartphys.2004.11.006

[151] M. de Jong, discussion in git issue. URL https://git.km3net.de/working_groups/simulations/-/issues/39#note_24158

[152] T. Golan, C. Juszczak, J. T. Sobczyk, *Final State Interactions Effects in Neutrino-Nucleus Interactions*, Phys. Rev. C 86 (2012) 015505, Preprint 1202.4197.

[153] M. Unland Elorrieta, R. Busse, L. Classen, A. Kappes, *Homogeneity of the photocathode in the Hamamatsu R15458-02 photomultiplier tube*, Journal of Instrumentation 16 (11) (2021) P11038. URL http:

//dx.doi.org/10.1088/1748-0221/16/11/P11038

[154] H. Moritz, *Geodetic reference system 1980*, Journal of Geodesy 74 (1) (2000) 128–133. URL https://doi.org/10.1007/s001900050278

[155] A. Kurakin, I. Goodfellow, S. Bengio, *Adversarial examples in the physical world*, arXiv e-prints (2016) arXiv:1607.02533 Preprint 1607.02533.

[156] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, *Domain-Adversarial Training of Neural Networks*, arXiv e-prints (2015) arXiv:1505.07818 Preprint 1505.07818.

Acknowledgements

I would like to thank Prof. Dr. Alexander Kappes very much for giving me the opportunity to do my PhD in his group! I am thankful for the trust he put in me to do what I thought helped most the KM3NeT experiment. He was always open to my problems and questions and we could always figure out a way to solve them.

Furthermore, I would like to thank Associate Prof. Dr. Dorothea Samtleben, who not only gave me the first introduction to KM3NeT, was so kind to take over the duties of second examiner, but also accompanied me throughout my whole time as PhD. From that, many fruitful studies and discussions emerged that did not only help me but also the experiment progress.

Then, there are additional people in KM3NeT that helped me significantly: Stefan Reck with the implementation of the deep learning algorithms and Michael Moser for laying the groundwork and introducing me to that topic. Tamás Gal developed various software pieces that I was able to profit from directly. Maarten de Jong, who also developed software I was utilizing heavily and who responded quickly to my numerous questions. Jannik Hofestädt, whom I consulted about pretty much every aspect of the experiment. Luigi Fusco, who produced additional training data for me and with whom I did a lot of MC testing. And João Coelho, who helped me a lot with the oscillation studies, especially with statistics questions. Further contributions came from Zineb Aly, Agustín Sánchez Losa, Valentin Pestel, Alba Domi and Lodewijk Nauta.

Next, a thank you goes out to the local group in Münster, who created a friendly and jovial environment I was delighted to spend a few years of my life in. In particular my long standing PhD colleagues Raffiela Busse, Markus Dittmer, Cristian Lozano, Berit Schlüter and Martin Unland, as well as our post doc Dr. Lew Classen. Together, we also had the privilege to spend many unforgettable moments outside work.

Finally, I guess I would like to thank my parents, as without them I would have never been born - probably.

Lebenslauf

Persönliche Daten

VORNAME / NACHNAME	Daniel Guderian
ADRESSE	Steinfurter Straße 14, 48149 Münster
TEL.	0176 96583388
E-MAIL	daniel.guderian@uni-muenster.de
GEBURTSDATUM	26.10.1989
GEBURTSORT	Ahlen

Schule	Gymnasium Laurentianum Warendorf
8/2000 - 04/2010	Abitur, Note 2.5

Zivildienst	Amt für Grünflächen und Umweltschutz Münster
8/2000 - 04/2010	

Studium	an der Westfälischen-Wilhelms-Universität Münster
10/2010 - 09/2011	Geowissenschaften
10/2011 - 03/2015	Physik (Abschluss: Bachelor, Note 2.4)
04/2015 - 01/2018	Physik (Abschluss: Master, Note 1.5)
02/2018 - 05/2022	Physik Promotion am Institut für Kernphysik, AG Kappes