Bachelor's Thesis in Physics

Studies on systematic influences of the detector calibration on the reconstruction performance of Graph Neural Networks in KM3NeT

Submitted by

Lukas Hennig

April 16, 2021

Erlangen Centre for Astroparticle Physics
Department of Physics
Friedrich-Alexander University Erlangen-Nürnberg

First reviewer: PD Dr. Thomas Eberl
Second reviewer: Prof. Dr. Uli Katz

Abstract

Graph Neural Networks (GNNs) showed promising results in reconstructing physical properties of neutrino events detected by the KM3NeT detector. This thesis aims at testing how robust the predictions of these GNNs are in the context of direction reconstruction if the input is subject to systematic miscalibrations. For this, input data generated by a Monte Carlo simulation, which simulated typical events and the traces they leave in the KM3NeT detector, were used. Information about the simulated particles causing the events, especially the "true" directions of the particles, was also provided by the simulation. Several different types of miscalibrations were applied to this input data, and the predictions of the GNN were compared to the true directions by calculating the opening angle between the predicted and the true direction. A short introduction to neutrino physics, the KM3NeT detector, Deep Learning, and GNNs is presented at the beginning of this thesis. The results are compactly summarized in the last chapter.

Contents

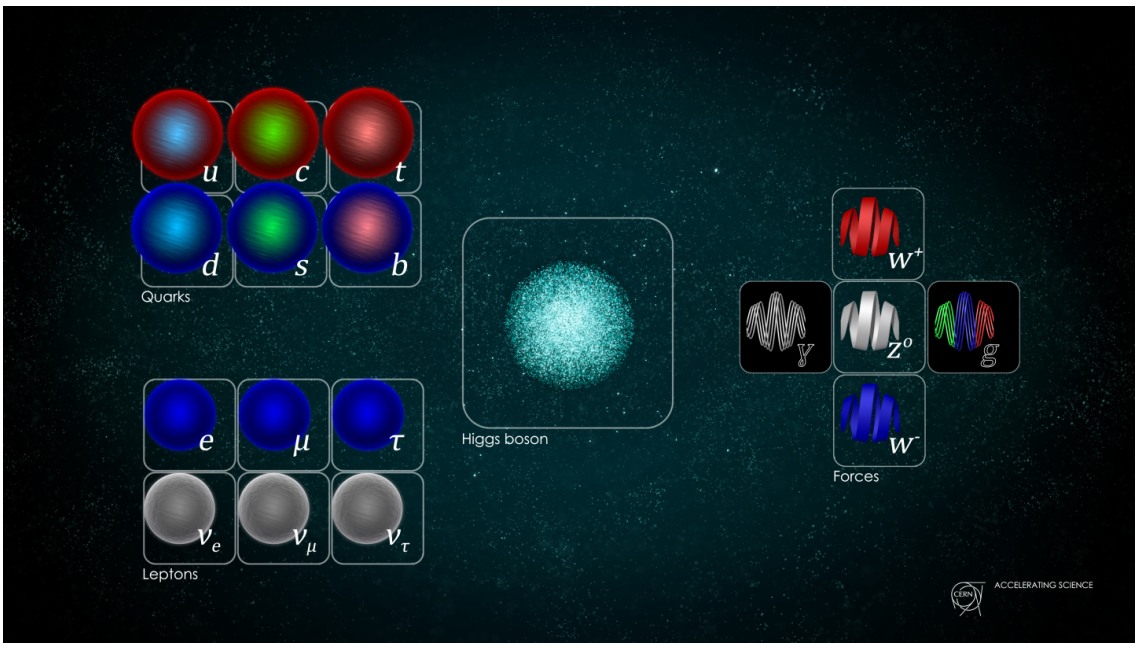

Figure 1: All the particles and the carriers of the fundamental forces in the Standard Model of particle physics [12].

1 Introduction

1.1 Neutrino physics

1.1.1 Why neutrino physics?

Neutrinos are elementary particles that gave rise to a lot of interesting questions in the recent past. In the "Standard Model of particle physics" (see figure 1), the currently most accepted theory for describing the building blocks of our universe and three of the four fundamental forces, the neutrino is included as a lepton appearing in three different flavors called ν_e, ν_μ and ν_τ. Around 20 years ago, it was discovered that a neutrino of one flavor can turn into a neutrino of another flavor - a phenomenon called neutrino oscillations. The 2015 Noble Prize in physics was awarded to Takaaki Kajita and Arthur B. McDonald for the experimental discovery of this effect [3]. This phenomenon implies that neutrinos must have a mass, which is very interesting because neutrinos are included as massless in the Standard Model. Thus neutrino physics is a promising area in physics that may guide the way to an improved version of the Standard Model.

1.1.2 The mass hierarchy problem

To explain the so-called neutrino mass hierarchy problem, we have to explain the implications of neutrino oscillations a little bit further. Quantum mechanically spoken, we call ν_e, ν_μ and ν_τ flavor eigenstates of the neutrino. A neutrino will be in a state obtained by a superposition of those three eigenstates, which means that the three flavor eigenstates build a basis of the Hilbert space of possible neutrino states. Therefore the state $|\psi\rangle$ of a neutrino can be described as

$$|\psi\rangle = c_e \, |\nu_e\rangle + c_\mu \, |\nu_\mu\rangle + c_\tau \, |\nu_\tau\rangle \tag{1}$$

where the constants c_e, c_μ and c_τ are normalized complex numbers, i.e., $\langle\psi|\psi\rangle = 1$. However, the discovery of neutrino oscillations shows that a neutrino in one flavor eigenstate (e.g., after it was projected onto a flavor eigenstate due to a measurement) can change into another flavor eigenstate after some time. This behavior is only possible if the three flavor eigenstates are not eigenstates of the Hamiltonian that determines the time evolution of a non-interacting neutrino state. There must be three other eigenstates called the mass eigenstates ν_1, ν_2 and ν_3 which are a different basis for our Hilbert space [16]. These are related to the flavor eigenstates by multiplication with a matrix $U_{\text{PMNS}} \in \text{SU}(3)$, i.e.,

$$\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \\ |\nu_\tau\rangle \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \\ |\nu_3\rangle \end{pmatrix} . \tag{2}$$

Experimental measurements of the so-called "MSW-Effect" for solar neutrinos revealed that one of these eigenstates is heavier than another one [13]. The lighter of both is named ν_1, and the heavier is named ν_2. However, until today we do not know whether the third mass eigenstate ν_3 is heavier (called the normal hierarchy) or lighter (inverted hierarchy) than both of the other states - an unsolved question called the neutrino mass hierarchy problem [7]. The situation is illustrated in figure 2.

1.1.3 Detection and classification of neutrino events

The cross section for interactions between neutrinos and other particles is very small. Each second about 100 trillion neutrinos go through our body unnoticed [20]. Most of them even go through the whole earth without participating in any reaction, which

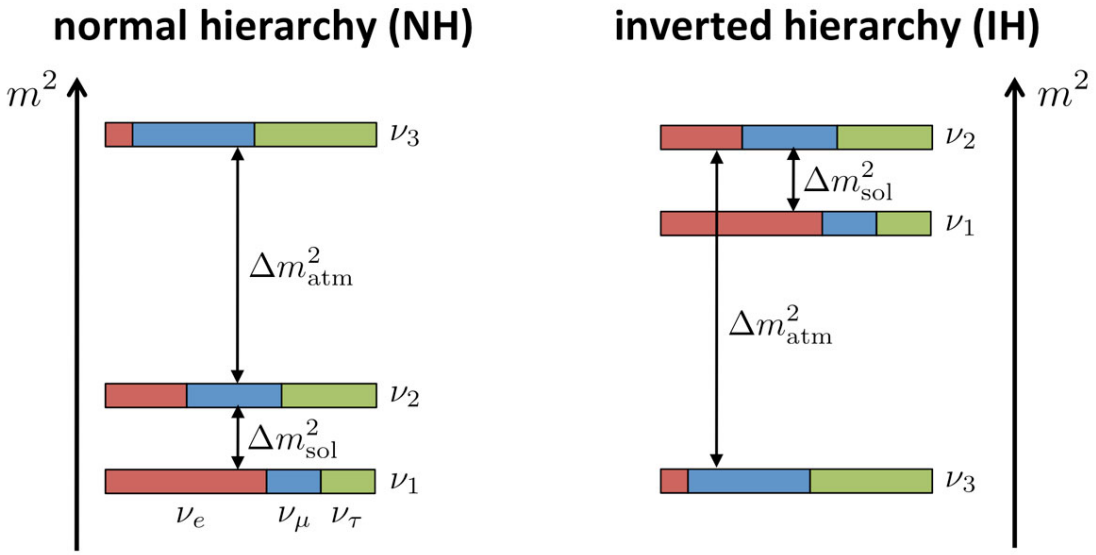

Figure 2: There exist two possibilities for the hierarchy of the mass states: ν_3 is either the heaviest (left) or the lightest (right) mass eigenstate [11].

makes them very difficult to detect. In the following, we go through one possible way of detecting them, which is used in the KM3NeT detector that we will discuss later in section 1.2.

From the three fundamental forces of the standard model, the neutrinos only interact via the weak interaction. The most frequent interactions between neutrinos and other particles (e.g., the molecules in the seawater) in the energy region larger than GeV are sketched in figure 3. In this picture, we can see three so-called charged current interactions and three neutral current interactions. Note that the neutral current interactions are also three different interactions since the letter x in ν_x can take on the values e, μ, and τ, respectively.

We will start by discussing the top part of each of these six different interactions in figure 3. In the diagrams for the three charged current interactions, a ν_x-neutrino approaches a nucleon, exchanges an electrically charged W-boson with it (hence the name charged current), and then an electrically charged x-particle together with a hadronic jet leaves the vertex. On the other hand, in the three neutral current interactions, an electromagnetically neutral Z-boson is exchanged between the approaching ν_x-neutrino and a nucleon. Because no charge is exchanged (hence the name neutral current), there

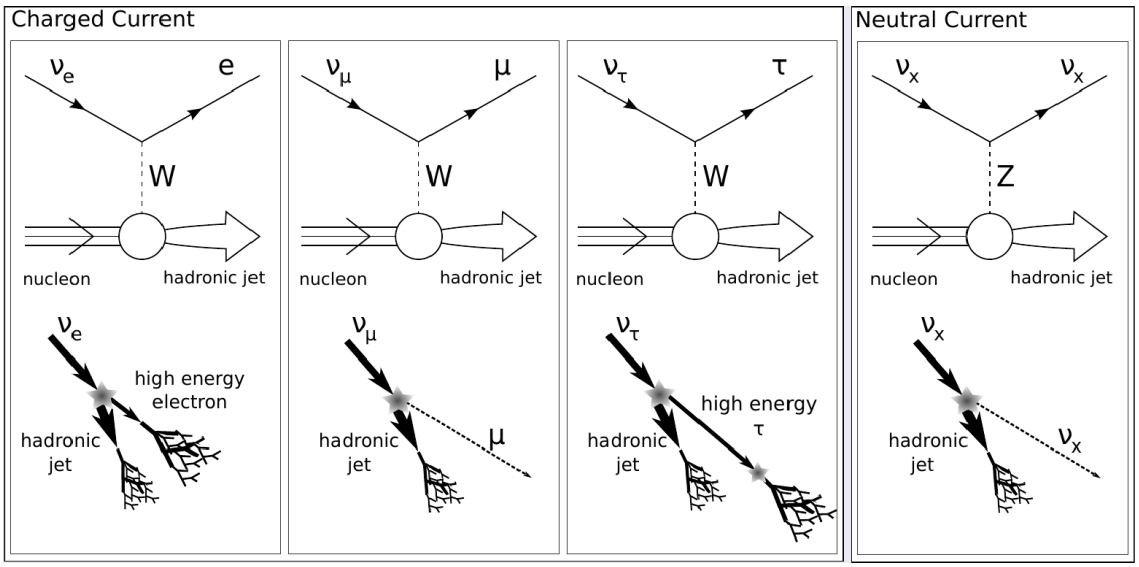

Figure 3: Sketch of three frequent charged current and three neutral current interactions between a neutrino and a nucleon for energies in the region larger than GeV. The interaction itself is sketched at the top, and at the bottom the traces of these interactions in the detector can be seen [19].

can not be a charged particle going out of the interaction alongside the hadronic jet - therefore, from respecting conservation laws, a ν_x-neutrino must go out of the reaction together with a hadronic jet.

We can use these rare interactions between neutrinos and nucleons to detect the neutrino. For this, we use the emission of Cherenkov light, which can be seen in figure 4. In this picture, a charged particle is emitted to the right from an interaction with a neutrino. As long as it is faster than the speed of light in the medium (water in the case of the KM3NeT detector, see section 1.2), the charged particle will induce the emission of a Cherenkov light cone by the particles in the surrounding dielectric medium. The emitted cone is registered by a detector depicted as an array of yellow cells, and the orange cells are the ones detecting the cone. From the cone's opening angle and the cone's geometric orientation, one can deduce the speed and direction of the charged particle that created the cone. These deducted quantities give us the ability to calculate physical quantities like energy and direction of the original neutrino that caused the reaction.

Now we can discuss how the different events in figure 3 will be detectable by their Cherenkov light. If we look at the bottom parts of figure 3, we can see that neutral current interactions emit a neutrino and a hadronic jet. The neutrino is not electrically

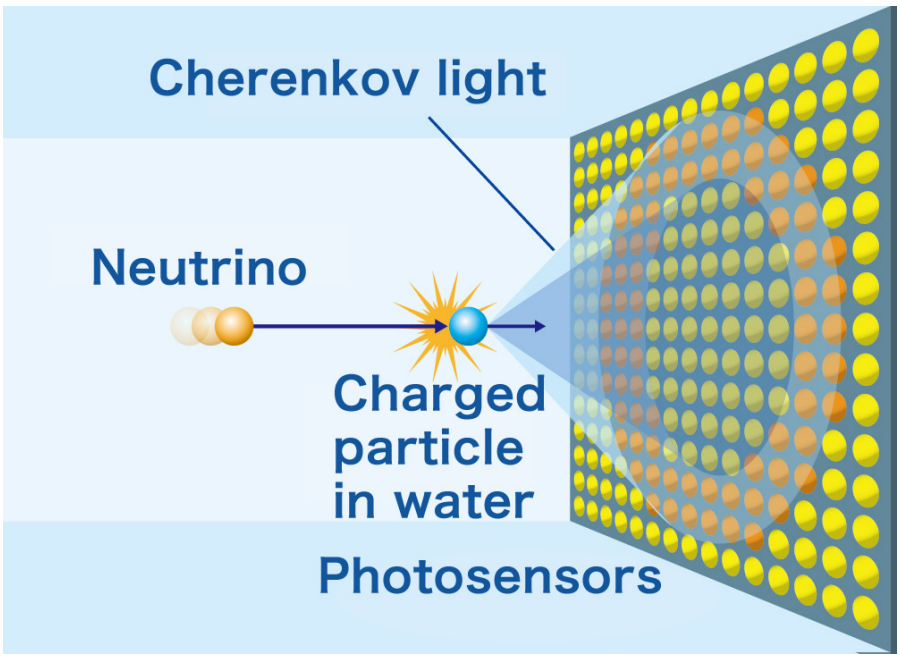

Figure 4: A neutrino interaction emits a charged particle in a dielectric medium. As long as this particle is faster than light in the medium, it emits Cherenkov light which gets registered by a detector [15].

charged and will therefore not induce the emission of Cherenkov light. The hadronic jet, on the other hand, can consist of charged particles like protons and pions. Furthermore, they can decay into other charged particles. This behavior results in the emission of a lot of Cherenkov radiation in different directions, making it more challenging to reconstruct directions. An event where lots of particles decay, creating Cherenkov light in different directions, will be called a shower event from now on.

On the other hand, charged currents produce electrons, muons, or tauons that are electrically charged. Electrons will quickly interact with other particles as indicated in figure 3, creating a shower event. Muons, on the other hand, will travel a relatively long way through the detector before reacting with other particles, which makes them create a clear Cherenkov light cone suitable for direction reconstruction. Events where such a light cone is created will be called track events from now on. Tauon reactions are very rare compared to electrons and muons, and they have many decay channels which can be track- or shower-like. This is the reason why we do not consider them any further in this thesis. All three charged current reactions also have a hadronic jet with them, creating shower-like light signals.

Neutrinos are not the only particles that can reach the deep-sea with enough energy to trigger a reaction. Muons created in the earth's atmosphere can also travel through the detector, creating track-like events with no hadronic jet. These events are a background for the study of neutrinos and their sources, but we will not ignore them because they help researchers in the field of muon physics.

To summarize, we will classify the events as follows:

- **Track-events:** Muon neutrino creating muon in charged current interaction.
- **Shower-events:** Electron neutrino creating electron in charged current interaction and all neutral current interactions.
- **Atmospheric muons (AMs):** Muons from the earth's atmosphere traveling through the detector.

1.2 The KM3NeT detector

KM3NeT is a distributed research infrastructure of neutrino telescopes that run in the Mediterranean deep-sea [8]. Currently, two detectors called "ARCA" and "ORCA" are under construction. ARCA aims at detecting high-energy neutrinos from the cosmos with the ultimate goal to clarify the sources of cosmic rays [6]. On the other hand, ORCA will detect GeV-energy neutrinos created in the earth's atmosphere by reactions with cosmic rays [7]. These cosmic rays consist of protons (95%), alpha particles ($< 5\%$), and heavier nuclei and electrons ($< 1\%$) [2]. As a result of collisions between these particles and nuclei in the atmosphere, new hadrons leave the vertex. The subsequent possible decays of these hadrons then create neutrinos suitable for investigating the neutrino mass hierarchy problem explained in section 1.1.2.

Both ARCA and ORCA are almost built from the same devices. In the following, we will discuss some of these building blocks of the detector, which are necessary for this thesis. These are summarized from [9], where also further information about the detector's unmentioned devices can be found.

1.2.1 Digital Optical Modules (DOMs)

DOMs are glass spheres thick enough to resist the immense water pressure in the deep-sea [10]. Figure 5 shows the cross section of a DOM. 31 photomultipliers (PMTs) are installed in each DOM. They face radially outwards to detect Cherenkov light from neutrino reactions (see section 1.1.3). Several instruments are built into each DOM that track its position and geometrical orientation in the water.

Each photomultiplier in the DOM waits for hits. A hit consists of one or more photons inducing a signal in the PMT, which must be over a certain threshold to count as a hit. Each PMT in each DOM collects data about the position, direction, brightness, and the arrival time of the light reaching it, which is passed on to a central processing logic where it is saved and regularly sent to a shore station.

1.2.2 Detection Units (DUs)

A DU is a collection of 18 DOMs connected by two ropes as shown in figure 6 to align the DOMs vertically in the water [10]. The DOMs are connected via an electro-optical cable which supplies the DOMs with power and makes data transmission via light signals possible. At the bottom, an anchor attaches the DU to the ground, and the top is made buoyant to minimize deviations from the vertical alignment caused by sea currents.

One detector consists of a set of DUs as shown in figure 7. Reactions between neutrinos and seawater in or near the detector induce Cherenkov light, which is detected by the PMTs in the DOMs, and a signal distribution as indicated by the colors in figure 7 is recorded by the 3-dimensional array of DOMs. With the collected data of such an event, physical quantities like the original neutrino's direction and energy can be reconstructed.

The difference between the ARCA and the ORCA detector lies in the distance of the DUs among each other [9]. The ARCA detector is supposed to detect high-energy neutrinos, which create more light than low-energy neutrinos, and therefore the DUs can be distributed more sparsely. The ORCA detector has a much denser DOM distribution to detect low-energy neutrinos, which causes the ORCA detector to span only a 250 times smaller volume than ARCA [10].

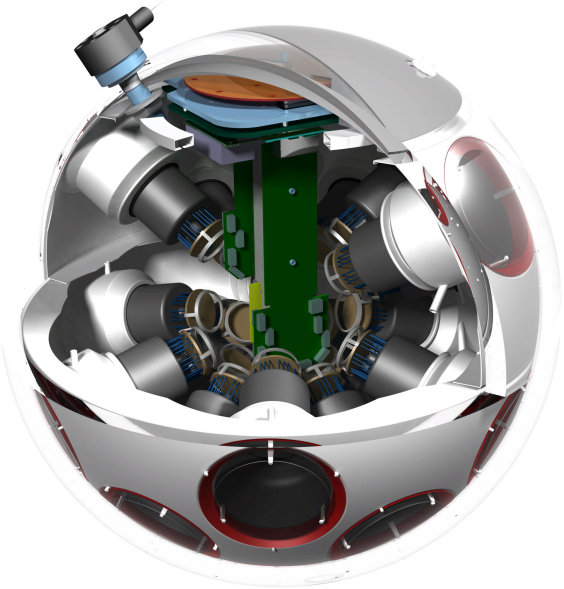

Figure 5: Cross section of a DOM [4]. The black circles on the outside are the photomultipliers facing outwards.

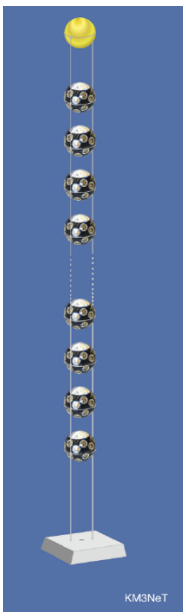

Figure 6: A DU consisting of 18 DOMs connected vertically [1].

1.3 Deep Learning and Graph Neural Networks

1.3.1 Deep Learning

Deep Learning (DL) is a subfield of Machine Learning that gained much popularity in a variety of applications, including physics, over the last years. Machine Learning explores how to construct programs that can automatically improve in a given task with experience [17], i.e., with training. Training is a rather abstract term because there are different ways to train such a program. We will talk about one of them, namely "supervised learning", at the end of this section.

Programs in DL have a specific structure that distinguishes them from other subfields of Machine Learning. This structure can be seen in figure 8. We will call the programs constructed in the field of DL "neural networks" from now on. A simple deep neural network consists of several neurons which hold one real number each. The left layer of neurons is called the "input layer" which consists of three neurons in this example. This means that the neural network receives three real numbers as input. These numbers

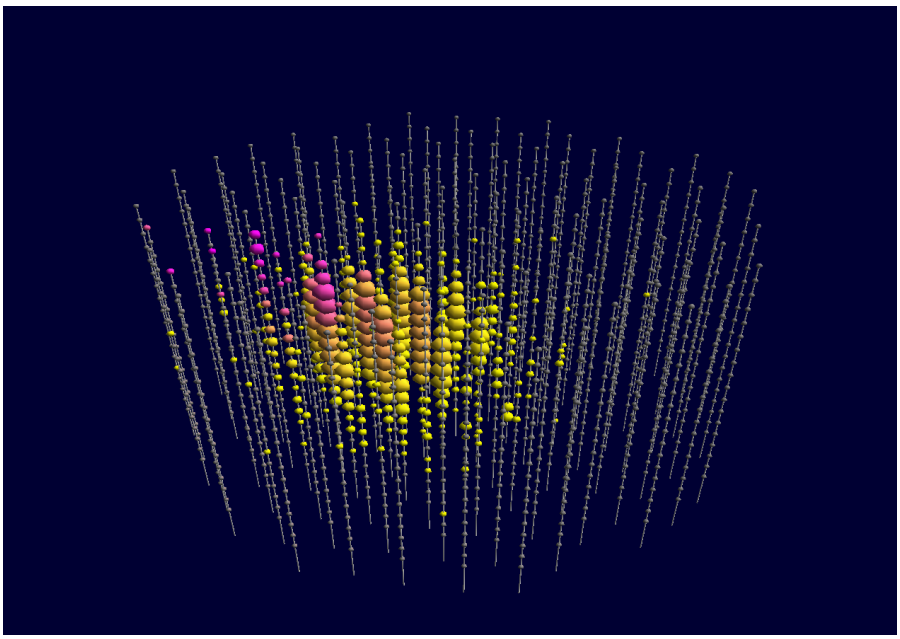

Figure 7: Sketch of an event detected by a KM3NeT detector. The DOMs detect a signal distribution caused by a neutrino event as indicated by the different colors and the different sizes of the colored spheres [5].

could represent anything in the real world, and the neural network does not need to know what these numbers mean - it is the task of the constructor of the neural network to find a suitable way to represent the data by some real numbers which the neural network can take as input. One could, e.g., represent a black and white image consisting of 90×90 pixels by a 90×90 array of real numbers describing how bright each pixel is. This array can be fed to a neural network with 8100 input neurons, and the neural network can perform a previously learned task on the image.

The layer on the right is called the output layer. Similarly to the input layer, the neural network does not need to know the meaning of its output - this is again something the neural network's constructor must have thought about before building the neural network. Characteristic for deep learning is the use of multiple hidden layers. The neural network's constructor does not need to know what the neuron values in the hidden layers mean - it is the neural network's task to find a way to use these hidden layers to perform the specified task. For this, it uses so-called "weights", which are also real numbers. These are represented in figure 8 as the edges connecting the neurons.

Furthermore, each neuron holds one additional weight, called the "bias", which is not

shown in figure 8. Each neuron value of layer i (except the input neurons' values) is calculated from neuron values of the previous layer $i-1$, from the weights connecting our neuron with the neurons from layer $i-1$, and the bias of our neuron in a nonlinear way. The exact method of how this works is not necessary for understanding this thesis. In the beginning, the weights and biases are initialized randomly, i.e., the neural network's output is meaningless before training. During training, the neural network changes its weights and biases such that it gets better and better at performing the specified tasks.

The type of training mainly used for DL is called "supervised learning" (SL). In SL, the neural network's constructor has the role of a teacher who gives the neural network a lot of training material. The training material consists of many typical input data and the corresponding correct solutions called "labels". "Correct" means in this context that it is the output that the neural network is supposed to produce if the corresponding input data is given. Suppose one, e.g., wants to build a neural network that receives pictures of either a dog or a cat as input and gives as output whether there is a dog or a cat in this picture. In that case, the neural network's constructor should give many typical pictures of dogs or cats and the correct labels. The neural network uses these samples of inputs, tries to predict the correct output, then compares its output to the correct label, and tunes its weights and biases such that its outputs are closer to the correct labels in the subsequent trial. A discussion of how these training algorithms work precisely can be found in [17].

A neural network that takes a fixed number of real values as an input is a very simple neural network architecture that can be surprisingly effective for many tasks. Nevertheless, researchers found more sophisticated architectures that accept more complex data structures as input in the recent past. For example, there exist so-called "Convolutional Neural Networks" (CNNs), which accept whole images as inputs. Of course, one could also feed an image into a simple neural network like in our example of dog and cat pictures above. However, our simple neural network's performance will probably be worse than the performance of a CNN, which is trained for the same task. This is because the CNN's architecture is constructed to be especially suitable for image tasks. Another class of specialized neural networks are so-called "Graph Neural Networks", which we will discuss in the next section.

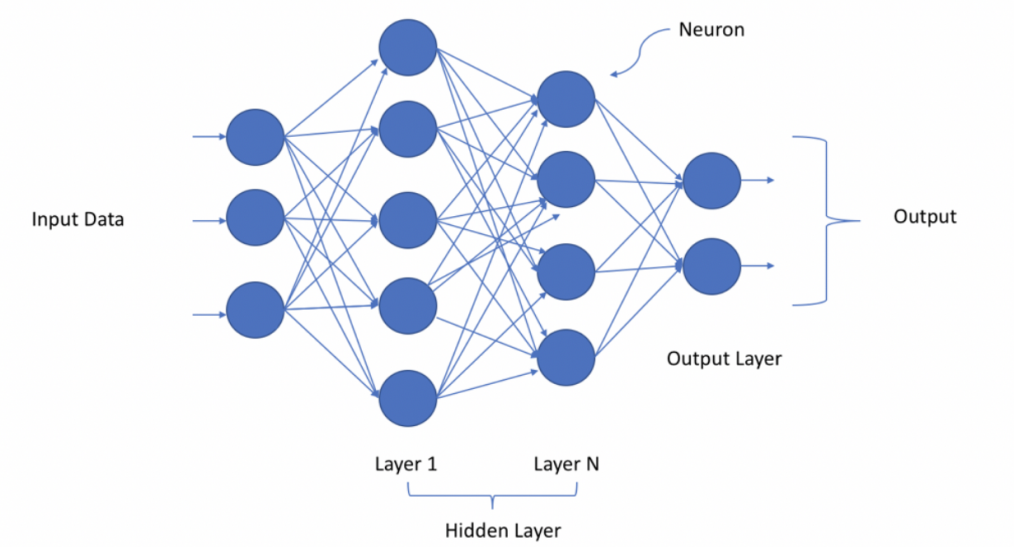

Figure 8: A simple deep neural network with an input layer, an output layer, and hidden layers [18].

1.3.2 Why Graph Neural Networks?

Graph Neural Networks (GNNs) are a current research topic in the deep learning community. With their use of graphs as inputs, they naturally generalize previous network architectures like Convolutional Neural Networks (CNNs), which only accept pictures as an input. Graphs are data structures consisting of vertices and edges connecting the vertices. The vertices represent objects, and the edges represent relationships among these vertices. An example of a graph is shown in figure 9.

KM3NeT is using methods for data representation and GNN architecture that are based on [21]. In the remaining part of this section, we will discuss all information about data representation and GNNs necessary for this thesis. Further information can be found in [21].

1.3.3 Data representation

An already existing Monte Carlo simulation is able to simulate typical events and their traces in the KM3NeT detector. The simulation also provides information about the particle that caused the event. We are especially interested in the "true" direction of this simulated particle since we want to compare the predicted direction of the GNN

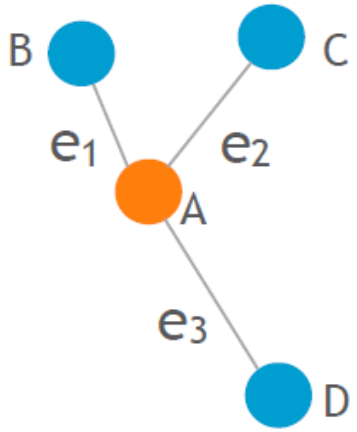

Figure 9: A simple example of a graph. We see four vertices A, B, C, D and three edges e_1, e_2 and e_3 [14].

	track	shower	AM
total number of events	194359	27366	264506
number of files	380	224	186
average events per file	511	122	1422
energy range	1 GeV - 5000 GeV	1 GeV - 500 GeV	$>$ 80 GeV
hits per event	20-1000	20-120	20-1300

Table 1: Information about the input data which was used to test the GNNs performance after applying miscalibrations.

with the true direction to measure the GNN's performance. Throughout this thesis, all the input data used for the GNN was calculated by this Monte Carlo simulation. The real detector currently consists of 6 DUs, but the Monte Carlo simulations calculated the events when only 5 DUs were deployed. However, DU number 1 was damaged at that time, and therefore no data was collected by it. In the Monte Carlo simulation, only the four at that time working DUs were simulated to approximate the real data collected by the detector. The four DUs' positions relative to each other are shown in figure 10 where we can see that the detector approximately forms a rhombus.

We have a certain number of files containing event information available for each event type. In table 1 informations about the used input data is summarized. These will be useful later when we discuss the procedure of applying miscalibrations to the input data.

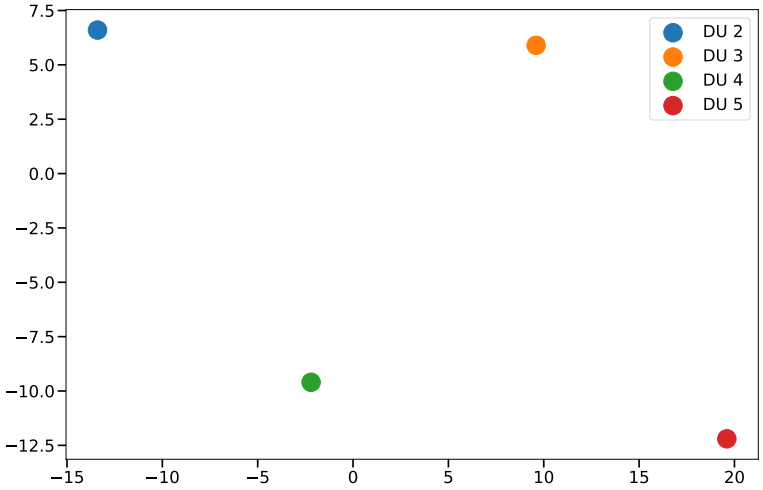

Figure 10: The position of the four DUs relative to each other.

Each event is represented as one graph. We define each node to represent exactly one hit, and the features of each node are the time of the hit, the position of the PMT that detected the hit, and the direction in which the PMT is oriented. These are seven features in total, i.e., $\{t, x, y, z, \hat{x}, \hat{y}, \hat{z}\}$. Each node has directed connections to its k nearest neighbors, where $k = 40$ for the GNN used throughout this thesis. The distance measure between two nodes A and B is defined as

$$d = \sqrt{\sum_i (A_i - B_i)^2} \tag{3}$$

where $i \in \{t, x, y, z\}$ and A_i is the feature i of the node A (analogously for B_i).

This representation is very suitable for our detector data. Reasons for this are, e.g., that graphs do not have a natural order of nodes, exactly as our hits do not have a natural order. Data structures like ordered lists would introduce an order that could unnecessarily influence the neural network's training process. CNNs are also not very suitable because we have seven-dimensional data, and CNNs are usually implemented for one-dimensional ordered lists or two-dimensional pictures. Furthermore, only a tiny fraction of the whole seven-dimensional feature space is nonempty, i.e., a hit was detected at that pixel. This would lead to an enormous waste of computational resources for

calculations on these empty pixels. On the other hand, graphs do not contain any more data than necessary because they have precisely one node for each hit in which only the seven features are stored.

1.3.4 Graph Neural Network architecture

In this section, we will discuss the data flow through the GNN, how this flow respects that there is no order between nodes, and which parts of the GNN contribute to its ability to learn a task.

In the beginning, we have a graph, and we choose a node i. With the definition of the distance measure in equation 3 we can find the k nearest neighbors which have indices i_j with $j \in \{1, 2, \ldots, k\}$. We denote the feature vector of node i with $\mathbf{x}_i \in \mathbb{R}^{\mathrm{F}}$, where $\mathrm{F} \in \mathbb{N}$ is the number of features. Node i then aggregates information from its k nearest neighbors by performing the operation

$$\mathbf{x}'_i = \frac{1}{k} \sum_{j=1}^{k} \mathbf{h}_{\mathbf{\Theta}}(\, \mathbf{x}_i,\, \mathbf{x}_{i_j} - \mathbf{x}_i) \tag{4}$$

where $\mathbf{x}'_i$ is the new feature vector of node i and $\mathbf{h}_{\mathbf{\Theta}}(\, \mathbf{x}_i,\, \mathbf{x}_{i_j} - \mathbf{x}_i) : \mathbb{R}^{\mathrm{F}} \times \mathbb{R}^{\mathrm{F}} \to \mathbb{R}^{\mathrm{F}}$ is the so called "edge function". It depends on the learnable parameters $\mathbf{\Theta}$ which are called "weights". This function is implemented by a multi layer perceptron with $\mathbf{\Theta}$ being the weights and biases in the usual sense. Note that equation 4 respects that there is no order in a graph since the sum is commutative, i.e., it does not introduce an order of the nodes in the graph. We call this operation an "EdgeConv" block.

It is essential to mention that the same edge function $\mathbf{h}_{\mathbf{\Theta}}(\, \mathbf{x}_i,\, \mathbf{x}_{i_j} - \mathbf{x}_i)$ is used for all edges in the graph which drastically reduces the number of weights that must be learned. It also enables the GNN to accept a graph with an arbitrary number of nodes since there is no need to introduce new weights if a graph should have more nodes than usual. Furthermore, it also ensures that we do not have any order or orientation in the graph. If the GNN learned different weights for each edge (and therefore one edge function for each edge), we would have to pay attention that we always feed the correct $\mathbf{x}_i$ and $\mathbf{x}_{i_j} - \mathbf{x}_i$ into each edge function which would create an unnatural order in the graph.

We obtain a new graph with new feature vectors if we do this EdgeConv operation once

for all nodes in the graph. This graph can also be given as input to another EdgeConv operation, which we can repeat as often as we want. These repetitions are essential because if we would perform the operation only once, each node would only have access to its k nearest neighbors' information. If we perform this action multiple times, each node's information can travel further through the network. Therefore, we can adjust whether the GNN should base its decisions in this first step upon more local or more global regions of the graph.

After this aggregation phase, we want to feed all the updated features to another neural network, e.g., a dense neural network. This dense neural network outputs our desired energy and direction of the neutrino/muon, which caused the event. Because the number of input nodes of a dense neural network is fixed while the output of the aggregation phase can be variable-sized due to a different number of nodes for every input graph, it is necessary to ensure that the output of the aggregation phase is formatted correctly. This can be achieved, e.g., with pooling layers like average pooling.

The training of a GNN works in principle like for any other neural network. We have the weights Θ of the edge function and the dense neural network weights. The Monte Carlo simulation provides us with input data (the hits) and the labels (true energy and direction of the neutrino/muon) to use for supervised learning. Therefore, we have everything we need to update our weights with gradient descent.

1.3.5 Miscalibration of the detector geometry

In this thesis, we want to test how the GNN responds to miscalibrated input data. However, what does this exactly mean, and why is it helpful to do that? In section 1.2, we discussed that the KM3NeT detector consists of DUs, DOMs, and PMTs. Different calibration mechanisms exist to measure which systematic miscalibrations exist between each DU, between each DOM belonging to one DU, and between each PMT belonging to one DOM. For example, it could be suspected that each hit measured by DU X has a constant time offset of 1 s, which should not be there. To properly synchronize this DU with the other DUs, the offset of DU X is measured by one of these calibration mechanisms and subsequently subtracted from all the data measured by this DU. However, these calibration mechanisms are not perfect, and small systematic deviations between the different devices of the detector can still exist after applying the calibration mechanisms. If one wants to use GNN's as a reconstruction tool, it should be tested

how robust the predictions are to these systematic uncertainties. The following sections will discuss what systematic miscalibrations were performed to produce miscalibrated graphs and how this manipulated input data influenced the GNN's predictions.

2 Manipulation of entire DUs

2.1 Track events

2.1.1 Miscalibration of the time coordinate

This section aims at testing how the reconstruction of track events is influenced by a miscalibration of the time coordinate of an entire DU $m \in \{2, 3, 4, 5\}$ (see figure 10). In other words, we want to see the influence of a systematic deviation between the time coordinate of one DU relative to the other DUs on the reconstruction of track events. Because of the different topologies of the different event types explained in section 1.1.3, we will treat each event type in its own section.

We perform this test for each DU $m \in \{2, 3, 4, 5\}$. We will denote the value of the time miscalibration, which is applied to DU m, with the letter n. We will start with choosing some $n \in [-30\,\text{ns}, 30\,\text{ns}]$ because these are possible, realistic values for systematic deviations between the DUs in the real detector [14]. As an example, the value $n = 10\,\text{ns}$ means that DU m detected a hit at time t in reality, but because of the miscalibration n, the central processing logic stores that DU m registered a hit at time $t + 10\,\text{ns}$. After applying the miscalibration to the DU m and creating a graph of an event k with miscalibrated feature vectors, we let the GNN predict the direction of the neutrino that caused the reaction. To compare the predicted direction with the real one, we will calculate the opening angle α between both directions. In the optimal case (parallel alignment), it would be $\alpha = 0°$ and in the worst case (antiparallel alignment) $\alpha = 180°$. We repeat this procedure for all DUs m, for all miscalibrations n, and for all track events k. In the end, this gives us $m \cdot n$ distributions of opening angles, which we denote by $\alpha_{mn}(k)$.

These distributions of opening angles can be seen in figure 11. Each of the four diagrams in this figure has the opening angle α in degree on the y-axis. On the x-axis are the miscalibrations of the time coordinate on DU m. The DU number m has a different value in each of the four diagrams, which can be read from each diagram's title. In each diagram, we can find the different values of the miscalibration n on the x-axis. For fixed m and n, the distribution $\alpha_{mn}(k)$ is sketched by its 16%, 84% and 50% quantile, where the latter is called the "Median". This means, e.g., that for a given DU m and miscalibration n, 16% of the opening angles of the distribution $\alpha_{mn}(k)$ are smaller than the 16% quantile in figure 11. Analogous for the 50% and the 84% quantiles. These kind

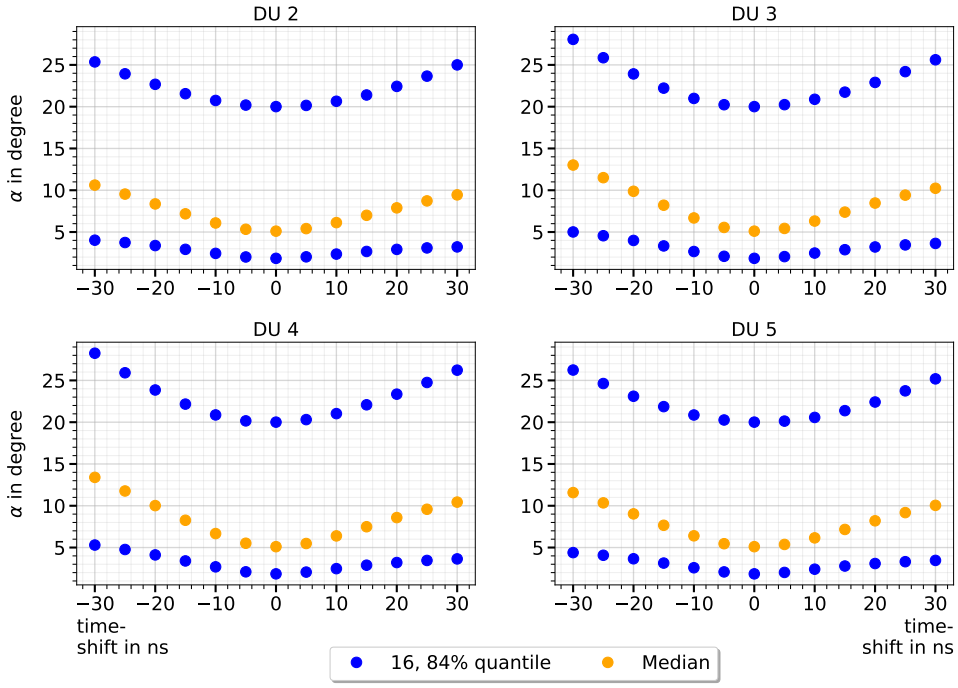

Figure 11: The quantiles of the opening angle distributions $\alpha_{mn}(k)$ for track events are plotted over different values for the time miscalibrations n. In diagram "DU m", the miscalibration acts on the time coordinate of PMTs belonging to DU m. The values of n are chosen to be in the range of possible real calibration errors of the time coordinate.

of plots therefore enable us to directly compare the distributions $\alpha_{mn}(k)$ for different m and n.

From figure 11, we can read off some interesting effects:

- As expected, the case without a miscalibration, i.e., the case with $n = 0\,\text{ns}$, has smaller quantiles than the $n \neq 0\,\text{ns}$ cases. This means that the best predictions are made for the case of no miscalibrations.
- If we denote the $X\%$ quantile ($X \in \{16, 50, 84\}$) of DU m and miscalibration n as Q^X_{mn}, then the relation $Q^X_{mn} \leq Q^X_{m,-n}$ approximately holds for most $n > 0$, X, and m. This means that negative timeshifts impair the prediction of the GNN more than positive timeshifts.
- The 16% quantile increases slower for $|n| > 10\,\text{ns}$, i.e., the graph has a negative

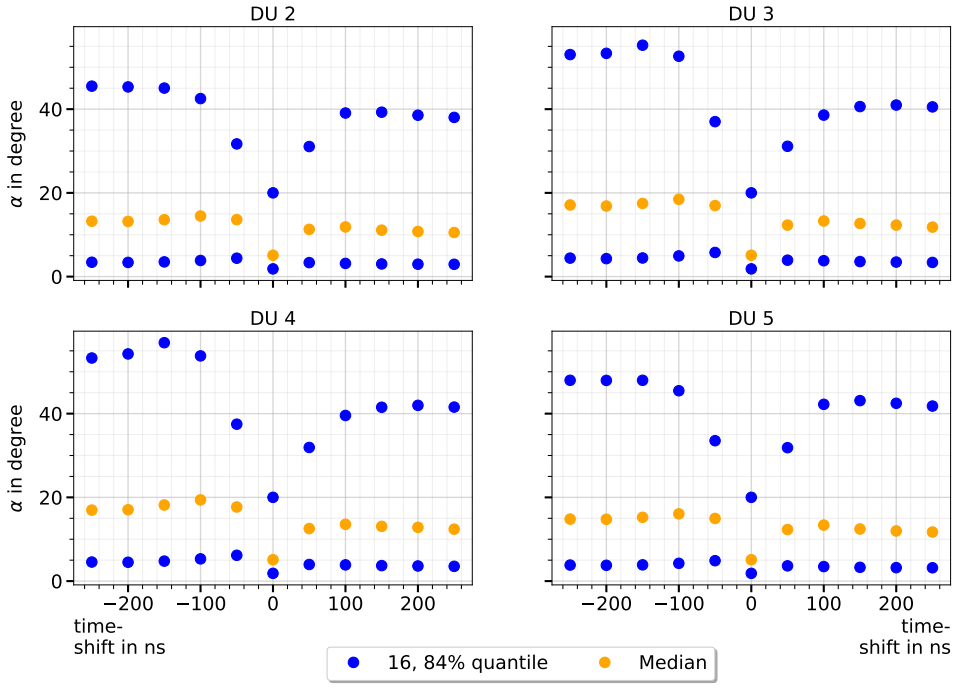

Figure 12: The quantiles of the opening angle distributions $\alpha_{m,n}(k)$ for track events are plotted over different values for the time miscalibrations n. In diagram "DU m", the miscalibration acts on the time coordinate of PMTs belonging to DU m. The range of values of n are chosen to be larger than usual real calibration errors of the time coordinate.

curvature. However, the median approximately increases with a constant slope for increasing $|n|$, although it has a tendency towards a negative curvature for the largest values of $|n|$ in the chosen range. The 84% quantile increases faster for increasing $|n|$, i.e., it has a positive curvature. This suggests that the best predictions are not as impaired by the miscalibrations as the worst predictions.

In figure 12 we can see a plot analogous to figure 11 but with a larger range of $n \in \{-250\,\text{ns}, 250\,\text{ns}\}$. Because this range consists largely of rather unrealistic miscalibration values and because these graphs show the same qualitative new effects for every event class, we will not discuss this kind of plot for every case. Nevertheless, we can observe some interesting behaviour of the GNN:

- The distributions $\alpha_{2,n}(k)$ and $\alpha_{5,n}(k)$ look qualitatively very similar. The same holds for $\alpha_{3,n}(k)$ and $\alpha_{4,n}(k)$. This is explainable by looking at the positions of

the DUs in figure 10. There we can see that the DUs 2 and 5 are the outer vertices of an approximate rhombus, and DUs 3 and 4 are the inner vertices. If we have an event $\tilde{k}$ that produces the angle $\alpha_{2,n}(\tilde{k})$, then the same event rotated with an angle of 180° around the axis parallel to the DUs and going through the centroid of the rhombus would create the angle $\alpha_{5,n}(\tilde{k}) = \alpha_{2,n}(\tilde{k})$. Because both the event $\tilde{k}$ and the rotated event are equally probable, both distributions should look approximately the same. Analogous for $m = 3$ and $m = 4$. This suggests that the GNN learned this symmetry successfully.

- For a sufficiently large absolute value $|n|$ of the time miscalibration, the 16% quantile and the median get smaller as $|n|$ increases. This effect also happens for the 84% quantile if we only consider positive miscalibrations $n > 0$. If we just look at negative miscalibrations $n < 0$, the 84% quantile slowly increases or remains constant for larger $|n|$ if $m = 2$ or $m = 5$, but it gets noticeably smaller for $m = 3$ or $m = 4$. This behavior is rather surprising since one would expect that larger miscalibrations imply worse predictions.

2.1.2 Miscalibration of the orientation

The procedure to apply the miscalibration is the same as in section 2.1.1 with the difference that we do not miscalibrate the time coordinate, but the angle φ of each DOM belonging to DU m. If we imagine a cartesian coordinate system with an origin lying in the center of the DOM and a z-axis perpendicular to the seabed, then the angle φ is defined as the angle between a point in the x-y plane and the x-axis. In other words, φ is the azimuth if one uses standard spherical coordinates (θ, φ) in the above-mentioned coordinate system. In the following sections, the words "azimuth", "orientation", and even sometimes the word "angle" are used interchangeably.

In figure 13 we can see the quantiles of the distributions $\alpha_{mn}(k)$. The plots are analogous to the ones before, except that the φ miscalibration is on the x-axis of the four diagrams instead of the time miscalibration. The range of miscalibrations $n \in (-90°, 90°)$ is chosen such that it includes possible real miscalibrations which could be in the real data the detector is collecting now. It may be surprising that the range goes all up to $|n| = 90°$. We must note here that the detector is still under construction. We only have the deployment data available right now, i.e., the positions and directions of the DOMs were measured by an RC submarine that deployed the DUs in the water. There will be

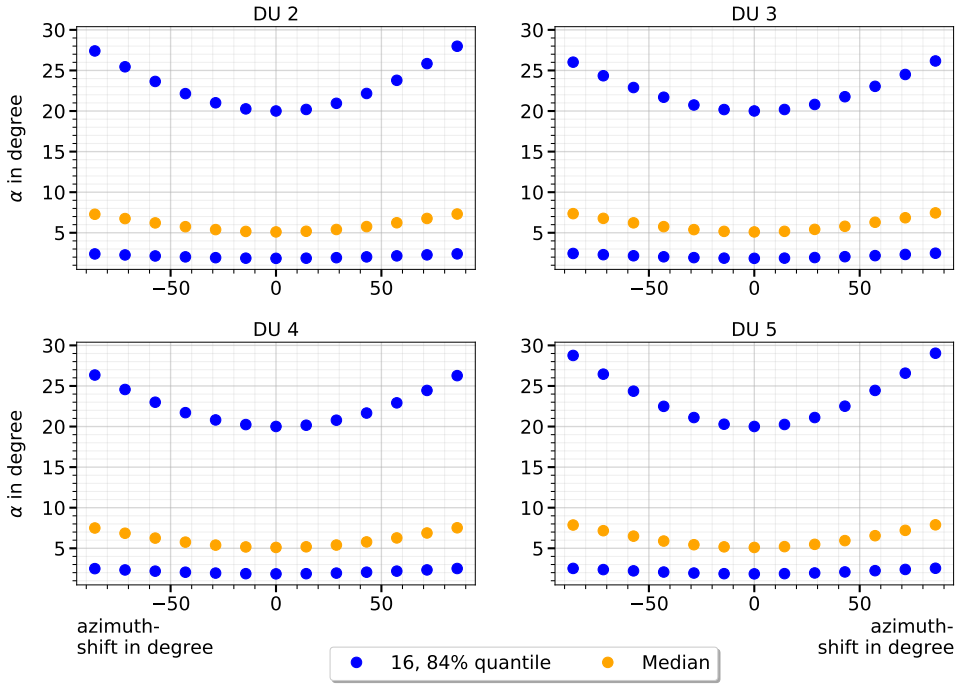

Figure 13: The quantiles of the opening angle distributions $\alpha_{mn}(k)$ for track events are plotted over different values for the φ miscalibrations n. In diagram "DU m", the miscalibration acts on the PMTs belonging to DU m. The values of n are chosen to be in the range of possible real calibration errors of the azimuth φ.

a much finer resolution of the position and direction of the DOMs when the construction phase is finished, and finer calibrations were made [14].

We can observe some interesting effects in figure 13:

- As expected, we see that the quantiles are increasing for increasing $|n|$. However, we can see that the 16% quantile is increasing very slowly compared to the other quantiles, even if we have a miscalibration of $\approx \pm 87°$.
- If we compare the boundary cases $n = \pm 30\,\text{ns}$ for the time case in figure 11 and $n' \approx \pm 87°$ for the angle case, we obtain the relation $Q^X_{mn} \geq Q^X_{mn'}$ for $X \in \{16, 50\}$. This means that the worst realistic time miscalibration impairs the prediction more than the worst possible angle miscalibration, at least for the best 50 % of the predictions. Similar results were already found if one uses the standard reconstruction of KM3NeT which uses a maximum likelihood approach [14].

- The equation $Q^X_{mn} = Q^X_{m,-n}$ approximately holds for $X \in \{16, 50\}$. For $X = 84$, it almost holds. This is more symmetric than for the time case in figure 11.

2.1.3 Comparison of both miscalibrations

In this section, we want to compare timeshifts and azimuthshifts more closely, and we want to see which of these impair the predictions of the GNN more in the real detector. Because there is no obvious way to compare a shift in time with a shift in an angle, we have to make a reasonable choice of values for both shifts here. Suppose we assume that the systematic deviations of the time coordinate and the azimuth φ are Gaussian distributed. In that case, we could estimate the standard deviation σ of both distributions, i.e., we could estimate the interval $[-\sigma, +\sigma]$ in which $\approx 68\%$ of the measurement values fall in (assumption: mean μ of the distributions vanishes). If we denote the standard deviation for timeshifts as σ and the standard deviation for azimuthshifts as $\bar{\sigma}$, then we can compare the opening angle distributions $\alpha_{\tilde{m},\sigma}(k)$ for timeshifts and $\alpha_{\tilde{m},\bar{\sigma}}(k)$ for azimuthshifts with each other, where $\tilde{m}$ is a fixed DU number.

Without any deep reason behind it, we choose $\tilde{m} = 4$ for our comparison. We estimate the standard deviations to be $\sigma = 4\,\text{ns}$ and $\bar{\sigma} = 20°$ [14]. In figure 14 the distributions for a timeshift $\sigma = 4\,\text{ns}$ and an azimuthshift $\bar{\sigma} = 20°$ are plotted via two histograms. To compare these two cases to the case without any miscalibration, the black histogram in figure 14 shows the opening angle distribution for the case with no miscalibration. We can see that the histogram for azimuthshifts has about $0.6\,\%$ more counts in its first bin, i.e., in the bin in which the best predictions from 0° up to 2° are, than the first bin of the time case. As expected, the case without a miscalibration has the highest bins for the best predictions in the range $\alpha \in [0°, 4°]$. The rest of the distributions look qualitatively the same. In the time case, the best predictions are mainly pushed into the range $\alpha \in [4°, 8°]$, and the differences in the distributions get smaller for higher values of α. Summarized, we can say that timeshifts impair the best predictions of the GNN more than azimuthshifts for typical values of the systematic deviations which occur in the real detector, at least for track events.

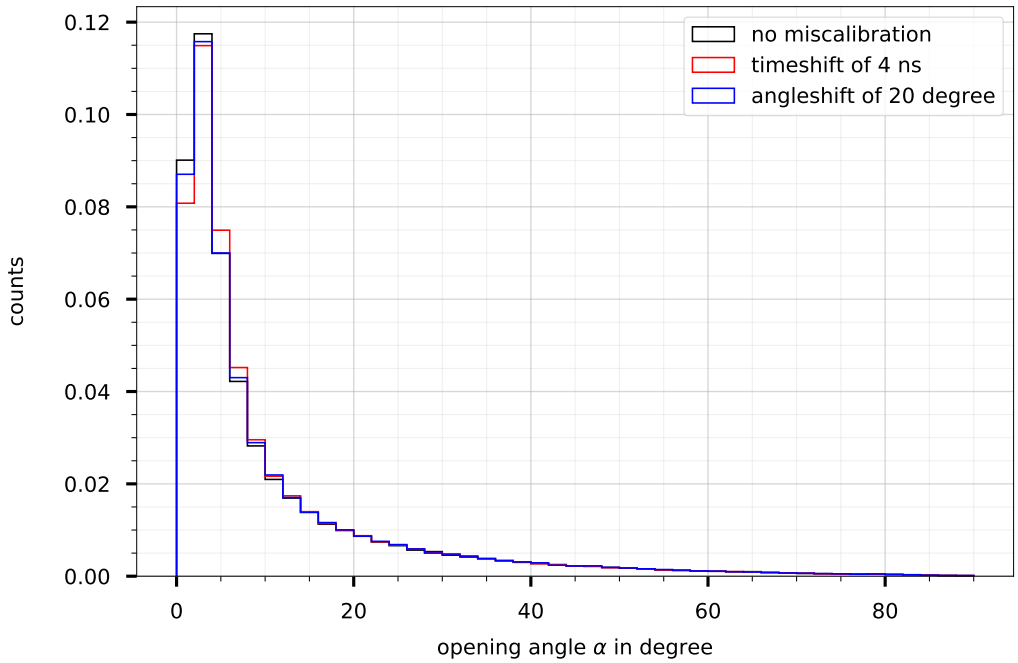

Figure 14: This figure consists of three histograms which can be distinguished by their color. The red histogram represents the opening angle distribution for track events after applying a time miscalibration of $\sigma = 4\,\text{ns}$. Analogous for the blue histogram, but with an angle miscalibration of $\bar{\sigma} = 20°$. The black histogram is the opening angle distribution without a miscalibration. The binsize of all histograms is 2° and they are normalized.

2.2 Shower events

2.2.1 Miscalibration of the time coordinate

In this section, we will repeat the procedure in section 2.1 with the difference that we will now use shower events instead of track events. This results in the distributions $\alpha_{m,n}(k)$ shown in figure 15.

We can make the following observations:

- The quantiles are much larger than in figure 11. This result is expected because shower events are more difficult to reconstruct than track events. We can see, e.g., that 50% of the predicted directions differ by more than 26° from the true direction

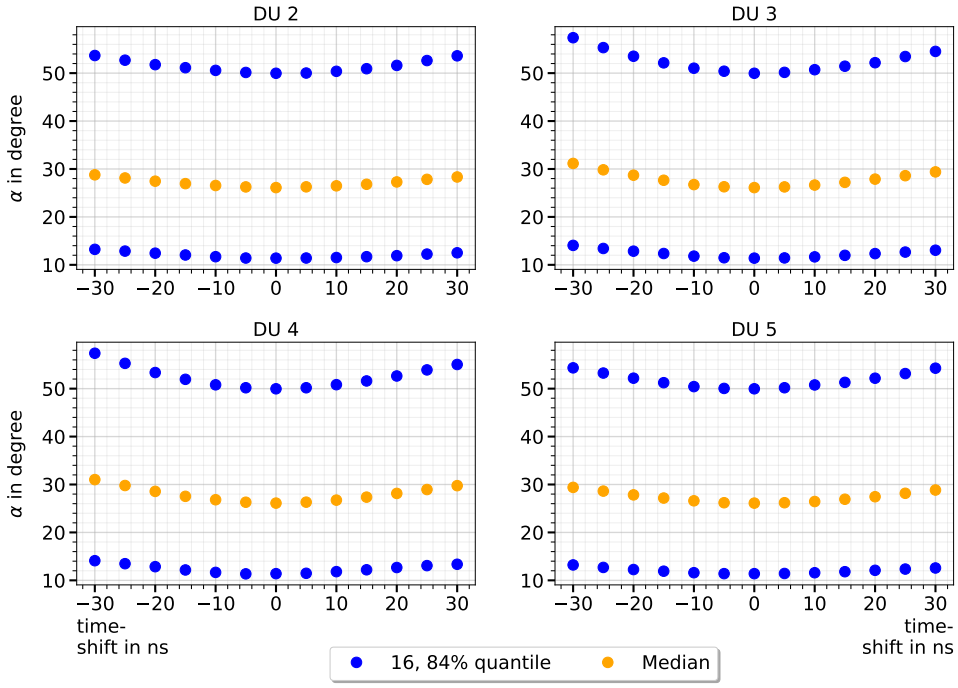

Figure 15: The quantiles of the opening angle distributions $\alpha_{m,n}(k)$ for shower events are plotted over different values for the time miscalibrations n. In diagram "DU m", the miscalibration acts on the time coordinate of PMTs belonging to DU m. The values of n are chosen to be in the range of possible real calibration errors of the time coordinate.

for the $n = 0\,\text{ns}$ case. For tracks, this value was about 5°.

- The 16% quantile increases slower for larger values of $|n|$ than the other quantiles, i.e., the best 16 % of shower predictions are not as impaired by the miscalibrations than, e.g., the worst 16 % predictions. This behavior can be observed for all diagrams of this kind, regardless of the miscalibration type or the event type. To avoid repetition, we will not mention this property any further in the subsequent cases.

- We see that the relation $Q^X_{m,n} \leq Q^X_{m,-n}$ for $n > 0$ approximately holds again for the most values of m and n. Negative timeshifts seem to impair the GNNs performance more than positive timeshifts.

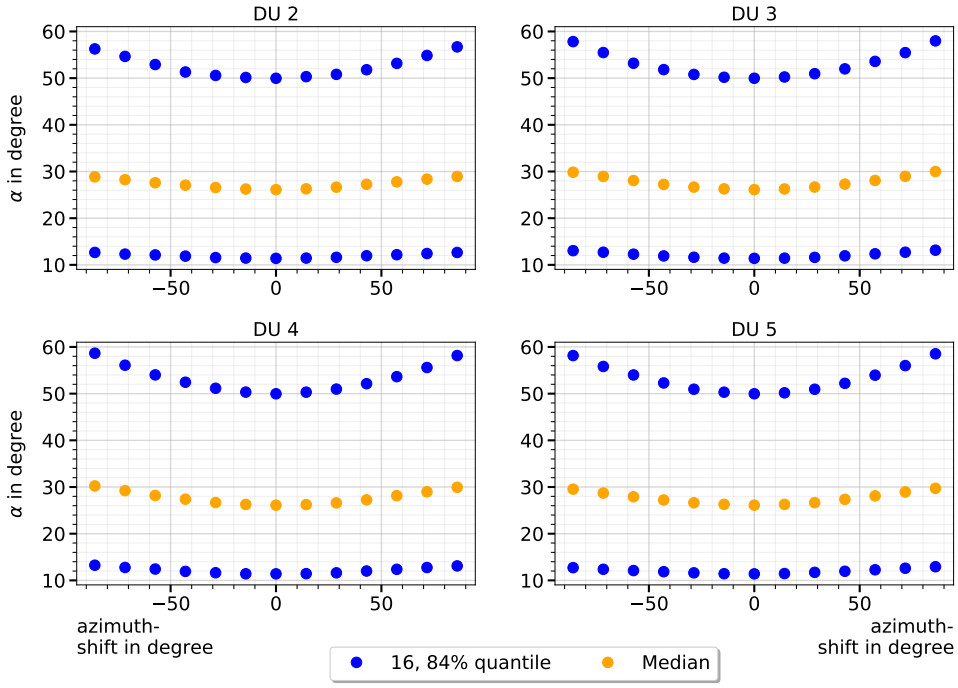

Figure 16: The quantiles of the opening angle distributions $\alpha_{m,n}(k)$ for shower events are plotted over different values for the φ miscalibrations n. In diagram "DU m", the miscalibration acts on the PMTs belonging to DU m. The values of n are chosen to be in the range of possible real calibration errors of the azimuth φ.

2.2.2 Miscalibration of the orientation

Now we will apply the same procedure as for track events in section 2.1.2, i.e., we will miscalibrate the azimuth φ for each DOM belonging to DU m. The results can be seen in figure 16.

The effects from the analogous case for track events, i.e., the effects from section 2.1.2, appear here qualitatively, too, except for the statement that the quantiles in the boundary time case are larger than for the angle boundary case. This statement is not always true for shower events. Besides that, no interesting new properties emerge in this case.

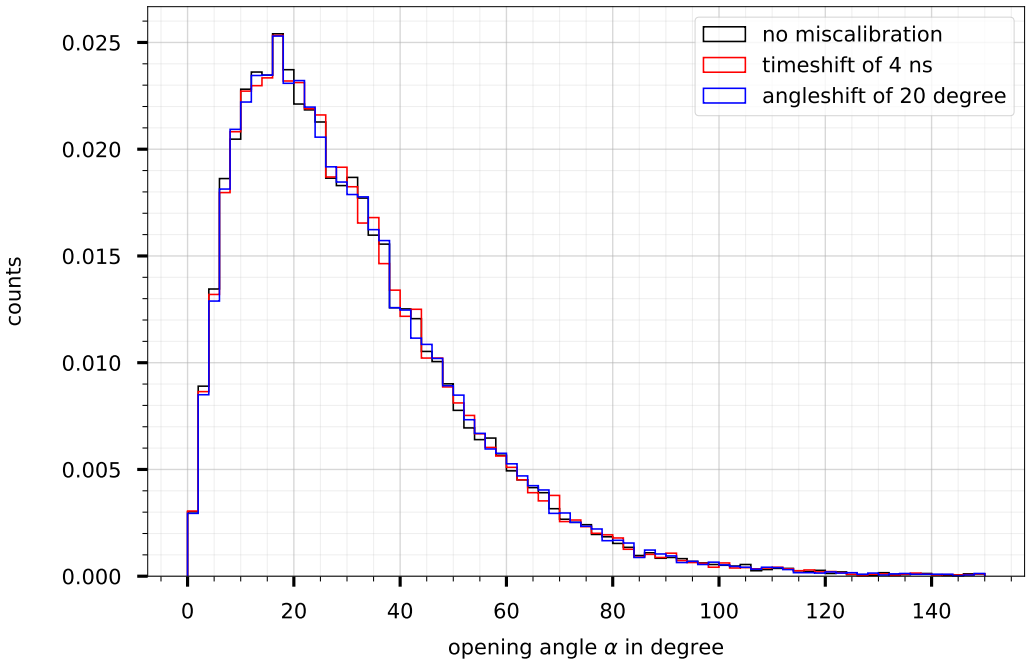

Figure 17: The red histogram represents the opening angle distribution for shower events after applying a time miscalibration of $\sigma = 4\,\text{ns}$. Analogous for the blue histogram, but with an angle miscalibration of $\bar{\sigma} = 20°$. The black histogram is the opening angle distribution without a miscalibration. The binsize of all histograms is 2° and they are normalized.

2.2.3 Comparison of both miscalibrations

This section is analogous to section 2.1.3, where we compared the distributions of opening angles after time miscalibrations and angle miscalibrations for track events. Now we repeat the same procedure for shower events. We choose again $\tilde{m} = 4$, $\sigma = 4\,\text{ns}$ as miscalibration for the time coordinate and $\bar{\sigma} = 20°$ as φ miscalibration. The distributions are shown in figure 17.

We can see that both miscalibrated distributions look both qualitatively and quantitatively very similar. The absolute differences in bin height between the time and the azimuth case are now $\leq 0.14\,\%$, which is smaller than the value for track events where the differences were $\leq 0.63\,\%$. For the largest bin $\alpha \in [16°, 18°]$, almost no differences between the three histograms is observable. Interestingly, for the best predictions in the

range $\alpha \in [0°, 6°]$, the angle miscalibrations impair the predictions of the GNN more than the time miscalibrations. The absolute differences are rather small, namely $\leq 0.031\,\%$. For the first bin, the time case has even a higher bin than the case for no miscalibration. However, the absolute difference is only $\approx 0.004\,\%$. This small value means that this is probably just a random fluctuation with no further meaning. Summarized, the impact of time and angle miscalibrations is not as different as for track events in figure 14. Suppose the difference between the second blue and the second red bin is not only a random fluctuations. Analogously for the third blue and the third red bin. Then, it seems that the angle miscalibrations impair the best predictions of the GNN worse than time miscalibrations, which is different than for track events.

2.3 Atmospheric muon events

2.3.1 Miscalibration of the time coordinate

In this section we will repeat the procedure explained in section 2.1.1 for the atmospheric muon (AM) events. The results can be seen in figure 18.

A lot of effects are qualitatively the same as for track events in section 2.1.1. Some noteworthy properties are:

- The quantiles are much smaller for AMs than for track events. This result is not surprising because AM events not only create the most hits during an event, but they also travel relatively far through the detector, which makes them induce long Cherenkov light cones. For the $n = 0\,\text{ns}$ case, the best 16% of predicted directions deviate by $\leq 1°$ from the true direction. For track events, this value was $\leq 2°$. The median for AMs is $\approx 2.25°$, where it was $\approx 5°$ for the track events. Most noticeably, the 16% worst predictions have opening angles $\geq 6.75°$, while the worst predictions are $\geq 20°$ for track events.
- The already familiar relation $Q^X_{mn} \leq Q^X_{m,-n}$ for $n > 0$ can be clearly seen here. Compared to track events, the effect has substantially more influence here.
- The qualitative symmetry between DUs 2 and 5 and between DUs 3 and 4, which was already mentioned for track events for the case of an unrealistically large range of n, can be seen here again. The qualitative differences between the previous diagrams were not strong enough to see this symmetry.

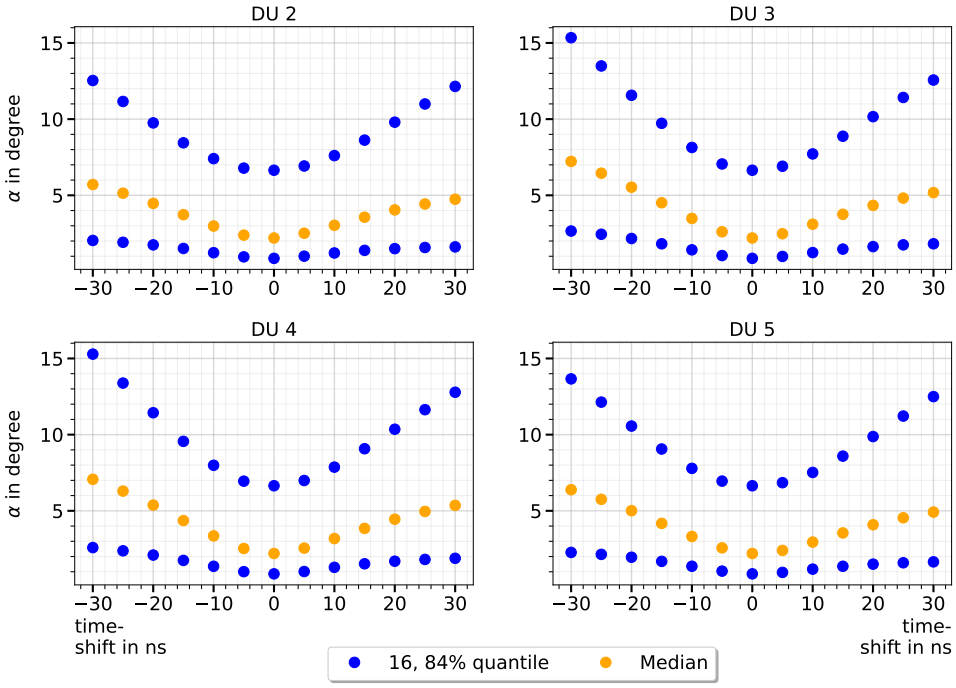

Figure 18: The quantiles of the opening angle distributions $\alpha_{m,n}(k)$ for AM events are plotted over different values for the time miscalibrations n. In diagram "DU m", the miscalibration acts on the time coordinate of PMTs belonging to DU m. The values of n are chosen to be in the range of possible real calibration errors of the time coordinate.

2.3.2 Miscalibration of the orientation

In figure 19, we can see the opening angle distributions $\alpha_{m,n}(k)$ for AM events with miscalibrations of the azimuth φ.

The most properties are qualitatively the same as for track events with φ miscalibration. Therefore, we refer to section 2.1.2. The only mentionable property is that the equation $Q^X_{m,n} = Q^X_{m,-n}$ approximately holds for all $X \in \{16, 50, 84\}$. Therefore, this is the most symmetric case.

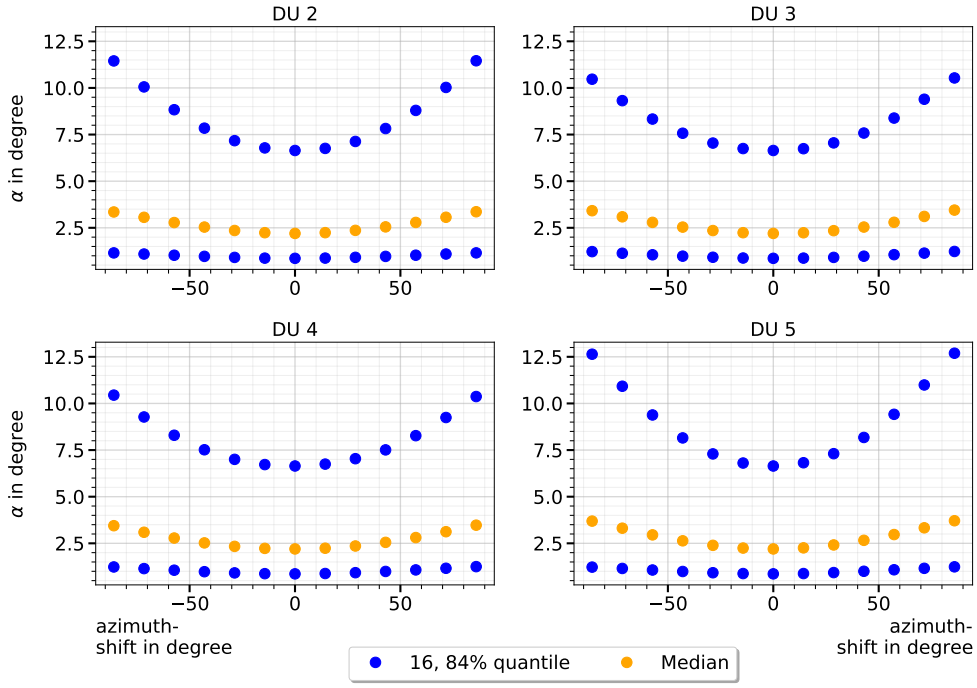

Figure 19: The quantiles of the opening angle distributions $\alpha_{m,n}(k)$ for AM events are plotted over different values for the φ miscalibrations n. In diagram "DU m", the miscalibration acts on the PMTs belonging to DU m. The values of n are chosen to be in the range of possible real calibration errors of the azimuth φ.

2.3.3 Comparison of both miscalibrations

This section is analogous to section 2.1.3, but for AM events. We choose the same values for $\tilde{m}$, σ and $\bar{\sigma}$ as before. The resulting distributions can be seen in figure 20.

We can see that the bin with the best predictions, i.e., $\alpha \in [0°, 2°]$, is the highest bin of the distribution. This result is different from the track events in figure 14, and especially from the shower events in figure 17. The absolute differences between the time and the azimuth case are now $\leq 1.6\,\%$, which is substantially larger than the value $0.63\,\%$ for track events and $0.14\,\%$ for shower events. The first bin with the best predictions is $\approx 1.55\,\%$ larger for angle miscalibrations than for time miscalibrations, which means that time miscalibrations impair the best predictions of the GNN substantially more than the angle miscalibrations. This was also observed for track events in section 2.1.3. The bins with values $\alpha \in [2°, 8°]$ have higher bins for time miscalibrations than for the

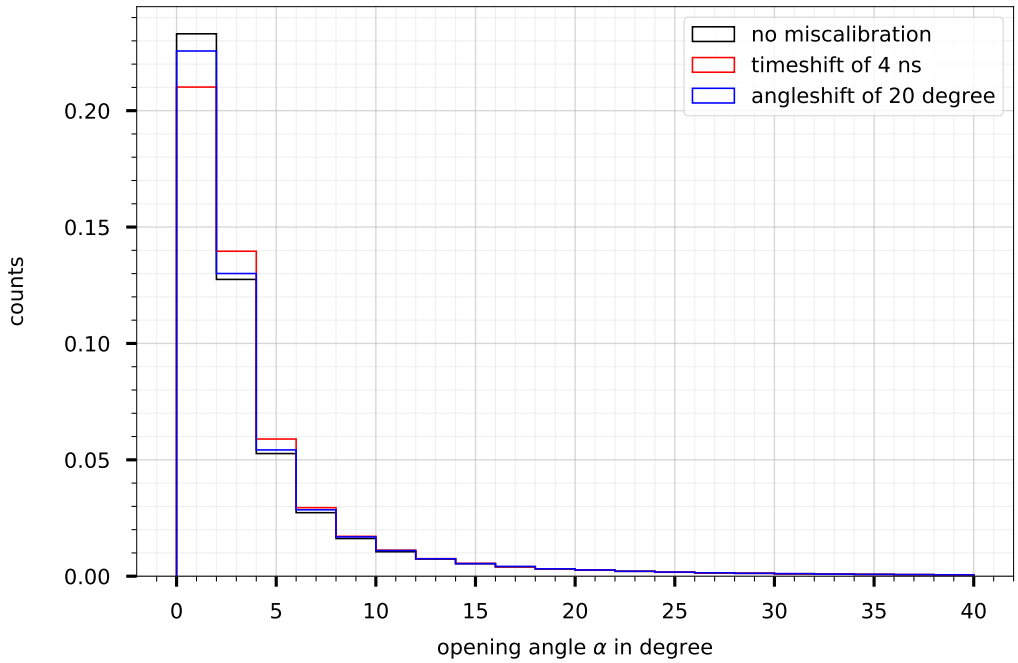

Figure 20: The red histogram represents the opening angle distribution for AM events after applying a time miscalibration of $\sigma = 4\,\text{ns}$. Analogous for the blue histogram, but with an angle miscalibration of $\bar{\sigma} = 20°$. The black histogram is the opening angle distribution without a miscalibration. The binsize of all histograms is 2° and they are normalized.

other two cases. This is because the best predictions from the first bin are pushed into these bins for the time case. Summarized, the results are qualitatively very similar to the results for track events.

3 Probabilistic manipulation of individual DOMs

3.1 Track events

3.1.1 Influence of the calibration quality

This section will go one level deeper and discuss several miscalibrations between individual DOMs belonging to the same DU. In section 2, we chose one DU m and applied the same miscalibration to all DOMs belonging to DU m. We will now apply miscalibrations to all DOMs in the detector, but we make sure that the effects from chapter 2 do not interfere with the inter-DOM effects we want to explore in this section.

Before we discuss how exactly we apply the miscalibrations to produce the graphs of our events, we want to clarify which miscalibrations are applied in this section. In section 2 the time coordinate and the azimuth φ were miscalibrated, and we will use both of these miscalibrations again in this section. Furthermore, the z-coordinate of each DOM will be miscalibrated in this section to compare the impact of a spatial position with the previous miscalibrations.

There is no need to distinguish the DUs m because every individual DU will be treated in the same way in this section. We will use the index m to distinguish the type of miscalibration that we applied. Therefore, we now have $m \in \{t, \varphi, z\}$ for the miscalibration of the time coordinate, the azimuth φ, and the z-coordinate. We will switch out the index n with the index σ. In this section, we will not choose exact miscalibration values, which are then applied. Instead, we sample the miscalibration values from a Gaussian distribution with vanishing mean μ and standard deviation σ. Larger standard deviations imply more significant systematic deviations between the measurements of DOMs belonging to the same DU, and therefore we expect the GNN to perform worse for large σ. To compare the opening angle distributions obtained by sampling from Gaussian distributions with different standard deviations, we will use the index σ to denote the value of the standard deviation used for sampling the miscalibrations. An example of this notation is the expression $\alpha'_{\varphi,90°}(k)$, which describes the opening angle between predicted and true direction of the particle causing event k after a miscalibration of the azimuth φ sampled from a Gaussian distribution with standard deviation 90° was applied. Note that we write α' now instead of α to emphasize that primed opening angles belong to inter-DOM effects and unprimed opening angles to inter-DU effects.

Several metadata about the used input data can be found in table 1. This table informs us that a certain number of files was used for each event class, where each file contained a different number of events. This distinction is crucial since we will not sample and apply a new miscalibration for each event. Instead, a new miscalibration was sampled and applied for each file, which should be a reasonable approximation to the random experiment "Sample one miscalibration per DOM per event from a Gaussian, apply it to one event, let the GNN predict and calculate the opening angle between predicted and true direction". After each file got its own random miscalibration and subsequent creation of graphs from the files, we let the GNN predict on all graphs from all files to get an opening angle distribution $\alpha'_{m,\sigma}(k)$ which approximates the probability distribution of above mentioned random experiment.

We already mentioned in the beginning of this section that we do not want to let the inter-DU effects from chapter 2 interfer with the inter-DOM effects that we want to explore in this chapter. Let x_{ij} be a miscalibration which we want to apply to the j-th DOM belonging to DU i, i.e., $i \in \{2, 3, 4, 5\}$ and $j \in \{1, 2, \ldots, 18\}$ because there are 18 DOMs per DU. Note that all x_{ij} are sampled from the same Gaussian distribution. The mean μ_i is defined as

$$\mu_i = \frac{1}{18} \sum_{j=1}^{18} x_{ij}. \tag{5}$$

To make sure that all DUs have the same average miscalibration, i.e., to make sure that there are no inter-DU effects like in chapter 2, we define a new quantity

$$\tilde{x}_{ij} = x_{ij} - \mu_i. \tag{6}$$

If we calculate the mean $\tilde{\mu}_i$ of these new quantities $\tilde{x}_{ij}$ analogous to equation 5, we realize that $\tilde{\mu}_i = 0$ for all DUs i. Therefore, we use the quantity $\tilde{x}_{ij}$ as miscalibration for the j-th DOM of DU i to make sure that we explore only inter-DOM effects. It is important to note that the sample standard deviation $\tilde{\sigma}_i$ of the quantity $\tilde{x}_{ij}$ is equal to the sample standard deviation σ_i of x_{ij}:

$$\tilde{\sigma}_i = \sqrt{\frac{1}{18-1} \sum_{j=1}^{18} \left(\tilde{x}_{ij} - \tilde{\mu}_i\right)^2} = \sqrt{\frac{1}{18-1} \sum_{j=1}^{18} \left(x_{ij} - \mu_i\right)^2} = \sigma_i \tag{7}$$

where we used $\tilde{\mu}_i = 0$ and equation 6 to obtain the second equality sign. This equality ensures that we can still use the standard deviation σ from which we sampled the x_{ij} as a control parameter that we will vary in the following.

The resulting opening angle distributions are shown in figure 21. Each of the three diagrams represents a different m, i.e., a different type of miscalibration that can be read off from the diagram's title. The different values σ of the standard deviations of the Gaussians can be found on the x-axis of each diagram. As in chapter 2, we sketch the resulting opening angle distributions $\alpha'_{m\sigma}(k)$ by its $X\%$ quantiles denoted by $Q'^{X}_{m\sigma}$, where $X \in \{16, 50, 84\}$. The range of the values of σ includes possible realistic values for the standard deviation of the fluctuations, which could be in the real data the detector is collecting at the moment [14].

We can observe the following effects from figure 21:

- As mentioned above, the chosen ranges for n include possible realistic values of the standard deviation. If we compare the time miscalibration with the angle miscalibration in the chosen ranges, we see that the predictions are tendential worse for typical time miscalibrations than for typical angle miscalibrations. This observation agrees with the result of section 2.1 for the inter-DU case. A more precise comparison of the time and angle case for a realistic estimate of σ can be seen in the next section.

- A rather surprising result is that the miscalibration of the z coordinate up to a standard deviation of $\sigma = 0.4\,\mathrm{m}$ almost does not influence the GNNs performance. Therefore, typical z miscalibrations are not as bad as typical values for the time and azimuth miscalibration. In table 2, the quantiles' values in the z case are shown because one can not see a difference in the diagram. An upward trend is identifiable for the quantiles, although some values behave contrary, e.g., the 84% quantile of $\sigma = 0.4\,\mathrm{m}$ is smaller than the 84% quantile of $\sigma = 0.35\,\mathrm{m}$. This can be due to approximation errors that happen because of our usage of one miscalibration per file instead of one miscalibration per event (see explanation at the beginning of this section) and because we use only a limited number of files.

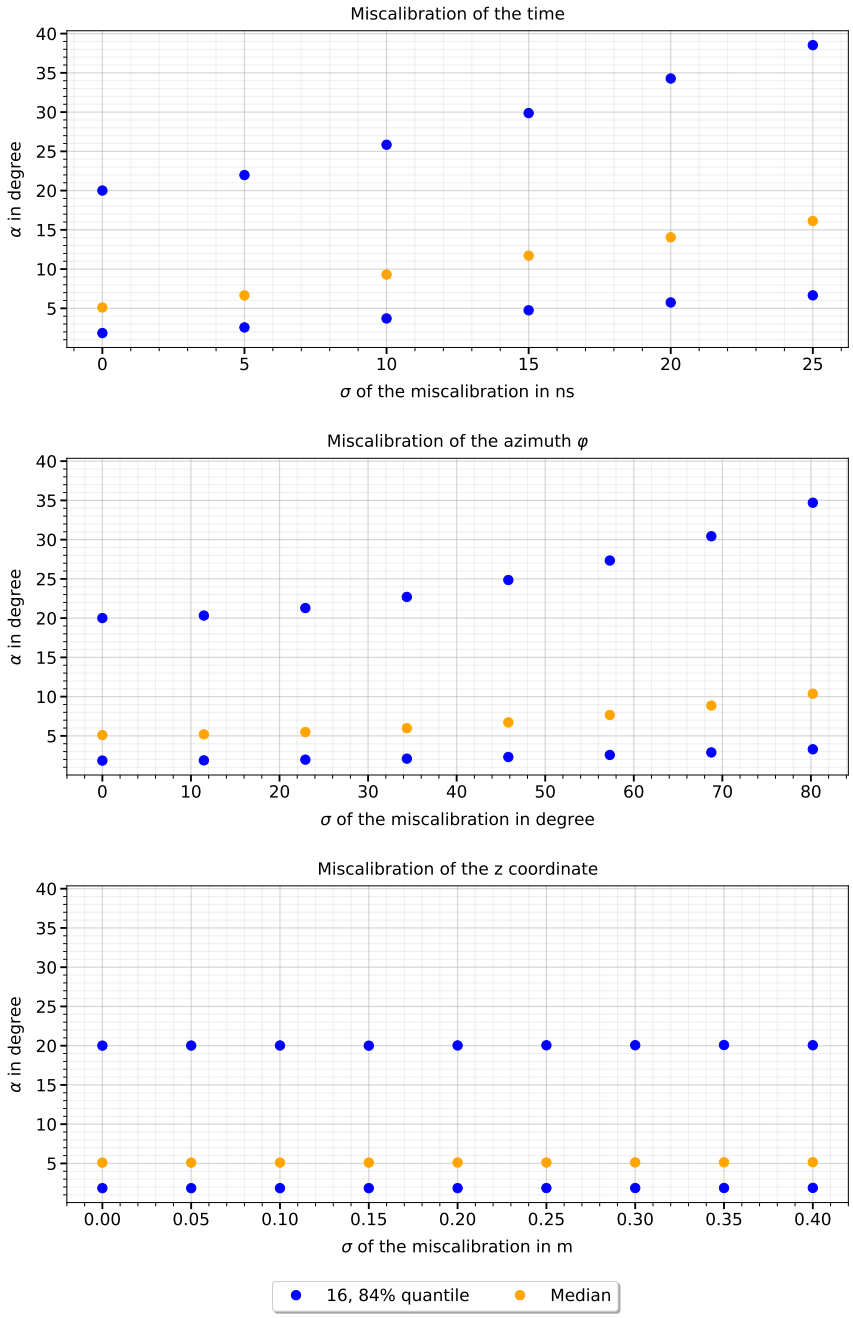

Figure 21: The opening angle distributions $\alpha'_{m\sigma}(k)$ for track events are sketched via their quantiles over the standard deviations σ that belong to the Gaussian distribution used for sampling the miscalibrations. The type of miscalibration can be found in the title of each diagram.

	16% quantile in °	Median in °	84% quantile in °
$\sigma = 0.00\,\mathrm{m}$	1.8499	5.1045	20.0078
$\sigma = 0.05\,\mathrm{m}$	1.8494	5.1019	20.0166
$\sigma = 0.10\,\mathrm{m}$	1.8518	5.1071	20.0250
$\sigma = 0.15\,\mathrm{m}$	1.8510	5.1075	20.0014
$\sigma = 0.20\,\mathrm{m}$	1.8572	5.1194	20.0298
$\sigma = 0.25\,\mathrm{m}$	1.8646	5.1162	20.0524
$\sigma = 0.30\,\mathrm{m}$	1.8648	5.1260	20.0628
$\sigma = 0.35\,\mathrm{m}$	1.8723	5.1448	20.0841
$\sigma = 0.40\,\mathrm{m}$	1.8852	5.1563	20.0602

Table 2: This table shows the quantile values for the z miscalibration in figure 21 for different standard deviations σ. The values of the quantiles are rounded to four decimal places.

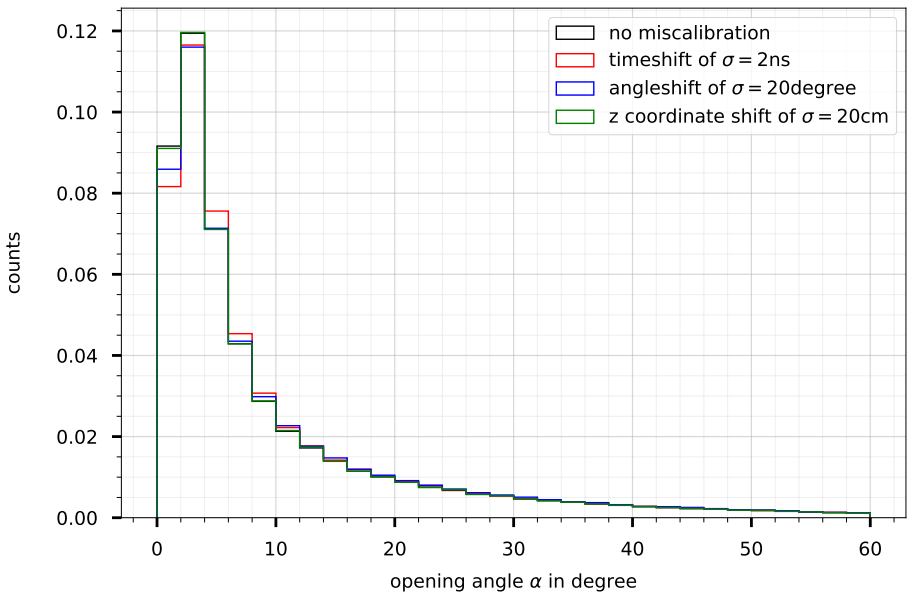

Figure 22: In each of the four histograms, the opening angle distribution $\alpha'_{m\sigma}(k)$ for the miscalibration type m, the standard deviation σ, and the track event k is shown. Both labels m and σ can be found in the legend for each histogram. The binsize is 2° and the histograms are normalized.

3.1.2 Comparing the performance for a specific calibration quality

In this section we want to compare the opening angle distributions $\alpha'_{m\sigma}(k)$ directly for different miscalibration types m. There is no obvious way how to compare, e.g., a time miscalibration of $\sigma = 10\,\text{ns}$ with a z coordinate miscalibration of $\sigma = 0.2\,\text{m}$. However, we could repeat the procedure in section 2.1.3 and estimate the standard deviation of systematic deviations between different DOMs belonging to the same DU in the real detector. Then we can sample and apply the miscalibrations as explained in section 3.1.1 with the estimated σ and compare the resulting distributions $\alpha'_{m\sigma}(k)$ directly to see which of these impairs the GNN's performance the most.

For time miscalibrations, we use the estimate $\sigma = 2\,\text{ns}$. The difference to the value used in section 2.1.3 results because we expect inter-DOM time miscalibrations to be smaller than inter-DU time miscalibrations. We estimate $\sigma = 20°$ for the azimuth miscalibrations, and $\sigma = 0.2\,\text{m}$ for the z coordinate miscalibrations [14].

In figure 22 the opening angle distributions $\alpha'_{m\sigma}(k)$ are shown. The black histogram is the opening angle distribution for the case of no miscalibration. The other three diagrams belong to different $m \in \{t, \varphi, z\}$ which can be read off from the diagram's legend. We can observe the following differences between the four histograms:

- The azimuth case has $\approx 0.43\,\%$ more counts in the first bin, i.e., the bin with values between 0° and 2°, than the time case. This result agrees with our observation from section 3.1.1 that typical angle miscalibrations have less influence on the performance than typical time miscalibrations.
- The z case has $\approx 0.51\,\%$ more counts in its first bin than the azimuth case.
- The case with no miscalibration has the highest bin. It is $\approx 0.059\,\%$ higher than the bin for the z case. This small value also agrees with the results of section 3.1.1 where we saw that the z miscalibrations almost have no impact on the resulting distribution $\alpha_{zn}(k)$.
- There is no significant qualitative difference between the three cases, e.g., there are no new maxima created in any of these cases.

3.2 Shower events

3.2.1 Influence of the calibration quality

In this section we will repeat the procedure from section 3.1.1 with shower events. We can read off the following effects from figure 23:

- The 84% quantiles for the time case and the azimuth case are more similar now than for track events in figure 21, although the 84% quantile for the time case still has larger quantiles than the azimuth case. The value for the highest σ in our range of typical values is even almost equal. This observation suggests that the difference between time and angle miscalibrations is not as crucial as for track events for the worst 16 % of predictions.
- The same holds for the 16% quantile and the median, except that the quantiles for the highest σ in our chosen range are not almost equal. The time case still has higher quantiles, i.e., a worse impact on the GNNs performance than the azimuth case over the chosen ranges.
- Again, no reaction to the z miscalibration is identifiable in the figure. This is because the differences between the values for different σ are again very small.

3.2.2 Comparing the performance for a specific calibration quality

In this section we will look again at the histograms for the estimated 1σ values in the real detector. The histograms can be seen in figure 24. The following effects are observed:

- The height of the first bin $\alpha \in [0°, 2°]$ in figure 22 for track events was the highest for the case of no miscalibration, then the z case, the angle case, and finally the time case. We denote this shortly as none $>$ z $>$ angle $>$ time. The order for shower events is different, namely z $>$ time $>$ none $>$ angle.
- The difference between the z case and the time case is $\approx 0.0054\,\%$.
- On the other hand, the difference between the time case and the none case is $\approx 0.015\,\%$.
- The difference between the none case and the azimuth case is $\approx 0.018\,\%$.

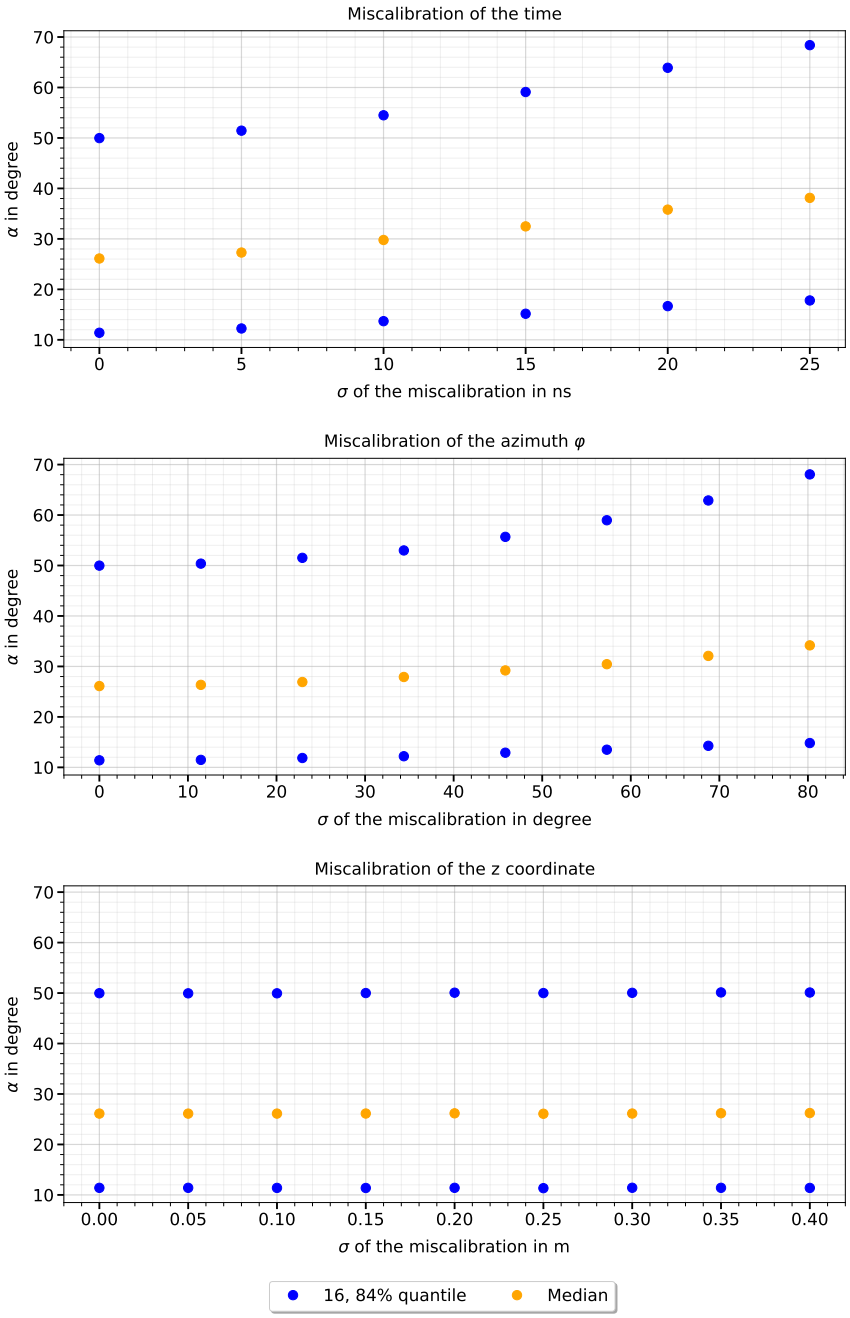

Figure 23: The opening angle distributions $\alpha'_{m\sigma}(k)$ for shower events are sketched via their quantiles over the standard deviations σ that belong to the Gaussian distribution used for sampling the miscalibrations. The type of miscalibration can be found in the title of each diagram.

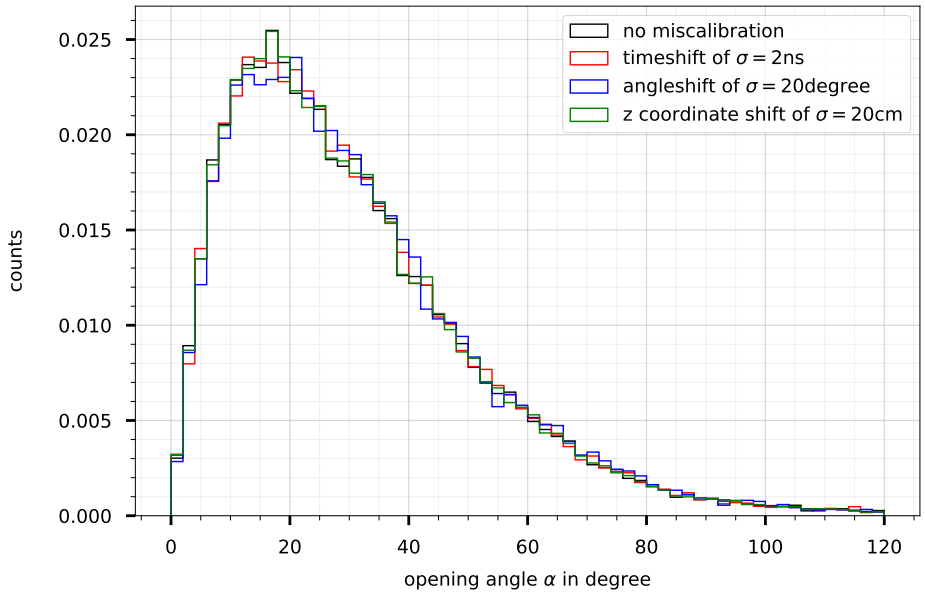

Figure 24: In each of the four histograms, the opening angle distribution $\alpha'_{m\sigma}(k)$ for the miscalibration type m, the standard deviation σ, and the shower event k is shown. Both labels m and σ can be found in the legend for each histogram. The binsize is 2° and the histograms are normalized.

- These values are one order of magnitude smaller than for the track case. These may be random fluctuations with no further meaning.
- It is difficult to make further statements for this figure because no obvious patterns emerge. Each type of miscalibration changes the original distribution seemingly randomly for each bin.
- These results confirm our observation in section 2.2.3, where we saw that the type of miscalibration is not as important for the best predictions of shower events as for the best predictions of track events.

3.3 Atmospheric muon events

3.3.1 Influence of the calibration quality

The procedure explained in section 3.1.1 is repeated here for AM events to see whether there are differences between the track and the shower case. The opening angle distributions are shown in figure 25. We dispense a further discussion of the effects because they are qualitatively the same as for track events in section 3.1.1. Only the quantiles' values are much smaller than for track events, which is expected because AM events generate more hits per event than track events.

3.3.2 Comparing the performance for a specific calibration quality

The procedure in this section is analogous to section 3.1.2, but for AM events. The histograms are shown in figure 26, which show us the following properties:

- For all miscalibration cases, the first bin $\alpha \in [0°, 2°]$ is the highest bin, respectively. In general, we see that bin n is larger than bin $n+1$ for the first several bins, which was different for track events and especially different for shower events. This result shows again that the best predictions are made for AM events.
- If we compare which of the miscalibration leads to the highest bin $\alpha \in [0°, 2°]$, then we obtain the order none $>$ z $>$ angle $>$ time. This order is the same as for track events.
- The none case produces a $\approx 0.14\,\%$ higher bin than the z case. This value is larger than for track events with $\approx 0.059\,\%$. The z miscalibration impairs the best predictions for AM events more than for track events.
- The z case has an $\approx 0.99\,\%$ higher bin than the azimuth case. This value is also larger than the one for track events, which was $\approx 0.51\,\%$.
- Furthermore, the azimuth case produces a bin which is $\approx 1.07\,\%$ larger than for the time miscalibrations. Again, this value is larger than for track events where we obtained the value $\approx 0.43\,\%$.
- Summarized, the relative order of the impact of the different miscalibrations on the best predictions is the same as for track events. However, the miscalibrations'

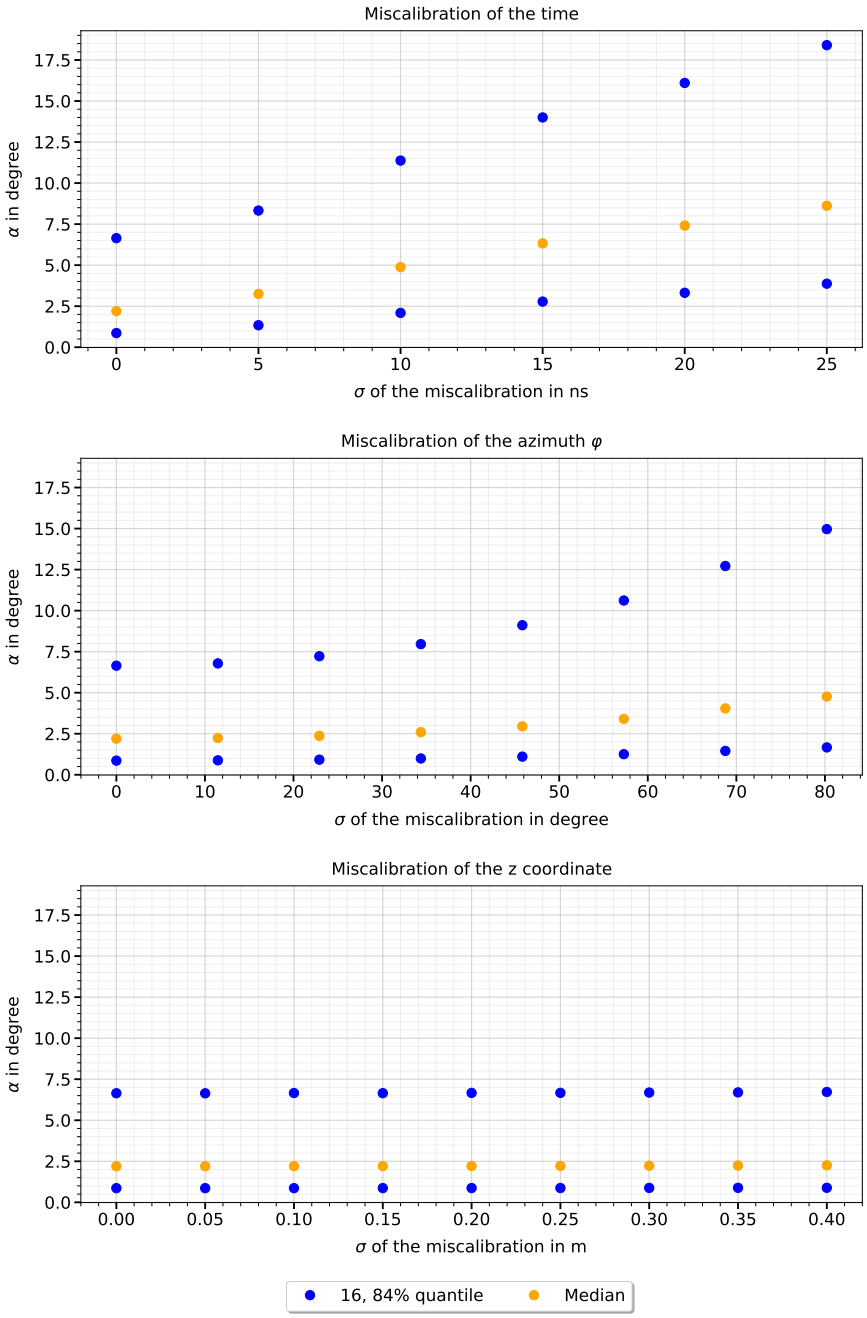

Figure 25: The opening angle distributions $\alpha'_{m,\sigma}(k)$ for AM events are sketched via their quantiles over the standard deviations σ that belong to the Gaussian distribution used for sampling the miscalibrations. The type of miscalibration can be found in the title of each diagram.

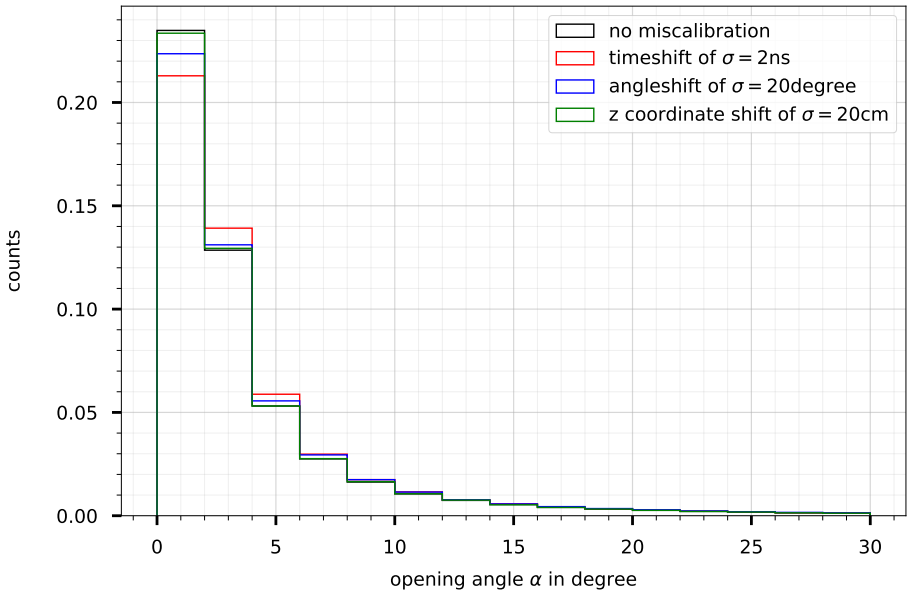

Figure 26: In each of the four histograms, the opening angle distribution $\alpha'_{m\sigma}(k)$ for the miscalibration type m, the standard deviation σ, and the AM event k is shown. Both labels m and σ can be found in the legend for each histogram. The binsize is 2° and the histograms are normalized.

impact is clearly larger for the AM case than for the track case.

4 Probabilistic manipulation of individual PMTs

4.1 Track events

4.1.1 Influence of the calibration quality

In this section, we will go one level deeper and explore inter-PMT effects., i.e., miscalibrations between different PMTs belonging to the same DOM. This means that we sample a miscalibration for each PMT of the detector and apply it. However, the inter-DU and inter-DOM effects of the previous chapter will be carefully separated from the inter-PMT effects during the manipulation of the detector geometry because we want to see the effects independently from each other. In the rest of this section, we will use the phrases "inter-PMT effects" and "PMT effects" as synonyms. Analogous for inter-DOM effects.

The only reasonable miscalibration type of this chapter will be the time miscalibration because the azimuth and the z coordinate of all PMTs belonging to the same DOM are not independent of each other. The aim is to compare the performance of the GNN mainly to the inter-DOM effects caused by a time miscalibration in chapter 3 to see whether inter-DOM or inter-PMT miscalibrations affect the GNN's performance worse.

Because we only use time as a miscalibration in this section, we can drop the label m from the previous sections. The label σ denotes, precisely like in chapter 3, the standard deviation of the Gaussian distribution from which we sample the time miscalibrations. The label k denotes, as always, one track event (or a shower/AM event in later sections). An example of our notation would be the expression $\alpha_{10\,\mathrm{ns}}(k)$, which denotes the opening angle between the predicted and the true direction of the particle causing track event k after a miscalibration was applied to each PMT, which was sampled from a Gaussian with a standard deviation $\sigma = 10\,\mathrm{ns}$. Because we do not compare the PMT effects to the inter-DU effects from chapter 2 in this section, we are free to reuse the symbol α for the opening angles suffering from inter-PMT effects, while α' remain the opening angles for the inter-DOM effects. For the same reasons as explained in section 3.1.1, we will not produce one detector miscalibration per event but per file. This method should give us a sufficient approximation to the random experiment "Sample one time miscalibration per PMT per event from a Gaussian, produce a miscalibrated graph, feed it into the GNN, and calculate the opening angle between the predicted and the true direction".

At the beginning of this section, we mentioned that we do not want to let inter-DU and inter-DOM effects interfere with the inter-PMT effects that we are interested in in this section. To achieve this, we use the following method: let x_{ijk} be the sampled miscalibration of the k-th PMT of DOM j which belongs to DU i. Therefore, $i \in \{2, 3, 4, 5\}$, $j \in \{1, \ldots, 18\}$, and $k \in \{1, \ldots, 31\}$ since there are 31 PMTs built into each DOM. Let us denote the mean of all miscalibrations of DOM j belonging to DU i by μ_{ij}, i.e.,

$$\mu_{ij} = \frac{1}{31} \sum_{k=1}^{31} x_{ijk}. \tag{8}$$

Analogous to section 3.1.1, we define a new quantity $\tilde{x}_{ijk} = x_{ijk} - \mu_{ij}$. The mean of this quantity over all k, denoted as $\tilde{\mu}_{ij}$, is equal to zero. As shown in section 3.1.1, the sample standard deviation $\tilde{\sigma}_{ij}$ of the quantity $\tilde{x}_{ijk}$ is equal to the sample standard deviation σ_{ij} of quantity x_{ijk}. Furthermore, if we calculate the average miscalibration over the whole DU i, we obtain

$$\frac{1}{18 \cdot 31} \sum_{j=1}^{18} \sum_{k=1}^{31} \tilde{x}_{ijk} = \frac{1}{18} \left[\frac{1}{31} \sum_{k=1}^{31} \tilde{x}_{i,1,k} + \cdots + \frac{1}{31} \sum_{k=1}^{31} \tilde{x}_{i,18,k} \right] = \tag{9}$$

$$= \frac{1}{18} \left[\tilde{\mu}_{i,1} + \cdots + \tilde{\mu}_{i,18} \right] = 0, \tag{10}$$

where we used $\tilde{\mu}_{ij} = 0$ in the last step. This result means that the quantity $\tilde{x}_{ijk}$ also eliminates all inter-DU effects, which leaves only our wanted inter-PMT effects.

The opening angle distributions obtained after applying the above procedure are shown in figure 27. Let us denote the quantiles in figure 21 by $Q^X_{\mathrm{DOM},\sigma}$ and the quantiles in figure 27 by $Q^X_{\mathrm{PMT},\sigma}$, where $X \in \{16, 50, 84\}$ as before, σ is the standard deviation used for sampling the miscalibrations, and the subscript DOM or PMT represents whether we use the quantile in the inter-DOM plot or in the inter-PMT plot. The following effects can be observed:

1. In a direct comparison with the inter-DOM effects in figure 21, we can see that the 16% quantiles of both diagrams have similar values.

2. However, it holds $Q^X_{\mathrm{DOM},\sigma} \leq Q^X_{\mathrm{PMT},\sigma}$ for almost all $X \in \{50, 84\}$ and almost all σ. The difference between both increases for larger σ in the chosen range. For

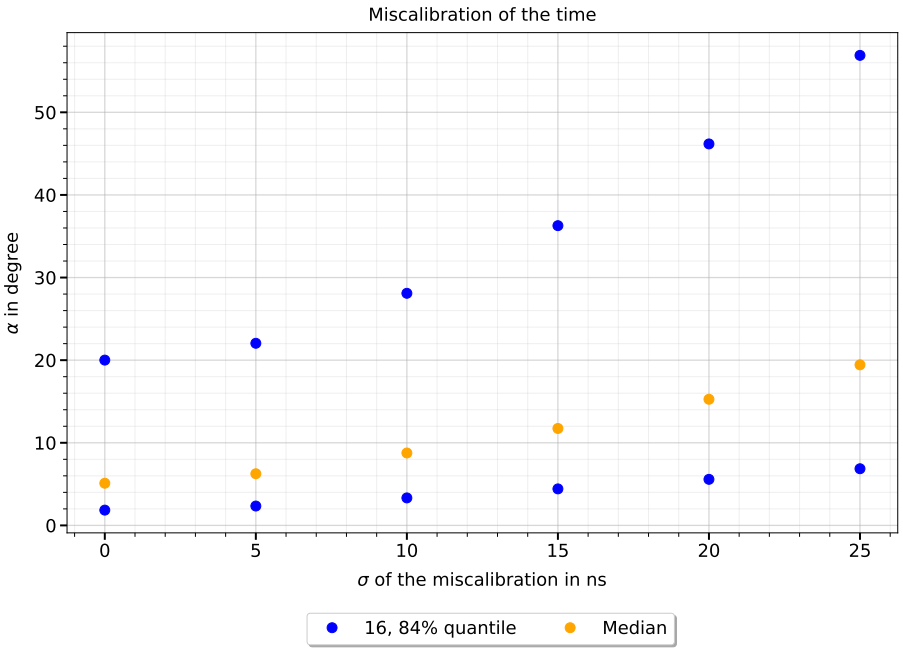

Figure 27: The opening angle distributions $\alpha_\sigma(k)$ for track events k are sketched via their quantiles. The quantity σ on the x-axis is the standard deviation used for sampling the time miscalibrations which were applied to the time coordinate of each PMT.

$\sigma \leq 5\,\text{ns}$ the equality case seems to hold approximately. This suggests that the GNN's predictions are more sensitive to inter-PMT effects than to inter-DOM effects for the worse 50 % of predictions and for the larger values of σ in the given range.

4.1.2 Comparing the performance for a specific calibration quality

The estimated standard deviation of the real detector for inter-PMT time miscalibrations is 2 ns [14]. Analogous to section 3.1.2, we visualize the distribution $\alpha_{2\,\text{ns}}(k)$ for the estimated real value of σ and compare it to the inter-DOM case in figure 22. The histogram of $\alpha_{2\,\text{ns}}(k)$ can be seen in figure 28. To make the comparison between inter-PMT effects and inter-DOM effects easier, we also plotted the inter-DOM time case from figure 22 into figure 28. As always, the opening angle distribution for the case of no miscalibrations is also shown.

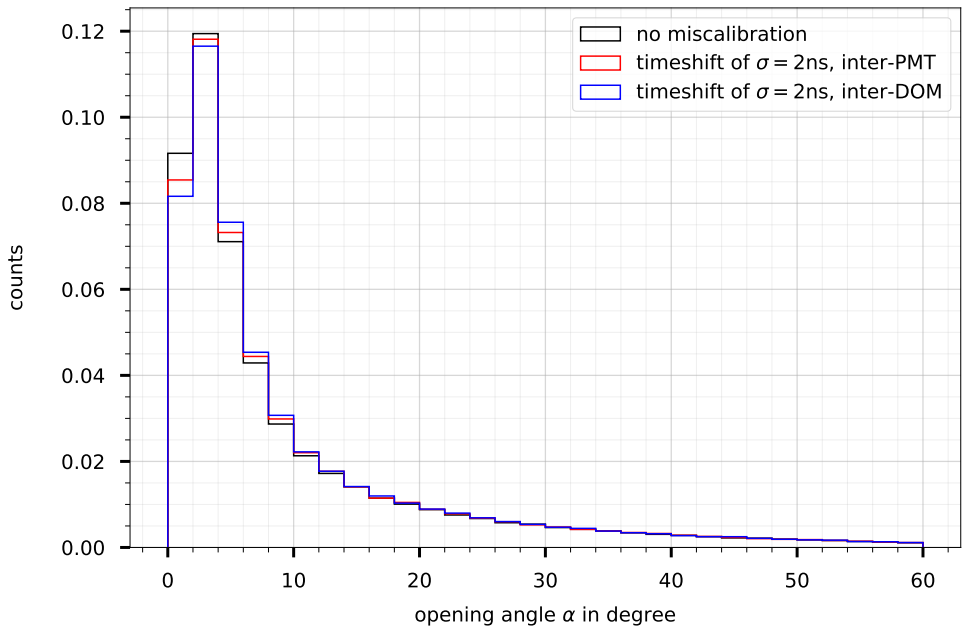

Figure 28: In this figure, three histograms of opening angle distributions are plotted. The red histogram is the distribution $\alpha_{\mathrm{PMT},\sigma}(k)$ where $\sigma = 2\,\mathrm{ns}$ and k is a track event. Analogous for the blue histogram with $\alpha_{\mathrm{DOM},\sigma}(k)$. To compare these distributions with the case of no miscalibration, the black histogram is also plotted again. The binsize is 2° and the histograms are normalized.

We can observe the following effects:

- For the first bin $\alpha \in [0^\circ, 2^\circ]$, the none case produces the largest bin. The PMT case has the second-highest bin and the DOM case the smallest. We denote this shortly as none $>$ PMT $>$ DOM.
- The first bin of the none case is $\approx 0.62\,\%$ higher than the bin of the PMT case.
- The difference between the first bins of the PMT case and the DOM case is smaller, namely $\approx 0.38\,\%$.
- The order of the second bin $\alpha \in [2^\circ, 4^\circ]$ is also none $>$ PMT $>$ DOM. We observe that inter-DOM effects impair the best predictions for track events more than PMT effects for our estimated $\sigma = 2\,\mathrm{ns}$. This is rather surprising, because we saw in figure 27 that PMT effects broaden the distribution more than DOM effects for sufficiently high σ. It seems that the opposite effect holds for the distribution of

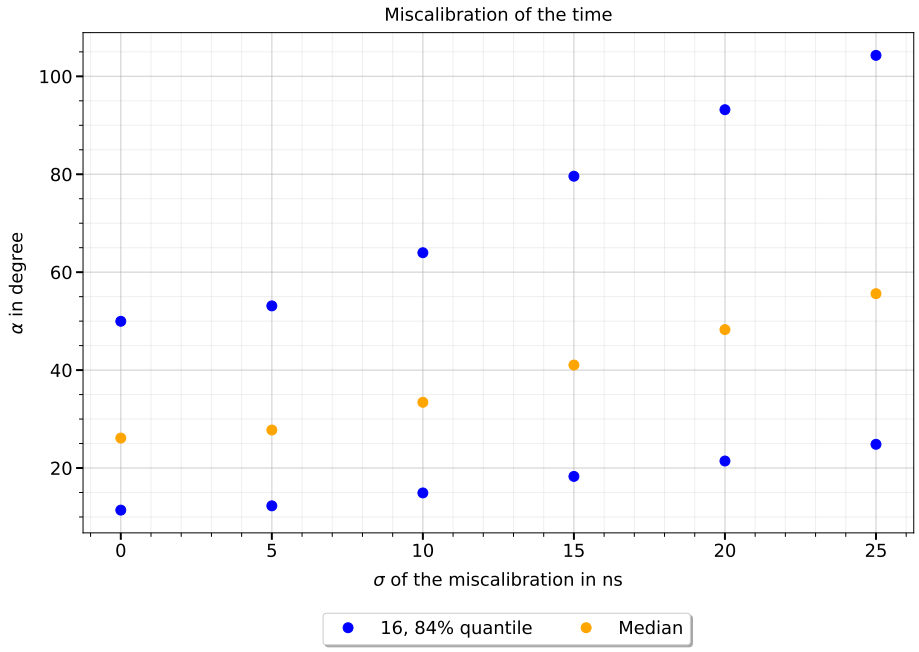

Figure 29: The opening angle distributions $\alpha_\sigma(k)$ for shower events k are sketched via their quantiles. The quantity σ on the x-axis is the standard deviation used for sampling the time miscalibrations which were applied to the time coordinate of each PMT.

the best 16 % of the predictions and for realistic, small values of σ.

4.2 Shower events

4.2.1 Influence of the calibration quality

This section is analogous to section 4.1.1, but for shower events. If we compare figure 23 and figure 29, we can see the following effects:

- The relation $Q^X_{\text{PMT},\sigma} \geq Q^X_{\text{DOM},\sigma}$ holds for the most values of X and σ, especially for larger values of σ in the chosen range.
- The 16% quantiles for the PMT and the DOM case are not as approximately equal any more as for track events in figure 27. Even on the best 16 % of predictions for shower events, we can see that PMT effects impair the GNNs performance more

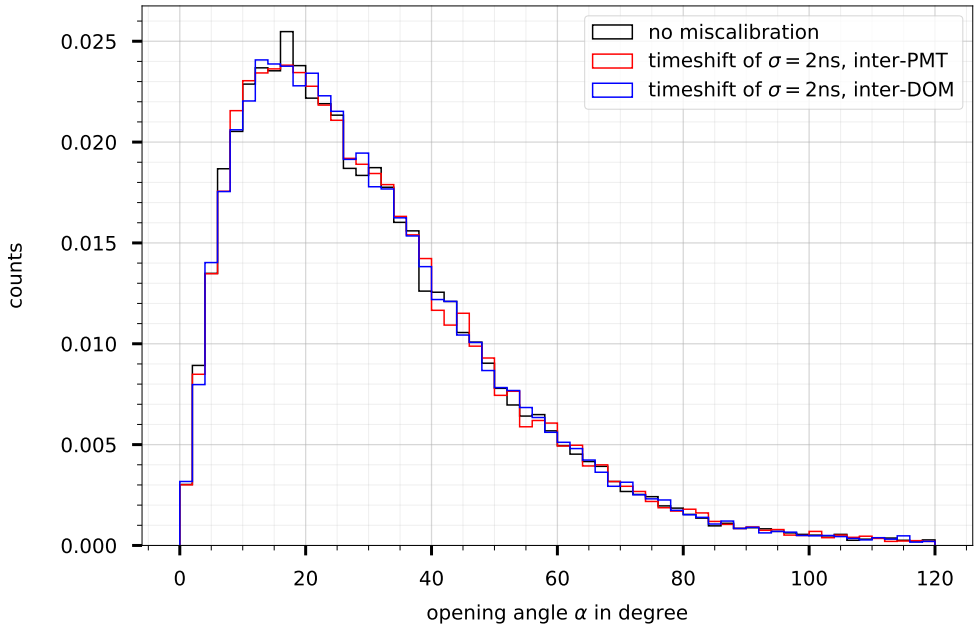

Figure 30: In this figure, three histograms of opening angle distributions are plotted. The red histogram is the distribution $\alpha_{\mathrm{PMT},\sigma}(k)$ where $\sigma = 2\,\mathrm{ns}$ and k is a shower event. Analogous for the blue histogram with $\alpha_{\mathrm{DOM},\sigma}(k)$. To compare these distributions with the case of no miscalibration, the black histogram is also plotted again. The binsize is $2°$ and the histograms are normalized.

than DOM effects in the chosen range of σ.

4.2.2 Comparing the performance for a specific calibration quality

In this section we visualize the distribution $\alpha_{2\,\mathrm{ns}}(k)$ for shower events. The histogram of this distribution can be seen in figure 30. To compare the PMT effects with the DOM effects, the time case of figure 24 is also plotted into figure 30.

The following observations can be made:

- The first bin $\alpha \in [0°, 2°]$ has the order DOM > none > PMT.
- The DOM case has a $\approx 0.015\,\%$ larger bin than the none case.
- The none case is a tiny bit larger than the PMT case, namely $\approx 0.0018\,\%$.

- The second bin $\alpha \in [2°, 4°]$ has the order none > PMT > DOM, which we already know from the track events. The differences are larger than for the first bin: $\approx 0.044\,\%$ for the difference between the none and the PMT case and $\approx 0.051\,\%$ for the difference between the PMT and the DOM case.
- The height difference between each bin for the different cases is relatively small compared to the differences observed for track events. It could be that most of the differences are just random fluctuations without further meaning. In this case, there would be no substantial difference between PMT and DOM effects for realistic values of σ.

4.3 Atmospheric muon events

4.3.1 Influence of the calibration quality

This section is again analogous to section 4.1.1, but for AM events. If we compare figure 25 and figure 31, we can see that the relation $Q^X_{\text{PMT},\sigma} \leq Q^X_{\text{DOM},\sigma}$ holds for the most values of X and σ, which is the exact opposite of the track and shower case. It seems that DOM effects impair the GNN's performance for reconstructing AM events more than PMT effects, at least for larger values of σ in the chosen range.

4.3.2 Comparing the performance for a specific calibration quality

In figure 32 the distribution $\alpha_{2\,\text{ns}}(k)$ is visualized for AM events. To compare the distribution for the PMT effects with the DOM effects, the distribution from the time case in 26 is also plotted. We can observe the following phenomena:

- The relative order of the first bin $\alpha \in [0°, 2°]$ is none > PMT > DOM. We already saw this order for track events in section 4.1.2.
- The first bin for the none case is $\approx 1.4\,\%$ higher than for the PMT case. This difference is larger than the one for the track case, which was $\approx 0.62\,\%$.
- Analogously, the first bin for the PMT case is $\approx 0.83\,\%$ higher than for the DOM case. This value is also higher than the one for track events, which was $\approx 0.38\,\%$.
- Summarized, we see that inter-DOM effects impair the best predictions for AM

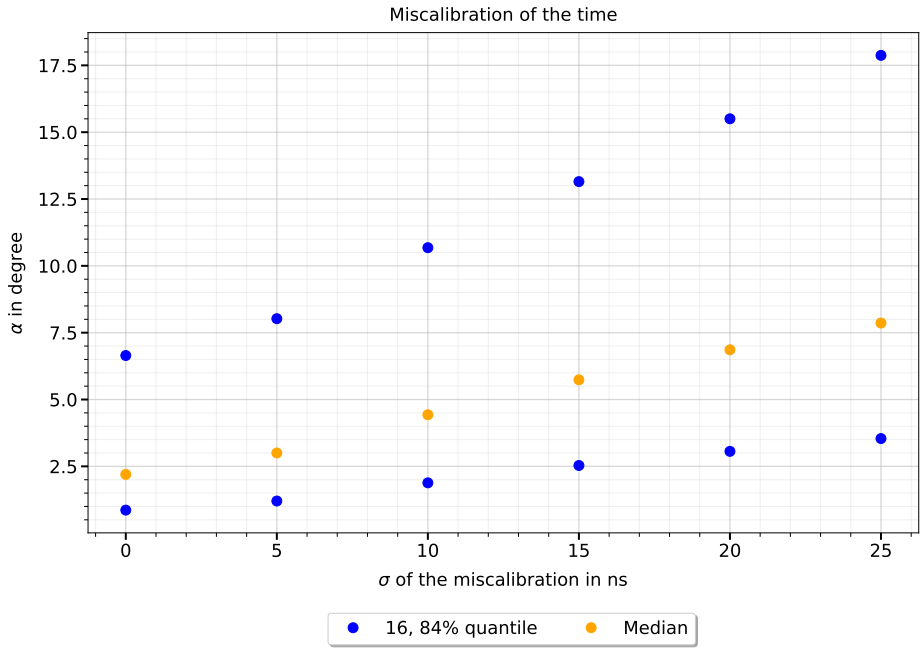

Figure 31: The opening angle distributions $\alpha_\sigma(k)$ for AM events k are sketched via their quantiles. The quantity σ on the x-axis is the standard deviation used for sampling the time miscalibrations which were applied to the time coordinate of each PMT.

events more than PMT effects. This observation was also made for track events, but the effect has a higher impact on AM events.

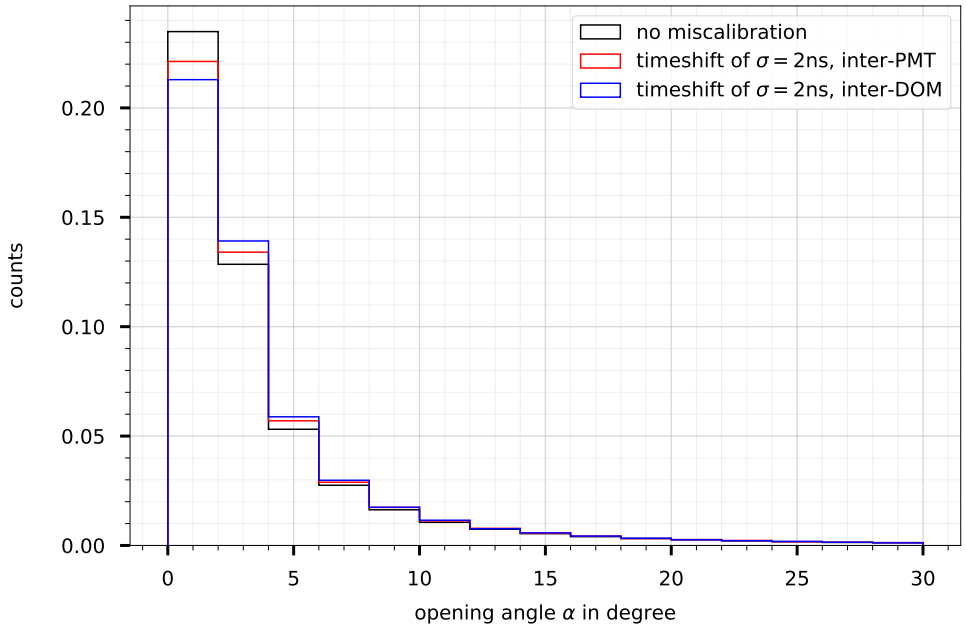

Figure 32: In this figure, three histograms of opening angle distributions are plotted. The red histogram is the distribution $\alpha_{\mathrm{PMT},\sigma}(k)$ where $\sigma = 2\,\mathrm{ns}$ and k is an AM event. Analogous for the blue histogram with $\alpha_{\mathrm{DOM},\sigma}(k)$. To compare these distributions with the case of no miscalibration, the black histogram is also plotted again. The binsize is 2° and the histograms are normalized.

5 Conclusions

5.1 Summary of the results

In chapter 2, time and azimuth miscalibrations n were applied to each DU m in a non-probabilistic manner. The miscalibrations n took on values that could be found in the real detector at this point of the construction phase. Furthermore, if we assume that the real detector's systematic miscalibrations are normal distributed, we estimated the standard deviation σ for both miscalibration types and compared the opening angle distributions. The following observations were made:

- The opening angle distributions were qualitatively similar for miscalibrations on DUs 2 and 5, and for miscalibrations on DUs 3 and 4. An explanation for this symmetry was found after looking at the detector geometry in figure 10.

- The 16% quantiles increase slower for increasing miscalibrations $|n|$ than the other quantiles. This observation means that the best predictions do not get as impaired as the worst predictions by increasing miscalibrations $|n|$.
- For time miscalibrations, the relation $Q^X_{mn} \leq Q^X_{m,-n}$ holds for the most values $n > 0\,\text{ns}$, m, and X. This means that negative time miscalibrations impair the performance of the GNN worse than positive time miscalibrations.
- For angle miscalibrations, the quantiles were more symmetric under an exchange $n \to -n$ than for the time case.
- If we apply unrealistically large miscalibrations $|n|$, the performance of the GNN gets better for increasing $|n|$ in some situations.
- For track and AM events, realistic time miscalibrations impaired the best predictions more than realistic azimuth miscalibrations.
- The opposite holds for shower events. However, the differences were so small in this case that they could be the result of random fluctuations with no further meaning.

In chapter 3, we used a normal distribution with vanishing mean μ and standard deviation σ to sample one miscalibration for each DOM. To avoid interference with inter-DU effects from chapter 2, the mean of the miscalibrations of all DOMs belonging to one DU were adjusted to equal zero. As miscalibration types, we used, additionally to time and azimuth miscalibrations, a spatial miscalibration in the z-direction. The following results were observed:

- In the diagrams with the quantiles, we could not identify any impact of the z miscalibration on the GNN's performance. Realistic z miscalibrations do not impair the predictions as much as time or azimuth miscalibrations.
- If we compare the quantiles for time and azimuth miscalibrations over their respective intervals of realistic miscalibrations, we saw that the time miscalibrations impaired the GNN's performance tendentially more than azimuth miscalibrations. This effect is more crucial for track and AM events than for shower events.
- If we estimate a realistic value for the standard deviation σ of all miscalibration types, respectively, we can compare their opening angle distributions among each other. For the height of the first bin $\alpha \in [0°, 2°]$, we found the relative order

none > z > angle > time for track and AM events. The inequality $X > Y$, where X and Y are two miscalibration types, means that the specified bin, e.g., the first bin, is smaller for the miscalibration Y than for X. The impairment of the best predictions $\alpha \in [0°, 2°]$ is worse for AM events than for track events.

- For shower events, we found the order z > time > none > angle. However, the differences were so small that they could result from random fluctuations without further meaning.

In chapter 4, we proceeded analogous to chapter 3, but we sampled a miscalibration for each PMT. To avoid interference with inter-DOM effects, we adjusted the miscalibrations such that all PMTs belonging to one DOM have a vanishing mean miscalibration. The only miscalibration type used in this chapter was the time miscalibration since the azimuth and the z coordinate of all PMTs belonging to the same DOM are not independent quantities. We compared the inter-PMT effects of this chapter with the inter-DOM effects from chapter 3 and obtained the following results:

- For track events, the 16% quantiles were similar for the PMT case and the DOM case.
- However, the relation $Q^X_{\mathrm{DOM},\sigma} \leq Q^X_{\mathrm{PMT},\sigma}$ holds for almost all σ and $X \in \{50, 84\}$. This effect was particularly observable for the larger σ in our chosen range of possible realistic values. This observation suggests that PMT effects have more impact than DOM effects, at least on the worst 50 % of predictions and for large σ.
- Surprisingly, we saw that for a small, realistic value of σ, the first bin $\alpha \in [0°, 2°]$ of a histogram had the relative order none > PMT > DOM. This order suggests that DOM effects impair the best predictions more than PMT effects for small values of σ.
- For shower events, the relation $Q^X_{\mathrm{DOM},\sigma} \leq Q^X_{\mathrm{PMT},\sigma}$ holds for almost all values of X and σ.
- For a small, realistic value of σ, the relative order of the first bin was DOM > none > PMT for shower events. The bins' height differences are relatively small, i.e., the differences could be due to random fluctuations with no further meaning.
- Surprisingly, for AM events, the opposite relation $Q^X_{\mathrm{DOM},\sigma} \geq Q^X_{\mathrm{PMT},\sigma}$ holds for most

values of X and σ. This is particularly true for larger values of σ in our chosen range. This suggests that DOM effects impair the predictions for AM events more than PMT effects, at least for sufficiently large values of σ.

- For a small realistic value of σ, a comparison of the histograms for PMT and DOM effects reveals that the first bin $\alpha \in [0°, 2°]$ has the relative order none > PMT > DOM. This order was also observed for track events, but the height differences are larger for AM events than for track events, suggesting that the miscalibrations' impact is even higher for AM events than for track events.

5.2 Outlook

There are many possible ways to continue the topic of this thesis. One could, e.g., use a different quantity than the opening angle α to test how well the GNN's predictions are. Especially for the neutrino mass hierarchy problem, the zenith angle of a neutrino causing an event is an important quantity [13].

The GNN not only predicts the direction and the energy of the particle causing the interaction but also the uncertainty of these quantities. This fact was not mentioned until now because this uncertainty did not play a role in this thesis. However, one could repeat our studies with this uncertainty instead of the opening angle α, which we used. If we use these GNNs on real data obtained by the detector, we have no access to the opening angle between the true direction and the predicted direction. However, we would have access to the uncertainties predicted by the GNN, even if we use real detector data. If one would investigate how these uncertainties behave after applying systematic miscalibrations, one could estimate the systematic errors in the real detector's measurements by comparing the predicted uncertainties from the real data to the investigated behavior [14].

The GNNs are a relatively new attempt to analyze the data obtained by the KM3NeT detector. There exists already a standard reconstruction that can be used to reconstruct the direction and the energy of the particle causing an event with a maximum likelihood approach [14]. One could repeat the studies in this thesis for the standard reconstruction to see how robust its reconstructions are after applying systematic errors. These results could be compared to the results obtained for GNNs in this thesis to see which method is more robust under systematic errors.

5.3 Acknowledgments

Finally, I want to thank everybody who supported me during the process of writing this thesis. Special thanks go to the following persons:

- PD Dr. Thomas Eberl for being my supervisor and for always helping me out whenever I had a question.
- Daniel Guderian, who constructed and trained the GNNs that I used, for always answering my questions and for the interesting discussions about GNNs and related topics.
- Prof. Dr. Uli Katz, who offered to be the second reviewer.

References

[1] S. Adrian-Martinez et al. "Letter of Intent for KM3NeT 2.0". In: *Journal of Physics G: Nuclear and Particle Physics* 43 (2016). DOI: 10.1088/0954-3899/43/8/084001.

[2] Dr. Steve Boyd. *Neutrino Detectors and Sources*. URL: https://warwick.ac.uk/fac/sci/physics/staff/academic/boyd/stuff/neutrinolectures/lec_neutrinodetectors_writeup.pdf.

[3] CERN. *Neutrinos: after the Nobel Prize, the hunt continues*. URL: https://home.cern/news/news/physics/neutrinos-after-nobel-prize-hunt-continues.

[4] KM3NeT collaboration. URL: https://www.km3net.org/wp-content/uploads/2015/07/KM3NeT-DOM-3D-Drawing.jpg.

[5] KM3NeT collaboration. URL: https://www.km3net.org/wp-content/uploads/2015/08/KM3NeT-event.png.

[6] KM3NeT collaboration. *Astroparticle physics with ARCA*. URL: https://www.km3net.org/research/physics/astronomy-with-arca/.

[7] KM3NeT collaboration. *Astroparticle physics with ORCA*. URL: https://www.km3net.org/research/physics/particle-physics-with-orca/.

[8] KM3NeT collaboration. *KM3NeT the next generation neutrino telescopes*. URL: https://www.km3net.org/.

[9] KM3NeT collaboration. *The detectors*. URL: https://www.km3net.org/research/detector/.

[10] KM3NeT collaboration. *The sensors*. URL: https://www.km3net.org/research/detector/sensors/.

[11] JUNO Collaboration/JGU-Mainz. URL: https://neutrinos.fnal.gov/wp-content/uploads/2018/04/mass-hierarchy-jgu-mainz-web.jpg.

[12] Daniel Dominguez/CERN. *Particles of the Standard Model of particle physics*. URL: https://cds.cern.ch/images/OPEN-PHO-CHART-2015-001-1/file?size=large.

[13] PD Dr. T. Eberl. Private communication.

[14] D. Guderian. Private communication.

[15] Hyper-Kamiokande. URL: http://www.hyper-k.org/en/img/cherenkov.jpg.

[16] Prof. Dr. Uli Katz. Lecture script from a particle physics lecture in the summer semester 2020.

[17] Tom M. Mitchell. *Machine Learning*. ISBN: 0070428077.

[18] Jojo John Moolayil. *A Layman's Guide to Deep Neural Networks*. URL: https://miro.medium.com/max/700/1*KHs1Chs6TCJDTIIQVyIJxg.png.

[19] Michael Moser. "Sensitivity studies on tau neutrino appearance with KM3NeT/ORCA using Deep Learning Techniques". PhD thesis. URL: https://ecap.nat.fau.de/wp-content/uploads/2021/03/phd_thesis_michael_moser_v2.pdf.

[20] Staff/IceCube Neutrino Observatory. *A first look at how the Earth stops high-energy neutrinos in their tracks*. URL: https://icecube.wisc.edu/news/press-releases/2017/11/first-look-at-how-earth-stops-high-energy-neutrinos-in-their-tracks/.

[21] Huilin Qu and Loukas Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". In: *Phys. Rev. D 101, 056019* (2020). DOI: 10.1103/PhysRevD.101.056019.

Erklärung

Hiermit bestätige ich, dass ich diese Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe.

Fürth, den 16.04.2021

Lukas Hennig