Reconstruction of the energy of muons and muon bundles via Graph Neural Networks in the ORCA detector

Bachelor's Thesis in Physics

Presented by
Julia Häfner
Erlangen, April 14, 2021

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

Supervisor: PD Dr. Thomas Eberl

Contents

1 Introduction

Concerning the fundamental properties of neutrinos, various experiments using solar, atmospheric, reactor, and accelerator neutrinos have provided valuable evidence for non-zero neutrino masses and neutrino oscillations. The latter indicate that the neutrino flavour states $(\nu_e, \nu_\mu, \nu_\tau)$ are not directly coupled to the neutrino mass states (ν_1, ν_2, ν_3), but rather consist of a mixing, described by the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix. This matrix is a product of three rotation matrices depending on the three mixing angles θ_{12}, θ_{13} and θ_{23} and the complex CP phase δ, which is related to CP-violating processes. The oscillation experiments, however, are not sensitive to the absolute values of the neutrino masses, though they provide insight into the squared mass splittings $\Delta m^2_{ij} = m^2_i - m^2_j$. Assuming 3 neutrino flavours, the two independent squared mass splittings are the solar mass splitting $\Delta m^2_{sol} = \Delta^2_{21}$, which is responsible for oscillations observed in solar and long-baseline reactor experiments, and the atmospheric splitting $\Delta^2_{atm} = \Delta^2_{31}$, which concerns the atmospheric neutrino experiments. The values of the mixing angles and squared mass splittings were determined with a precision better than 15 %, but the ordering of the neutrino mass eigenstates (and therefore the sign of Δ^2_{atm} is still unknown: either the Normal Hierarchy (NH) with $\nu_1 < \nu_2 < \nu_3$ or the Inverted Hierarchy (IH) with $\nu_3 < \nu_1 < \nu_2$ is true. To investigate the neutrino mass hierarchy (NMH), neutrino oscillation experiments with long baselines and matter effects can be exploited, as transitions in vacuum are insensitive to the sign of Δ^2_{atm}. These matter effects are probed in the ν_μ survival probability and in the rate of $\nu_\mu \leftrightarrow \nu_e$ appearance. [1]

Next to atmospheric experiments like ICAL at INO, HyperKamiokande or PINGU, the KM3NeT-collaboration is constructing the Mton-scale undersea detector ORCA (Oscillation Research with Cosmics in the Abyss) to measure the before mentioned oscillation probabilities in order to determine the NMH. This requires a precise estimation of the neutrino energy and the zenith angle, as well as high purity of the event samples. One major problem to the identification of neutrino events are atmospheric muons, as even at the ocean depth of more than 2000 m, the number of muons being detected is larger than that of neutrinos by several orders of magnitude. [1]

In order to minimize the effect of the muon background, only upward-going events are taken into account, since upward-going muons are shielded by the earth. However, the Cherenkov light from downward-going muons might still produce a hit pattern that is reconstructed as an upward-going neutrino event. Especially muon bundles, which are multiple muons parallel to the shower axis, perpendicular to the same plane at the same time, and produced from high energy cosmic rays, can mimic neutrino events. For this reason, a Random Decision Forest (RDF) classifies the events as muon-like, ν-track-like and ν-shower-like by exploiting the information contained in the light emissions, so the muon background can be rejected. [1]

One input feature for the RDF classification is the reconstructed particle energy. So far, the so-called JGandalf reconstruction chain provides an estimation of the muon energy. This thesis will now introduce another way of reconstructing the muon energy, based on machine

learning via dynamic graph convolutional neural networks. First, the geometry of the ORCA detector will be presented, followed by the physical properties of the Monte-Carlo event generation used for training the network. Then, the working principle of the network itself will be explained, and finally, the network results will be analyzed, including a comparison of the machine learning vs. JGandalf results and a reconstruction of the muon energy with an additional parameter σ.

2 The ORCA detector

The ORCA detector is located 40 km offshore from Toulon (France), at a sea bottom depth of 2450 m. The detector volume encompasses about 8 Mton, including 64,000 PMTs (Photo Multiplier Tubes) for Cherenkov light detection. Respectively 31 PMTs of 80 mm diameter are combined to a so-called DOM (Digital Optical Module) inside a pressure resistant glass sphere. A series of 18 DOMs connected by an electro-optical cable and arranged in a flexible, vertical string then form a DU (Detection Unit). The DU is anchored to the sea bottom by a dead weight, and a buoy is connected at the top to keep it vertical. Finally, 115 DUs form a building block. While ARCA consists of two building blocks, ORCA consists of only one. As ORCA is supposed to measure the oscillation probabilities of atmospheric neutrinos and therefore detect neutrinos in the range of 1-100 GeV, it has a denser configuration than ARCA - the average DOM distance, the total unit height and the DU distance amount to about 9 m, 250 m and 20 m, respectively. The DUs are arranged in a circular shape with a radius of about 106 m, as can be seen in figure 2.1, which shows a footprint of the ORCA detector. [1]

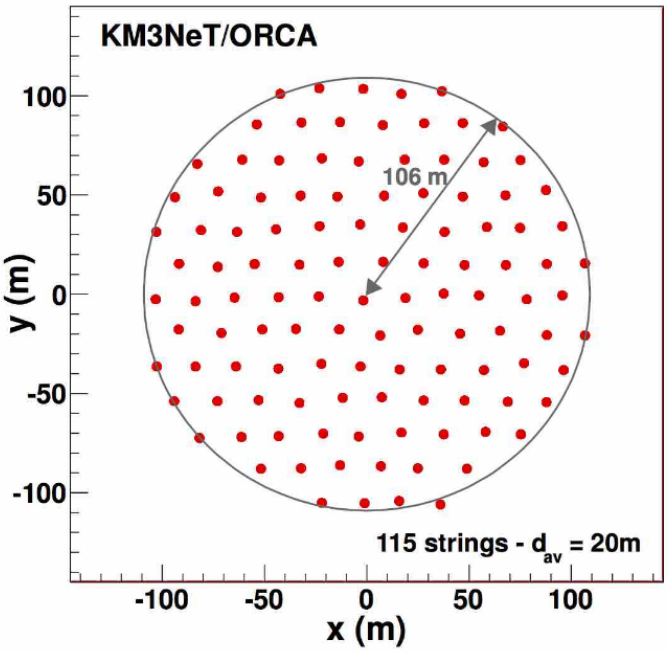

Figure 2.1: Footprint of the ORCA detector. The red dots represent the DUs from a top view. They each are about 20 m apart and form a circular shape with a radius of about 106 m. [1]

3 Muon MC event simulation

3.1 Energy loss of atmospheric muons

The primary component of the cosmic rays (CRs) arriving at the outer layer of the atmosphere includes all stable charged particles and nuclei. Atmospheric muons (and muon neutrinos) are produced by the decay of charged mesons, mainly pions and kaons [3]:

$$\pi^+(\mathrm{K}^+) \quad \rightarrow \quad \nu_\mu + \mu^+ \tag{3.1.1}$$

$$\pi^-(\mathrm{K}^-) \quad \rightarrow \quad \bar{\nu}_\mu + \mu^- \tag{3.1.2}$$

When the muons propagate through matter, they lose energy due to interaction with the traversed medium. In the low energy regime, ionization and excitation of the atoms prevail. In this case, the average energy loss dE per length dx of a particle traversing a medium is given by the *Bethe-Bloch formula* [2]

$$-\frac{\mathrm{dE}}{\mathrm{dx}} = 4\pi \mathrm{N_A} \mathrm{r_e^2} \mathrm{m_e} \mathrm{c^2} \mathrm{z^2} \frac{\mathrm{Z}}{\mathrm{A}} \frac{1}{\beta^2} \left(\ln \frac{2\mathrm{m_e c^2} \gamma^2 \beta^2}{\mathrm{I}} - \beta^2 - \frac{\delta}{2} \right), \tag{3.1.3}$$

where $\mathrm{N_A}$ is Avogadro's constant, $\mathrm{r_e}$ the classical radius of the electron, $\mathrm{m_e}$ the electron mass, c the velocity of light, z the charge of the incident particle, Z and A the atomic number and atomic weight of the absorber, $\gamma = \frac{\mathrm{E}}{\mathrm{m_0 c^2}}$ the Lorentz factor, $\beta = \frac{\mathrm{v}}{\mathrm{c}}$ the relative velocity, I the mean exitation energy of the absorber and δ a density effect parameter [2].
In the low energy domain, the loss decreases like $1/\beta^2$ and reaches a broad minimum, where the incident particles are referred to as MIPs (Minimum Ionizing Particles). For relative particles ($\beta \approx 1$) and light absorber materials ($\mathrm{Z/A} \approx 0.5$), the minimum is located at $\gamma \approx 4$, and the loss is approximately $-\frac{\mathrm{dE}}{\mathrm{dx}}_{\mathrm{min}} \approx 2 \frac{\mathrm{MeV}}{\mathrm{g/cm^2}}$, which means that the loss does not depend on the particle energy. For $\gamma > 4$, the loss increases again like roughly $2\ln\gamma$. [2]
At higher energies, radiation losses prevail over ionization and exitation. Processes in which the incident particle radiates photons include *bremsstrahlung* (deceleration in the Coulomb fields of the atoms of the traversed medium), *pair production* via virtual photons and *photonuclear reactions* via virtual gauge bosons that interact with the absorber nuclei. The loss due to these high energy processes is proportional to the energy of the incident particle. [2]
The transition between the two regimes depends on the absorber material. It is roughly located at 10 TeV in hydrogen [2]. In figure 3.1, the energy loss of a muon in water is plotted over the muon energy. The total loss is denoted with *c*, *a* refers to the contribution of ionization and exitation, *b* to the contribution of radiation processes. It is clearly visible that the slope of the loss in the high energy regime is much steeper than that of the low energy regime, which in comparison can roughly be approximated as constant. Also, the transition between the regimes lies at about 1 TeV, which is comparable to the transition in hydrogen. therefore, relativistic muons in water can be treated as MIPs for energies up to 1 TeV.

3.2 Simulation of atmospheric muons with MUPAGE

In order to simulate atmospheric muon paths in an underwater detector, various physical processes must be taken into account, such as muon production mechanisms and muon energy

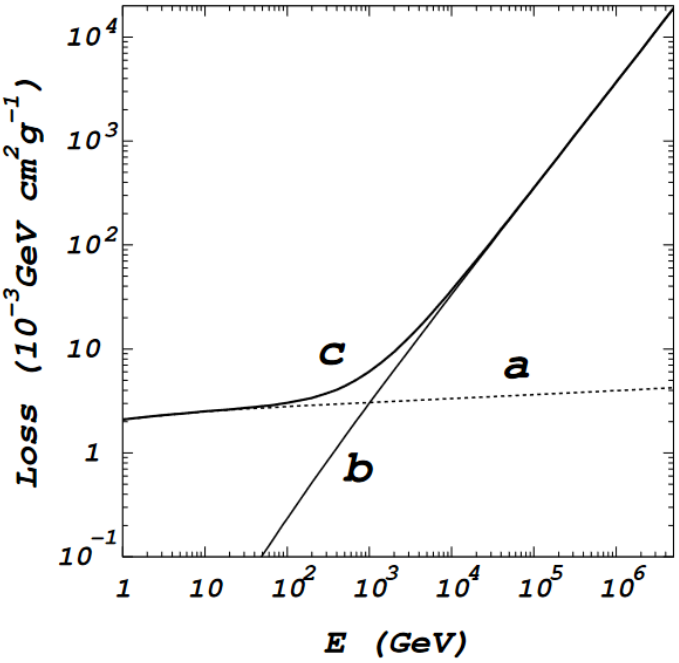

Figure 3.1: The loss of a muon in water as a function of the muon energy E. The contributions of ionization and exitation (a) and that of radiation processes (b) add up to the total energy loss (c). [4]

loss. Full MC simulations start with the HEMAS code, which takes particle cross sections as input. The first interaction points of the primary CRs are computed, as well as the propagation of electromagnetic and hadronic components of showers through the atmosphere, considering the mean free paths of the particles, the deflection by the geomagnetic field, and hadronic interactions in the atmosphere. The propagation through water is calculated via the MUSIC code, following the physical processes described in the previous section. [5]
With these results, parametric formulas for the muon flux Φ, the muon energy spectrum and the radial distance R of each muon in a bundle with respect to the bundle axis are obtained. The energy physically depends on the vertical depth h, the zenith angle θ and the multiplicity m of a bundle. Since the most energetic muons in a bundle are expected to arrive closer to the shower axis, the radial distance R from the axis must also be taken into account. The flux then follows [5]

$$\Phi(\mathrm{h},\theta,\mathrm{m}) = \frac{\mathrm{K}(\mathrm{h},\theta)}{\mathrm{m}^{\nu(\mathrm{h},\theta)}} \tag{3.2.1}$$

where K is the flux of single muons and ν is a free parameter. The differential energy spectrum is described as [5]:

$$\frac{\mathrm{dN}}{\mathrm{d}(\log_{10}\mathrm{E})} = \mathrm{G}\cdot\mathrm{E}_\mu \mathrm{e}^{\beta\mathrm{X}(1-\gamma)}[\mathrm{E}_\mu + \epsilon(1-\mathrm{e}^{-\beta\mathrm{X}})]^{-\gamma} \tag{3.2.2}$$

This assumes a power law for the initial CR energy. $\mathrm{X} = \frac{\mathrm{h}}{\cos\theta}$ denotes the slant depth, $\epsilon = \epsilon(\mathrm{h},\theta,\mathrm{R})$ and $\gamma = \gamma(\mathrm{h},\mathrm{m},\mathrm{R})$ are free parameters, β is a constant, $\mathrm{G} = \mathrm{G}(\epsilon,\gamma)$ is a normalization factor. [5]
MUPAGE uses these parametric formulas to calculate the flux and angular distribution of underwater muon bundles, taking into account the multiplicity and energy spectrum. It finally outputs a table of kinematics of events at the surface of a *can*, which is an imaginary cylinder

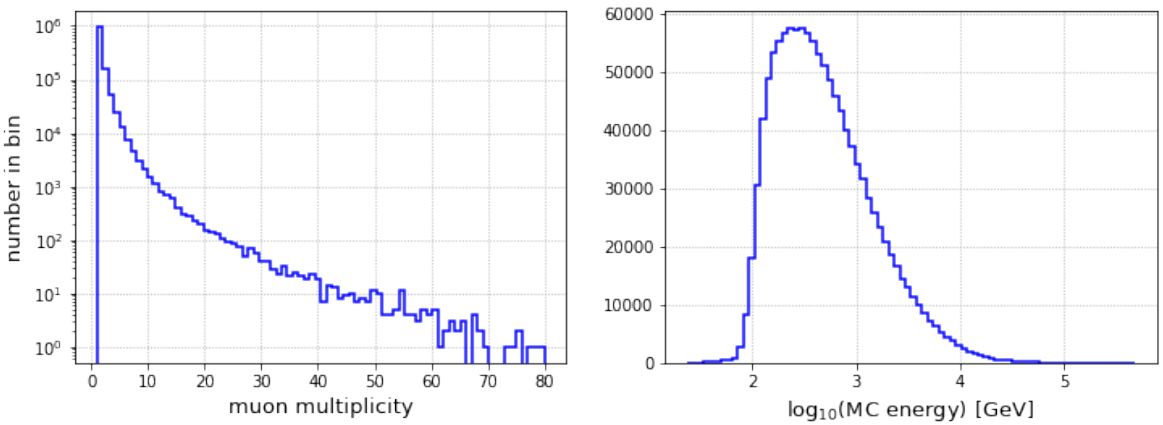

Figure 3.2: **Left:** histogram of the multiplicity distribution of the data. The y-axis is plotted logarithmically. **Right:** histogram of the energy distribution of the data. The number of bins in the left histogram has been set to 80.

surrounding the active volume of a generic detector. Bundles are generated uniformly, and a hit-or-miss method is used to sample the impact point on the *can* and to reproduce the correct dependence on the physical parameters. The muons are then propagated inside the *can* and Cherenkov light is generated. [5]

3.3 Investigation of the simulated data

The data arrays from the output of MUPAGE used in this thesis contain 1220389 simulated muon events. Only four detection lines of the ORCA detector were in the ocean at the time the simulations were produced, so they also only use four detection lines. The imaginary *can* where the events were generated exceeds the detector radius by three absorption lengths [1]. The muon multiplicity of an event is defined as the number of muons with more than 10 Monte-Carlo hits. The multiplicity distribution (figure 3.2, left), the energy distribution (figure 3.2, right) and the angular distribution (figure 3.3) are shown below. The multiplicity approximately follows a power law. In the resulting distribution, single muon events are most abundant, and the event number decreases with increasing multiplicity, reaching up to 80 muons. Since the CR energy also follows a power law and higher multiplicity is correlated with higher CR energy, this also reflects in the multiplicity distribution.
The energy distribution follows equation 3.2.2 at the sea-level depth of the ORCA detector for all multiplicities and zenith angles. The Monte-Carlo energy is defined as the sum of all the energies of the muons in a bundle. The most probable energy is roughly $\log_{10}(\mathrm{E/GeV}) \sim 2.5$. The zenith angle distribution in figure 3.3 shows that most events are found at around $\cos(\theta) = -1$ (the vertical direction), while the event number decreases, the more the direction approaches the horizon ($\cos(\theta) = 0$).
Figure 3.4 shows properties of the muon bundles. On the left, histograms of the ratio $\frac{\mathrm{E_{max}}}{\mathrm{E_{tot}}}$ for multiplicities $\mathrm{m} = 3, 7$ and 11 are shown, where $\mathrm{E_{max}}$ is the energy value of the most energetic muon of the bundle, and $\mathrm{E_{tot}}$ is the sum over all muon energies in the bundle. At $\mathrm{m} = 3$, the ratio maximum is located at around 0.5, while for $\mathrm{m} = 11$, it is close to 0.2. The energy for bundles with a high multiplicity is therefore more evenly split between the muons, and outliers where one muon carries most of the bundle energy are rarer. The right plot shows histograms of the average distance of all muons in a bundle to the detector centre for

multiplicities m = 3, 7 and 11. This is also a measure for the lateral spread of the bundle. The most common average distance for all multiplicities lies around 50 m, which is about half the detector radius. While for m = 3 and 7 some outlier bundles have an average spread of over 200 m (larger than the detector), these outliers are missing in bundles with m = 11, though more of them have an average distance of up to 100 m.

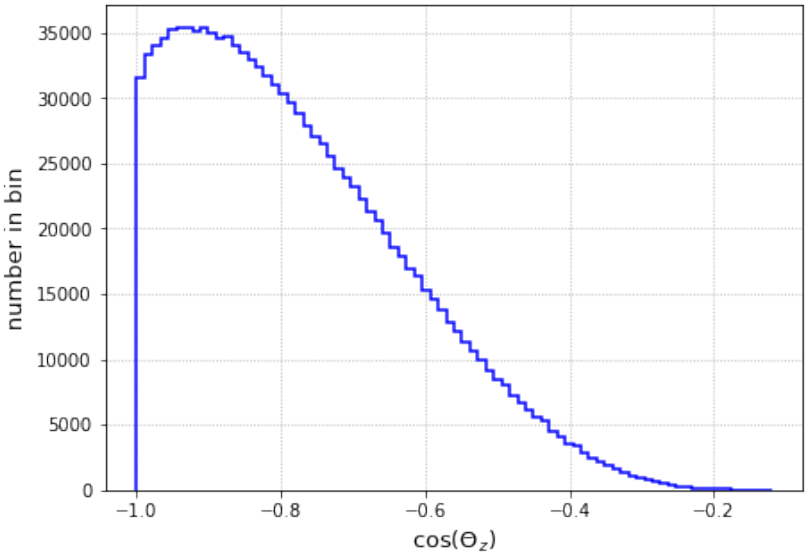

Figure 3.3: Histogram of the zenith distribution of the data. The cosine of the angle is plotted on the x-axis. Here, -1 is the vertical downgoing direction, 0 the horizontal. The number of bins has been set to 80.

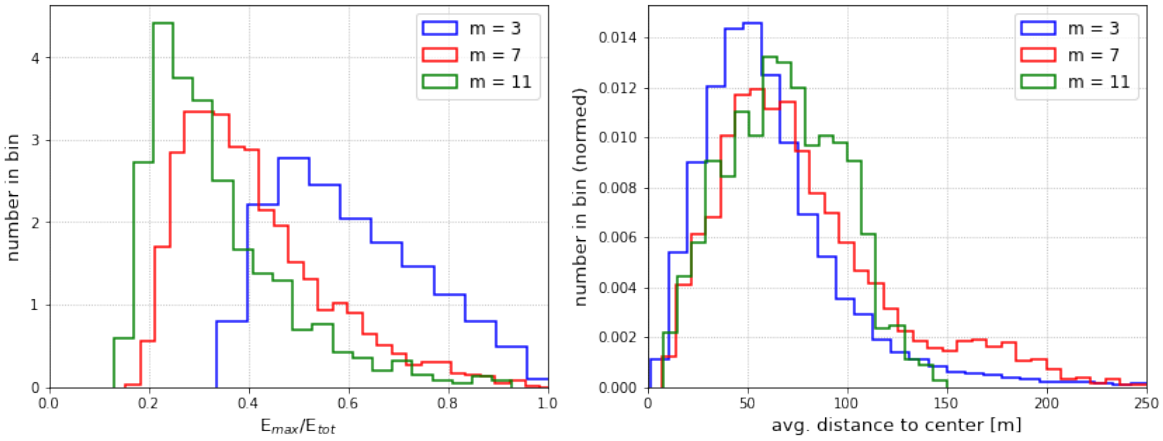

Figure 3.4: **Left:** distributions of the ratio between the largest energy value in a bundle and the total summed up bundle energy. **Right:** distributions of the average distance of the muons in a bundle to the detector centre. Both plots show bundles with multiplicities m = 3, 7 and 11. The histograms are normed for better shape comparison, since bundles with a high multiplicity are much less abundant than those with a low multiplicity. The number of bins has been set to 50.

4 Dynamic Graph Convolutional Neural Networks

A neural network describes a complicated nonlinear function F_ω with input y^{in} and output y_{out} that depends on all the weights and biases ω that map the outcome of a neuron (the smallest unit in the network) to the next:

$$y^{out} = F_\omega \left(y^{in}\right) \tag{4.0.1}$$

The neuron itself sums up all previous weighted inputs and passes them through an activation function. Multiple hidden layers of neurons can connect the input layer to the output layer. The network is trained by providing a large amount of training examples with already given outcomes (labels) and adjusting the weights and biases in such a manner that the network reproduces the training outcomes best. In other words, a target function F for a certain output is desired:

$$y^{out} = F \left(y^{in}\right) \tag{4.0.2}$$

As this is basically a problem of fitting a model to a set of data, a cost function can be defined, which describes the difference between the current nonlinear function and the target function. During training, this cost function is minimized. In a conventional neural network, all neurons are connected with every other neuron in the adjoining layers. To reduce the number of weights and therefore the computing power, a so-called filter kernel can be used, which consists only of a few neurons and is scanned across the input. This is called a Convolutional Neural Network (CNN). As the name suggests, the output is then calculated from a convolution of the input and the filter kernel $K(x - x')$. As this depends also on the difference of the original point x' and the target x, translational invariance can be exploited and the network recognizes features independent of their position in the input.

Since standard CNNs require binning the data, they work well on dense inputs like photos. The ORCA data however is rather sparse, and binning would result in a large amount of empty pixels. For a better use of the data structure, the network used in this thesis was originally created for learning on irregular point clouds. the filters are not scanned over the single event points but rather over the connecting lines (edges) between local neighbors, thus making the network a Graph Neural Network (GNN). For an n-dimensional point cloud, a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $V = \{1, ..., n\}$ and edges $\mathcal{E} = \mathcal{V} \times \mathcal{V}$ is computed, along with edge features $e_{ij} = (x_i, x_j - x_i)$ of a local patch. To calculate the update vectors, the edge features are given to the kernel network $\overrightarrow{h_\Theta}$, which is a nonlinear function consisting of three fully connected layers with a set of learnable parameters $\Theta = (\theta_1, ..., \theta_m)$ (encoding the weights of m different filters). The output of the so-called EdgeConv operation at the i-th vertex then is

$$x'_i = \frac{1}{k} \sum_{j:(i,j)\in\mathcal{E}} \overrightarrow{h_\Theta}(x_i, x_j - x_i) \tag{4.0.3}$$

where k is the number of the nearest neighboring points. This way, the new value for the central point is the average over the neighboring feature vectors. The variable choice for h

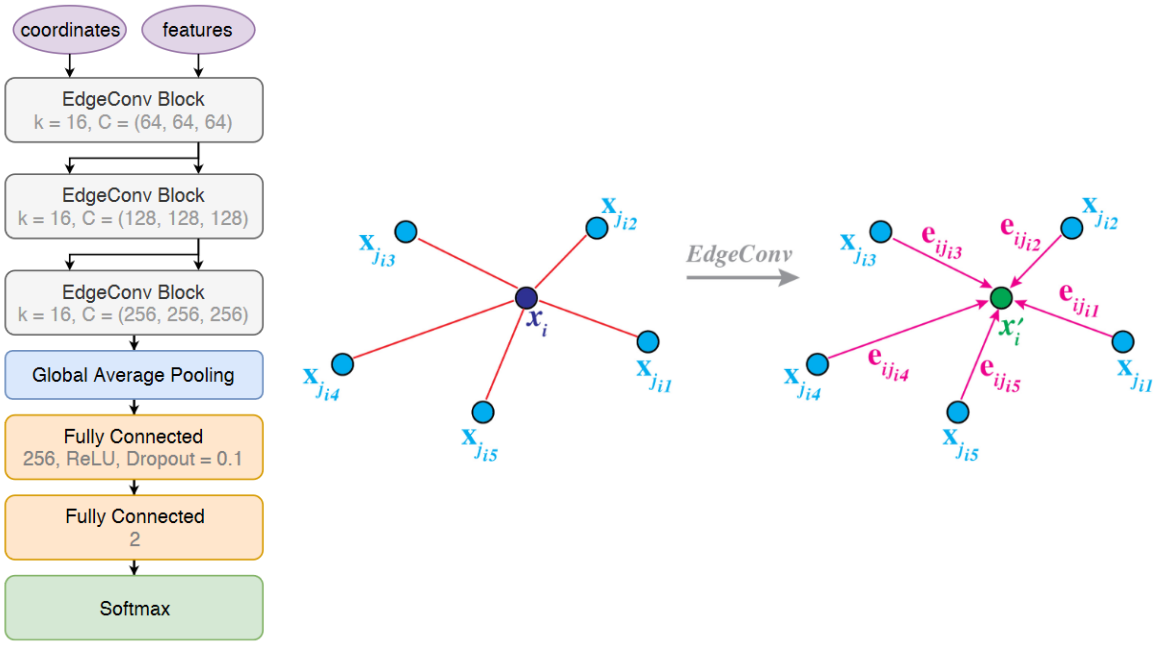

Figure 4.1: **Left:** Architecture of ParticleNet, proposed by [6], which the net for reconstructing the muon energy is based on. **Right:** working principle of the EdgeConv operation. It aggregates the edge features of all edges from each connected vertex for the output x_i' [7].

combines global shape structure (by patch centres x_i and local neighborhood information (by distances $x_j - x_i$). The choice for the type of activation functions in h is the Rectified Linear Unit function (ReLU). The averaged updates then are added onto the previous node features. Finally, the feature vectors e_{ij} learned by EdgeConv can be viewed as new coordinates for the original points in a latent space, where the distance can be recalculated after each convolutional layer. This dynamic graph update approach allows the receptive field of the network to be as large as the diameter of the cloud, while still being sparse. [7]
When a muon traverses the ORCA detector, it radiates Cherenkov light cones along its path. The light cones then hit the PMTs and the EdgeConv operation takes the hit coordinates (x,y,z) as input to compute the distances and the time t as well as the directions (pmt_x,pmt_y,pmt_z) that the hit PMTs were facing as features for the nodes. The two hyperparameters that can be changed for each layer are the number of neighbors k and the number of channels (kernels) $C = (C_1, C_2, C_3)$ corresponding to the number of units in each of the three linear transformation layers the EdgeConv block consists of. The network is trained with simulated logarithmic muon energies from MUPAGE data sets as labels. Like ParticleNet (architecture shown in figure 4.1, left) proposed by [6], the network in this thesis consists of three EdgeConv blocks with (64,64,64), (128,128,128) and (256,256,256) channels respectively, followed by a global average pooling layer to aggregate the learned features from all particles, and a fully connected layer with 256 units and the ReLU activation function. Since ParticleNet was constructed to tag particle jets, the output is generated by a fully connected layer with two units followed by a softmax function, allowing for a binary classification. The network in this thesis only outputs the logarithm of the muon energy and therefore has one neuron at the end (or two neurons for mean and variance of the energy).

5 Energy reconstruction with Graph Neural Networks

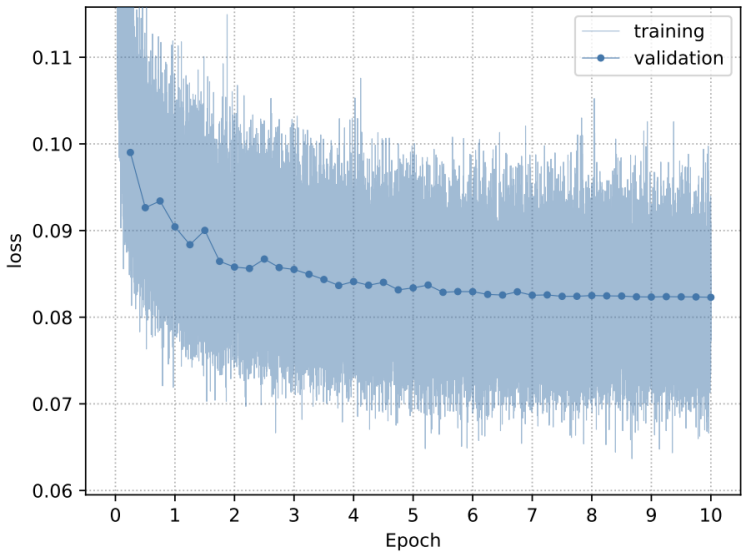

Figure 5.1: Loss for training and validation over the course of 10 epochs.

5.1 Training

The network was trained on four MUPAGE files, each containing 148565 simulated events with a maximum of 2000 hits. The labels were the logarithmic energies of single muons and summed up logarithmic energies of muon bundles. One different file was also used for validation and inference, respectively. The training was carried out over 10 epochs, where the learning rate started at 10^{-3} for file one in epoch one and decreased at a 10% rate for each following file. As a loss function, the mean squared error (MSE) was employed:

$$\text{loss} = \frac{1}{n}\sum_{i=1}^{n}(E_{\text{reco}}^{(i)} - E_{\text{true}}^{(i)})^2 \tag{5.1.1}$$

The keras optimizer *Adam* was used for training [8]. The loss function over the course of 10 epochs can be seen in figure 5.1. The pale blue line shows the training loss, the darker line with dots shows the validation loss. Even though the validation loss seems a bit unstable up to epoch three, at epoch 10 it reaches a plateau and therefore shows that the network has reached its best results.

5.2 Results

As a first result, the predicted energy is plotted over the true MC energy in a 2D-histogram in figure 5.2. Each MC energy bin in the histogram is normed, so the color of the bins represents

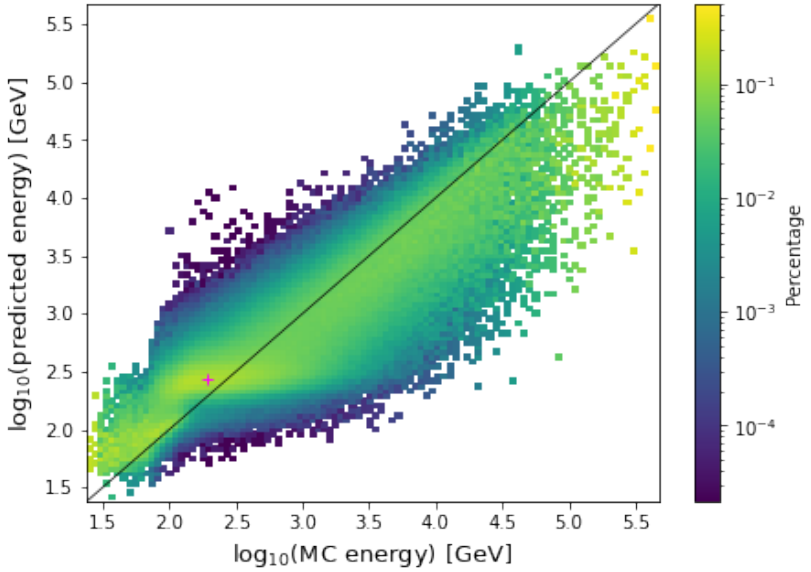

Figure 5.2: 2D-histogram of the predicted energy over the true MC energy. The logarithm of the energies is plotted, as this is the quantity the network reconstructs. The black line represents a perfect linear dependency with slope 1. The pink cross denotes the maximum of the distribution.

the percentage of the events within a true energy interval contributing to a certain predicted energy interval. This way, the large difference between the number of events along the energy range doesn't effect the distribution and features at energies with less events are better visible. The shape of the plot shows an overall linear dependency, as it follows the black line representing the ideal linear slope where predicted values equal true values. At low MC energies, the network tends to overestimate, as more predictions are above the black line than below. This might be due to some muons with low enough energies that are not able to traverse the detector completely and decay still inside, and therefore produce more light than a muon track alone. Proceeding to higher energies, the shape widens. This means that the predictions become worse, both for under- and overestimating. At even higher energies, the predictions become slightly better again until the events thin out. Here, underestimations are more present as more events are below the line. This might be because the instrumented detector volume is too small to register the Cherenkov light of high energy muons completely, so a fraction of light might escape unseen. Looking at figure 3.4, the bundles with an average lateral spread larger than the detector radius of 100 m likely contribute to these underpredictions. A quantification of these phenomena will follow in section 7.
The number of events peaks at $\log_{10}(E_{true}/GeV) = 2.28$ and $\log_{10}(E_{reco}/GeV) = 2.44$. The most event-abundant parts form a beam shape parallel to the x-axis, where the network assigns the MC energy range of about 10^2 GeV to 10^3 GeV to the same predicted value (denoted with a pink cross in figure 5.2). This is reminiscent of the cut between energy regimes from section 3.1. As was concluded there, muons in water can be treated as MIPs for energies up to 1 TeV (or 10^3 GeV). Looking at the 2D-plot, this energy is roughly where the shape starts to narrow again and the beam shape ends. Muons in the lower energy regime all deposit the same amount of light in the detector due to their non-energy dependent energy loss, and the network therefore cannot deduce their initial energy properly. In the upper energy regime beyond 10^3 GeV, the correlation between muon energy and energy loss is linear, so the network can distinguish between the initial energies.

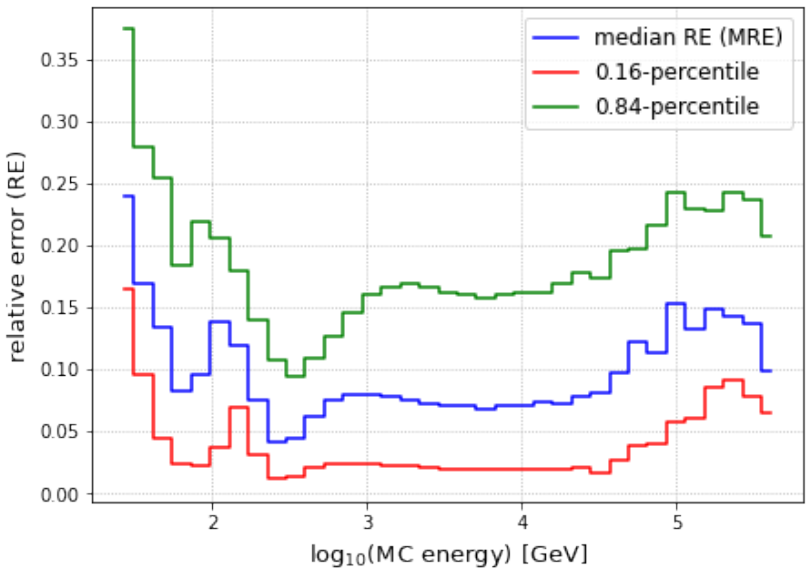

Figure 5.3: Median Relative Error (MRE) with 0.16- and 0.84-percentiles for the predicted energy distribution plotted over the true MC energy.

To quantify how well the network performs in the respective regimes, the Median Relative Error (MRE) is defined:

$$\mathrm{MRE} = \mathrm{Median}\left(\left|\frac{\mathrm{E_{true}} - \mathrm{E_{reco}}}{\mathrm{E_{true}}}\right|\right) \tag{5.2.1}$$

This quantity describes the median difference between the true and the reconstructed energy with respect to the true value. A lower relative error means a better resolution. The median is preferred over the mean because for quantifying the overall performance, outliers can be neglected. To compute the MRE, the true energy values were merged into 35 truth bins. The MRE together with the 0.26- and 0.84-percentile to showcase the shape of the distribution is plotted in figure 5.3. Below 10^2 GeV and above $10^{4.5}$ GeV, the event number is too low to draw meaningful statistics from. Between 10^2 GeV and $10^{2.5}$ GeV, the plots show a significant dip, where the resolution of the network is as low as 5%, although the beam structure at this energies should lead to a bad resolution. This might arise because the event number in this area is so high (76 % of events have a true energy of less than 10^3 GeV) and the maximum of events, as seen in figure 5.2, is located near the diagonal, which gives the impression of many successful predictions.
Form a range of 10^3 to almost $10^{4.5}$ GeV, the network keeps a steady resolution of about 7.5 %. This is where the correlation is linear and the energy prediction works best.

In figure 5.4, only the reconstructed energy distributions for multiplicities $\mathrm{m} = 1, 2, 3, 4$ are shown. For a single muon, the energy distribution looks similar to the distribution for all multiplicities, except the event density at high energies is lower. For higher multiplicities, the minimum energy increases, as the multiplicity acts as a cut to the energy of the cosmic ray event - only very energetic CRs can produce multiple muons. Also, the arrow shape for $\mathrm{m} = 1$ becomes more and more elliptic for higher multiplicities, and a bright strip containing most of the events follows the diagonal, indicating that most predictions are good predictions. The beam shape is only visible for $\mathrm{m} = 1$. As can be seen on the MRE plots in figure 5.4, the

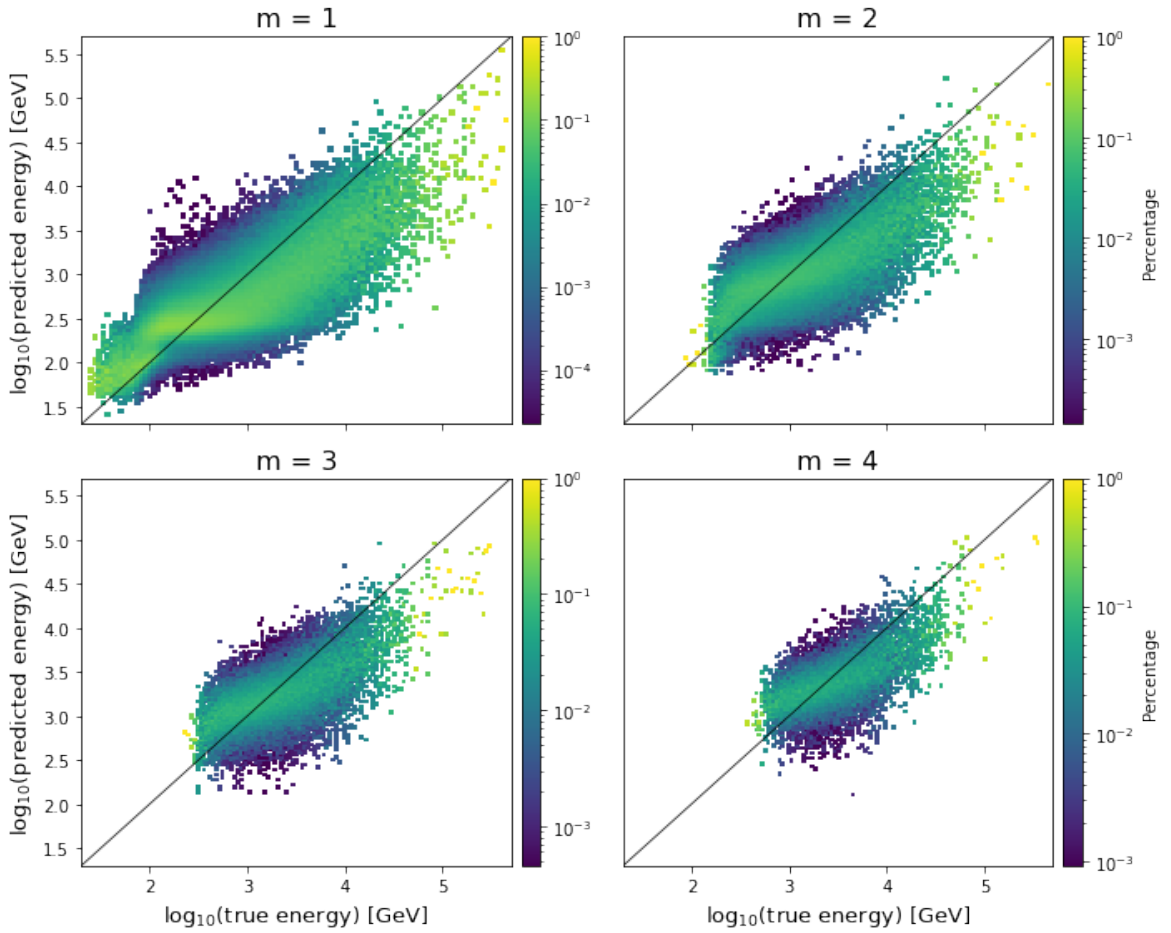

Figure 5.4: Predicted 2D-energy-histograms for multiplicities m = 1, 2, 3, 4. The diagonal line again represents perfect correlation.

dip, where most events are located near the diagonal, migrates towards higher energies with increasing multiplicity. Unlike m = 1 though, with $\log_{10}(E/GeV) > 3$ it doesn't fall into the MIP regime. Therefore, the resolution of less than 5 % might actually be significant. Although for higher multiplicities there are less events, the network generally gives better predictions than for single muons. The fact that in bundles with higher multiplicities, the energy is spread more evenly between the muons (as seen in figure 3.4) also contributes to this. For improving the predictions, single muon events therefore should be cut.

In figure 5.6, the MRE with percentiles as a function of the muon multiplicity (on the left) and the cosine of the zenith angle (on the right) is shown. For the multiplicity, it again becomes clear that the resolution is worst for single muons. It continuously becomes better for increasing multiplicity until about m = 10, less so until m = 30 and stays constant until m = 70. Beyond that, the network doesn't work well anymore. So, to improve the performance of the network upon roughly 3 %, events with up to m = 10 can be cut. This would cut 99.38 % of the whole distribution though.

For the zenith angle, the MRE shows a more continuous improvement towards -1 (vertical direction) over about 2.5 %. This might arise because there are less events for increasing inclination (as seen in figure 3.4). Cutting for example all events with $\theta_z > -0.5$ only improves the resolution upon roughly 1 % though, so the whole angle range will be kept in the following.

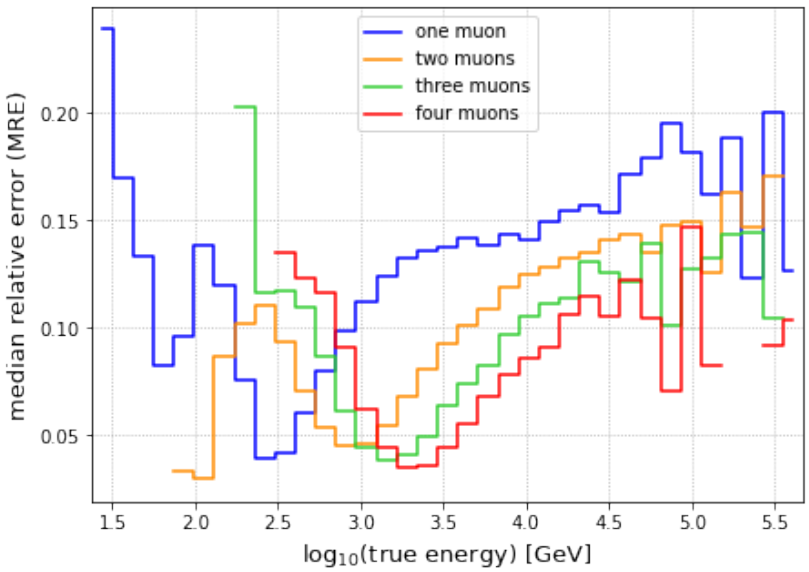

Figure 5.5: Median Relative Error (MRE) on the true MC energy for multiplicities $m = 1, 2, 3, 4$.

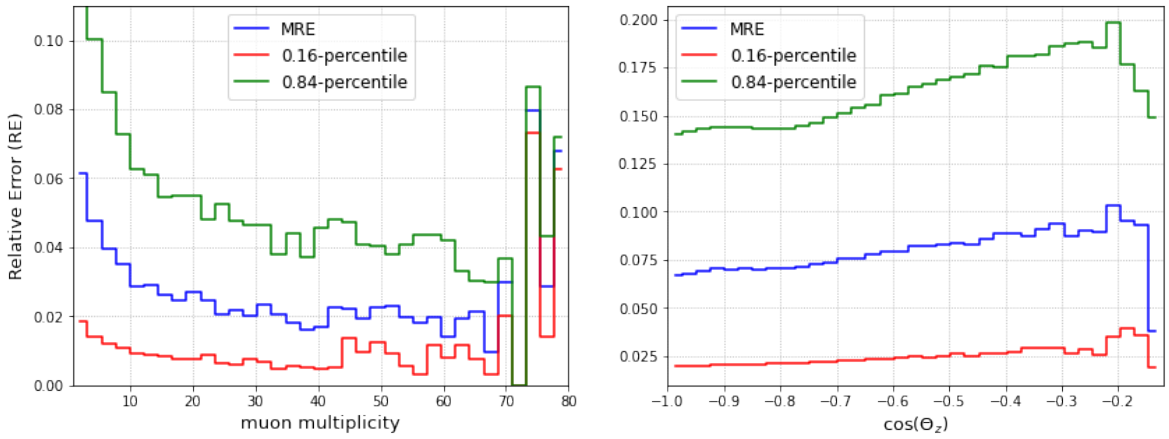

Figure 5.6: Median Relative Error (MRE) with percentiles as a function of the multiplicity m and the cosine of the zenith angle θ_z.

6 Comparison to standard KM3NeT event reconstruction

6.1 JGandalf reconstruction chain

The conventional method to reconstruct muon energies is the JGandalf reconstruction chain. Like the network, JGandalf takes the Cherenkov light signals caused by muons passing through the detector as input and outputs a fitted energy and direction of the muon track. The chain of reconstruction steps follows:

$$\text{JPrefit} \rightarrow \text{JGandalf} \rightarrow \text{JStart} \rightarrow \text{JEnergy}$$

The single steps are the following [9]:

- **JPrefit** performs an initial fit in the detector. It assumes N track direction hypotheses in all directions via scanning the sky in 5° intervals. Then it performs a χ^2 fit for the position relative to the PMT and the time. The N best-fit test direction (evaluated by a quality parameter Q) are handed to JGandalf.

- **JGandalf** scans over the JPrefit fits to carry out the full maximum likelihood fit (with all 5 parameters: 3D-position, direction, time) via

$$\mathcal{L} = \prod_{\text{hitPMTs}} \left[\frac{\partial \mathrm{P}}{\partial \mathrm{t}} (\mathrm{R_i}, \theta_\mathrm{i}, \phi_\mathrm{i}, \Delta \mathrm{t}) \right], \tag{6.1.1}$$

  a set of PDFs to describe the PMT response, taking bremsstrahlung, the optical background, light dispersion and PMT properties into account. $\mathrm{R_i}$ is the distance of closest approach, θ and ϕ describe the PMT orientation, Δt is the difference between the expected and the measured arrival time.

- **JStart** determines the observed start position of the muon trajectory and therefore the track length by back-projecting onto the track unter the Cherenkov angle.

- **JEnergy** uses only the PMTs within a predefined roadwidth around the muon trajectory. It then compares the calculated hit probability for a PMT to the actual occurrences of hits. This is followed by a one-parameter likelihood fit of log(E) and therefore provides a reconstruction of the muon energy.

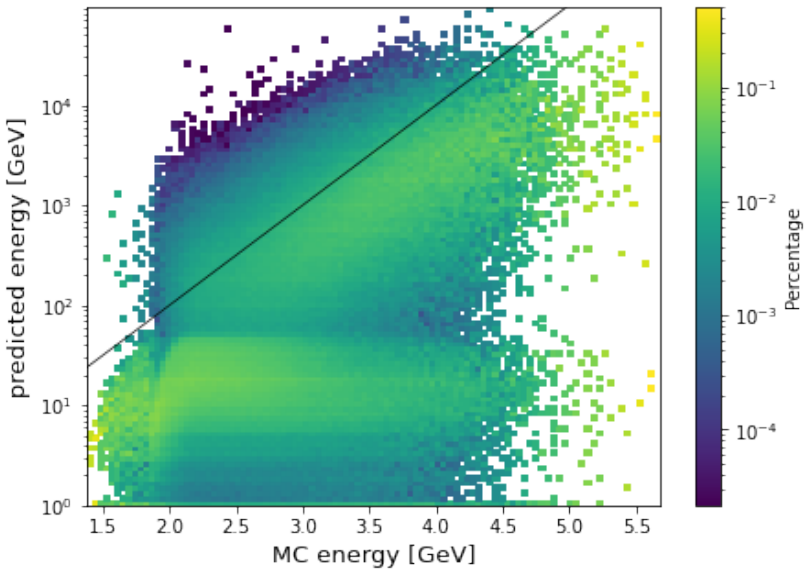

Figure 6.1: 2D-histogram of the energies predicted by the JGandalf algorithm plotted over the MC energies. The solid line represents perfect correlation. Below $E_{reco} \sim 50\,GeV$, the reconstructed values form a large area of wrong predictions.

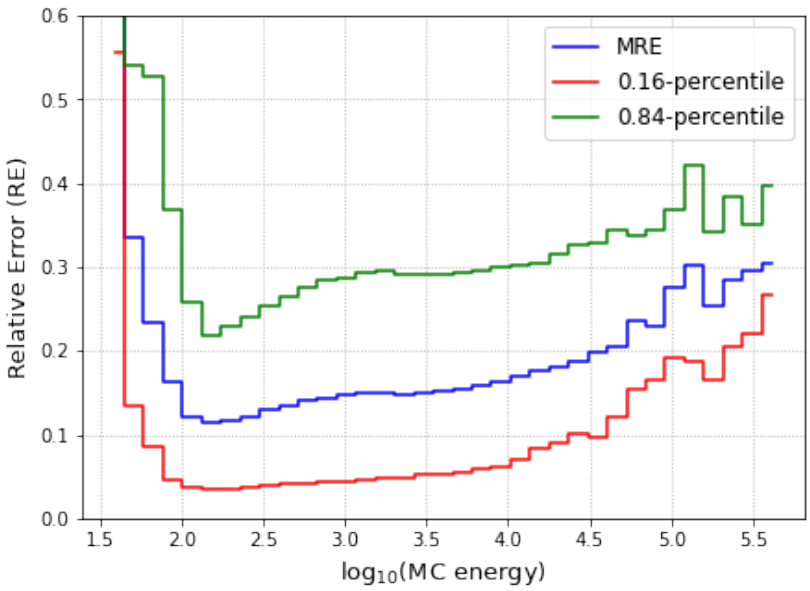

Figure 6.2: The MRE of the JGandalf results with 0.16- and 0.84-percentiles. The reconstructed energy has been cut at 50 GeV.

6.2 JGandalf results

The 2D-energy-distribution from the JGandalf reconstruction chain is plotted in figure 6.1. Below a rather sharp edge at roughly 50 GeV predicted energy, a large area of underestimations is visible. This leads back to which kind of energy is reconstructed by JGandalf, and which by the network. The GNN reconstructs the energy of the simulated muons that are generated at the *can*, as described in section 3.2. The MC energy also specifies the energy of the

generated muons at the can surface. JGandalf however includes the additional energy loss of the muons during the propagation from the can surface to the detector. For higher energies, this additional loss is not significant, though it is for energies in the same range of magnitude as the energy loss. This can easily be estimated with the position of the beam structure: while found at $\sim 10^{2.5}\,\mathrm{GeV} \approx 300\,\mathrm{GeV}$ in the GNN reconstruction, for JGandalf, its upper edge is located at $\sim 50\,\mathrm{GeV}$. Assuming the MIP energies are always reconstructed with the same values, this gives an energy loss of around 250 GeV. For plotting the reconstructed energy over the MC energy, these lower energy events are therefore misleading. For the MRE and the percentiles in figure 6.2, the predicted energy has been cut at 50 GeV, so only comparable energy reconstructions are examined, assuming the additional energy loss of the events becomes significant at the border of the MIP regime. Like in figure 5.3, the true energy range of $2 < \log_{10}(\mathrm{E_{true}}/\mathrm{GeV}) < 4.5$ yields the greatest statistical importance. Right after $10^2\,\mathrm{GeV}$, a dip is visible too, though not as pronounced, as the (green) area of the highest event number is much broader than the (yellow) beam shape in figure 5.2. Most of the MRE lies between 10 % and 20 %. Between $10^3\,\mathrm{GeV}$ and $10^{4.5}\,\mathrm{GeV}$, where the GNN reconstruction works best, the JGandalf-MRE increases almost linearly.

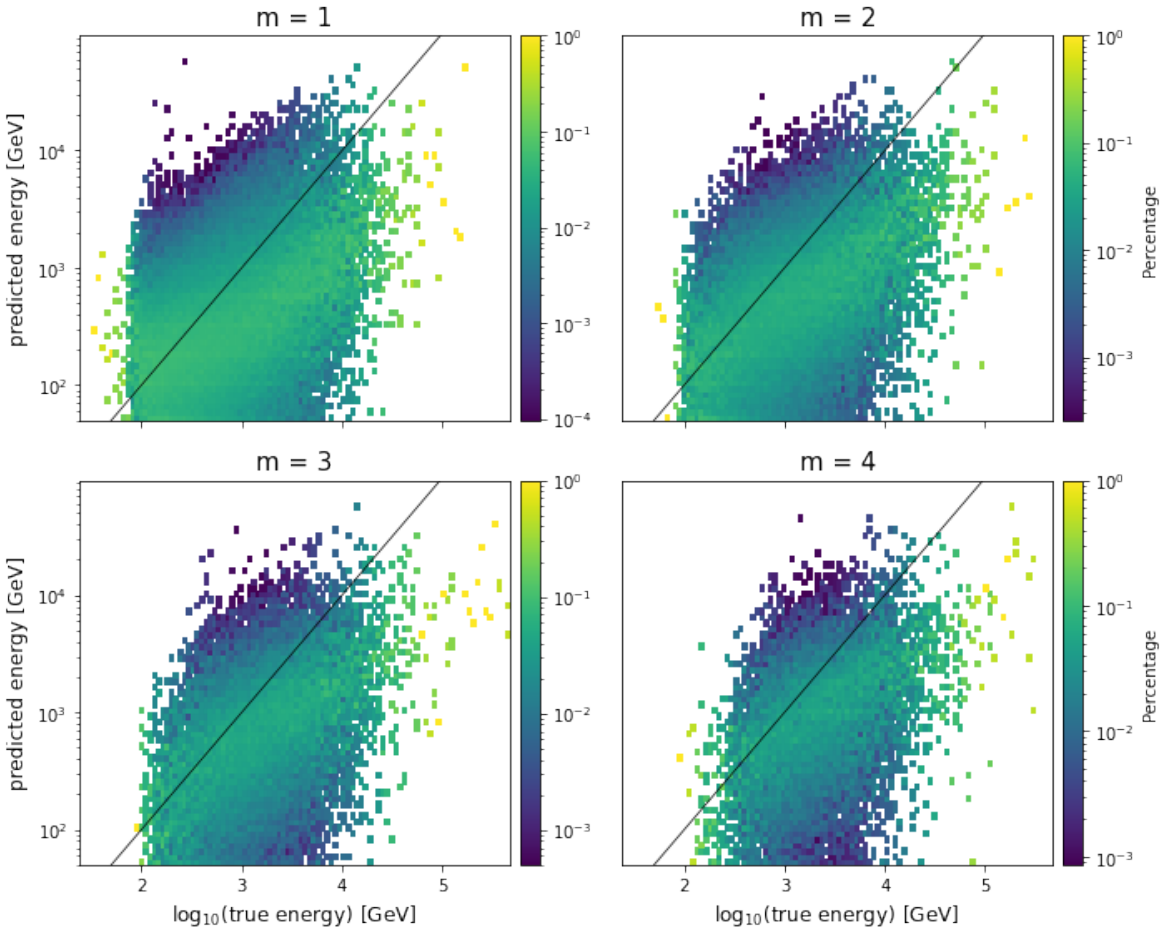

Figure 6.3: 2D-histograms of the predicted energy distribution from the JGandalf reconstruction for multiplicities $\mathrm{m} = 1, 2, 3, 4$. The reconstructed energy has been cut an 50 GeV, as above. The black line represents perfect correlation.

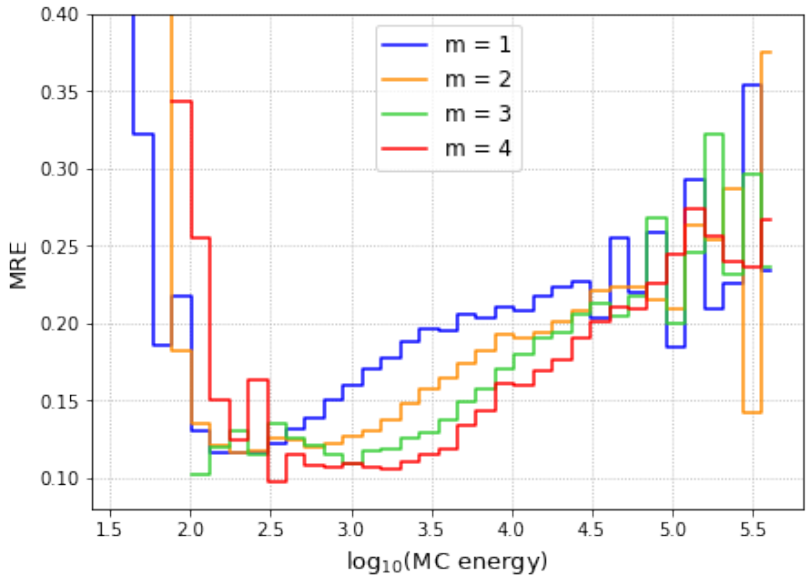

Figure 6.4: MRE of the JGandalf-reconstruction with $E_{reco} > 50\,GeV$ for multiplicities $m = 1, 2, 3, 4$.

The multiplicity plots in figure 6.3 were also cut at $E_{reco} = 50\,GeV$. Unlike the network reconstruction, JGandalf seems to perform worse, as the distribution becomes more and more unsymmetrical with respect to the diagonal for increasing multiplicity. At $m = 4$, a majority of the predictions are underestimations. Also the bright strip as seen in figure 5.4 is not so well visible and seems to diverge from the diagonal at higher energies for all multiplicities, indicating that the predictions become worse for higher energies.

Looking at the MRE plots in figure 6.4, the dip wanders to higher energies with increasing multiplicity like in figure 5.5, though it never reaches below 10 % resolution. The dip for $m = 4$ barely reaches beyond $10^3\,GeV$, where the loss-energy relation is linear. According to this, JGandalf doesn't perform significantly better for higher multiplicities.

6.3 Comparison to GNN results

For a meaningful comparison of the GNN and JGandalf results, the energy range where the reconstructions work best are most important, namely the linear regime beyond $10^3\,GeV$. In the following, the true MC energies below this value were cut. The properties of these cut events are shown in figure 6.5: On the left, the muon multiplicity of the events is plotted over their respective energy in a 2D-histogram. On the right, the same is plotted for the cosine of the zenith angle. Since these plots don't show predictions, the MC energy bins are not normed. The red line indicates the cut at $10^3\,GeV$, all events below are discarded for the comparison. The cut events make up 76 % of the whole distribution. In regards to the multiplicity, only bundles with less than 10 muons are cut as the events with higher multiplicities have higher energies. This also reinforces that the bundle multiplicity acts as a cut on the CR energy. The zenith angle on the other hand is rather evenly distributed with the energy. Although the cut events span the whole angle range, the events that remain have approximately the same range. Only the cosines close to zero (the vertical direction) are slightly less represented, since the farther the muons travel through water, the more energy is lost.

The 2D-energy histograms of both cut distributions can be seen in figure 6.6. It is apparent

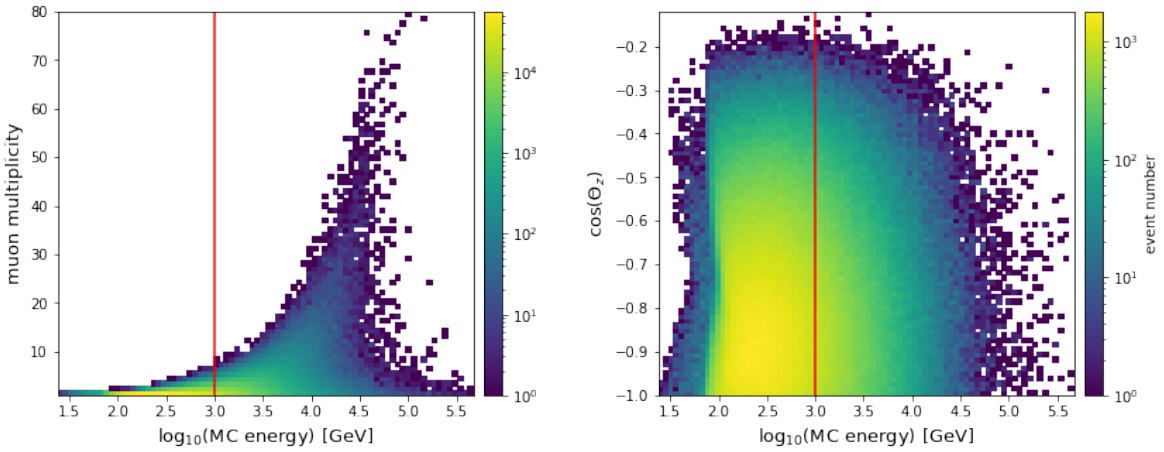

Figure 6.5: **Left:** 2D-histogram of the muon multiplicity vs. the logarithm of the MC energy. **Right:** 2D-histogram of the cosine of the zenith angle θ_z vs. the logarithm of the MC energy. The red vertical lines represent the cut at 10^3 GeV true energy, below which the events are discarded for the comparison.

that the JGandalf reconstruction on the left is broader than the GNN reconstruction on the right. Both appear to underestimate more than overestimate. Like in the multiplicity plots, the bright strip in the GNN distribution follows the diagonal, whereas in the JGandalf distribution, it diverges at high energies towards underpredictions.

The MREs of both distributions as well as the relative deviation $\frac{\text{MRE}_{\text{JGandalf}} - \text{MRE}_{\text{GNN}}}{\text{MRE}_{\text{JGandalf}}}$ between them are shown in figure 6.7. While the JGandalf reconstruction decreases from a resolution of 50 % at 10^3 GeV to almost 20 %, the GNN reconstruction remains below 10 % resolution until $10^{4.5}\,GeV$. At all energies, the GNN graph lies below the JGandalf graph, which shows that the GNN reconstruction generally works better. The relative deviation in the lower panel shows that at 10^3 GeV, the GNN reconstruction is 80 % better than the JGandalf reconstruction, and still about 65 % better at $10^{4.5}$ GeV. Although the network performs worse at higher energies, it still works at least 40 % better than JGandalf.

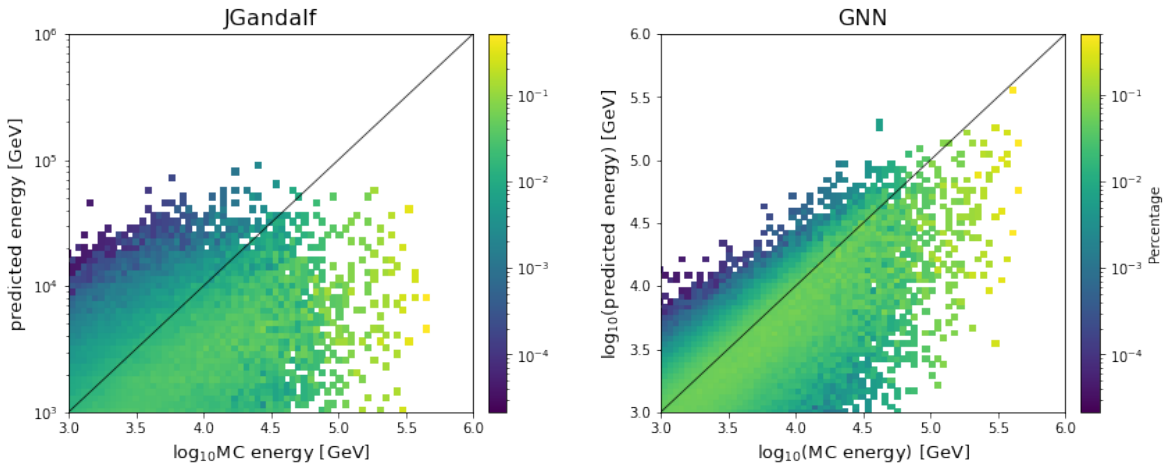

Figure 6.6: Comparison of the 2D-histograms of the JGandalf reconstruction (left) and the GNN reconstruction (right). Both MC and reconstructed energy were cut at 10^3 GeV.

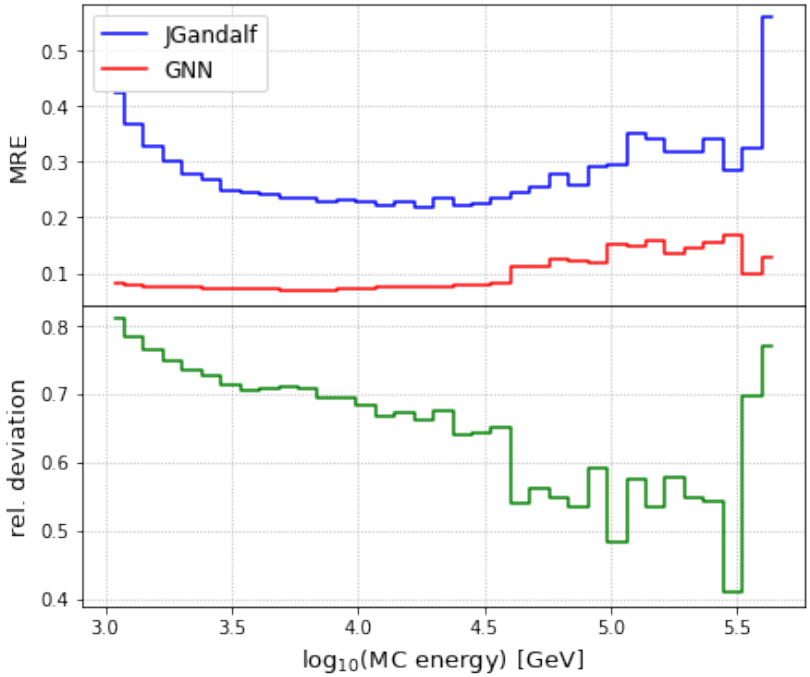

Figure 6.7: **Upper panel:** Comparison of the MRE for the cut JGandalf and GNN reconstruction. **Lower panel:** Relative deviation between the two MRE curves from the upper panel.

7 GNN error reconstruction

7.1 Training

The loss in equation 5.1.1 is the argument of the natural logarithm of a normal distribution with the variance set to 1, disregarding constants. This way, only the mean of a Gaussian is fitted to the data. To also give an estimation of the variance and therefore the error of the prediction, a negative log-likelihood loss of the normal distribution with mean and variance variables is employed:

$$\text{loss} = -\log \prod_{i=1}^{n} N(E_{\text{true}}^{(i)} | E_{\text{reco}}^{(i)}, \sigma^{(i)}) \tag{7.1.1}$$

The training was conducted on the same data sets as for the previous training. The process of the loss function for training and validation can be seen in figure 7.1. Despite some local instabilities, the loss reaches a stable plateau after the sixth epoch. With ~ 0.12, it is higher than the loss for only the mean (with ~ 0.082), indicating that the additional parameter σ complicates the determination of the best predictions.

7.2 Results

The predicted event energies can now be cut with respect to the corresponding variance. In figure 7.2, the original distribution is shown on the upper left, the events with the smallest 50 %, 25 % and 12.5 % follow on the top left, the bottom right and the bottom left. With increasingly small variance, The distribution at higher energies becomes narrower, which means that the reconstruction is better, as is expected.
Since the beam structure at lower energies corresponds to the MIP regime of muons where the energy loss is not energy dependent, one could expect that the reconstruction in that regime would have a large variance. In this case, the beam would vanish at smaller variances. But actually the opposite is true: the beam shape stays and becomes more pronounced in comparison to the rest of the distribution up to the top 25 % of events. For the top 12.5 %, it also begins to fade. So seemingly, the network classifies the MIP events as good reconstructions. A reason for this might be a fundamental lack of information in this regime, like it was discussed for the beam shape in figure 5.2, which causes the network to not work properly. In figure 7.3, the MREs of the improved distributions are shown. The resolution in the range between 10^3 GeV and $10^{4.5}$ GeV, where the network should work best, is vastly better and lies below 5 % most of the time, even for only the best 50 %. This seems to be the best cut option as the smaller variance distributions in comparison show no significant improvement and the event number is not significantly smaller compared to the original distribution. However, a peak forms at about $\log_{10}(E/\text{GeV}) = 2.6$ for the improved distributions. This is likely an indicator of the beam structure that loses its number of events more slowly than the local surroundings in a truth bin with decreasing σ.
In conclusion, the best reconstruction is possible for events with more than one muon (as shown in section 5.2) and with the smallest 50 % variance. This final distribution is shown in figure 7.4 on the right. 89.9 % of the original distribution were cut. To examine it even

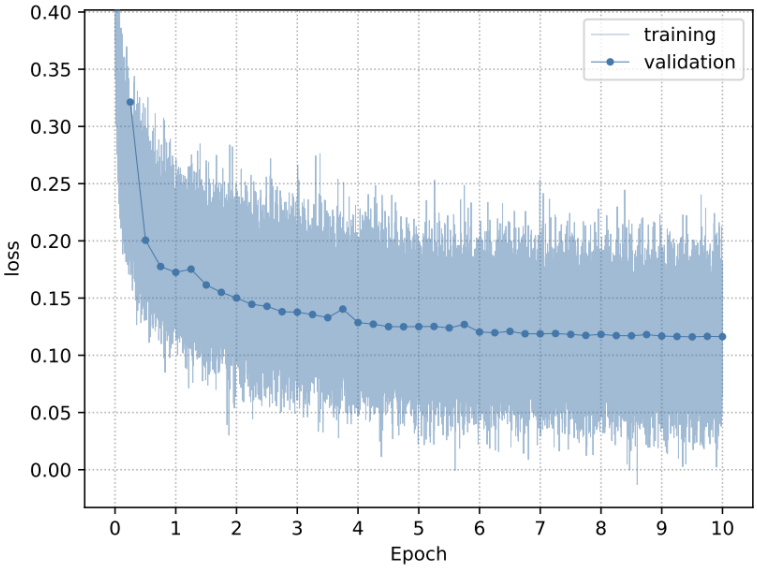

Figure 7.1: Training and validation loss over the course of 10 epochs for the training with an additional σ parameter.

further, the so-called *pull distribution* is defined:

$$\text{pull} = \frac{\text{E}_{\text{true}} - \text{E}_{\text{reco}}}{\sigma}. \tag{7.2.1}$$

In the ideal case where E_{reco} is reconstructed with a Gaussian distribution with mean E_{true} and variance σ, the pull of many predictions would be distributed as a standard Gaussian with mean zero and unit width. As can be seen in figure 7.4 on the left, this is not exactly the case - the pull of the best distribution showcases a tail at positive x-values and a deviation from the ideal Gaussian width at negative x-values. For comparison, a Gaussian curve

$$\text{G}(\text{x}|\text{A}, \mu, \sigma) = \text{A} \cdot \text{e}^{\frac{(\text{x}-\mu)^2}{\sigma^2}} \tag{7.2.2}$$

is fitted onto the histogram data. The distribution is therefore slightly askew. Nevertheless, the fitted mean ($\mu = -0.044 \pm 0.003$) and variance ($\sigma = 0.995 \pm 0.003$) are very close to zero and one, indicating that the reconstruction generally works well.

To look further into where the skewness comes from, the pull of the energy range divided into four intervals is plotted in figure 7.5. The original interval edges $\log_{10}(\text{E/GeV}) = (1.92529142, 2.86258772, 3.79988402, 4.73718032, 5.67447662)$ are rounded in the figure titles. They are also plotted as dashed lines in figure 7.4 on the left. Out of the four intervals, the lowest (upper left) shows the largest deviation from the standard variance, as the fitted variance is only $\sigma = 0.553 \pm 0.004$ and the tail at the right and deviation at the left are clearly visible. Since $\sigma = 1$ is the ideal Gaussian variance, a smaller variance indicates that the network is not working correctly due to a lack of information in the MIP regime. Looking at the 2D-energy distribution at low energies, there is still a remnant of the beam structure in which the events were wrongly classified as good reconstructions. These events following a different slope than the diagonal account for the skewness of the pull, and more precisely for a tail in positive direction as the beam structure events are underpredictions. The mean is shifted to $\mu = -0.828 \pm 0.004$ because the majority of events is located above the diagonal in the lowest energy interval. To lessen the effect of the beam structure, events with $m = 2$ or higher could also be cut.

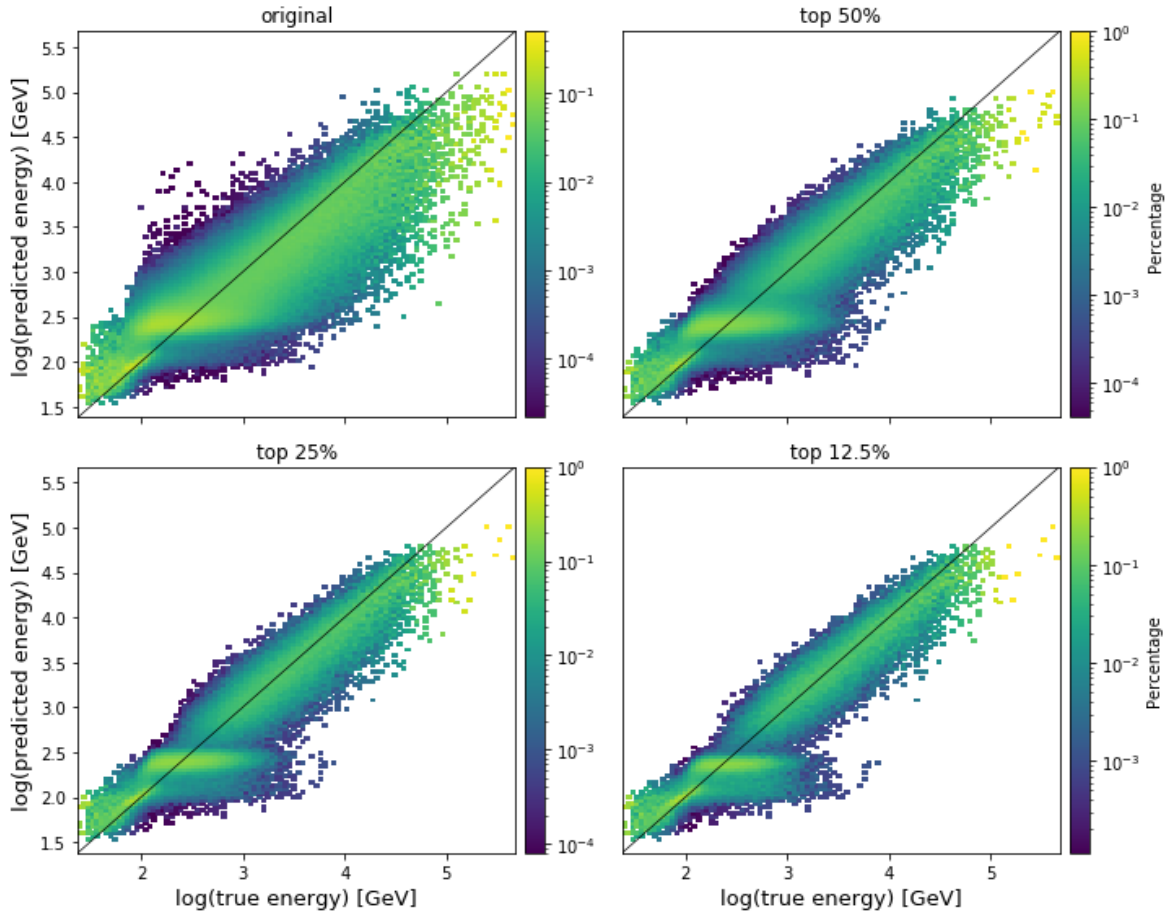

Figure 7.2: 2D-histograms of the original GNN-predicted energy distribution (top left) in comparison with the sigma filtered distributions. The top 50 % distribution shows the 50 % smallest sigma energies, and so on.

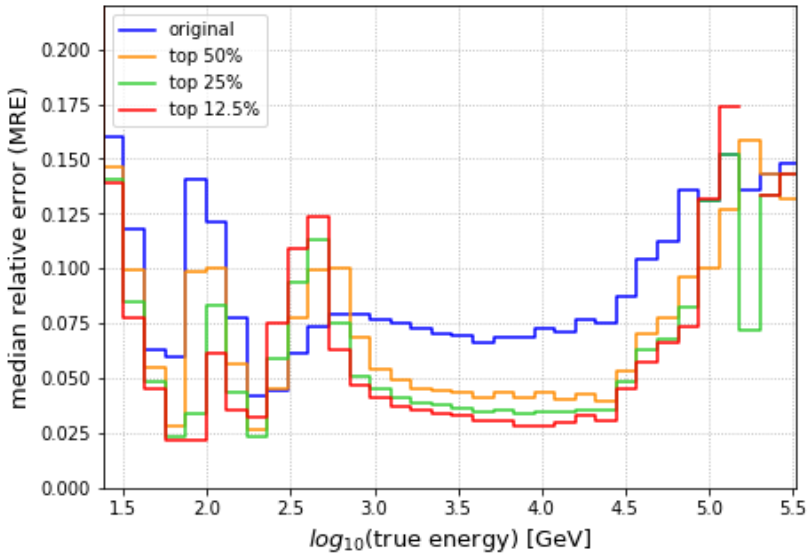

Figure 7.3: MRE curves for the sigma filtered distributions from figure 7.2.

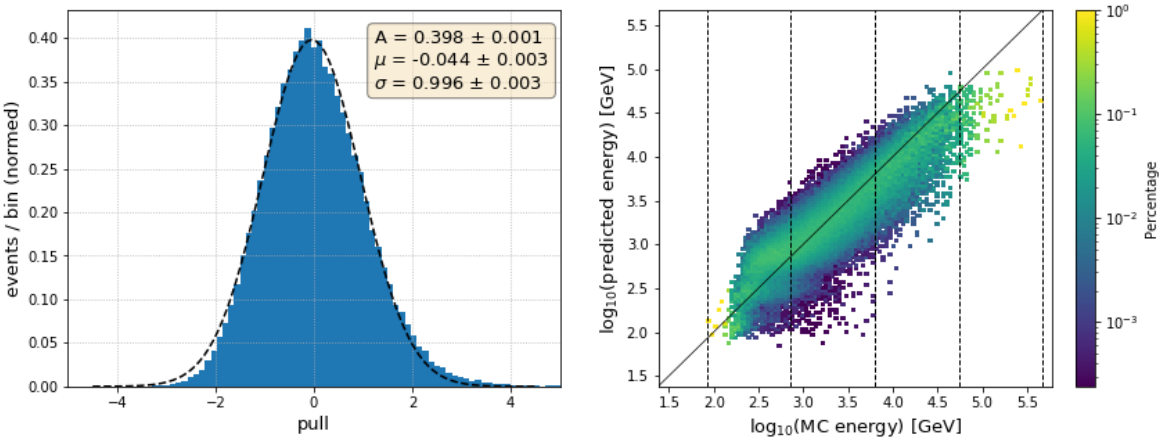

Figure 7.4: **Right:** 2D-histogram of the GNN-predicted energy distribution with single muon events and 50 % worst sigma events cut out, evenly divided into four true energy intervals by the dashed lines. The solid line represents perfect correlation. **Left:** normed histogram of the pull distribution of the improved reconstruction, as seen on the right. The fit values of the Gaussian fit (black dashed curve) are shown in the box.

The highest energy interval (lower right) contains too few events to be statistically significant. The remaining two middle intervals show almost perfect Gaussian shapes with negligible tails. The smaller of the two (upper right) has the closest to zero mean of $\mu = 0.318 \pm 0.003$ and an acceptable variance of $\sigma = 0.868 \pm 0.003$, while the higher interval (bottom left) shows the best variance of $\sigma = 1.067 \pm 0.002$, but a shifted mean of $\mu = 0.907 \pm 0.002$. A shift of the mean without changing the variance means shifting the reconstructed energies by a constant. The positive mean therefore comes from the bias towards underpredictions in the two middle intervals. The variance close to 1 though indicates that the network recognizes the underpredictions as such. This then can be improved by cutting at smaller variances. In conclusion, the $2.9 < \log_{10}(\mathrm{E/GeV}) < 3.8$ interval is probably the best choice for the interval of best reconstruction, since it is closest to the standard Gaussian. Regarding the border between the MIP and the linear regime at 10^3 GeV, this seems physically reasonable.

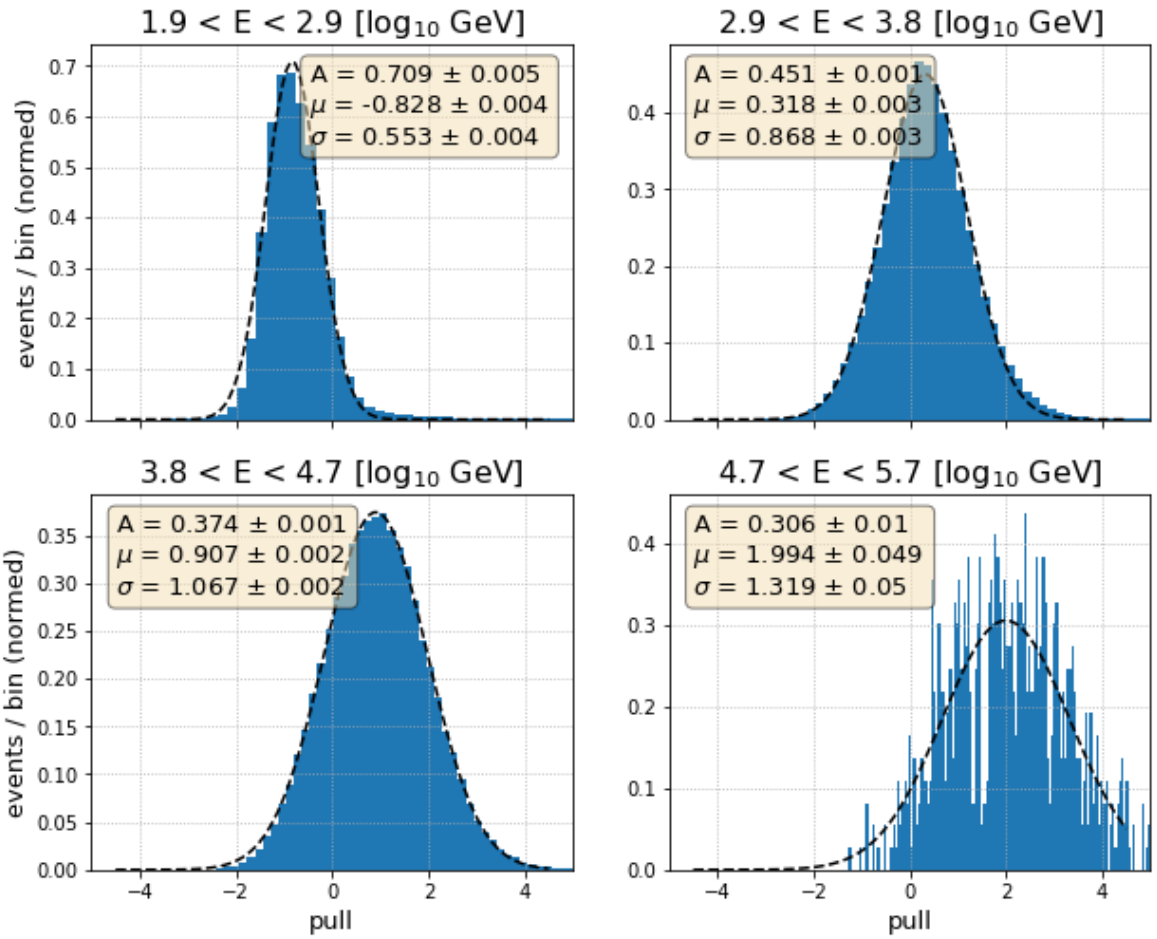

Figure 7.5: Normed pull distributions of the improved reconstruction divided into four true energy intervals. The respective fit values of the Gaussian fits are shown in the boxes.

8 Discussion and outlook

In conclusion, the best GNN reconstructions are those with m > 1 and top 50 % variance in the MC energy interval between roughly $2.9 < \log_{10}(\mathrm{E/GeV}) < 3.8$. Since mostly muon bundles mimic neutrino events, the single muon events are not important for neutrino event classification. The resolution still can be improved by cutting higher multiplicity events and lower σ values, as long as it is reasonable to discard even more than 89.9 % of events. Nevertheless, only cutting the energy at 10^3 GeV still provides good reconstructions that are superior to the JGandalf reconstructions. This clearly shows that the network recognizes even more correlations in and can extract more information from the Cherenkov light signals in the detector than the PDFs in the likelihood function from equation 6.1.1.It should however be noted here that JGandalf was originally developed with a larger detector in mind, so the results for more than four detection lines may be better than the results presented here. And while the end result surely is what matters most for discerning muon and neutrino events, a disadvantage of the GNN reconstruction might also be that the additional physical processes the network draws information from are not clear.
Despite of this huge improvement over the classical reconstruction, there are still problems that the network faces: At low energies, some muons already decay inside the detector and release more light than a regular muon track would, resulting in the network overestimating the energy. At high energies, the muon tracks might produce more light than can be contained by the detector, leading to an underestimation of the energy. Both these phenomena are intrinsic problems of the detector physics and cannot be overcome by the network. Also, the fact that the network classifies the wrong estimations of energies from the MIP regime as good reconstructions leads back to a lack of information in the light signals and is intrinsic to the physics.

For further investigations, the performance of the GNN reconstruction could be tested as a function of the pairwise distance of muons in a bundle, the track length in the detector, the minimal distance from the detector centre, and other quantities. Also, a Monte-Carlo data comparison could bring more detailed insight into the network performance.

9 Bibliography

[1] KM3NeT 2.0: Letter of Intent for ARCA and ORCA. Astroparticle & Oscillation Research with Cosmics in the Abyss (2016)

[2] C. Grupen, *Particle Detectors*, Cambridge University Press (2008)

[3] M. Spurio, *Probes of Multimessenger Astrophysics. Charged Cosmic Rays, Neutrinos, γ-Rays and Gravitational Waves*, Springer (2018)

[4] S.I. Klimushin et. al., *Precise parametrizations of muon energy losses in water*, Contribution to the 27th ICRC (2001)

[5] G. Carminati et. al., *Atmospheric MUons from PArametric formulas: a fast GEnerator for neutrino telescopes (MUPAGE)*, Comput.Phys.Commun.179:915-923 (2008)

[6] H. Qu, *Jet Tagging via Particle Clouds*, Phys. Rev. D 101, 056019 (2020)

[7] Y. Wang et. al., *Dynamic graph cnn for learning on point clouds*, ACM Trans. Graph. 38, 146 (2019)

[8] D.P. Kingma, *Adam: A Method for Stochastic Optimization*, 3rd International Conference for Learning Representations, San Diego (2015)

[9] B. O Fearraigh, *Track Reconstruction in KM3NeT*, 2021

Acknowledgements

I want to give many thanks to Thomas Eberl for the provision of this interesting thesis topic, to Stefan Reck for the supervision and his patience with my neverending questions, and to everyone in the Erlangen KM3NeT working group for the friendly and welcoming atmosphere, despite of the current pandemic situation and the fact that mostly only online meetings were possible.

Special thanks go to my roommates Julia and Bastian and my friends Felix and Anna for their soothing words and help in my battle against Python. Many thanks also to my family, who always supports me.

Statutory Declaration

I declare that I have developed and written the enclosed thesis entirely by myself and have not used sources or means without declaration in the text. Any thoughts or quotations which were inferred from these sources are clearly marked as such. This thesis was not submitted in the same or in a substantially similar version, not even partially, to any other authority to achieve an academic grading and was not published elsewhere.

.....

Julia Häfner Erlangen, April 14, 2021