Event Classification and Energy Reconstruction for ANTARES using Convolutional Neural Networks

Master's Thesis in Physics

Presented by
Nicole Geißelbrecht
April 30, 2021

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

ERLANGEN CENTRE
FOR ASTROPARTICLE
PHYSICS

Supervisor: PD Dr. Thomas Eberl

Abstract

ANTARES is a very-large-volume water Cherenkov detector in the Mediterranean Sea that has been taking data continously since 2007. It consists of a three-dimensional matrix of several hundred highly-sensitive photosensors. Its main goal is the search for high-energy neutrinos from astrophysical point sources, the detection of diffuse cosmic and atmospheric neutrino fluxes, and the indirect search for dark matter.
Deep learning is a rapidly developing field with a broad range of applications. Convolutional neural networks (CNNs) are currently state-of-the-art in deep learning image recognition and have already been applied successfully in other neutrino experiments such as KM3NeT. This work presents the first application of CNNs, and of deep learning methods in general, to the analysis of data recorded with the ANTARES neutrino telescope. Results obtained with classification and regression networks trained on Monte Carlo simulations and their application to simulated and measured data are reported.
The CNN-based background classifier is trained to reject atmospheric muons, which constitute the dominant source of background in the search for cosmic neutrinos. This network is able to extract neutrino interaction events from the data successfully, in particular if the neutrino energy and arrival direction differ sufficiently from that of atmospheric muons. The network performance is evaluated for different data-taking periods and environmental conditions in the deep sea. It could be shown that the network performs best in periods of low optical background noise. Even though around 25 % less atmospheric neutrino candidates can be identified in measured data compared to simulation, which is within the expectation given the systematic uncertainty in our knowledge of the atmospheric neutrino flux, the resulting energy and arrival direction distributions of the identified neutrino candidates agree well between data and simulation.
A second network is presented that has been developed for the reconstruction of the energy of neutrinos, for both main event types, i.e. track- and shower-like events. The performance is found to be significantly better for shower-like events and can be improved further for both event types by constraining the position of the neutrino interaction vertex within the detector volume. Furthermore, a network output has been designed and evaluated that successfully estimates the uncertainty of the energy reconstruction and offers the possibility to reject badly reconstructed events, which is essential for the sensitivity of cosmic neutrino searches.
Both networks that have been developed and presented in this work show promising results and indicate that the application of CNNs can contribute to future analyses of ANTARES data.

Contents

1 High-energy neutrino astronomy **7**
- 1.1 Basic neutrino properties . . . 7
- 1.2 Neutrino sources . . . 8
  - 1.2.1 Cosmic neutrinos . . . 9
  - 1.2.2 Atmospheric neutrinos . . . 10
- 1.3 Detection principle . . . 11
  - 1.3.1 Neutrino interactions . . . 11
  - 1.3.2 Cherenkov light . . . 12
- 1.4 ANTARES . . . 14
  - 1.4.1 General information . . . 14
  - 1.4.2 Detector layout . . . 14
  - 1.4.3 Signal and background in ANTARES . . . 14
  - 1.4.4 Monte Carlo Simulations . . . 15

2 Deep Learning **17**
- 2.1 Supervised learning . . . 17
- 2.2 Neural networks . . . 17
- 2.3 Loss functions . . . 19
- 2.4 Activation functions . . . 20
- 2.5 Optimizers and Backpropagation . . . 21
  - 2.5.1 Gradient descent . . . 21
  - 2.5.2 Adam . . . 22
  - 2.5.3 Backpropagation algorithm . . . 23
- 2.6 Weight initialization . . . 24
- 2.7 Dropout . . . 25
- 2.8 Batch Normalization . . . 26
- 2.9 Convolutional Neural Networks . . . 26

3 Preprocessing **29**
- 3.1 Data conversion . . . 29
- 3.2 Image generation . . . 29
  - 3.2.1 Spatial binning . . . 29
  - 3.2.2 Temporal binning . . . 31
- 3.3 Event selection . . . 33
- 3.4 Organizing the training with OrcaNet . . . 34

4 Background classifier **37**
- 4.1 Training set . . . 37

4.2 Network architecture . . . 38
4.3 Performance . . . 39
4.3.1 Loss and accuracy . . . 39
4.3.2 Probability distribution . . . 40
4.3.3 Muon contamination and neutrino efficiency . . . 42
4.3.4 Data . . . 50
4.3.5 Detector condition studies . . . 58

5 Energy reconstruction **70**
5.1 Training set . . . 70
5.2 Visible energy . . . 71
5.3 Loss function . . . 72
5.4 Network architecture . . . 72
5.5 Performance . . . 73
5.5.1 Loss . . . 73
5.5.2 Reconstructed energy . . . 75
5.5.3 Reconstructed uncertainty . . . 80
5.5.4 Comparison with standard reconstructions . . . 86
5.5.5 Data . . . 88

6 Summary and Outlook **92**

7 Bibliography **95**

A Appendix **i**
A.1 Background classifier . . . i
A.2 Energy reconstruction . . . x

1 High-energy neutrino astronomy

The neutrino was first postulated by Wolfgang Pauli in 1930 [1] in order to resolve the observation of the continuous beta spectrum. Therefore, he proposed a new electrically neutral particle with spin 1/2 and a small mass to ensure energy, momentum, and agular momentum conservation. In 1956 the first experimental evidence for the existence of the neutrino followed by Cowan and Reines via the inverse beta decay [2]. This effort also opened the possibility for neutrino astronomy like the detection of solar neutrinos by Davis and Bahcall [3] and even extragalactic neutrinos from the supernova SN1978A among others by Super Kamiokande [4] and IMB [5]. Today neutrino astronomy is a large field of research with many different experiments.
This chapter will cover the basic neutrino properties giving rise to the unique possibilities in astronomy as well as the resulting challenges. Further, possible sources are resumed and the detection methods are discussed. Finally, ANTARES will be demonstrated as an example for a neutrino telescope performing high energy neutrino astronomy.

1.1 Basic neutrino properties

Today the neutrino is part of the Standard Model of particle physics. It belongs to the lepton sector and occurs in three different flavours as electron neutrino ν_e, muon neutrino ν_μ, and tau neutrino ν_τ (with corresponding antiparticle $\bar{\nu}_x$) originating from the charged lepton it can produce:

$$\nu_x + n \rightarrow p + x^- \qquad \text{with} \quad x = e,\ \mu,\ \tau \tag{1.1}$$

However, neutrinos are not bound to their flavour. This phenomenon is called neutrino oscillations and means that a neutrino can change its flavour while propagating. It also implies that neutrinos do have a finite mass even though it is very small. This is in contradiction with the Standard Model. Among numerous other projects, determining its mass is a current research area concerning the neutrino.
Neutrinos only participate in weak interaction processes that can be divided in two different interaction types. They are denoted as charged current (CC) and neutral current (NC) interactions. Since neutrinos don't interact via strong or electromagnetic interactions, their cross section is very small. Hence, it is difficult to detect them.

Combining all these properties shows the benefits as well as the challenges of neutrino astronomy. In a way, neutrinos are optimal messenger particles for astronomy. Unlike charged particles, they are not deflected by magnetic fields. Thus, they point

back to their source. In addition, due to their low interaction rate, neutrinos are not absorbed and they barely interact. Hence, a neutrino can escape even from dense production sites, travel cosmological distances, and still reach the Earth. Furthermore, the detection of neutrinos would be an evidence for hadronic acceleration as this process implies neutrino production.
However, the small neutrino cross section is also a big challenge for its detection. Therefore, neutrino experiments require huge detectors with an enormous amount of target material. In addition, a good shielding against other particles is essential. For this reason, the detectors are located underground or even in the deep sea or ice and they are optimized to look downwards. This way, the Earth is used as a shield against upward going particles as most of them can't make it through.

1.2 Neutrino sources

Neutrinos are produced in a variety of different sources. Depending on the neutrino origin, the energies range from energies below meV for cosmological neutrinos (analogous to the CMB) up to the EeV range for cosmogenic neutrinos (see figure 1.1).

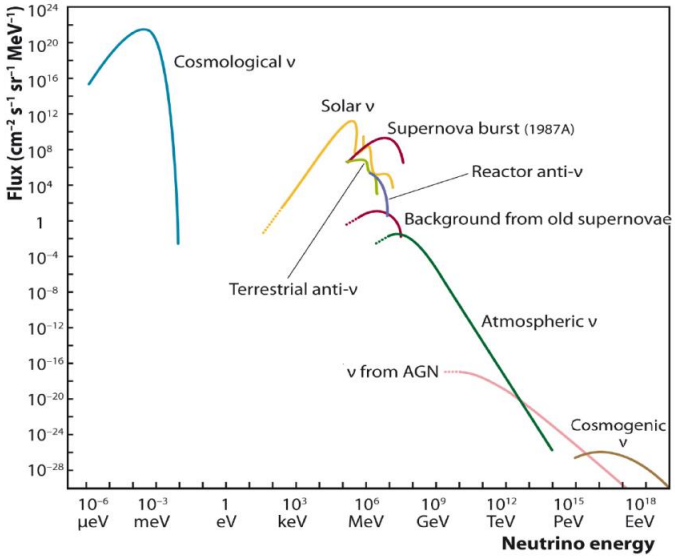


Figure 1.1: Measured and expected energy spectra of neutrinos from different sources. [6]

In the lower energy regime around a few MeV, some astrophysical neutrino sources like the Sun or supernovae are already known and observed. At higher energies, some cosmic accelerators are assumed to be possible sources but aren't proven so far. Only the existence of atmospheric neutrinos is certain at GeV energies and above.

Regardless of their production site, neutrinos are generated from the decay of

charged pions $\pi^{\pm}$:

$$\pi^{\pm} \quad \rightarrow \quad \mu^{\pm} + \nu_{\mu}\,(\bar{\nu}_{\mu}) \tag{1.2}$$

The produced muon can then further decay and give rise to two more neutrinos:

$$\mu^{\pm} \quad \rightarrow \quad e^{\pm} + \nu_{e}\,(\bar{\nu}_{e}) + \bar{\nu}_{\mu}\,(\nu_{\mu}) \tag{1.3}$$

Assuming all particles decay, this decay chain leads to a flavour ratio $\nu_e : \nu_\mu : \nu_\tau$ of $1 : 2 : 1$ at the neutrino source.

1.2.1 Cosmic neutrinos

The general assumed process for cosmic neutrinos is that cosmic rays interact at their production site leading to a neutrino generation like described in equation 1.2. The pions mainly result from $p\gamma$ or pp reactions that are the most likely processes for their production.
So far, there is only a small amount of astrophysical neutrinos that could be assigned to a high energy source. However, there are both galactic and extragalactic accelerators that are assumed to produce such neutrinos. In general, these are the same sources as for cosmic rays if non-thermal emission is involved. The first high energy neutrino that could be associated with a source was detected in September 2017 by IceCube. A coincident flare in gamma rays as well as an investigation of earlier data confirmed the blazar TXS 0506+056 as first high energy neutrino source [7, 8]. Additionally, a track-like neutrino event detected by IceCube in October 2019 is associated with the tidal disruption event AT2019dsg [9, 10].
As cosmic neutrinos travel large distances before reaching the Earth, neutrino oscillations become important and the flavour ratio converts into $1 : 1 : 1$ [11].
The following galactic and extragalactic sources are taken from [12].

Galactic neutrinos

An example for a potential galactic source are Supernova remnants (SNR). As for other candidates, particles are assumed to be accelerated in shock fronts. For SNRs the believed mechanism is that the supernova's ejecta collides with the interstellar medium.
Pulsars are fast rotating neutron stars. Their rotational and magnetic field axis do not coincide. Particle acceleration is assumed to happen due to their large magnetic fields in the order of magnitude of $\mathrm{B} \sim 10^{12}\,G$ [12].
Another possible site for particle acceleration including neutrino emission are binary systems with one partner being a neutron star or a black hole like X-ray Binaries. At a certain point the compact object starts accreting mass from its companion. The increased energy leads to X-ray emission and eventually also to the formation of jets. If the system forms two jets, it is referred to as microquasar. These jets are considered as potential regions of shock acceleration and thus neutrino emission.

Extragalactic neutrinos

However, the observation of high energetic cosmic rays with $E > Z \cdot 10^{17}$ eV, where Z is the charge of the particle, can't be explained by galactic sources [12]. On the one hand, galactic accelerators don't provide enough energy to produce such particles, but even if they could, these particle would escape from the galaxy because of their gyromagnetic-ratio exceeding the diameter of the galaxy. In addition, the observed isotropic arrival direction for such high energies can only be explained by very long propagation lengths that can only result from extragalactic sources.

A very promising extragalactic neutrino source are active galactic nuclei (AGN). These are very bright and strongly radiating core regions, that can be found in a small amount of galaxies. They are powered by a spinning supermassive black hole at their centre. It is surrounded by an accretion disk (in some cases) giving rise to two jets perpendicular to it. Some regions inside the jets are believed to build shock environments and thus lead to particle acceleration.
Another extragalactic candidate are gamma-ray bursts (GRB). These phenomena can be divided in short and long bursts. Short bursts are believed to result from the merger of compact binaries like two neutron stars or a neutron star and a black hole. Long bursts, on the other hand, are assumed to be caused by the collapse of a massive star into a black hole. A commonly accepted explanation is the fireball model. According to this theory, masses of material are emitted in shells and shock fronts can be built either by collision of two shells (internal shock) or the collision of a shell with the interstellar medium (external shock). As for SNRs and AGNs, these shock fronts lead to particle acceleration.
Starburst galaxies are another potential neutrino source. These are galaxies with a very high number of star formation and a high density of cosmic rays. The existence of relativistic electrons is highly suspected. Assuming an additional acceleration of protons, leads to the assumption of neutrino emission.

1.2.2 Atmospheric neutrinos

Atmospheric neutrinos are secondary particles from interactions of cosmic rays with the atmosphere. They originate mostly from the pion decay described in equation 1.2. In case of higher energies, the analogous decay of kaons $K^{\pm}$ is the more dominant process for the generation of muons and neutrinos.
For most neutrino astronomy experiments, atmospheric neutrinos are an inevitable background source. In contrast to the background due to atmospheric muons, which are only entering from above, atmospheric neutrinos can enter the detector from all directions as they can traverse the Earth even though the zenith angle is not uniformly distributed. A distinction from real astrophysical neutrinos can be done via the energy spectrum that is harder for cosmic neutrinos. Thus, above a certain threshold energy, the flux of cosmic neutrinos exceeds the one from atmospheric ones (see figure 1.1). [13]

1.3 Detection principle

The detection of a neutrino is quite challenging as it is not possible to measure it directly. Instead, one has to measure the charged particles originating from neutrino interactions. How to detect the secondary particles depends on the energy range. However, all neutrino experiments need a large amount of target material as the cross section is so small. Therefore, good shielding against background is crucial for such detectors.

1.3.1 Neutrino interactions

As already mentioned, neutrinos only participate in weak interactions. For neutrinos with energies above 10 GeV, the dominant interaction is deep inelastic scattering on nucleons. Therefore, there are two types of interaction: Neutral current (NC)

$$\nu_x + N \rightarrow \nu_x + \text{hadronic jet} \qquad \text{with} \quad x = e,\ \mu,\ \tau \tag{1.4}$$

and charged current (CC)

$$\nu_x + N \rightarrow x + \text{hadronic jet} \qquad \text{with} \quad x = e,\ \mu,\ \tau \tag{1.5}$$

These interactions are named after the charge of the respective exchange boson, i.e. $W^{\pm}$ for CC and Z^0 for NC reactions.
Depending on the type of interaction and the flavour of the participating neutrino, one can further distinguish between two different event signatures (see figure 1.2).

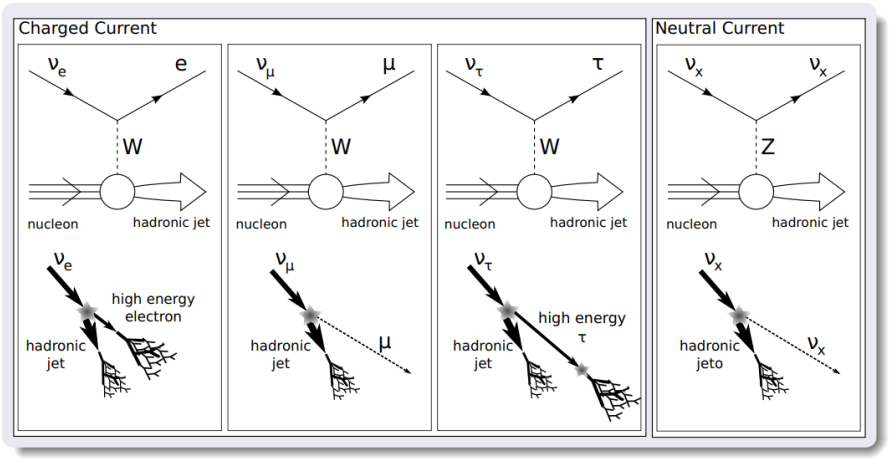


Figure 1.2: Illustration of neutrino-nucleon interactions. Neutral current and charged current ν_e interactions exclusively lead to a shower-like signature. A track-like signature is given by charged current ν_μ interactions and eventually also by charged current ν_τ reactions. [14]

Track-like events originate from charged current interactions of muon (anti)neutrinos

(ν_μ CC). The muon produced in this interaction can travel a long distance before it decays. Therefore, it leaves a track signature in the detector. This event type is best suited for direction reconstruction and hence for the search of point sources as the performance is of the order of magnitude of less than $1\,^\circ$ [15]. At high energies, this direction corresponds approximately to that one of the primary neutrino.
Unlike the angular resolution, the energy reconstruction is much more difficult and can only be determined statistically. One reason is that the amount of energy deposited inside the detector depends on whether the vertex is inside the detected volume and if the muon only passes through or ends inside the target material. Additionally, the energy loss of the muon depends on its energy. The energy loss by ionization is nearly constant and is about $2\,\frac{\mathrm{MeV}}{\mathrm{cm}}$ in water [6]. At higher energies, radiation processes dominate. The critical muon energy in water is about $E_{\mu,c} = 600\,\mathrm{GeV}$ [15]. This means that the energy loss profile is quite different above and below $E_{\mu,c}$. For low energies, it is nearly continuous but at higher energies the losses become more stochastically caused by local bursts along the trajectory.

The second signature are cascades or shower-like events. They originate from charged current interactions of electron (anti)neutrinos (ν_e CC) and from all flavour neutral current interactions (ν_e NC , ν_μ NC and ν_τ NC) within or at least close to the detector as a typical shower length of a few metres in water [15] is much shorter than a track.
Compared to track-like events the angular resolution is much worse but on the other hand energy resolution is better. In case of ν_e CC interactions, a hadronic shower will be produced as well as an electron that in turn produces an electromagnetic shower. Hence, all the energy is deposited near the interaction vertex. This does not hold for NC events. Here, most of the energy stays invisible as it is carried away by the neutrino. The visible part of this interaction is given by the hadronic cascade. The fraction of the neutrino energy that is transferred to the hadronic system is given by the Bjorken scaling variable y and depends on the neutrino energy as well as the fact if the primary was a neutrino or an antineutrino (see figure 1.3).

Charged current ν_τ interactions have to be considered separately as they can initiate both tracks and cascades. In the latter case, it might be not possible to resolve the hadronic cascade from the shower created by the decay of the tau. Only if the energy is high enough, normally in the PeV range, decay and interaction vertex can be separated and hence create the characteristic double bang signature for the tau neutrino. However, ν_τ reactions are not relevant for this work and therefore are not discussed further.

1.3.2 Cherenkov light

The detection principle depends on the neutrino energy. For high energy neutrino astronomy, the common way to measure neutrinos is via the Cherenkov effect [17]. The following explanation follows mainly [18].
A charged particle, like the secondary particle from a neutrino interaction, travelling

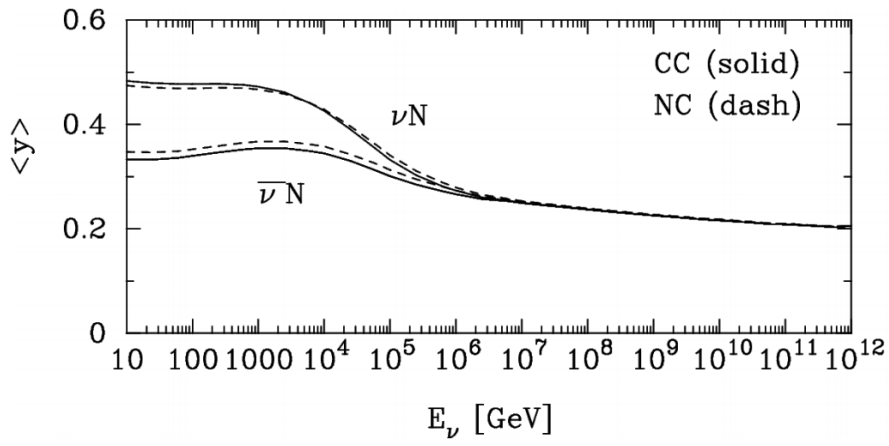


Figure 1.3: Mean value of the Bjorken scaling variable y illustrating the amount of (anti)neutrino energy transferred to the hadronic system for charged and neutral current interactions. [16]

in a dielectric, such as water, polarizes the medium for a short time at every point it passes. If the particle velocity is small, the polarization is symmetric and no resulting net field or emission can be detected at larger distances. For high velocities, the polarization field becomes asymmetric leading to a resulting field. In general, still no net radiation can be detected as the single wavelets interfere destructively. Only if the particle velocity is greater than the speed of light in the medium c_n, the interference becomes coherent and thus gives rise to wavefronts (see figure 1.4).

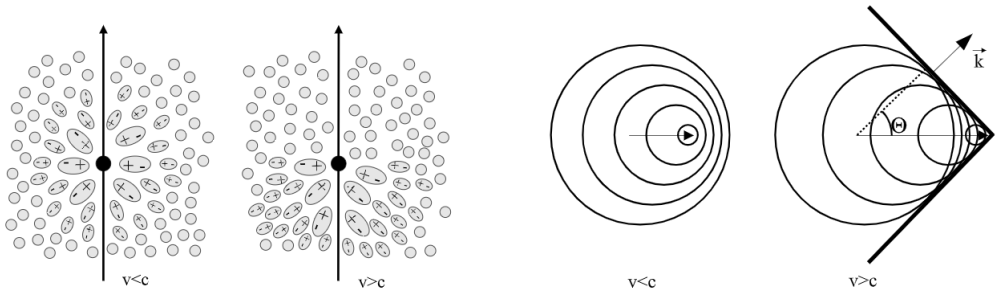


Figure 1.4: Illustration of Cherenkov radiation. *Left:* Symmetric and antisymmetric polarization for particles with a speed smaller and larger than the speed of light. *Right:* Constructive interference of the single wavelets, leading to the emission of Cherenkov radiation in the shape of a cone, is only possible in case of v > c. [19]

This so-called Cherenkov light is emitted in the shape of a cone with an opening angle that can be calculated with

$$\theta_C = \arccos\left(\frac{c}{vn}\right) = \arccos\left(\frac{1}{\beta n}\right) \tag{1.6}$$

For a particle with $\beta \approx 1$ the opening angle in water is about 43° [6].
The spectrum of Cherenkov radiation is dominated by shorter wavelengths and thus, the emitted light is blue.

1.4 ANTARES

1.4.1 General information

ANTARES (Astronomy with a Neutrino Telescope and Abyss environmental RESearch) is a neutrino telescope located in the Mediterranean Sea 40 km offshore Toulon at a depth of 2475 meters. With its completion in May 2008, it was the first operating undersea neutrino telescope.
The main goal of ANTARES is the search for cosmic neutrinos, but it also offers the possibility for research in marine and Earth science.
Due to its location, it allows for the observation of the central region of the galaxy and in general supplements the observed region of the Universe monitored by the IceCube experiment.
For more detailed information see [20].

1.4.2 Detector layout

The layout of ANTARES is illustrated in figure 1.5. The detector consists of 12 vertical lines with a spacing of about 60 meters. The strings are attached to the seabed by an anchor and kept in a vertical position by a buoy at the top. Except for one line, every string consists of 25 so-called storeys (black nodes in figure 1.5) with a vertical distance of 14.5 metres. The last line only holds 20 storeys but additional acoustic detection sensors. All storeys consist of three optical modules (OMs) each containing one photomultiplier tube (PMT) so that in total ANTARES is a three dimensional array of 885 PMTs. All PMTs are looking downwards at an angle of 45 degrees. This configuration leads to an improved sensitivity to upwards going particles and hence to a suppression of the background.

1.4.3 Signal and background in ANTARES

ANTARES aims for the detection of high energetic neutrinos from cosmic sources. However, this is by far not the dominant part of measured data. The main amount of recorded events results mainly from down-going atmospheric muons leading to a trigger rate of $5 - 10\,\mathrm{Hz}$ while only a few neutrinos per day are expected [20] . The cosmic neutrino signal can be distinguished from this background by rejecting all down-going events and by a separation via the energy spectrum that is in general harder for astrophysical neutrinos. This leads to an excess over atmospheric neutrinos above a certain threshold energy.
Additionally, there are some noise sources in ANTARES [21]. For one thing, bioluminescene from bacteria or other living organisms leads to an increase of the detector

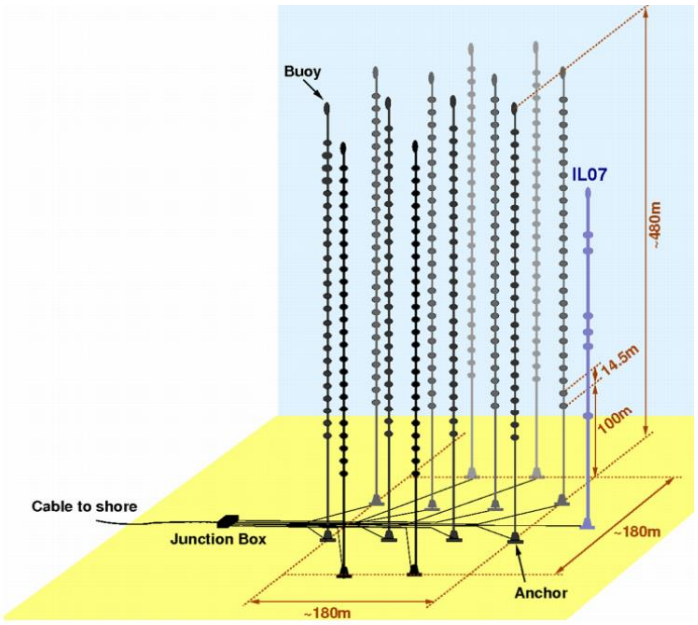


Figure 1.5: Illustration of the detector layout of ANTARES. [20]

counting rate. Besides, the constant radioactive decay of Potassium-40 ^{40}K via

$$\begin{aligned} {}^{40}K &\rightarrow {}^{40}Ca + e^- + \bar{\nu}_e \\ {}^{40}K \rightarrow {}^{40}Ar^* (+\bar{\nu}_e) &\rightarrow {}^{40}Ar + \gamma \end{aligned} \tag{1.7}$$

leads to further Cherenkov radiation, that is detected.
In general, the noise can be quantified by two parameters [22]. The first one is the baseline, that is the average photon counting rate for all working PMTs in the detector for a certain time. It builds the homogeneous and continuous part of the noise and is composed of photomultiplier noise, radiation from the ^{40}K decay, and bioluminescence from bacteria. The second parameter results from short light bursts induced by living organisms. Therefore, it is called burst fraction. It quantifies the fraction of entries in the counting rate distribution with a rate larger than $1.2 \cdot \mu$ with μ being the mean of the Gaussian fit with mean at the maximum of the distribution.

1.4.4 Monte Carlo Simulations

The simulations for ANTARES are generated in a so-called *run-by-run* approach. This strategy takes into account that the conditions for the detector might change over its lifetime or only for a short time period. Therefore, each run is simulated with the corresponding parameters that are directly extracted from the raw data. This includes changes in the detector layout like missing PMTs as well as environmental conditions like bioluminescence or baseline.
Simulations are generated for CC and NC interactions of all neutrino flavours using GENHEN (GENerator of High Energy Neutrinos). The energy spectrum ranges

from a few GeV up to 10^9 GeV and particles from all arrival directions are produced. The statistics of simulated neutrino events is much larger than the number of expected neutrinos. Depending on the neutrino source and spectrum, a reasonable weighting needs to be done when doing a physics analysis.
Atmospheric muons are simulated with MUPAGE [23]. Only down-going muons with an energy between 20 GeV and 50 TeV are available. Unlike for neutrinos, only 1/3 of the expected number of muons is simulated.
The event generation is followed by particle and light generation which includes the generation of Cherenkov light, determining if the photons cause a hit, and the particle propagation in water. Finally, the data acquisition simulation takes place. During this step, not only optical background is added but also the environmental conditions are taken into account. In order to be competitive with the real data, read-out and trigger algorithms are also applied in the simulation chain.
For more details see [24].

2 Deep Learning

Deep learning belongs to the field of machine learning. Its benefit compared to the classical approach is that the network learns the features, i.e. characteristics from the raw data that are used as actual input, by itself. Hence, it is able to learn directly from the unprocessed data without human interference. In this way, deep learning can be very time saving as the process of feature engineering can be skipped. Furthermore, the performance of deep learning exceeds the one of earlier techniques if the dataset is large enough. In times of Big Data, large amounts of data are now often easily accessible. In addition, hardware and software components improved over the last years leading to the latest success of deep learning.
This chapter covers the fundamental basics of deep learning with artificial neural networks including the most important parameters, hyperparameters, and techniques. Special attention is paid to convolutional neural networks.

2.1 Supervised learning

Like in classical machine learning, the learning process can be supervised or unsupervised. However, this work is restricted to supervised learning.
In this approach the networks are trained with labelled data. This training set is used to fit the free parameters. All samples are successively fed into the network. When the total set has been used, one epoch is completed and the training inputs are used again. Therefore, it might happen that the network gets too well adapted to these samples. This might lead to a poor performance on an independent dataset. In order to avoid this, the validation set is used. The network does not learn from these samples but it is applied to it in order to get an unbiased performance estimate. However, this set is used to optimize the hyperparameters of the model. These are non-learnable parameters that have to be set manually and which can affect the success of training and learning enormously. In the end, it is therefore recommended to test the performance on an unseen dataset. This is called test set.

2.2 Neural networks

The main goal of deep learning is to determine a function $F(x, \theta)$ which maps the input x to the correct output $F^*(x) = y_{true}$. This is usually achieved by using neural networks.
The basic component of such a neural network is a neuron. It receives several inputs x_i that can be combined as the input vector $\vec{x}$. These input values are then multiplied by their according weights w_i, with corresponding weight vector $\vec{w}$, and summed up.

Eventually, a constant bias term b can be included to this so-called weighted sum z. Finally, a non-linear activation function f is applied to z in order to determine the neuron's output value y. In total, the mathematical operation performed by a neuron can be described by

$$y = f(z) = f(b + \sum_i w_i x_i) = f(b + \vec{w} \cdot \vec{x}) \tag{2.1}$$

and is illustrated in figure 2.1.

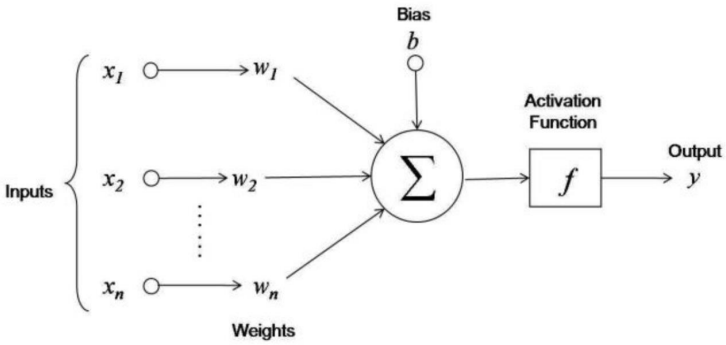


Figure 2.1: Illustration of the mathematical operation performed by a single neuron. [25]

Neural networks consist of several layers each containing one or more of such neurons. In a feed-forward network or multilayer preceptron these layers are connected in such a way that the output of the neurons of one layer is the input to the neurons of the following layer.
The first and last layer are input and output layer, the ones in between are referred to as hidden layers. A network with two or more hidden layers is called deep. An example can be seen in figure 2.2.

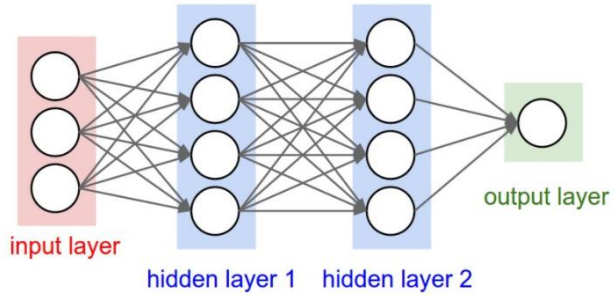


Figure 2.2: Deep feed forward neural network with two hidden layers (adapted from [26]).

In this case, the output of one layer can be displayed in a similar but slightly more complicated way than equation 2.1. The output of the n-th layer transforms into a vector $\vec{y}^{\,n}$ where each component y_i^n is given by the output of neuron i in this layer.

The weight vector turns into a weight matrix $\mathbf{W}^n$ with entries w_{ij}^n referring to the weight for the connection between neuron i in layer n and neuron j in layer $n-1$. Finally, the bias, which is now designated $\vec{b}$, also turns into a vector. In total, the output of the n-th layer is given by

$$\vec{y}^{\,n} = f(\vec{b}^{\,n} + \mathbf{W}^n \cdot \vec{x}^{\,n}) = f(\vec{b}^{\,n} + \mathbf{W}^n \cdot \vec{y}^{\,n-1}) \tag{2.2}$$

In this way, the approximated function $F(x, \theta)$ depends on all the n layers in the network.
The output of the last layer is the output of the network. Depending on the type of network, the prediction can be a class score (classification) or a real value (regression).

2.3 Loss functions

During supervised learning, the network produces an output (prediction) y_{pred} for each given labelled input. The goal of the learning process is to adapt the weights and biases of the network in such a way that this prediction is equal or at least as close as possible to the true label of the sample y_{true}. In order to monitor how well this is achieved, the loss or cost function is introduced. It depends on the learnable parameters of the network.
In this sense, the loss is a quantity to determine the performance of the network and is therefore one of its most important hyperparameters. It is a measure for the deviation between predicted value and label. The higher the difference, the more it gets penalized. It is monitored during the training process in order to update the parameters properly (see section 2.5).
The cost function is computed for all N training samples. The total loss is then given by the average of the individual losses:

$$L = \frac{1}{N} \sum_i L_i \tag{2.3}$$

Depending on the type of network, there are several common loss functions.
For classification tasks there are usually j different outputs each referring to a score for one of the j different categories. It is very common to use this so-called one-hot encoding. That means that if sample i belongs to class k, the k-th component of the label $y_{i,k}$ is 1 and all the others $y_{i,j}$ are equal to 0.
A very common loss function for classification tasks that has been used for this work is cross entropy. This cost function is usually connected to a special activation function (see section 2.4) in the last layer called softmax activation function:

$$f(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \tag{2.4}$$

This way, each class score gets mapped to a value between 0 and 1 and additionally

they all sum up to 1. Hence, the resulting class scores of sample i can be interpreted as normed probabilities $p_{i,j}$ to belong to the corresponding class j.
The actual cross entropy loss is then defined as:

$$L_i = -\sum_j y_{i,j} \log\left(p_{i,j}\right) \tag{2.5}$$

In case of regression, the network's output is not discrete but takes on continuous values. Therefore, a typical approach to penalize deviations from the true value is measuring the distance between label and prediction. This can be done via the mean squared error (MSE)

$$L_i = (y_{true,i} - y_{pred,i})^2 \tag{2.6}$$

or very similar with the mean absolute error (MAE)

$$L_i = |y_{true,i} - y_{pred,i}| \tag{2.7}$$

The choice of loss function depends very much on the network's task and has to be selected carefully. Sometimes it might be reasonable to calculate the loss of the logarithm of y_{pred} and y_{true} rather than the real value. An example is the energy reconstruction of neutrinos where the mean absolute error penalizes deviations for high energies much more even if the relative error is quite small.

2.4 Activation functions

There are various common activation functions but they all got in common that they are non-linear functions. This is very important as for a network that only uses linear activation functions the output of the network also is just a linear function of the input. This could also be realized by a single layer. Therefore, such a network is restricted in its ability to model more complex functions.
On the other hand, choosing a non-linear function means that any arbitrary continuous function can be approximated as well as desired with a single hidden layer if enough neurons are available [27, 28]. This is known as the universal approximation theorem.
In general one can choose out of several common options [26]: One activation function, which was mainly used for many years, is the sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.8}$$

The use of the sigmoid function is mainly substantiated with its analogy to biology and offers a simple interpretation. Unfortunately, it suffers from the vanishing gradient problem for very low or large values which prevents the network from learning as the parameter updates become very small (see section 2.5). Additionally, it is

not zero-centred which means that later layers in the network will not receive zero-centred input.
This problem can be fixed by using the tangens hyperbolicus:

$$\tanh(x) = 2\sigma(2x) - 1 \tag{2.9}$$

but still this activation doesn't solve the vanishing gradient problem which makes it unsuitable to large networks.
Therefore, a very common choice nowadays is ReLU (Rectified Linear Unit) which is given by:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{2.10}$$

This function is very easy and fast to compute and, unlike sigmoid and tangens hyperbolicus, it does not saturate. Although it suffers from a problem called dying ReLU that leads to no weight updates, it is currently the state of art for activation functions.
For this work ReLU has been used for most of the layers. An exception is only made for the output layers.
However, in practice there are some cases where slight modifications of ReLU might be more successful, like for example leaky ReLU. An example that has been used for this work is the softplus function:

$$f(x) = \log(1 + \exp(x)) \tag{2.11}$$

It is a smooth version of ReLU and might be the activation function of choice if a strictly positive output is desired.

2.5 Optimizers and Backpropagation

As described in section 2.3, the goal of training is to adapt the free parameters w_{ij} and b in such a way that the loss becomes minimal. Therefore, weights and biases need to be updated regularly. The optimizer is the method which is used to achieve this goal.

2.5.1 Gradient descent

There are several options that can be used but most of them somehow use gradient descent. The loss is a function of weights and biases which means that for finding the optimal parameters one has to find the minimum of the cost function. Therefore, a very common approach is to go along its negative gradient.
The free parameters are then updated according to

$$\theta_j \to \theta_j - \eta \frac{\partial L}{\partial \theta_j} \tag{2.12}$$

where θ_j can denote any weight or a bias.
The parameter η is called the stepsize or learning rate. It is a very important hyperparameter that has to be chosen carefully. If it is set too small, the training

process can be very slow, inefficient, and can get stuck in a local instead of the global minimum of the loss function. On the other hand, if the learning rate is very high, the system can converge too early before reaching the optimal region or even not converge at all as it jumps around arbitrarily. Usually, the learning rate isn't even a fixed value but is reduced via step decay during training. However, there are also some suggestions to use cyclic learning rates [29].

Unfortunately, gradient descent is not always very practicable as it can be computationally very costly. As defined in equation 2.3, the loss is the average value of the individual losses of each training sample. This way, calculating the gradient means calculating the gradient for each sample separately as well and then evaluating the average. The computation of one parameter update therefore requires going through the complete set of training inputs. Obviously, for a large set this can slow down the learning process enormously.
A common solution to tackle this is called stochastic gradient descent (SGD). In this approach only a random subset of training samples is used (sometimes referred to as mini-batch) in order to calculate an average gradient. The idea behind this algorithm is that the average gradient of a subset is approximately equal to the average gradient of the total training set. This way, the parameters get updated after each mini-batch.

2.5.2 Adam

Stochastic gradient descent was a very popular optimizer for many years. However, over the years several modified variants have been developed aiming to speed up the learning process mainly by adapting the learning rate for the individual parameters separately. One of the currently most popular optimizers was presented by Kingma et al [30]. The name *adam* stands for adaptive moment estimation referring to the use of first and second momentum of the gradients.
For every timestep t, the exponential moving averages m_t and v_t of the gradient g_t and the elementwise squared gradient g_t^2 get updated in the following way:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{2.13}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{2.14}$$

The exponential decay is regulated by β_1 and β_2. These newly introduced hyperparameters set the main focus on the gradients calculated only a few steps before. The authors propose default values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$.
Both moving averages are initialized with zero and thus lead to a bias towards zero. Therefore, a bias correction is performed:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{2.15}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{2.16}$$

Finally, the weight updates are given by

$$\theta_t \rightarrow \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{2.17}$$

with η being the learning rate and ϵ being a small value preventing a vanishing denominator.
All networks trained for this thesis use the adam optimizer. However, the settings for the hyperparameters differ from those suggested by the authors.

2.5.3 Backpropagation algorithm

Of course there are several additional options to chose the optimizer. But as already seen here for gradient descent and adam, both of them make use of the loss function's gradient. Usually, it is calculated by using the backpropagation algorithm.
The name results in the fact that, in contrast to the output that is calculated in the forward pass from the first to the last layer, the gradients have to be computed the other way around.
The loss is a function of the output $\vec{y}^{\,n}$ which can be calculated using equation 2.2. Therefore, calculating the derivative with respect to an arbitrary weight from the last layer w_{ij}^n connecting neuron j from layer $n-1$ to neuron i from layer n is straight forward:

$$\begin{aligned}\frac{\partial L}{\partial w_{ij}^n} &= \sum_p \frac{\partial L}{\partial y_p^n}\frac{\partial y_p^n}{\partial w_{ij}^n} = \sum_p \frac{\partial L}{\partial y_p^n} f'(z_p^n)\frac{\partial z_p^n}{\partial w_{ij}^n} = \\ &= \sum_p \frac{\partial L}{\partial y_p^n} f'(z_p^n) y_j^{n-1}\delta_{i,p} = \frac{\partial L}{\partial y_i^n} f'(z_i^n) y_j^{n-1}\end{aligned} \tag{2.18}$$

The sum takes into account all p neurons in the last layer. If the derivative of a bias from the last layer is used, the expression changes slightly:

$$\frac{\partial L}{\partial b_i^n} = \sum_p \frac{\partial L}{\partial y_p^n} f'(z_p^n)\delta_{i,p} = \frac{\partial L}{\partial y_i^n} f'(z_i^n) \tag{2.19}$$

In case of a weight from an earlier layer w_{ij}^k, the term $\frac{\partial z_p^n}{\partial w_{ij}^k}$ from equation 2.18 has to be evaluated more carefully as z_p^n does not directly depend on w_{ij}^k:

$$\begin{aligned}\frac{\partial L}{\partial w_{ij}^k} &= \sum_p \frac{\partial L}{\partial y_p^n}\frac{\partial y_p^n}{\partial w_{ij}^k} = \sum_p \frac{\partial L}{\partial y_p^n} f'(z_p^n)\frac{\partial z_p^n}{\partial w_{ij}^k} = \\ &= \sum_p \frac{\partial L}{\partial y_p^n} f'(z_p^n) \sum_s \frac{\partial z_p^n}{\partial y_s^{n-1}}\frac{\partial y_s^{n-1}}{\partial w_{ij}^k} = \\ &= \sum_p \frac{\partial L}{\partial y_p^n} f'(z_p^n) \sum_s w_{p,s}^n f'(z_s^{n-1})\frac{\partial z_s^{n-1}}{\partial w_{ij}^k}\end{aligned} \tag{2.20}$$

The additional sum refers to the s neurons in layer $n-1$.
This procedure is now repeated until the term $\frac{\partial z_l^k}{\partial w_{ij}^k}$ appears. This can now be evaluated as

$$\frac{\partial z_l^k}{\partial w_{ij}^k} = y_j^{k-1}\delta_{l,i} \tag{2.21}$$

or respectively in the case of a bias:

$$\frac{\partial z_l^k}{\partial b_i^k} = \delta_{l,i} \tag{2.22}$$

In this manner any derivative for weights and biases of earlier layers can be derived from using the derivatives from all weights and biases from the later layers.
In summary, the gradients can be calculated layer after layer but compared to the forward pass this is done in a backward pass from the last layer to the first one. In other words, the corrections for the later layers have to be done first.

2.6 Weight initialization

The way the free parameters get updated was already explained in section 2.5. However, this does not include how to choose reasonable weights in the first step. This procedure is called weight initialization.
Choosing a starting value for the free parameters can have a large impact on the training process. For example, setting all weights to zero will cause all neurons to predict the same output which obviously prevents the network from learning effectively.
However, due to data preprocessing, it is still reasonable to assume that the mean value of all weights is equal to zero. Hence, initializing the weights with values from a zero-centred gaussian distribution is intuitive. Unfortunately, the variance of a single neuron's output scales with its input dimension leading to a worse performance due to a shift in the activation distribution for later layers. The desired goal however, are similar activation distributions for all layers, not only the first ones.
There are different approaches to combat this issue by restricting the variance of the weights to a certain value. In general, the strategy is to use smaller weights for neurons with a larger number of inputs.
A solution that works in a linear regime (activation function zero-centred) was presented by Glorot et al. [31]:

$$Var(w) = \frac{2}{n_{in} + n_{out}} \tag{2.23}$$

with n_{in} and n_{out} being the number of units in the previous and next layers.
Unfortunately, ReLU is not a zero-centred activation function. Hence, some of the assumptions leading to equation 2.23 do not hold. However, He et al. [32] developed a restriction for ReLU:

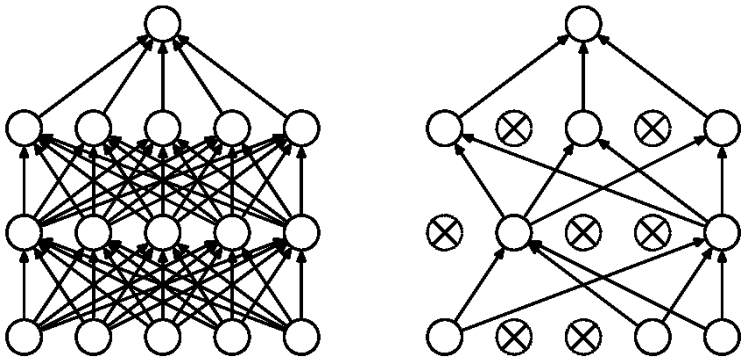
$$Var(w) = \frac{2}{n} \tag{2.24}$$

Hereby n denotes the input dimension.
As this work mainly uses ReLU as activation function, the weight initialization follows [32].

2.7 Dropout

A danger when training neural networks is overfitting. This term refers to the phenomenon that a network gets adapted too well to the training data and gets trained on the underlying noise rather than the important features. As a consequence, the performance on the unseen test data becomes very poor which is of course not desirable.
Therefore, some regularization techniques can be exploited in order to combat overfitting. Some very simple methods provide more training data to prevent the network from memorizing the samples. Another approach is the early stopping criteria, i.e. terminating the training process before overfitting occurs.

A very common and simple regularization technique, that is used for this work, is dropout [33]. For this method a fixed dropout rate p is chosen. This percentage of neurons is then switched off during the training process, meaning that their value is set to zero. This way, they don't contribute to the network (see figure 2.3). The neurons which are dropped out are chosen randomly for every new batch. In this sense, dropout acts like introducing noise to the network leading to a more robust and unbiased performance.
In other words, each batch of training samples is fed into a different thinned representation of the network. For testing no neuron is dropped out. Instead, the weights are multiplied by p in order to ensure that the output for the test samples is equal to the expected output from training with dropout.



(a) Standard Neural Net (b) After applying dropout.

Figure 2.3: Schematic illsutration of droput. Neural network without (left) and with (right) dropout. Circles with a cross denote dropped out neurons. Hence, the connections are removed. [33]

2.8 Batch Normalization

Another technique which is not always directly referred to as regularization but contributes to it is batch normalization [34]. It was introduced to combat the so-called internal covariant shift. This denotes the change of distribution for each input layer during training as it depends on the previous layers. By applying batch normalization each layer's input gets normalized and follows a Gaussian distribution. Therefore, it is possible to use higher learning rates and accelerate the training process.

The normalization process is usually done over a batch B of m samples x_i. In order to get a uniform distribution, mean μ_B and variance σ_B of the mini-batch are calculated in the standard way:

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{2.25}$$

$$\sigma_B = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \tag{2.26}$$

The actual output of the batch normalization is then given by

$$y_i = \gamma\frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \tag{2.27}$$

with γ and β being learnable parameters that ensure that the batch normalization operation can represent the identity transformation. The parameter ϵ is a constant value used for numerical stability.

2.9 Convolutional Neural Networks

Convolutional neural networks (CNNs) use a special network architecture that makes them particularly powerful for analysing images. In general, they can be applied to any kind of data that can be brought into an image-like shape. Usually, they consist of different layer types.
The core type is the **convolutional layer**. The main difference to an ordinary network layer is the fact that the neurons in a convolutional layer are not connected to all the neurons in the previous layer but only to a small part of them in its vicinity. In this way, CNNs don't only greatly reduce the number of weights but also maintain the spatial structure of the input.
The classical weights are therefore arranged in filters with the same number of dimension as the input image. Their size is usually smaller than the one of the image with the exception of the last dimension that extends to the full size of the input. The filter, or kernel size, determines the size of the region of influence of the previous layer.

The filter is slided over the input image in a grid fashion, as it is illustrated in figure 2.4. The stepsize or stride is a hyperparameter that has to be chosen. It determines the output volume. At every point of this grid the kernel is applied to the respective field calculating the dot product and hence determining the output value in the corresponding pixel. This way, a feature can be detected anywhere in the image. The full output of one filter is called activation map. Typically one layer consists of several of such filters each representing a different feature. The filters in the lower layers usually represent more basic features like edges or diagonals while filters from later layers learn more complex structures. All the produced activation maps from one layer get stacked on each other. This way, the number of filters in one layer determines the (last) output dimension of the resulting image.
Figure 2.4 also illustrates the concept of zero-padding. This is a technique mostly aiming for preserving the spatial volume of the input by setting all pixels outside the actual image to zero.

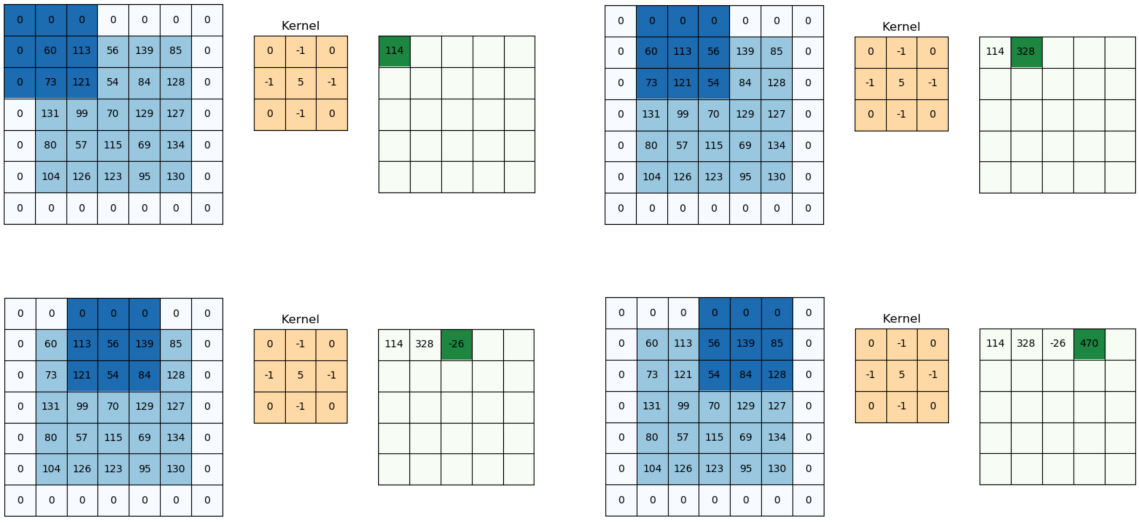


Figure 2.4: Illustration of the operation done by a convolutional layer. Only the first four steps are shown, but the operation continues analogously (adapted from [35]).

After the covolutional layers usually the activation function is applied.

However, there is a second important layer in CNNs which is called **pooling layer**. It is applied in regular intervals after some convolutional layers. The purpose of such layers is downsampling so that the number of parameters gets reduced and overfitting can be prevented.
There are different options of pooling but currently the most used one is max pooling where only the highest value of a certain region is kept (see figure 2.5). This way, max pooling focusses on the dominant features.

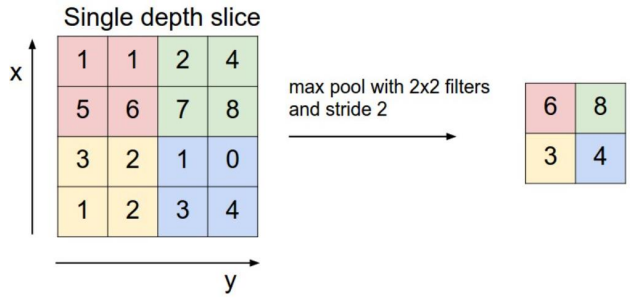


Figure 2.5: Illustration of the operation performed by a max pooling layer (adapted from [36]).

Finally, at a certain point there is no more use of exploiting translational invariance. The last layers in a CNN are therefore usually **fully connected layers** that compute the output. Neurons from these layers are connected to all neurons from the previous layer.

In 2012, CNNs had their breakthrough when winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [37]. Since then they are state-of-the-art for image recognition but they are also applied in text and speech recognition.

3 Preprocessing

Before starting the training of the networks, some preprocessing steps have to be done that aim for bringing the input data into the right shape for the given network type. In case of CNNs, as used for this work, these are images. The main generation process is adopted from the methods developed for KM3NeT/ORCA [38] but has to be adapted slightly to ANTARES in some points. Furthermore, the events for the training have to be chosen. This chapter explains the procedure as well as the event selection.

3.1 Data conversion

The very first step in preprocessing is bringing the data into the desired file format. In order to generate images using the same approach as for ORCA, the required file type is hdf5. Therefore, a python tool has been developed in order to convert the ANTARES data from the i3-format to hdf5. This converter is now integrated in *km3pipe* [39], a framework built for working with KM3NeT data.

3.2 Image generation

As it was mentioned before, the input for CNNs are images. Currently, the maximum available dimension of convolution is three, which limits the maximum input image dimension to four. All the input images for this work hence are four dimensional. They contain the spatial, i.e. x, y, and z position, and temporal information of the hits from an event.
The main process of image generation is done with *OrcaSong* [40], an open source framework for preparing deep learning data of KM3NeT.

3.2.1 Spatial binning

The choice of spatial binning is completely determined by the detector layout (see section 1.4.2).
An ideal binning would provide one PMT per bin. However, this very fine binning poses several challenges. First of all, it requires a high number of at least 885 bins. Even for an event with a high number of hits, the major part of the bins would be empty causing an unnecessary high amount of data and thus computational cost. In addition, a distinction between the three PMTs of one storey means a very small bin size in x and y direction. For a fixed detector geometry, this is no problem. However, due to the sea current, the positions of the lines and hence the individual

OMs are not completely fixed as they move or even rotate slightly. In case of small bins, this means that a PMT can easily leave its foreseen bin due to this movement so that one PMT could contribute to different bins. This could be difficult for the network.
As a compromise between sufficient local resolution and a reasonable image size, the binning is chosen in such a way that every storey is in one bin. Thus, the three PMTs of one storey share one bin.
Choosing the binning for the vertical component is straight forward. The number of bins is set to 25 which is the number of storeys per line. Even though one line is moderately offset regarding all other strings, one bin per 14.5 metres is sufficient. As the horinzontal displacement of the topmost storey is of the order of magnitude of a few metres [41], the storey stays in its foreseen bin even if the lines tilt in the sea current.
The x and y component are a slightly more difficult challenge, depending on the angle of top view. Since the initial detector layout (as it is stored in the original i3-files and therefore also in the hdf5 files) makes it difficult to choose a suitable grid, the complete detector layout is rotated for every event before applying the binning (see figure 3.1). By applying a grid of 4×4, every line is in one bin with one bin covering 60 metres in x and y direction, respectively.
In summary, the spatial binning is chosen to be $4 \times 4 \times 25$. Thus, the number of spatial bins is reduced from 885 (number of PMTs) to 400.

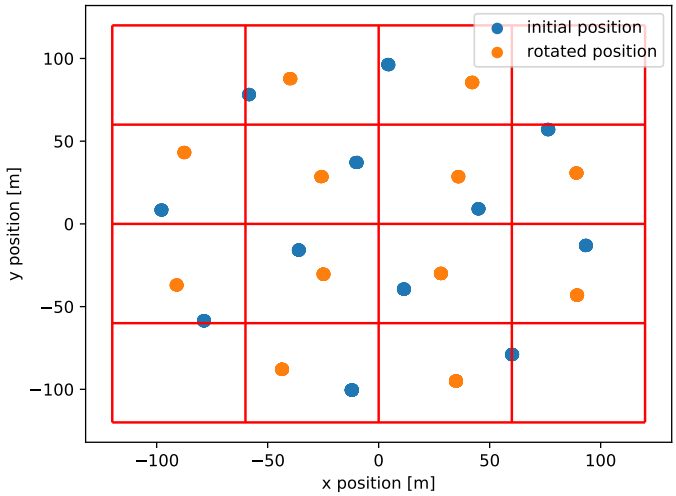


Figure 3.1: Top view of the ANTARES detector layout before (blue) and after (orange) applying the rotation. The dots represent all individual PMT positions. The chosen xy-binning is illustrated by the red grid.

3.2.2 Temporal binning

Similar to the challenges for the spatial binning, temporal binning also needs to be well considered. A large number of bins leads to a very good time resolution but also the number of empty bins and along with it image size increases.
Therefore, the first thing to be done is to apply a time cut to every event. Analysing the Monte Carlo signal time distribution shows that the main part of relevant hits lies in an asymmetric time range of about 2000 ns around the first triggered hit (see figure 3.2). Hits from outside this selection (before −400 ns and after 1600 ns) are not included into the images. This analysis is done for the longest event type in ANTARES, i.e. for muon events. Neutrino events, especially shower-like events, are much shorter so that a suitable cut for muon events is definitely sufficient for neutrino events as well.
By setting the number of temporal bins to 100, a time resolution of 20 ns is achieved. For light, this is equivalent to a distance of 4.5 m in water. Considering the fact that the three PMTs of one storey are in one bin anyway and taking into account the minimal distance between two neighbouring storeys of 14.5 m, this means that the chosen time binning is sufficient.
An example of all possible two dimensional projections for an input image is shown in figure 3.3.

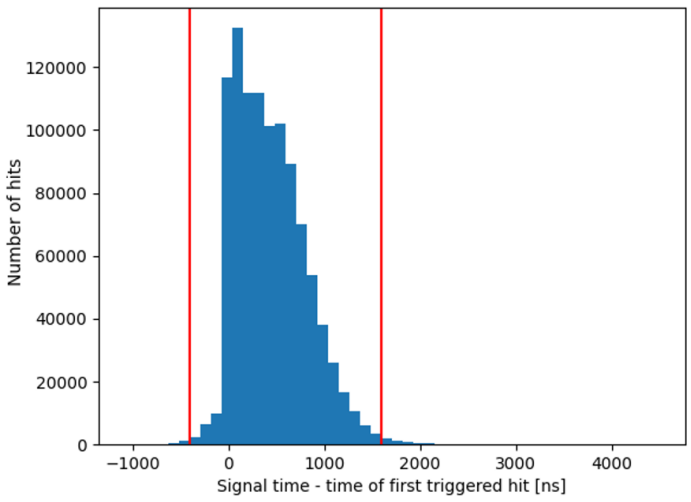
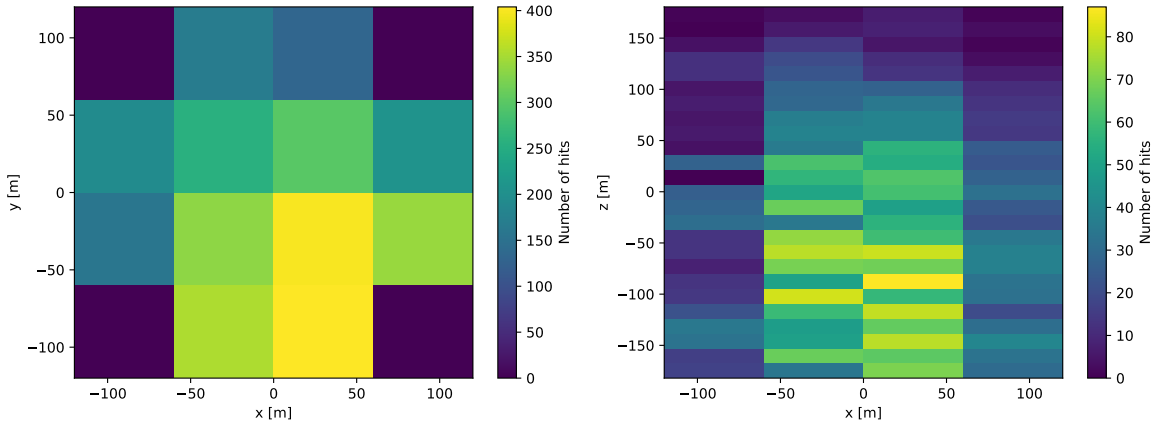
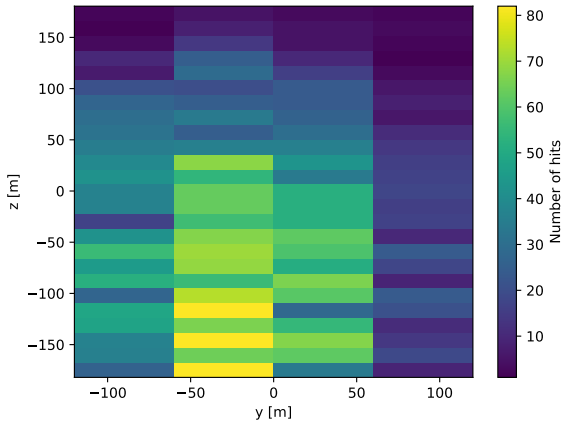


Figure 3.2: Time distribution of signal hits relative to the first triggered hit of about 5000 muon events. The red lines indicate the chosen timecuts.

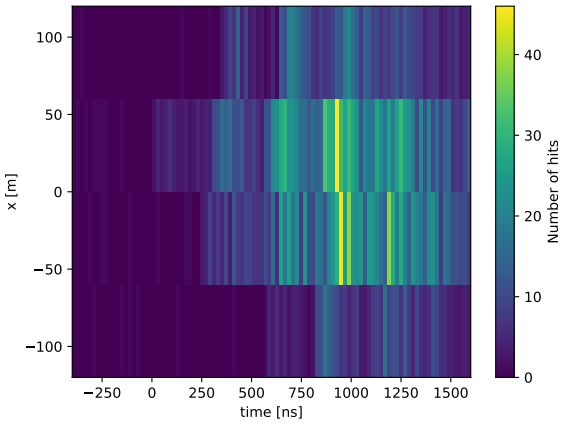


(a) xy-projection

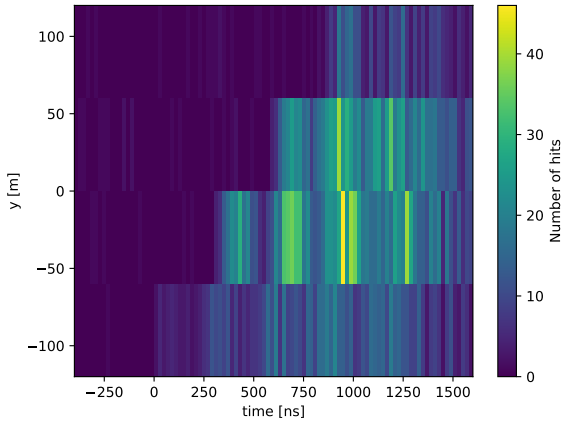
(b) xz-projection



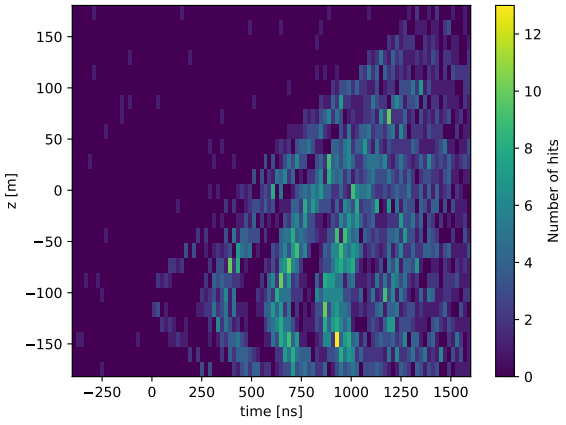
(c) yz-projection



(d) xt-projection



(e) yt-projection



(f) zt-projection

Figure 3.3: All possible two dimensional projections of an exemplary input image from a $\bar{\nu}_\mu$ NC interaction (run 36514, event 47242, $E = 5.2 \cdot 10^7$ GeV, dir z $= -0.16$). The time is relative to the time of the first triggered hit.

3.3 Event selection

In order to train a network properly, the training set has to be chosen carefully. This includes considerations like the number of events for different classes, e.g. the different neutrino flavours and muons as well as a reasonable split into training, validation and test sets (see section 2.1). Included event types, balancing and split into the different sets will be explained in section 4.1 and section 5.1 for the respective networks. However, especially in terms of deep learning, a large amount of training data is required. This section explains the considerations of choosing the simulations (and also data) runs that are used for the training.

One of the benefits of ANTARES is that it is constantly taking data since 2008, thus offering not only the Monte Carlo run-by-run simulations to train the network but also measured data to test the network's performance. However, the long lifetime of the detector also offers some challenges like unstable data taking conditions. One reason are seasonal variations like fluctuating bioluminescence rates. On the other hand, ANTARES suffers from a general sensitivity decrease over the years. Therefore, single PMTs or even complete lines might be temporarily or permanently missing. Thus, recorded data (and also simulations) for different episodes show significant differences. In terms of the input images of a CNN, missing lines or OMs can lead to complete dead pixels. However, the network can't distinguish whether an empty bin results from no hit or a missing PMT. This way, it is unclear how the network can deal with these changing parameters.
A solution could be training several networks for different detector configurations. However, it is not straight forward according to which subdivisions the different training sets have to be selected in order to optimize the performance of the network. Possible categories that could be taken into account are parameters like the respective detector geometry as well as the amount of expected noise, quantified by bioluminescence, baseline or burst fraction.
The simplest and most intuitive approach is to choose a connected period as stable as possible while offering enough data for training. Thereby, systematics and problems of the deep learning for ANTARES can be detected before addressing more complicated challenges like the sensitivity decrease. Therefore, the main training period of this work is half a year from September 2008 to February 2009. The networks presented in this work will mainly be trained on this time.
Apart from this main training set, some later periods have been used to perform a detector condition analysis for the background classifier. Therefore, three additional time frames have been chosen: July 2010, February 2012 and September 2014.
Finally, a second training period is chosen. It includes simulations from 2010 in order to compare how the performance on the different episodes changes.
The following is a summary of the named episodes:

Training 2008/2009 (run 35151 to 39389): This is the main training period of this work comprising six months from September 2008 to February 2009. It is an early phase of the detector mostly running in its full configuration with all 12 lines.

Since this six month period is during winter, rates and other important parameters are quite stable. In total, this episode can be considered as an ideal data taking period regarding internal as well as external conditions. The performance on data is tested exclusively on this time period.

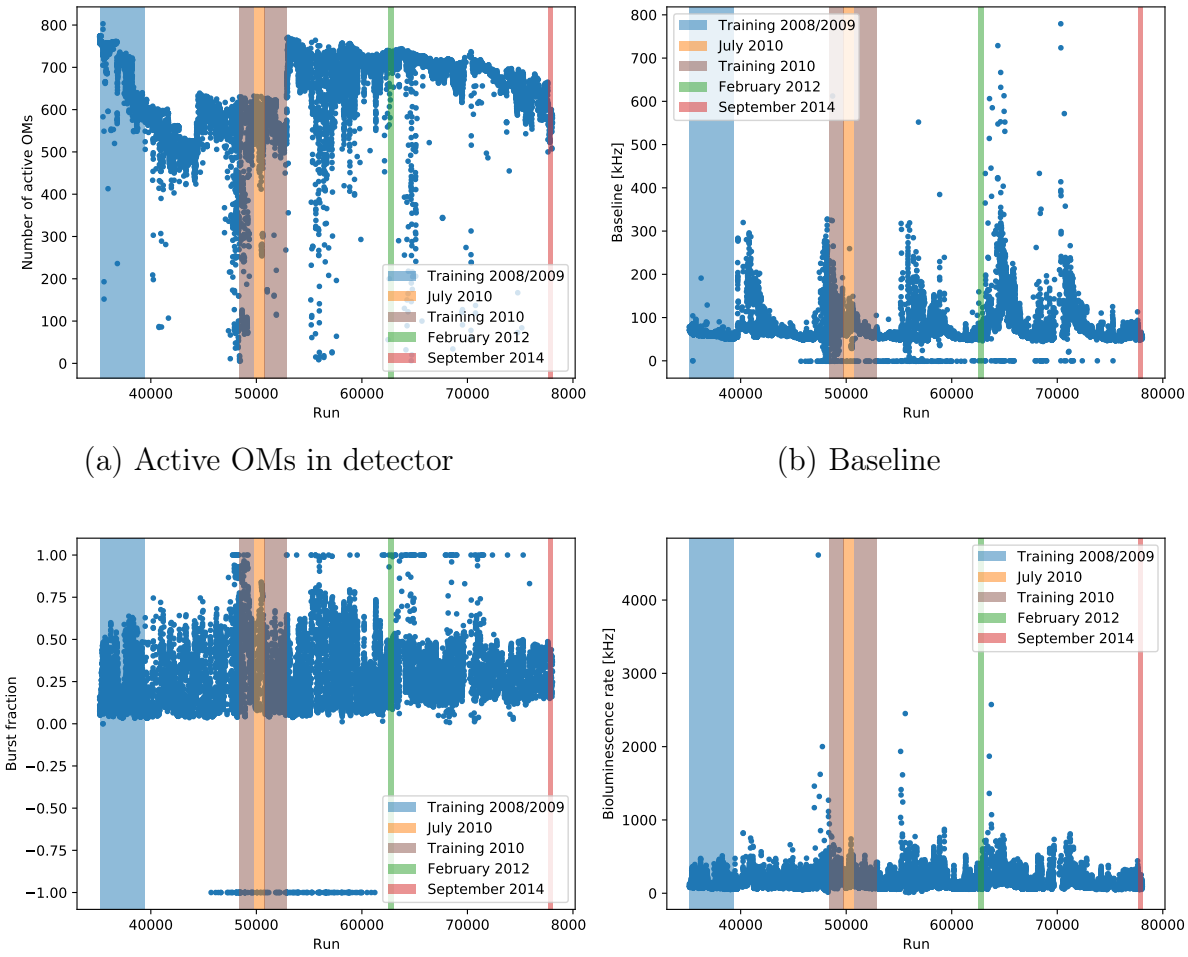
July 2010 (run 49770 to 50759): Unlike the main training period, this is a summer month. Consequently the rates are higher. In addition two lines are completely missing.
February 2012 (run 62505 to 63030): This period offers a good comparison as lots of parameters including the time of the year are similar to the initial training set.
September 2014 (run 77636 to 78062): This episode will be denoted as September 2014 even though the events are used from end of August 2014 to end of September 2014. During this episode one line was missing. Compared to July 2010 and February 2012 this period is slightly shorter as the number of lines changed before and after this time frame.

Training 2010 (run 48376 – 52878): This period has been chosen to analyse if a different training episode can improve the performance on data from the given period compared to a different training period. For this set, events from May 2010 to October 2010 are used. In order to get a good comparison with the original training set, the test set for the 2010 network is set to July 2010. In total, the number of events is smaller than for 2008/2009 but still sufficient for a good performance.

Some important parameters that can be used to quantify the detector situation for each run can be seen in figure 3.4. These plots include all runs even if it is not a physics run or excluded from the data. This might happen for runs with insufficient data quality, e.g. for runs with only very short lifetimes. For some runs, the figure presents a value of -1 for baseline and burst fraction. These are runs for which these parameters could not be computed, for example if the run duration is smaller than 15 minutes.
Figure 3.4 illustrates nicely that the individual parameters are partly highly variable even at very small time scales. Hence, even for the chosen training and test episodes, there are quite some remarkable differences that could affect the performance of the network.

3.4 Organizing the training with OrcaNet

All networks from this work were trained with *OrcaNet* [42], a deep learning framework to organize the training of neural networks, originally developed for KM3NeT. For this work, only CNNs have been used. Hence, the architecture of both networks is based on the convolutional blocks introduced in *OrcaNet*. It consists of the actual convolutional layer, followed by a batch normalization and the activation layer. Eventually, a max pooling layer is added. Finally, if needed, a dropout layer



(a) Active OMs in detector

(b) Baseline

(c) Burst fraction

(d) Bioluminescence

Figure 3.4: Several internal and external parameters for runs 30000 to 80000. All runs are shown, including non-physics runs. The selected periods used for training and testing for this work are highlighted.

is applied. The general composition of these blocks is introduced in table 3.1.
Many settings are adopted from the default settings of *OrcaNet* including the weight initialization that follows He et al. [32], the parameters used for the batch normalization layer as proposed in [34], and the use of zero padding. Additionally, the batch size is equal to 64. All networks were trained using adam as optimizer with the standard settings from *OrcaNet* that showed significantly better results than the ones proposed in the original paper.
Some other settings are the same for both networks. This includes a kernel size of $2 \times 2 \times 3$, a stride of the convolution of 1 along all dimensions, and a stride of $1 \times 1 \times 2$ for pooling. In general, these quantities are similar to the ones for KM3NeT/ORCA [38]. However, the kernel size for x and y direction is slightly smaller since the corresponding input dimensions are also smaller compared to ORCA. For the same

Layer Type	Properties
Convolution	Kernel size = (2, 2, 3) zero-padding standard *OrcaNet* weight initialization (according to [32])
Batch normalization	standard settings *OrcaNet* (according to [34])
Activation	ReLu
Pooling	optional, max pooling, stride = (1, 1, 2)
Dropout rate δ	optional

Table 3.1: General architecture of a convolutional block

reason, the pooling stride of these dimensions is set to one. As the input dimension is only four, there is no need for down-sampling along these two dimensions. Moreover, except for the output layers, ReLU is used as activation function.
The remaining hyperparameters, like for example the number of filters, are chosen individually for the different networks and will be presented in the respective chapter.
All networks were trained at the TinyGPU cluster of the Erlangen Regional Computing Center (RRZE) [43].

4 Background classifier

The interaction rate of neutrinos is very low. Although ANTARES is located in the deep sea and uses the Earth as a shield, still atmospheric muons form the majority of recorded events. Hence, a good discrimination is crucial. This chapter introduces the CNN-based background classifier. Event selection and network architecture are presented. Furthermore, the performance is evaluated on Monte Carlo as well as on data. Finally, a small detector condition analysis is presented.

4.1 Training set

The task of the background classifier is to distinguish neutrino events from atmospheric muon events. Therefore, the network has to be trained on both event types. For neutrinos it is of course necessary to use all of the different event types, i.e. tracks and shower-like events. Additionally, the training set contains neutrinos as well as antineutrinos. The number of all available events of the different event types is listed in table 4.1. Energy and direction distribution of the training set can be found in the appendix in figure A.1.

Event type	Number of events	Fraction [%]
Atmospheric μ	17 266 789	46.4
ν_e CC	6 662 507	17.9
ν_e NC	3 900 411	10.5
ν_μ CC	5 492 773	14.8
ν_μ NC	3 902 124	10.5

Table 4.1: Number of events for each event type and percentage with respect to the complete dataset of 2008/2009 for the background classifier. For neutrinos, the event type includes particles and antiparticles in approximately equal parts.

For a classifier with n categories the fraction of each class in the training set should be $\frac{1}{n}$. As can be seen from table 4.1, the split between neutrinos and muons is already approximately 50 : 50.
Neutrino interactions can be divided into two signatures. Hence, a reasonable balance of 50 : 50 between track- and shower-like events is conceivable. However, this choice is limited by the number of track-like events and decreases the number of samples from about 37 million events to around 16.5 million events. A small analysis with such a balanced dataset showed no significant improvement. Therefore, no

further balancing has been done in order to keep as much data as possible.

Furthermore, a split into training, validation, and test set has to be chosen. For the background classifier, about 81.25 % ($30 \cdot 10^6$ events) of the events are used for the actual training of the network. Around 6.25 % are used for validation and the remaining 12.5 % are held back during training and used for testing the network's performance on unseen data.

4.2 Network architecture

The general structure of a CNN was already explained in section 2.9. The full network architecture of the background classifier can be seen in table 4.2. In general, this architecture is based on the architecture of the CNNs that have been designed for KM3NeT/ORCA [38]. The main part consists of the convolutional blocks introduced in table 3.1. A small dropout rate was used to prevent the network from overfitting. Subsequent to these blocks, a row of dense layers follows. The last layer uses the softmax activation function and produces a two dimensional output. These are neutrino probability p_ν and muon probability $p_\mu = 1 - p_\nu$. Effectively, this can be interpreted as one output. Hence, this work will mostly refer to p_ν.

Building block / layer	Output dimension
xyzt input	$4 \times 4 \times 25 \times 100$
Convolutional block (64 filters, $\delta = 0.05$)	$4 \times 4 \times 25 \times 64$
Convolutional block (64 filters, $\delta = 0.05$)	$4 \times 4 \times 25 \times 64$
Convolutional block (64 filters, $\delta = 0.05$)	$4 \times 4 \times 25 \times 64$
Convolutional block (64 filters, pooling, $\delta = 0.05$)	$4 \times 4 \times 12 \times 64$
Convolutional block (128 filters, $\delta = 0.05$)	$4 \times 4 \times 12 \times 128$
Convolutional block (128 filters, $\delta = 0.05$)	$4 \times 4 \times 12 \times 128$
Convolutional block (128 filters, $\delta = 0.05$)	$4 \times 4 \times 12 \times 128$
Convolutional block (128 filters, pooling, $\delta = 0.05$)	$4 \times 4 \times 6 \times 128$
Flatten	12288
Dense 256 + ReLu	256
Dense 128 + ReLu	128
Dense 32 + ReLu	32
Dense 2 + softmax	2

Table 4.2: Network architecture of the background classifier

The optimizer that was used is adam with the *OrcaNet* standard settings. The initial learning rate was set to 0.02 and decayed exponentially during the training. After the validation loss stagnated, the learning rate was manually reduced by an

order of magnitude. This procedure was repeated a second time. Finally, when no further improvement was achieved the training was stopped. The network was trained for 5 epochs which took about three and a half days using a Nvidia GTX 1080 (Ti) GPU.

4.3 Performance

4.3.1 Loss and accuracy

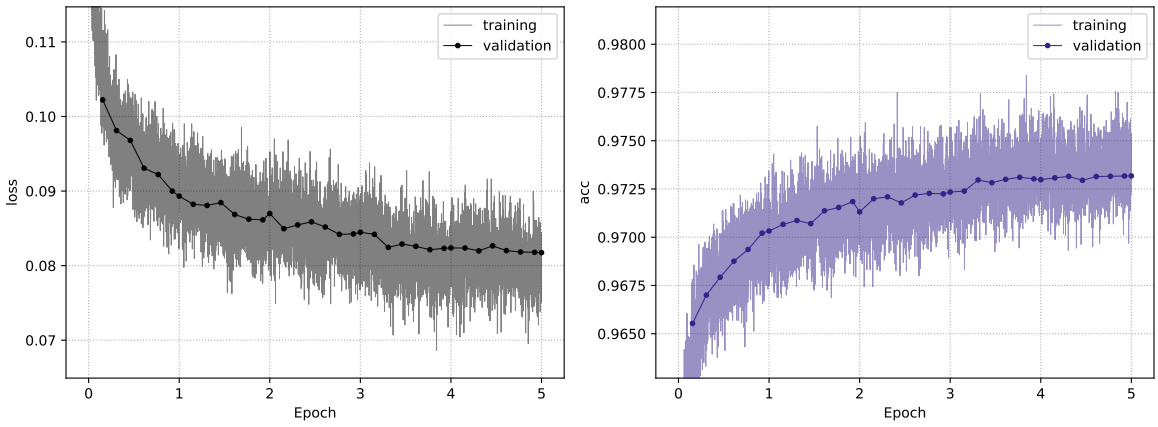
The first check on how well the network performs can already be done during the training process by monitoring the loss (see figure 4.1 (a)). For the background classifier, categorical cross entropy is used as cost function. Since it is a binary classification, the loss for one sample is calculated as:

$$L = -\left(y \cdot \log\left(p_\nu\right) + (1-y)\log\left(1-p_\nu\right)\right) \tag{4.1}$$

with y being equal to 1 for a neutrino event and 0 for a muon event.
In addition to the loss function, the accuracy was monitored as a further metric (see figure 4.1 (b)). It is given by

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}} \tag{4.2}$$

A correct prediction is given by an event where the correct class receives the highest probability. In case of the background classifier, these are neutrino events with $p_\nu > 0.5$ and muon events with $p_\nu < 0.5$.



(a) Loss (b) Accuracy

Figure 4.1: Loss and accuracy of the background classifier for training and validation during the training. For the training curve of both metrics, every data point is averaged over 250 batches, i.e. 250×64 images.

With this in mind, it is easy to see that the accuracy can be improved more easily.

Reaching 100 % accuracy is possible even if not all labels and predictions are equal. On the other hand, the accuracy might get stuck at a certain value even if the network is still improving. This is one of the reasons why the optimization is done with the loss.
The distribution of both metrics for training as well as for validation is shown in figure 4.1. The shape is as expected, i.e. a strong decrease (increase) for the loss (accuracy) in the beginning which corresponds to a fast learning. During the subsequent weaker improvement, more difficult features are learned. Finally, both metrics saturate for the validation set with slight overfitting.

4.3.2 Probability distribution

A good starting point to further analyse the performance of the background classifier is to study the actual network's output. Figure 4.2 shows the distribution of the neutrino probability for muon and neutrino events separately. In general, this distribution is typical for a binary classification. The prediction for the major fraction of neutrino (muon) events is a neutrino (muon) probability close to 1 with an exponentially decreasing number for a lower neutrino (muon) probability. However, a significant deviation from this behaviour can be stated for neutrino events with a very low neutrino probability. These are neutrino events for which the network is quite sure that they are actually muon events.

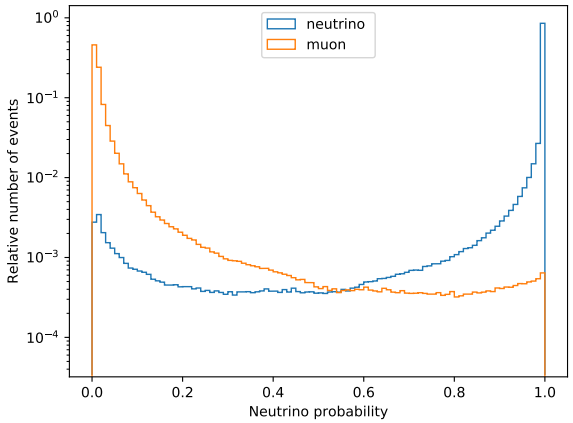


Figure 4.2: Neutrino probability distribution for muon and neutrino events. The number is normed per event type.

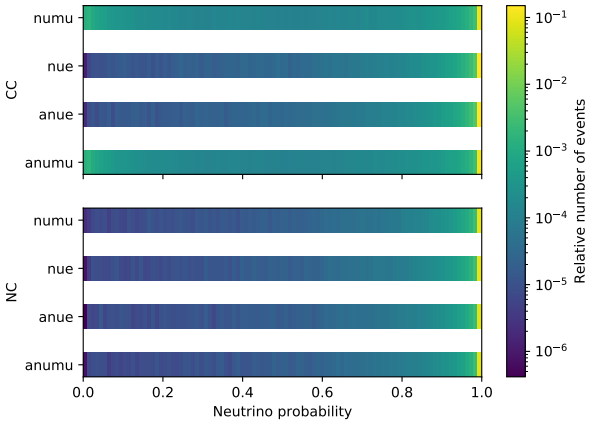
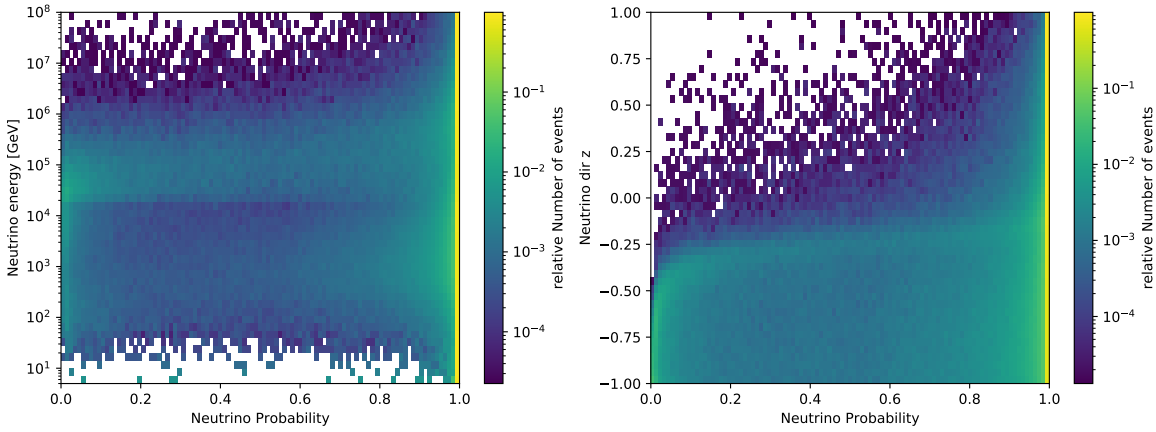


Figure 4.3: Neutrino probability distribution separately for the different neutrino flavours and interaction types. The number is normed per total number of neutrinos.

A separate analysis (see figure 4.3) of the probability distribution for the different flavours and interaction types shows that the increase is a result of misclassified

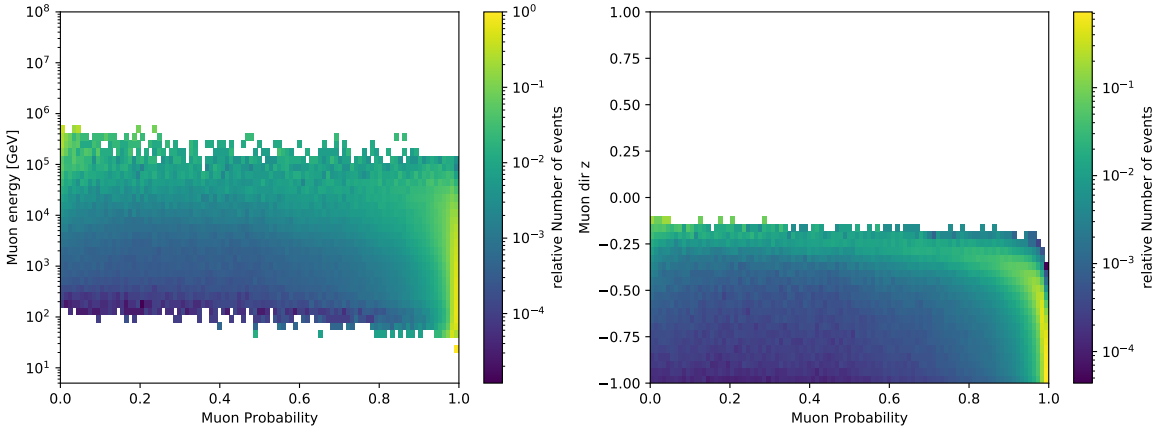
ν_μ CC and $\bar{\nu}_\mu$ CC events. This is not surprising as these track-like events are harder to distinguish from atmospheric muon events. However, this region of p_ν is not relevant for the later analysis as only events with a high neutrino probability will be used. Therefore, a cut has to be chosen. The choice will be explained in section 4.3.3.
Figure 4.4 demonstrates the relation between true neutrino energy and direction and neutrino probability. The same plot is done for muons in figure 4.5.



(a) Energy (b) Direction

Figure 4.4: Neutrino probability distribution dependent on the neutrino energy (left) and neutrino direction (right). The number is normed per energy and direction bin, respectively.

The direction is denoted as *dir z* and means the cosine of the zenith angle so that events with dir z = 1 are vertically up-going while events with dir z = −1 are vertically down-going.
Figure 4.4 illustrates nicely that the performance on up-going events is much better. Close to the horizon (dir z = 0) and especially with decreasing dir z the amount of neutrino events with low p_ν increases significantly. This is an expected result as atmospheric muon events are mainly down-going except near the horizon. Therefore, the classification of an up-going event should be much easier for the network.
Considering the neutrino energy, the lower and especially the medium energy range seem to be the most problematic ones. On the other hand, the major amount of misclassified muons are events with a high muon energy above 1 TeV and a direction close to the horizon (see figure 4.5). It should be mentioned that the transition in figure 4.4 (a) at 20 TeV is a result of the hard drop in the neutrino statistics of the training set at this energy as it can be seen in figure A.1.
By comparing figure 4.4 and 4.5, it is obvious that a good neutrino classification is more difficult in the energy-direction space where atmospheric muons have to be expected. As already pointed out, this holds primarily for track-like events. A further analysis of these relations for the different event signatures can be found in the appendix in figure A.2.



(a) Energy (b) Direction

Figure 4.5: Muon probability distribution dependent on the muon energy (left) and muon direction (right). The number is normed per energy and direction bin, respectively.

4.3.3 Muon contamination and neutrino efficiency

Figure 4.2 shows that the background classifier works fine on the test set. However, so far no weighting has been included. The number of simulated atmospheric muon events is smaller than the number of expected μ in ANTARES. On the contrary, the amount of simulated neutrino events is larger than the number of actually expected neutrinos. Depending on the analysis, a suitable weighting has to be chosen determined by the choice of an atmospheric neutrino flux or a cosmic point source spectrum.
Considering the last bin in figure 4.2 these are all events with $p_\nu > 0.99$. Even if only less than 0.01 % of the muon events contribute to this bin, this number still exceeds the amount given by more than 85 % of the neutrino events in this bin regardless which type of neutrinos (atmospheric or cosmic) is studied. Hence, a quite high probability cut value p_{cut} has to be chosen in order to reject as many muons as possible while keeping as many ν events as possible. Therefore, the relation between muon contamination and neutrino efficiency has to be analysed. The muon contamination is given by

$$\mu_{\text{cont}} = \frac{N_\mu \left(p_\nu > p_{\text{cut}}\right)}{N_{tot} \left(p_\nu > p_{\text{cut}}\right)} \tag{4.3}$$

indicating the amount of muons with a neutrino probability larger than p_{cut} with respect to the total number of events with p_ν above this threshold.
The neutrino efficiency is calculated with

$$\nu_{\text{eff}} = \frac{N_\nu \left(p_\nu > p_{\text{cut}}\right)}{N_{\nu,tot}} \tag{4.4}$$

with the number of neutrinos with a neutrino probability exceeding the cut value compared to the number of all neutrinos.
Figure 4.6 shows both quantities for different values of p_{cut} for an atmospheric neutrino spectrum. The chosen cut value ($p_\nu > 0.99996$) is highlighted in orange. It is chosen as reference point in order to establish a connection with the real contamination that is often present in the analyses. This way, the neutrino efficiency is just below 60 %. However, it must be noted that all events are included in this analysis. Usually, before applying a physics analysis, a pre-selection is done. This is not done here in order to study the network's performance without any restrictions. Thus, it seems likely that further cuts could improve the efficiency of the network. An example will be discussed later when comparing the performance to one of the standard reconstruction methods of ANTARES.

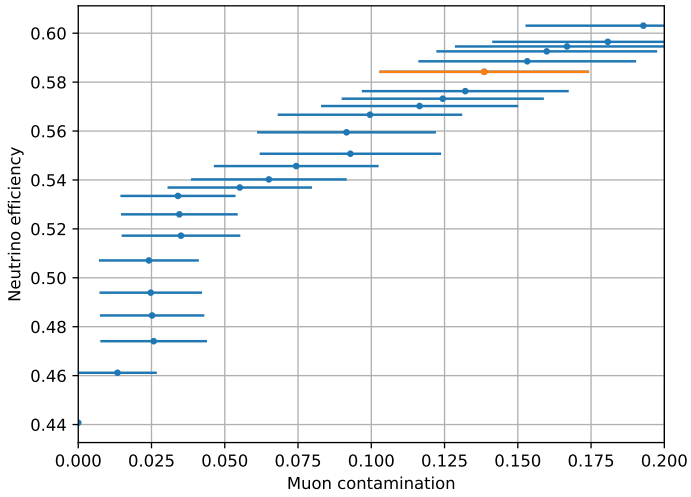
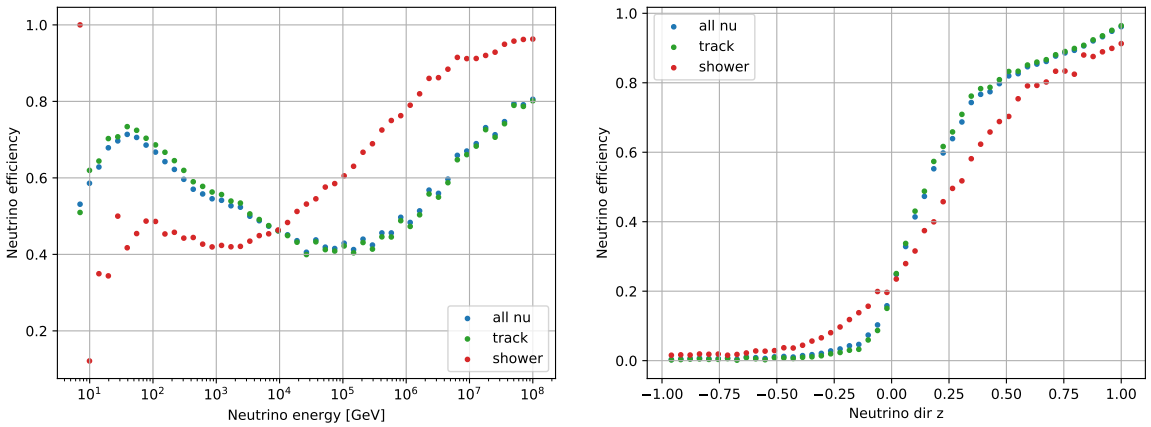


Figure 4.6: Neutrino efficiency and muon contamination for different values of p_ν. The highlighted point represents the chosen cut value ($p_\nu > 0.99996$).

Analysing the neutrino probability distribution in figure 4.2 showed a higher amount of misclassified neutrinos for track-like events. This leads to the suspicion that by choosing a cut value, a higher amount of tracks is rejected compared to shower-like events. However, a study of the neutrino efficiency shows that for p_{cut} chosen on the total neutrino set 60.1 % of all ν_μ CC but only 44.4 % of all shower-like events are kept. In general, it seems like that for tracks there are more misclassified (very small p_ν) events but also more events with a very high neutrino score.
As it was already indicated, a very big difference in the performance regarding muon contamination and neutrino efficiency can be found for a split into up- and down-going events. Only about 6 % of all down-going events are kept while the neutrino efficiency for up-going events is 75.4 %. The network clearly learns that muon events are down-going and thus rejects the major part of events from above. Maybe a direction classifier could therefore reject the atmospheric muon events even better and

improve the muon contamination. With this approach an even higher neutrino efficiency for up-going events seems possible.

In order to investigate the differences of the neutrino efficiency further, it is evaluated separately for track- and shower-like events over the whole energy and direction spectrum. The results are presented in figure 4.7. A combination of both quantities will be analysed in the following section.



(a) Energy (b) Direction

Figure 4.7: Neutrino efficiency for a fixed p_{cut} over energy (left) and direction (right) spectrum for an atmospheric neutrino spectrum. The distribution is demonstrated for all neutrinos *nu* and separately for track-like (green) and shower-like (red) events. The chosen cut value ($p_\nu > 0.99996$) is highlighted in figure 4.6.

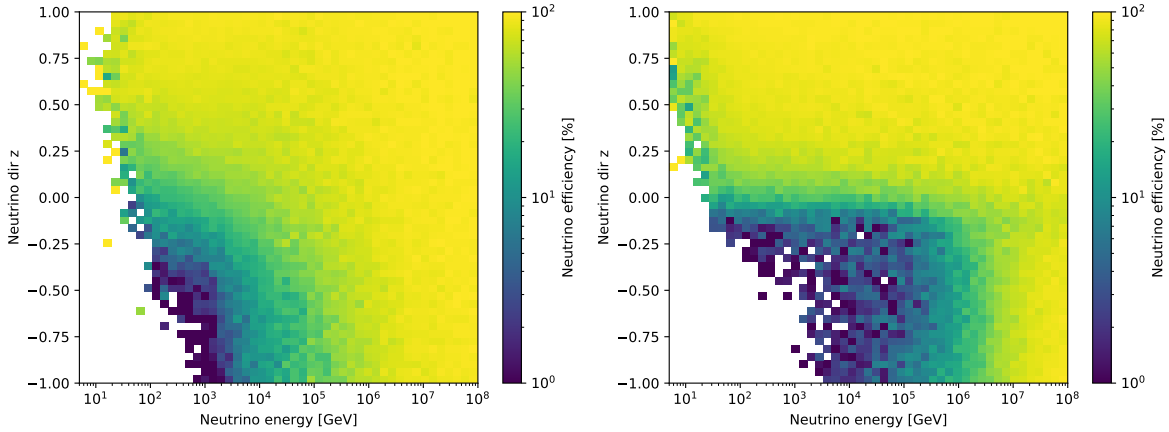
Regarding the energy, it has to be mentioned that for the lower end of the spectrum the statistics is low due to the small number of simulated events. This holds especially for shower events, probably leading to a worse efficiency.
The performance on shower-like events then increases up to an energy of about 100 GeV. This is approximately the energy from which on there are more muon than shower-like training events. The efficiency drops to 40 % until E = 1 TeV where the number of training muon events reaches its maximum. From this point on the neutrino efficiency increases up to about 95 %.
In general, the shape of the curve is similar for track-like events but it is slightly stretched. After the first maximum at E = 40 GeV the efficiency constantly decreases from 70 % to to 40 % at E = 20 TeV. This is mostly the energy range for which there are more muon training events than ν_μ CC events. At approximately 100 TeV, when no simulated muon events are left the performance starts to improve again. As for the shower-like events, the maximum of the neutrino efficiency is achieved for the highest energies but only reaches 80 %.

Regarding the neutrino direction the difference between shower- and track-like events

is less remarkable. The exact shape and efficiency depends of course whether one considers an atmospheric or astrophysical neutrino spectrum but regardless of this the neutrino efficiency is significantly worse for down-going events. The performance on events from above is better for shower events while the efficiency for up-going events is better for tracks. This holds for both neutrino spectra but for a cosmic spectrum the performance on down-going events is better for both signatures (around 10 % for tracks and 40 % for shower events, see figure 4.10).
In general, the reason for these distributions is probably a combination of the different event signatures and the number of training events in the respective energy and direction bins (see appendix A.1). Regarding the direction the major amount of training events with dir z < -0.2 are muon events. This can not be completely avoided as the total number of muon and neutrino events should be approximately equal and atmospheric muons only enter the detector from above. However, this way it seems logical that the network rejects all events from this direction. It is also possible that a more equally balanced training set in this energy and direction range regarding atmospheric muons and neutrinos could lead to a higher survival rate of neutrinos. However, this probably also results in a higher rate of misclassified muons.

Selected events

These results can be confirmed by analysing the finally selected neutrino events and comparing their energy and direction distribution to the complete neutrino set. Figure 4.8 presents the neutrino efficiency for the chosen cut for the atmospheric weighting per neutrino energy and direction separately for both event signatures. It is particularly noticeable that for down-going events with the lowest energies ($E < 1\,\mathrm{TeV}$) all real neutrino events are completely rejected. But even for higher energies below 50 PeV the percentage of selected down-going events is still poor, especially for track-like events. The survival rate is much better for up-going events, especially if neglecting very low energies below 10 GeV for which ANTARES is not sensitive. The further away from the horizon, the higher the chance to keep the neutrino.
Additionally, it has already been pointed out that regarding the neutrino signature there is a certain difference. While loosing almost all down-going ν_μ CC events for energies below several PeV, the energy range for shower-like events only extends up to several hundred TeV. Probably, it is easier to distinguish higher energetic shower-like events from atmospheric muons compared to track-like events. However, the performance on low energy events from below but close to the horizon is better for ν_μ CC events. The increase in the efficiency is quite sharp for track-like events around the horizon, whereas the transition to higher efficiencies is more continuous for shower-like events. As the direction reconstruction is more difficult for shower-like events, it is possible that events that are actually from below but close to the horizon are very similar to down-going events for the network.
Summarizing these observations, it seems plausible that the network learns the energy and direction distribution of atmospheric muons and thus rejects all events from



(a) Shower-like events (b) Track-like events

Figure 4.8: Neutrino efficiency for a fixed p_{cut} combined over energy and direction spectrum, split for shower-like (left) and track-like (right) events. The chosen cut value ($p_\nu > 0.99996$) is highlighted in figure 4.6.

this part of the phase space. As for a classical reconstruction, it seems not possible for the classifier to distinguish between a real atmospheric muon and a muon that was induced by a neutrino interaction. In fact, there is no prominent physical reason how the network could distinguish down-going ν_μ CC events from atmospheric muons. A possible feature to recognize a neutrino induced muon is given by the vertex shower. This is in general very difficult and impossible if the interaction vertex is not inside the detector. However, the network primarily rejects down-going events with energies from the atmospheric muon spectrum. Hence, there is a good chance to keep higher energetic neutrino events from above. This is very fortunate as astrophysical neutrinos, the main object of interest for ANTARES, usually are higher energetic. Therefore, the neutrino efficiency for the same probability cut is higher compared to an atmospheric neutrino spectrum. However, the number of expected cosmic neutrinos is much smaller. Hence, the muon contamination is higher.

Besides analysing the properties of the selected neutrinos, it is important to understand what kind of misclassified muons have to be expected when applying the background classifier. An analysis of the energy and direction distribution of the muon events that are kept after applying the probability cut as discussed in the previous section (see figure 4.6) compared to all muon events is presented in figure 4.9. As it was already indicated by the results of figure 4.5, most of the remaining muon events originate from a direction close to horizon. Additionally, the energy of the majority of these events is higher than 1 TeV. In general, it is easy to see that the main part of misclassified muons comes from the boundaries of the atmospheric muon spectrum.

It should be noted that these are results for the performance on all events. Usually,

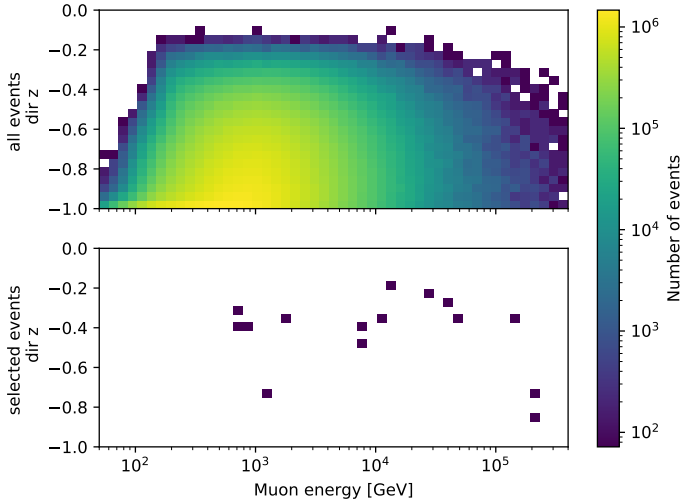


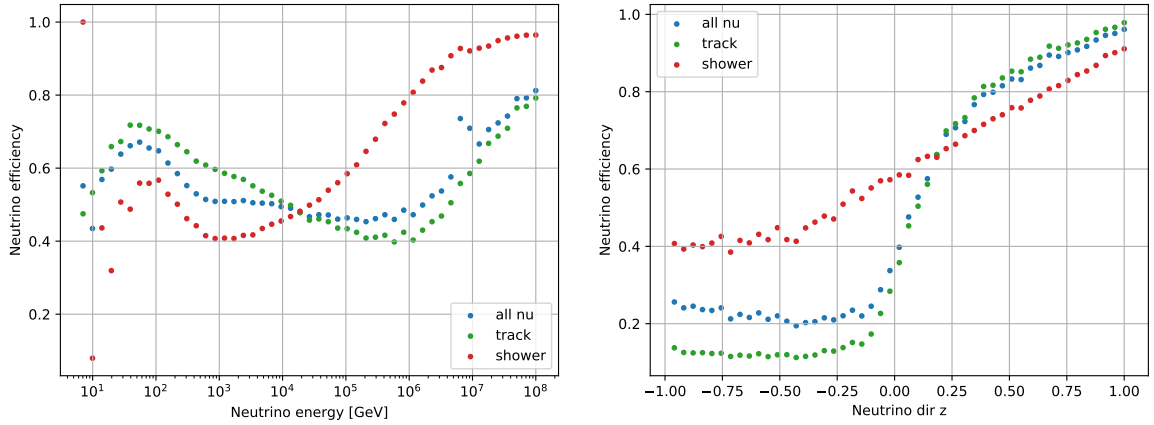
Figure 4.9: Energy and direction distribution of all muon events (above) and selected ($p_\nu > 0.99996$) muon events (below). The number refers to the expected number of events per year.

there are some criteria that are used in order to do a pre-selection of events. This could help to improve the performance further. A conventional approach is to reject all down-going reconstructed events. As already shown, this improves the performance of the classifier as well (up to more than 70 %). However, the advantage of using the background classifier without this cut is, that at least the high energetic down-going neutrinos could be kept. In summary, even without any pre-selection it can be seen that the deep learning approach works. However, an example for such a pre-selection will be presented in the following.

Point source spectrum and comparison with a standard reconstruction

The analysis so far was restricted to atmospheric neutrinos. However, the goal of ANTARES is the search for cosmic neutrinos. The assumed spectrum is proportional to E^{-2}. As the flux for these neutrino events is much smaller, the atmospheric muons need to be suppressed more. Hence, the neutrino probability cut p_{cut} needs to be stricter ($p_\nu > 0.999985$). It is chosen such that approximately one atmospheric muon per day is expected. This way, the neutrino efficiency on the total dataset is 50.0 %. The expected number of remaining atmospheric neutrinos per day is 16.9. This unselected set will be denoted as A_0. In contrast to an atmospheric neutrino spectrum, the integrated efficiency is better for shower-like events ($\nu_{\text{eff}} = 57.8\,\%$) compared to track-like events ($\nu_{\text{eff}} = 47.6\,\%$) when weighting the event sample with an astrophysical E^{-2} spectrum. The efficiency for up-going events is 76.0 % and therefore very similar as for atmospheric neutrinos. However, the performance on down-going events is much better ($\nu_{\text{eff}} = 23.8\,\%$) since the amount of high-energetic events, on which the classifier performs better, is larger for astrophysical neutrinos.

The bin-wise efficiency can be seen in figure 4.10. The corresponding two dimensional distributions for both event signatures can be found in the appendix (see figure A.3).



(a) Energy (b) Direction

Figure 4.10: Neutrino efficiency for a fixed p_{cut} over energy (left) and direction (right) spectrum for a cosmic neutrino spectrum $\sim E^{-2}$ for A_0. The distribution is demonstrated for all neutrinos *nu* and separately for track-like (green) and shower-like (red) events. The probability cut is chosen in such a way that approximately one atmospheric muon per day survives ($p_\nu > 0.999985$).

Finally, the performance is compared to a standard reconstruction of ANTARES. Since there is no background classifier including all event signatures, a pre-selection has to be done. The comparison is done with *Aafit* [44], a reconstruction for track-like events. Therefore, only ν_μ CC events are used for the actual comparison. However, the results will be presented for shower-like events as well.
As a first restriction for the pre-selection, only upwards going reconstructed events are used. This subset will be denoted as A_{up}. Additionally, events with an angular error estimate $\beta > 1°$ are excluded from the analysis. Finally, the cut parameter for the standard reconstruction is Λ, a variable related to the log-likelihood per degree of freedom. Therefore, the selection is restricted to events with a reconstructed Λ. This way, the selection of events is exactly the same for both reconstructions. It will be denoted as A^1_{up}. The results for the unselected (A_0) as well as the preselected sets A_{up} and A^1_{up} for both reconstructions are shown in table 4.3.
Based on the pre-selected event samples, the deep learning atmospheric muon suppression is applied such that the atmospheric muon rate is suppressed to approximately one muon per day ($p_\nu > 0.999984$ for A_{up} and $p_\nu > 0.99987$ for A^1_{up}). The same requirement is used for the Aafit reconstruction ($\Lambda > -5.205$ for A_{up} and $\Lambda > -5.3$ for A^1_{up}). The efficiency is given with respect to the corresponding pre-selection.

Reconstruction	N_μ/day	$N_{\nu_{atm}}$/day	ν_{eff} [%] total	**track**	shower
$\boldsymbol{A_0}$**:**					
Deep learning ($p_\nu > 0.999985$)	≈ 1	16.9	50.0	**47.6**	57.8
Aafit ($\Lambda > -4.05$)	≈ 1	0.04	0.6	**0.8**	0
$\boldsymbol{A_{up}}$**:**					
Deep learning ($p_\nu > 0.999984$)	≈ 1	16.5	69.4	**70.5**	66.1
Aafit ($\Lambda > -5.205$)	≈ 1	5.5	33.8	**44.3**	3.6
$\boldsymbol{A^1_{up}}$**:**					
Deep learning ($p_\nu > 0.99987$)	≈ 1	13.9	84.0	**85.0**	81.0
Aafit ($\Lambda > -5.3$)	≈ 1	6.0	48.6	**59.0**	4.8

Table 4.3: Number of atmospheric muons and neutrinos and signal efficiency for all neutrinos and the different event signatures for deep learning and a standard reconstruction, for all data A_0 and pre-selections A_{up} and A^1_{up}. The cuts for deep learning and the standard reconstruction are chosen in such a way that approximately one atmospheric muon per day survives for the respective selection. The efficiency is given with respect to the corresponding pre-selection.

Considering the track efficiency, one can see that the performance of the deep learning approach is significantly better, regardless whether a pre-selection is done or the whole dataset is used. However, the amount of atmospheric neutrinos is in all cases more than a factor 2 smaller for the standard reconstruction.
Additionally, it can be seen that the performance of the standard reconstruction is poor for shower-like events, as desired. For deep learning, the difference for both signatures is only a few percent, respectively.
For both reconstructions, the pre-selections lead to a higher track efficiency regarding the corresponding subset. For the standard reconstruction, the overall efficiency, i.e. the neutrino efficiency with respect to A_0, also increases. In case of the background classifier, the overall efficiency decreases for A_{up} as well as A^1_{up}, especially for shower-like events. However, the selections allow for a lower cut on p_ν and the efficiency with respect to the selections is very good. Additionally, due to the pre-selection, the number of expected atmospheric neutrinos per day is reduced slightly. In summary, a pre-selection is required for the standard reconstruction whereas the background classifier can also be applied to the total dataset.

It is remarkable that the restriction for up-going reconstructed events A_{up} does

not improve the performance for the deep learning approach on the total dataset. Usually, this cut is used to reject the background given by atmospheric muons. For the given test set, this reduces the amount to 5.9 % with respect to all muons while 80.5 % of the atmospheric and 59.0 % of the cosmic track-like neutrinos remain. In case of the background classifier, the same rate of atmospheric muons for A_0 is obtained for $p_\nu = 0.15$. With this very low cut, still more than 90 % of the track-like events from both considered spectra remain. But even with a still very loose cut of $p_\nu = 0.9$, 89.1 % of the atmospheric and 81.8 % of the cosmic track-like neutrinos remain. This is obviously still a higher efficiency than a cut on the reconstructed direction but the amount of atmospheric muons is further reduced to 0.5 % of the original number of muons. This is of course still far too high but demonstrates the power of the deep learning approach.

In summary, the background classifier clearly benefits from the fact that it does not require an additional direction reconstruction as it seems to determine the direction internally. This way, it can reduce the main part of atmospheric muons on its own and is still able to keep down-going neutrino events with the highest energies. The performance is significantly better compared to the classical approach. However, it should be mentioned that the standard reconstruction does not only aim for the suppression of atmospheric muons but also shower-like events. Hence, an ultimate comparison with a standard reconstruction is not possible as all of these methods aim for either tracks or showers. This way, these reconstructions need to reject not only muon events but also the undesired neutrino events. For a more fair comparison, an additional track-shower classifier is needed. However, the results indicate that a CNN-based up-down classifier could contribute to the standard reconstructions.
Looking at the performance of the background classifier, the pre-selection A^1_{up} does not only lead to a higher neutrino efficiency regarding this subset. Due to the pre-selection, a smaller p_{cut} can be applied still leading to a reasonable muon contamination. Of course the pre-selection already costs a certain amount of neutrino events but on the other hand due to a lower probability cut, it is possible to keep some of the neutrino events that are rejected when using all events. This way, it seems possible that a well suited pre-selection could lead to an improved performance of the network. Depending on the goal of the respective analysis, pre-selection as well as p_{cut} can be adapted in order to obtain the best results.
In summary, it seems like deep learning techniques can keep up with classical methods.

4.3.4 Data

In order to evaluate the performance of the background classifier on data, all available data runs from the training period of September 2008 to February of 2009 have been used. For the first investigation no run has been excluded regardless of any quality parameters. This leads to a total run duration of about 127.6 days.

Probability distribution

The first check on the performance on data is to study the probability distribution of the network and compare it to the expected shape of the Monte Carlo simulations (see figure 4.11). In general, the distribution follows the expected shape. The main part of events is classified as muon which is of course expected. However, there is a higher amount of data events in the bins with the lowest neutrino energy and a smaller fraction for all $p_\nu > 0.02$. Additionally, the transition to the last bin is more sharp than expected. Considering the logarithmic scale of figure 4.11 these deviations are quite small. The maximum deviation is about 5 % for the bin with the lowest neutrino probability.

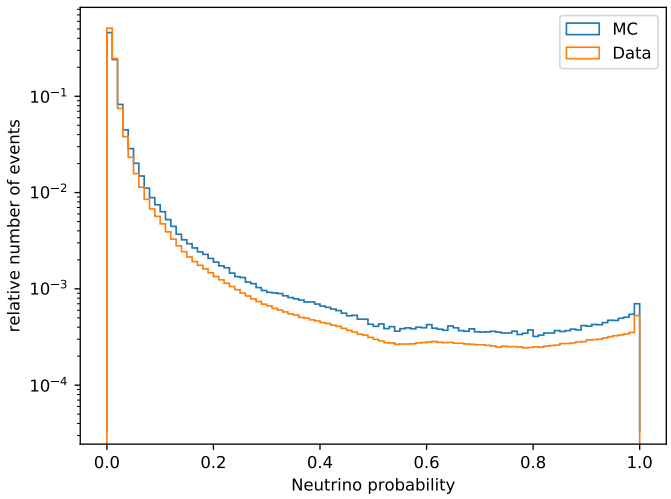
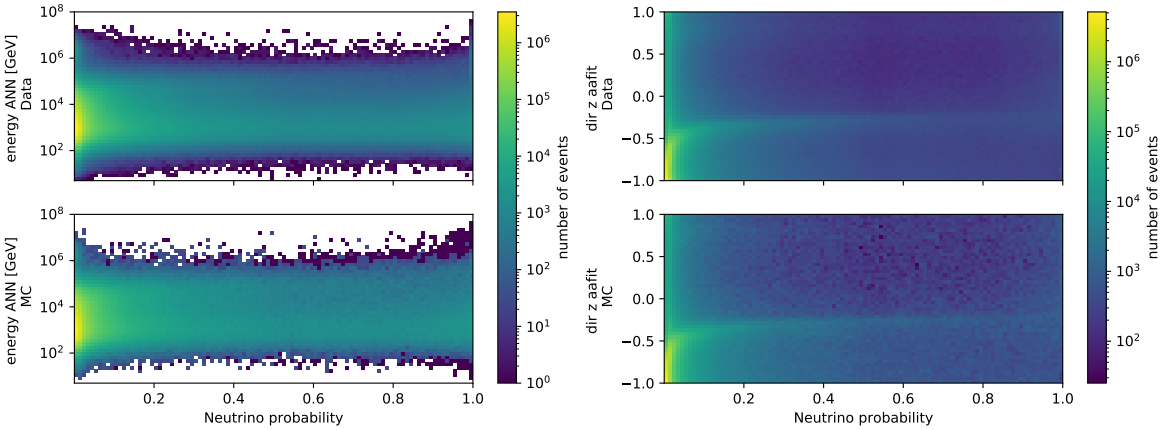


Figure 4.11: Neutrino probability distribution for Monte Carlo simulations and all data events from September 2008 to February of 2009.

A possible cause for these deviations could be pure noise events. This event type is not included in the simulations of ANTARES. Therefore, it is not present in the training process of the background classifier. Hence, it is not clear how the network performs on such events. Considering figure 4.11 it seems possible that such events receive a low neutrino probability leading to a higher fraction of events with low p_ν than expected. However, this is only a suspicion and should be investigated further.

Energy and direction distribution

As a next investigation of the agreement between data and simulations, energy and direction distribution are compared (see figure 4.12). Obviously, this can't be done by using the true energy and direction values. Therefore, two reconstructions are used.
For the direction the reconstruction algorithm *Aafit* [44] is used. The reconstructed energy uses the reconstructed value from the *ANN* [45] reconstruction. Both al-



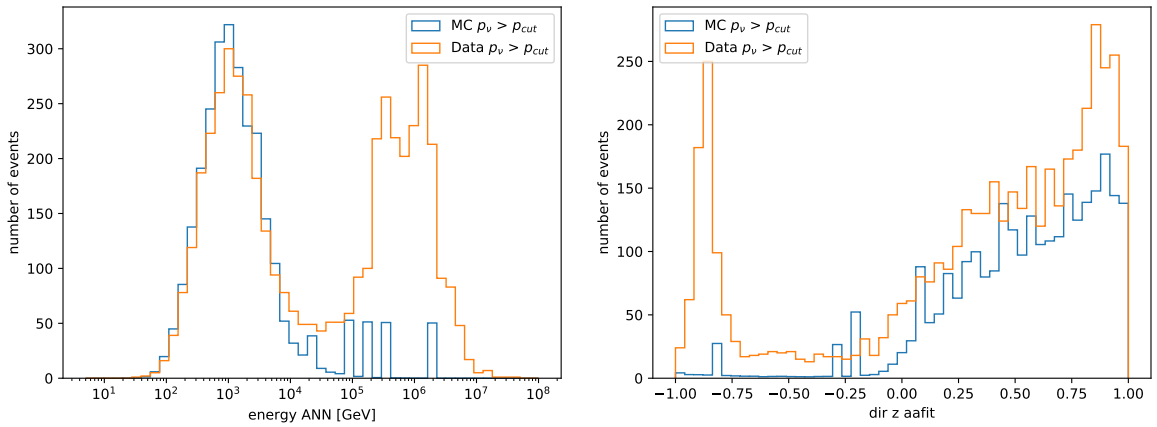
(a) Energy (b) Direction

Figure 4.12: Reconstructed energy (right) and direction (left) distribution for all data events from September 2008 to February 2009 (above) and Monte Carlo simulations (below). The color scale for the energy distribution is limited to 1.

gorithms are actually developed for track-like events. Hence, the performance on shower-like events might be poor or the quantities might not even be reconstructed. However, this also holds for their performance on simulations which justifies a comparison between data and Monte Carlo.
Obviously, a much better reconstruction performance could be achieved by applying some cuts from classical track or shower reconstruction algorithms. However, this analysis does not aim for a good reconstruction but for checking if the distribution of the background classifier on data matches the one of the simulations. This includes the performance on atmospheric muon events and justifies why no events are rejected.
Figure 4.12 demonstrates that events from Monte Carlo simulations and data show a good agreement for both reconstructed energy and direction. Due to a missing reconstruction, 3.7 % of the data events are missing for the energy and the direction reconstruction analysis, respectively.
A small study of the deviations between expected and detected distributions is presented in the appendix (see figure A.4). In general, these deviations are small and seem to be independent of the neutrino probability. The largest deviations can be found for the highest or lowest predicted energies. This could be a result of the lower statistics. However, for reconstructed energies above 10 TeV, the amount of events in data is larger whereas below this energy, the amount is larger for MC, except for the lowest energies.
Considering the reconstructed neutrino direction, larger deviations can be found for up-going reconstructed events. In this case, there are also more events in data than expected for almost all neutrino probabilities except for the lowest ones.

Identification of sparking runs

Finally, a neutrino study of the data runs is done. This investigation focusses on atmospheric neutrinos in order to have a higher statistics to evaluate the general performance of the network. Therefore, all data events with $p_\nu > p_{\text{cut}}$ are analysed. The expected number of (atmospheric) neutrinos (and misclassified atmospheric muons) is 2721.03 but the number of neutrinos found in data is 4692. Figure 4.13 demonstrates the energy and direction distribution for the selected events of data and simulations.



(a) Energy (b) Direction

Figure 4.13: Reconstructed energy and direction distribution for data and Monte Carlo simulation events that are selected after applying p_{cut}.

Obviously, there is a large amount of data events that does not match the results from the simulations. Apart from the expected events with a peak around 1 TeV, there is a second accumulation of events at a higher range of 100 TeV to 10 PeV.
Regarding the reconstructed direction, there are two striking excesses of data events compared to MC. These events are predicted to be almost vertically up-going or down-going. But even without these two peaks the number of events in data is almost consistently higher than in the simulations.
At least concerning the reconstructed direction close to dir z = 1, this is not a region for which atmospheric muon events are expected. Additionally, the energy range for atmospheric muons is lower than the position of the unexpected peak. Thus, an excess of atmospheric muons is not expected in this high energy range. Therefore, it seems unlikely that these events are misclassified atmospheric muon events.
Apart from that, the general probability distribution shows no significant anomalies compared to the expectation. This suggests that single runs might be the reason for these excesses rather than a systematic misclassification of the network.

In order to study this further, the neutrino probability distribution is evaluated run-wise. This is done separately for each month. There is no physical reason for

that but this way the plots are much more compact and therefore clearly arranged. Additionally, only the range 0.99 <= p_ν <= 1 is analysed because this seems to be the region with a significantly high amount of events. The result for October 2008 is illustrated exemplary in figure 4.14. All the other months can be found in figure A.5 in the appendix.

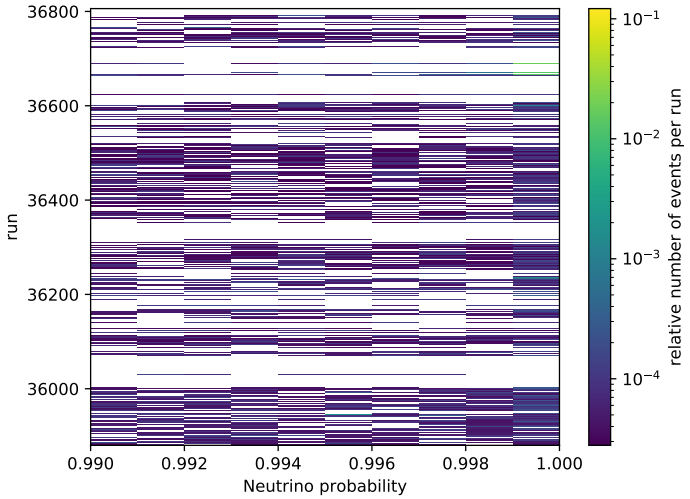


Figure 4.14: Neutrino probability distribution for $p_\nu > 0.99$ for the individual runs from October 2008.

Figure 4.14 shows a quite high suspicious fraction of events with $p_\nu > 0.999$ for a few runs ("isolated" runs between run 36600 and 36800) while the remaining runs show a similar distribution. However, the last bin includes still much more events than the ones that are selected by applying p_{cut}. In order to verify the assumption that these are runs with a very high number of events with $p_\nu > p_{\text{cut}}$, a further analysis is made. Therefore, the ratio R between the number of expected neutrinos for data and MC is calculated:

$$R = \frac{\mathrm{N}_{\text{data}}\left(p_\nu > p_{\text{cut}}\right)}{\mathrm{N}_{\text{MC}}\left(p_\nu > p_{\text{cut}}\right)} \tag{4.5}$$

N_{MC} is simply scaled for each run by the ratio between the individual run duration and the total duration of all runs. This way, the number of expected neutrinos depends only on the run duration, meaning that if the duration of run a is twice the duration of run b, the number of expected neutrinos in run a is also twice as high as for run b. Internal and external parameters that are actually influencing the data taking are neglected. However, this approach is a rough estimation that is, as will be seen, sufficient to detect odd runs. Besides, computing N_{MC} for a single run by predicting on the simulation of the corresponding run is not an option as it might be part of the training set. These MC runs are not suitable to test the performance. Figure 4.15 shows this ratio for each run of October 2010. The same plot for all the

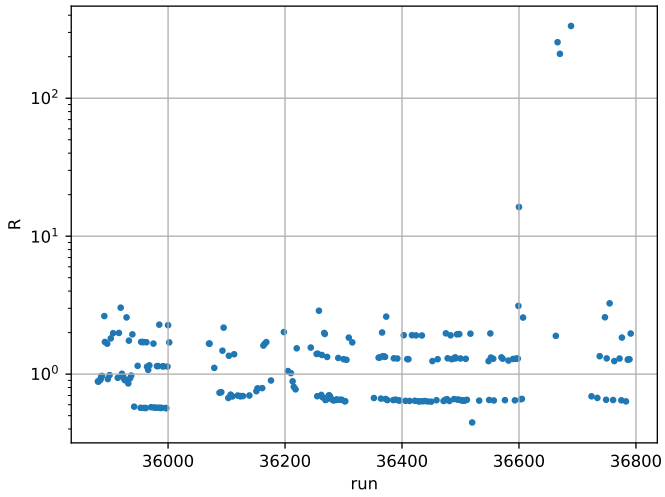
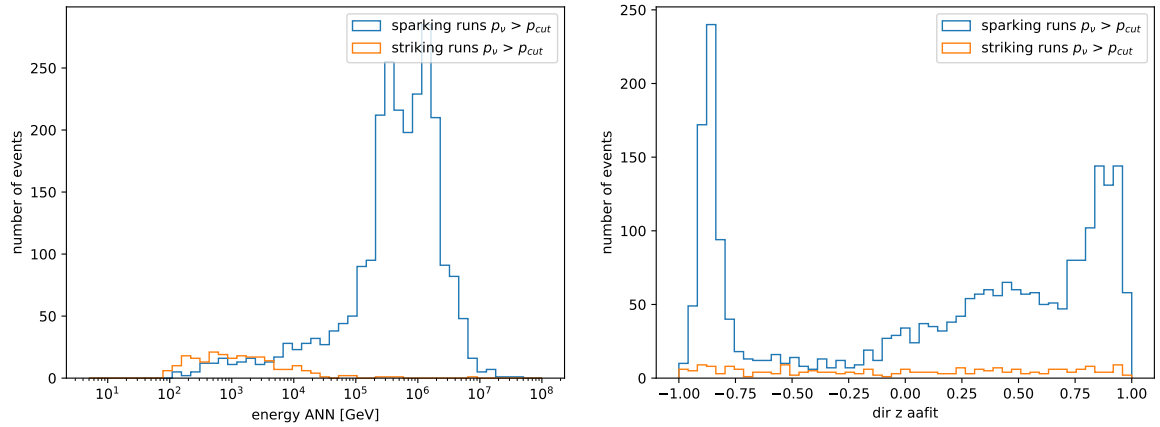


Figure 4.15: Ratio R for all runs from October 2008.

other months can be found in figure A.6 in the appendix.
From figure 4.15, three runs with a very high ratio ($R > 200$) and one run with a high ratio $R > 10$ can be identified. These are run numbers 36600, 36666, 36670 and 36689. All these runs are officially known to be so-called *sparking runs* and are usually excluded from physical studies. During such runs single PMTs can produce light flashes which then trigger the recording of events.
All runs from the official sparking run list (runs that are mentioned above and runs 38482 and 39192) were found to have a remarkable high ratio of at least 9.6. It is highly plausible that such events are falsely reconstructed as high energetic neutrinos explaining the second energy peak in figure 4.13.

However, there were also some runs that are not marked as sparking runs showing a high ratio for unknown reasons. These runs are denoted as *striking runs*. The run conditions seem to be stable for all these runs but the number of events with $p_\nu > p_{\text{cut}}$ still exceeds the one for some of of the sparking runs. There are four runs (35158, 35198, 35201 and 39339) that show ratios with $R > 9.0$ which is close to the minimal ratio of a sparking run ($R = 9.6$). It is remarkable that three out of those four runs were recorded within a few days. However, baseline, burst fraction, and bioluminescence are not worth mentioning and data quality is high, too. In summary, there is no obvious reason for the high amount of events with $p_\nu > p_{\text{cut}}$.
In total the number of such events from confirmed sparking runs is more than ten times larger than the one from striking runs but there is still another factor why the striking runs should not be classified as sparking runs. Both, energy and direction distribution, show a different behaviour (see figure 4.16) for the two run types.
The reconstructed energy for events from sparking runs are mainly around 1 PeV which is higher than the expected energy for atmospheric neutrinos. It is already known that sparking events in ANTARES are easily reconstructed as high ener-



(a) Energy (b) Direction

Figure 4.16: Reconstructed energy and direction distribution for officially confirmed sparking runs and striking runs.

getic neutrinos. The reconstructed direction of these runs is mainly vertically up- or down-going. This is also not surprising. The sparks caused by single PMTs should mostly affect the OMs in their vicinity. The closest ones are the ones in a storey below or above which is more or less exactly a direction of dir z $= 1$ or dir z $= -1$. Additionally, in sum there are more events that are reconstructed as up-going. This is probably caused by the fact that ANTARES is mainly sensitive to upwards going particles. Therefore, it is more likely that a light burst from a PMT below is detected.
On the other hand, the striking runs show basically a uniform distribution regarding the reconstructed direction and a reconstructed energy at the expected neutrino energy range for atmospheric neutrinos. The signature of these events thus is different and it seems unlikely that these are also sparking runs. Still, they don't coincide with the expectations.

In addition, a by-eye analysis of the zt-projection of the images has been done. A small study of the events with $p_\nu > p_{\text{cut}}$ of the sparking events shows two different signatures. The first one is the shape of a cone. This signature is observed dominantly for the sparking events. It clearly resembles shower-like events. The reconstructed energy for such events is mostly larger than 10^5 GeV leading to the large peak in the energy distribution. The reconstructed direction shows mostly values close to -0.9 or 0.9. The value seems to be similar for shortly successive events from one run.
The second signature is a horizontal band. Most of the hits are located within a vertical extension of 100-150 metres and only a few outside this region. The reconstructed energy for these events is mostly smaller compared to the first signature probably yielding the tail of the distribution. The reconstructed direction on the other hand is highly variable. Events with this band signature usually follow events

that show a cone signature with a very high reconstructed energy.
A cone signature is probably directly generated by a sparking event while a band signature could be the result of a so-called afterpulse. These are light pulses following the actual signal pulse due to luminous reactions or the ionization of the residual gas inside the PMT [46]. An indication for this can be found when analysing the xy-projection of the images. The dominant part of hits for such events is detected by a single or sometimes two neighbouring lines, meaning that the light does not propagate through the whole detector but seems to affect mainly the sector of one line.
Analysing the striking runs there is a large amount of images showing the horizontal band signature as well, whereas no cone structure was observed. Additionally, there are also some events that are not remarkable at all or even showing an up-going track. However, these events are not assumed to originate from afterpulses as the reconstructed energy of the previous event is mostly below 1 TeV often also showing the band structure with a very similar expansion and location regarding the height in the detector. Additionally, the xy-projection of the event images looks slightly different compared to the ones from sparking runs. In this case the hits are distributed over more than one line that is usually a block of close lines. This suggests that these events could be generated by a bioluminescence cloud travelling through the detector.
Exemplary plots for the zt- and xy-projection of sparking and striking runs can be found in figure A.7 in the appendix.

In summary, these observations lead to the conclusion that the here found striking runs are no sparking runs. However, the striking runs show definitely some unexpected characteristics. Therefore, all sparking runs (6) and the additional striking runs (4) are excluded from the analysis reducing the total used lifetime by 17.3 hours.

Final neutrino analysis

Table 4.4 shows the number of expected and detected atmospheric neutrino events within the different subsets. For the final neutrino analysis all runs that are neither officially sparking nor striking are used. The total run duration is then about 126.9 days.

Included runs	$N_{expected}$	$N_{detected}$
All runs	2721.03	4692
- sparking runs	2714.25	2240
- sparking runs and striking runs	2705.62	2012

Table 4.4: Number of expected and detected atmospheric neutrinos for all runs, after excluding only sparking runs and after excluding sparking and striking runs.

Although the deviation between prediction and experimentally determined neutrinos

is smallest for the subset only excluding the confirmed sparking runs, the striking runs are excluded as well. The number of events with $p_\nu > p_{\text{cut}}$ seems to be unusual high. While the energy distribution (see figure 4.17) matches the expectation very well, there are partially more down-going reconstructed events than predicted for the simulations (see figure 4.17). However, it can be expected that the distributions for data are more equally distributed than for Monte Carlo. For the simulations, the statistics of atmospheric muons are smaller than in reality leading to the observed peaks in figure 4.17. In case of doubt, the affected events could be analysed by eye. A combined analysis of reconstructed energy and direction of the final selection can be found in figure A.8 in the appendix.
In total, even though there are around 25 % less events found in data than expected, the agreement between data and simulations is still very good.

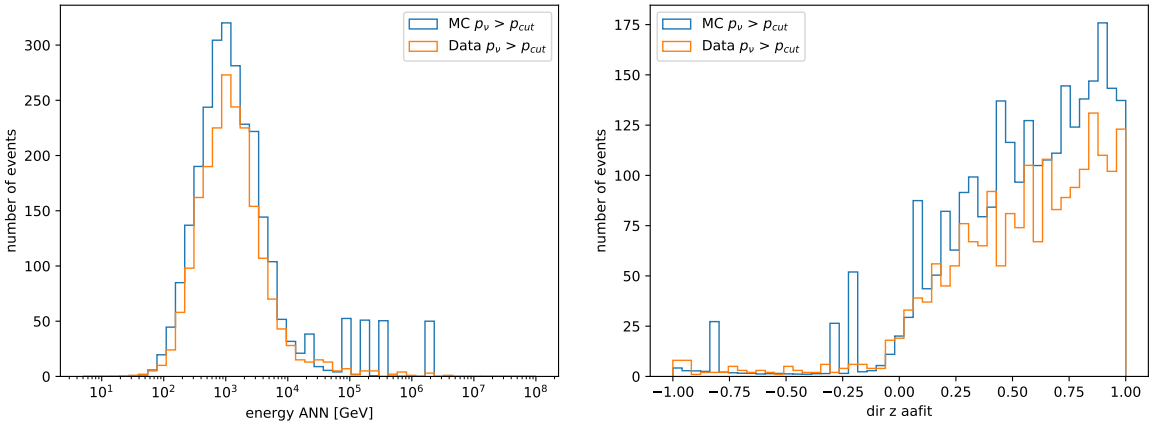


Figure 4.17: Reconstructed energy (left) and direction (right) distribution for the final run sets. All sparking and striking runs are excluded.

4.3.5 Detector condition studies

As it was already mentioned before, ANTARES is recording data since 2007 but internal and external conditions can be highly variable even within short time frames (see figure 3.4). This analysis deals with the question whether and how much different data taking conditions affect the performance of the background classifier.

Performance on MC simulations of other time periods

The easiest way to pursue this goal is to compare the performance on different episodes, i.e. time frames of subsequent runs. Therefore, the network is applied not only to its original test set, that consists of runs from the time it was trained on, but also on run-by-run simulations from later episodes. These are July 2010, February 2012 and September 2014. A brief description of the internal and external

conditions of these months is given in section 3.3. The composition of the individual test sets is shown in table A.1 in the appendix. As these events are not used for training, it is not necessary to balance these sets regarding the event type.
The first quantity that can be compared is the loss on the individual episodes. This, as well as the accuracy, is shown in table 4.5. As expected, the loss is lowest for the episode the network was trained on. July 2010 is the month with the highest loss. Comparing this to figure 4.1 (a) it can be seen that the loss curve of the training period has (at least on the validation set) passed this value already in epoch 1. The performance on the test set of February 2012 and September 2014 are somewhere in-between according to a validation loss reached before completing epoch 2.

Episode	Loss	Accuracy [%]
Training 2008/09	0.08078	97.36
July 2010	0.09333	96.84
February 2012	0.08656	97.12
September 2014	0.08833	97.09

Table 4.5: Loss and accuracy for different episodes

However, in this case the loss might not be the most meaningful quantity to judge the performance as it weights each sample the same, meaning that a different composition of the test set regarding the included event types certainly takes influence on the final loss. A more informative assertion is given by the loss for the individual event types but this does also not inevitably reflect the final performance when including an appropriate weighting. However, it gives a first impression that the performance on the different episodes might differ.

The general neutrino probability distribution for all four episodes is illustrated in figure 4.18. It can be seen that the general shape of the output distribution is the same for all four different test sets. This holds especially for the neutrino events for which this binning does not show any significant deviations particularly in the last bin with the highest and therefore most relevant neutrino probability. In contrast to this, the percentage of muon events in this bin shows some differences. The lowest fraction of falsely classified muon events can be found for the test set from the actual training period, whereas the samples from July 2010 and September 2014 show a slightly higher amount. However, the differences are very small. Still, this leads to the suspicion that the muon contamination might be worse for these two episodes. However, as already mentioned, the final probability cut is still higher than the last bin. In general, all distributions are very similar without any remarkable differences.

Therefore, a comparison of neutrino efficiency and muon contamination is done for atmospheric neutrinos (see figure 4.19). The first impression is that the results for the training period in 2008/2009 and February 2012 and the results for July 2010 and September 2014 are very similar, respectively. The training period and February 2012 have in general similar data taking conditions (see figure 3.4), whereas for

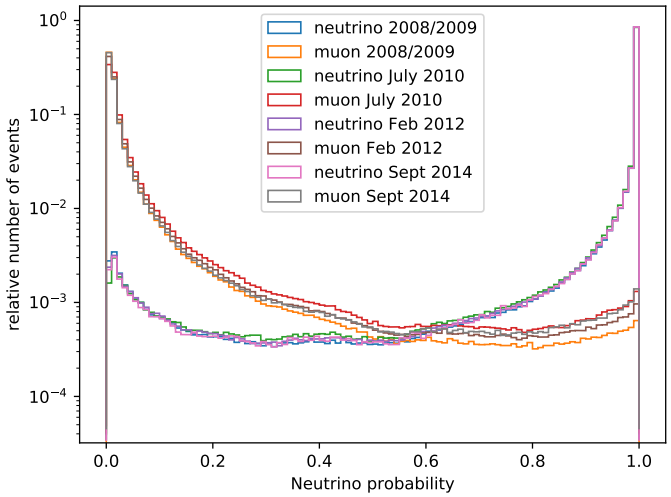


Figure 4.18: Neutrino probability p_ν distribution for the different test episodes separately for muon and neutrino events. The number is normed per event type.

July 2010 and September 2014 the detector configuration alone is different with two or one line permanently missing. It seems intuitive that, even without considering any further external parameters like noise, this might have negative influence on the performance of the background classifier. Additionally, the basic quality parameter for July 2010 and September 2014 is worse for most of the runs compared to 2008/2009. This quality parameter includes, besides other criteria, requirements concerning baseline and burst fraction.
In figure 4.19, the used neutrino probabilities are the same for all different episodes. With this in mind one can see that the neutrino efficiency for the same cut value is similar for all periods. The difference is only a few percent. This holds also for a separate analysis of track- and shower-like events as well as up- or down-going events. There is no significant improvement or worsening of the neutrino efficiency for the different episodes.
In contrast to this, the muon contamination is clearly differing. In terms of the training period, the deviations for February 2012 are relatively small but for the other two episodes the muon contamination is increased about 10 %.
This observation also holds for the chosen probability cut p_{cut} from section 4.3.3. The data points for this cut are highlighted for each episode. The choice of p_{cut} is usually determined by finding an acceptable value of muon contamination. This means that p_{cut} has to be adapted to every period individually. Regarding the chosen test sets used in this analysis, this means that July 2010 and September 2014 suffer from a decrease of the neutrino efficiency of about $5 - 10\,\%$. The neutrino efficiency for February 2012 could be increased about 1 %.

In case of cosmic neutrinos, p_{cut} is not determined by the muon contamination

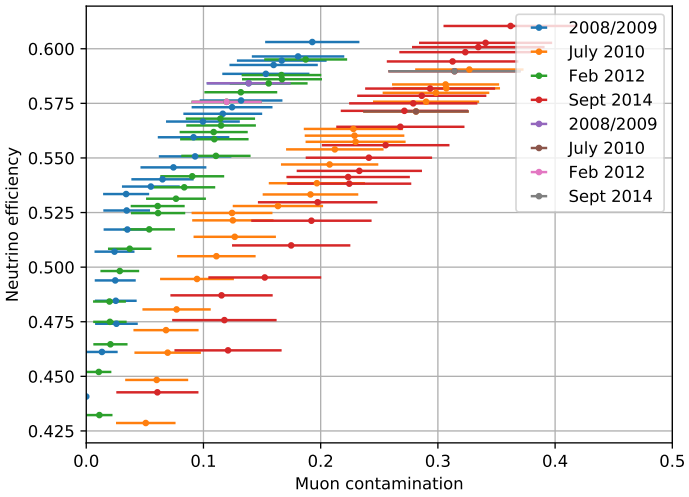


Figure 4.19: Muon contamination and neutrino efficiency for different values of p_ν on the different test sets. For all, the same values of p_ν have been used. The highlighted marker shows the quantities for p_{cut} as it was discussed in section 4.3.3.

but by the requirement of approximately one atmospheric muon per day. The neutrino efficiency for this cut as well as the expected number of remaining atmospheric neutrinos for the different test sets is shown in table 4.6. In general, the result is the same as for atmospheric neutrinos. The efficiency is basically equal for the original test set and February 2012, whereas it decreases for July 2010 and September 2014 around $6 - 10\,\%$. However, as already mentioned, the ratio between atmospheric muons and neutrinos might differ since the number of expected neutrinos is not equal for all episodes. Still, these observations confirm that the cut value has to be adapted to the respective episode.

Episode	N_μ/day	$N_{\nu_{atm}}$/day	ν_{eff} tot [%]
2008/2009	≈ 1	16.9	50.0
July 2010	≈ 1	11.5	44.0
Feb 2012	≈ 1	15.1	49.8
Sept 2014	≈ 1	7.9	41.6

Table 4.6: Number of atmospheric muons and neutrinos and signal efficiency for all neutrinos of the different episodes. The cuts on the neutrino probability are chosen in such a way that approximately one atmospheric muon per day survives for the respective time frame.

In summary, the results show a worse performance for those episodes with perma-

nently missing lines. It is however not clear if the different detector configuration is the reason for this. Still, it is obvious that the performance of the network on different episodes is not the same and has to be tested with Monte Carlo simulations in order to determine a proper cut before applying it to data.

However, even if such a test is done it is not clear for how long a chosen p_{cut} can be used reasonably. This opens up further questions. At some point, it might even be better to train a new network instead of just adapting the probability cut. This will be tested in the following.

Comparison with a classifier trained on July 2010

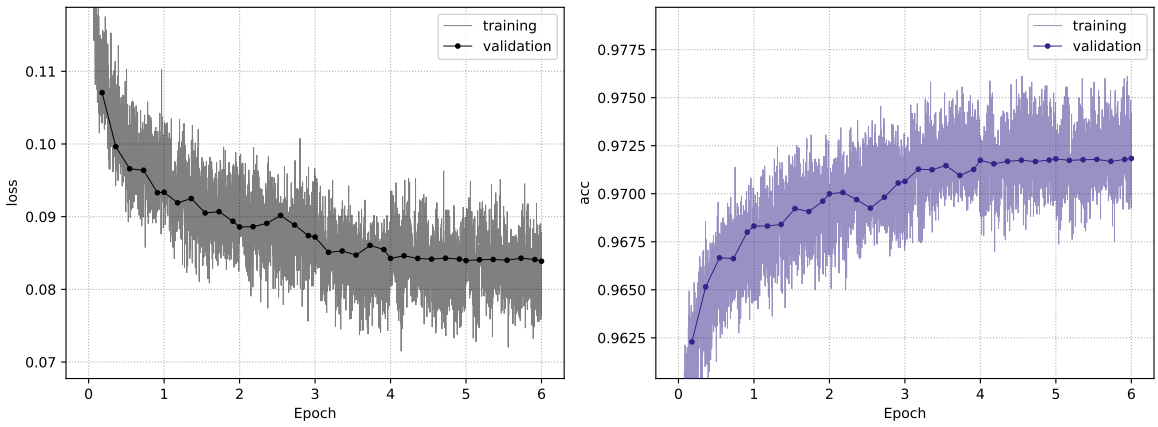
The results from the section before show that the performance of the background classifier depends on the episode it is applied to. While the performance on February 2012 is comparable to the one on 2008/2009, a deterioration can be observed for July 2010 and September 2014. For both of these periods, the detector geometry was different than during the training period. This way, events and therefore images have a different shape from what the network was adapted to.
In order to test if a network that is trained on a time around July 2010 achieves better performances, a second background classifier is trained. The training period is May 2010 to October 2010 excluding July 2010 that is, as for the original classifier, used as test set. The episode was chosen such that the detector geometry is approximately stable with two permanently missing lines.
Table 4.7 shows the amount of training (including validation) samples per event type including their percentage with respect to the training set. In total, the training set is smaller than for the original network but should still be sufficient. Additionally, the split between the different event types is slightly different, meaning that this time there are more muon events.

Event type	Number of events	Fraction [%]
Atmospheric μ	11 355 605	53.8
ν_e CC	3 339 171	15.8
ν_e NC	1 940 700	9.2
ν_μ CC	2 542 004	12.0
ν_μ NC	1 946 242	9.2

Table 4.7: Number of events for each event type and percentage with respect to the training and validation dataset of 2010 for the background classifier. For neutrinos, the event type includes particles and antiparticles in approximately equal parts.

Network architecture and hyperparameters are not modified compared to the original network. However, the network was trained for one epoch longer as the training set is smaller. The development of loss and accuracy is demonstrated in figure 4.20. Both quantities do not reach the performance of the original classifier (see figure

4.1). However, loss and accuracy are not evaluated on the same sample so that a direct comparison is not possible.
Therefore, the loss has to be evaluated on the same test sets. Comparing the loss of both networks on July 2010 shows an improved loss for the new classifier (0.08819). In contrast to this, the loss on the test set from 2008/2009 is worse (0.09434) (see table 4.5 for values of the original network). In general, both networks seem to perform better on their corresponding test samples.



(a) Loss (b) Accuracy

Figure 4.20: Loss and accuracy of the background classifier trained on Monte Carlo simulations from 2010 for training and validation during the training. For the training curve of both metrics, every data point is averaged over 250 batches, i.e. 250×64 images.

The distribution of the neutrino probability of the new classifier is very similar to the one of the original network (see figure A.9 in the appendix). In general, there are less events (muon and neutrino events) with $p_\nu > 0.99$ for the 2010 network but this alone is not a clue concerning the performance in terms of muon contamination and neutrino efficiency.

Figure 4.21 illustrates muon contamination and neutrino efficiency for atmospheric neutrinos for both networks on both test sets. It can be seen that the performance on the test sample from 2008/2009 is approximately equal for the corresponding 2008/2009 network (blue) and the 2010 network (red).
Regarding the performance on July 2010, the 2008/2009 network (green) even seems to achieve slightly better results than the one from 2010 (orange) over the whole range. That does also mean that the 2010 classifier performs better on a test set from a different episode.
Summarizing this means that both classifiers perform better on the 2008/2009 test set. Additionally, the 2008/2009 classifier performs even better on July 2010 than the 2010 network itself even though for this period the detector geometry is different with two missing lines. That is a situation the original network was not trained on,

whereas the 2010 network was exclusively trained with such events.

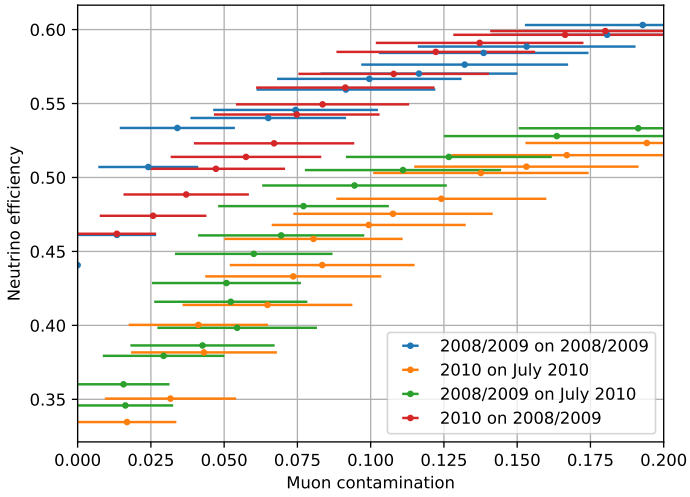


Figure 4.21: Muon contamination and neutrino efficiency for different values of p_ν on the test set of 2008/2009 and July 2010 for the network trained on 2008/2009 as well as the one trained on 2010.

It should be mentioned that the analysis for cosmic neutrinos shows slightly different results. In this case, both classifiers perform best on their corresponding test set and these results are better than the one from the other network. However, the differences are still quite small. The achieved efficiency for July 2010 only differs about 2 %. Again, it should me mentioned that the ratio between signal and noise is different for the used cuts.

In summary, it seems like the performance is limited by the test set itself. Hence, training a network with the detector geometry of the desired application episode is not sufficient to improve the performance on this time. Although it is probably a factor that has to be considered, there are obviously additional parameters that take influence on the performance. The most intuitive ones are external run conditions that are influenced by noise. A good starting point for this is the baseline.

In order to investigate this further, the neutrino efficiency for atmospheric neutrinos is evaluated run-wise on July 2010 for both networks and compared to the corresponding baseline (see figure 4.22). The probability cuts for the classifiers are chosen such that the muon contamination is on a comparable level.
In principle, both classifiers show the same shape following the baseline inversely, meaning that the efficiency drops for runs with higher baselines. This correlation is already known from classical reconstructions.
For most of the runs from July 2010, the 2008/2009 classifier yields a slightly better performance leading to the general better efficiency already observed in figure 4.21.

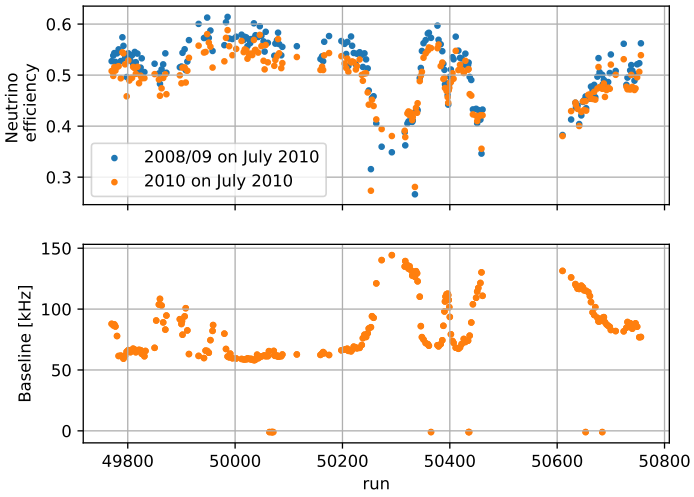


Figure 4.22: Neutrino efficiency on runs from July 2010 for the 2008/2009 network (blue) and the 2010 network (orange) as well as the corresponding baseline for each run.

However, the neutrino efficiency is worse for the runs with the highest baselines so that in total, there is a larger spread between the minimum and maximum achieved efficiency for this network compared to the 2010 classifier. This way, the performance of the 2010 network is in total worse but more robust in view of changing baseline rates. In total, these differences are very small but still remarkable.

An explanation for this can be found by analysing the baseline distributions of the different training sets. The 2008/2009 classifier is exclusively trained on runs with a baseline of less than 100 kHz mostly between 50 and 70 kHz (see figure 3.4 (b)). In total, the baseline is almost constant during training. This way, the network probably gets specialized to runs with rates in this range, whereas the performance on runs with higher rates is poor. The 2010 classifier on the other hand is trained with baseline rates from a broader spectrum up to more than 150 kHz. Therefore, it is not that specialized to a certain range but also more robust to fluctuations. This can also explain the good performance on the events from 2008/2009.
Additionally, it should be noted that for the 2008/2009 network both training and test set contain runs from the complete episode of September 2008 to February of 2009, meaning that all possible run conditions from that episode should be contained in the training as well as the test set. Hence, the external parameters in the training set should be representative for the test set.
This does not hold inevitably for the 2010 classifier where the test set is a connected period of one month the network was never trained on. As the run conditions can change within a very short time interval, this means that some conditions from July 2010 might not have been included in the training set. Comparing some of the external parameters from July 2010 and the training period shows that the distribution is

not completely the same, eventually leading to a worse performance. In July 2010, there are for example some runs with a higher burst fraction than any run from the training set of 2010. This could explain why the performance on July 2010 is clearly worse than on the test set from 2008/2009.

In general, the run-wise efficiency analysis of July 2010 for astrophysical neutrinos shows the same shape (see figure 4.23). However, in this case, the performance is consistently better for the 2010 network even though the improvement is very small. Still, some similar observations as for atmospheric neutrinos can be obtained. The difference between both networks is largest for runs with the highest rates. Hence, the spread between minimum and maximum efficiency is still larger for the 2008/2009 network, whereas the 2010 network seems to be more robust. In case of cosmic neutrinos, the network trained on an episode close to the desired application time leads to a minor improvement. Furthermore, external conditions definitely play a role and affect the performance of the network, regardless which neutrino spectrum is used.

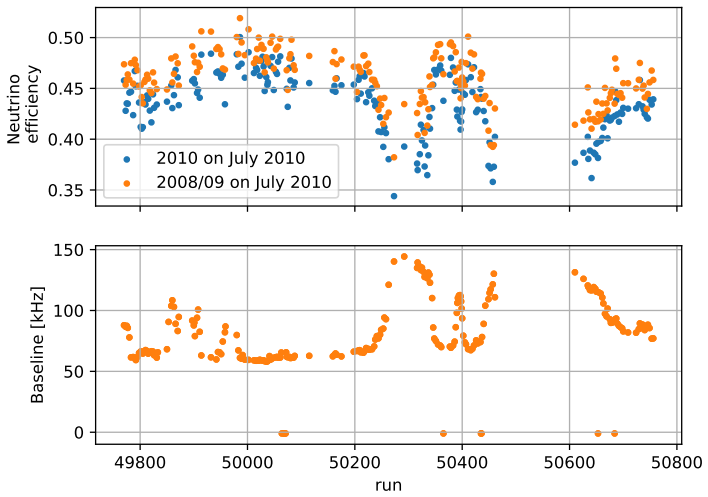


Figure 4.23: Neutrino efficiency on runs from July 2010 for the 2008/2009 network (blue) and the 2010 network (orange) for a neutrino spectrum $\sim E^{-2}$ as well as the corresponding baseline for each run.

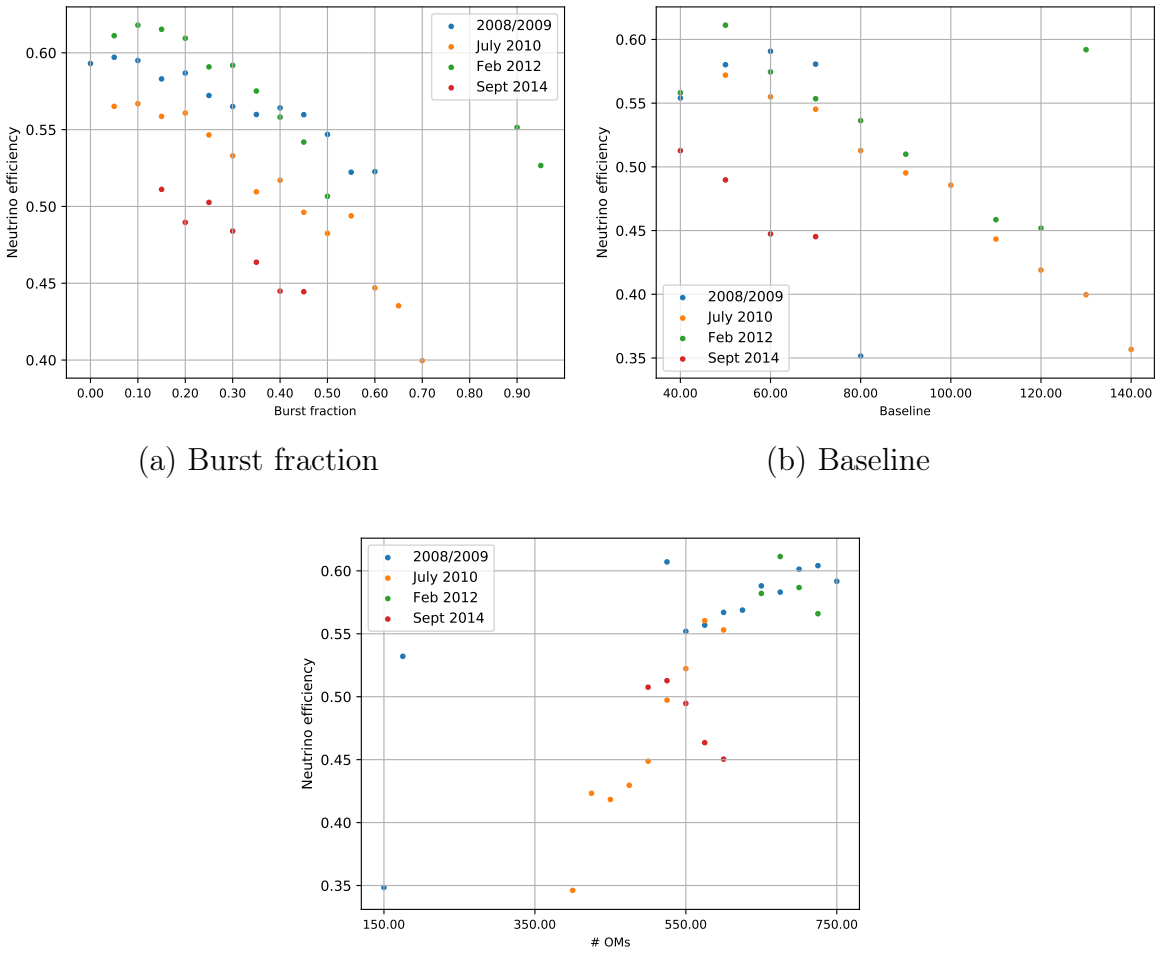
Summarizing the results from this section, it seems like training and evaluating the network on a connected period might not inevitably lead to the best possible performance. Instead, a possible strategy is to summarize runs with similar run conditions and train several classifiers for the individual categories.
Therefore, it is of course necessary to analyse the influence of different external parameters on the performance of the background classifier. A small investigation is done in the following section.

Noise dependency

The results from the section before show that training a network on a certain episode close to the desired application time does not necessarily improve the performance. It was already pointed out that this approach mainly focuses on a quite constant detector layout and therefore possibly includes a variety of different data taking conditions. With the observations from before, it seems plausible that some of these conditions might take more influence than the detector geometry.
In order to investigate which parameters take influence on the performance, a run-wise evaluation of the neutrino efficiency as it was already shown for the baseline before (see figure 4.22) was done. Some of the analysed parameters show a correlation that are mainly baseline and burst fraction. Both quantities characterize the noise situation in the detector. As this indicates that the noise mostly affects the network's performance, a small analysis was done. In addition to baseline and burst fraction, the number of active OMs in the detector was studied as this is also a quantity that is influenced by the noise.
The muon contamination was also analysed run-wise but no correlation was found.

For this analysis, the neutrino efficiency for atmospheric neutrinos is analysed bin-wise for the different parameters separately in figure 4.24. The results are presented separately for the different test sets. As before, the cut on p_ν is chosen such that the muon contamination is on a comparable level. Regarding the burst fraction, the exact value of the neutrino efficiency depends on the episode it is applied to. However, for all four test sets, the performance decreases with increasing burst fraction. The neutrino efficiency is reduced by $1-2\,\%$ for an increase of the burst fraction of about 0.1.
Analysing the baseline shows the same trend. The higher the baseline, the worse the efficiency. However, except for September 2014, all episodes show very similar results within the same baseline range. Mostly it seems like an increase of the baseline of about 10 kHz leads to a reduction of the efficiency of about $2-3\,\%$.
Analysing the number of active OMs in the detector is however not that clear. While a higher number leads to an improved performance for 2008/2009 as well as July 2010, it leads to a decrease for September 2014. However, the number of OMs alone does not include any information about the current detector geometry. It is unclear if a missing OM from the core region leads to the same influence as a missing OM from the edge of the detector. Moreover, a missing OM does not necessarily affect the performance at all if the events take place in another part of the instrumented volume.

Additionally, these quantities can be evaluated in a two dimensional histogram (see figure 4.25). It is for sure that a higher baseline and burst fraction lead to a lower efficiency respectively. However, there are still some differences regarding the different episodes. When analysing the same conditions, the efficiency is for example worse for September 2014 compared to February 2012. This indicates that these two parameters can't be the only ones affecting the performance. However, the influence



(a) Burst fraction

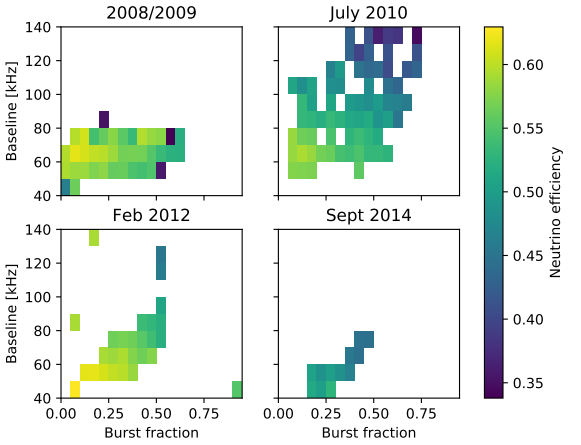
(b) Baseline

(c) Number of active OMs

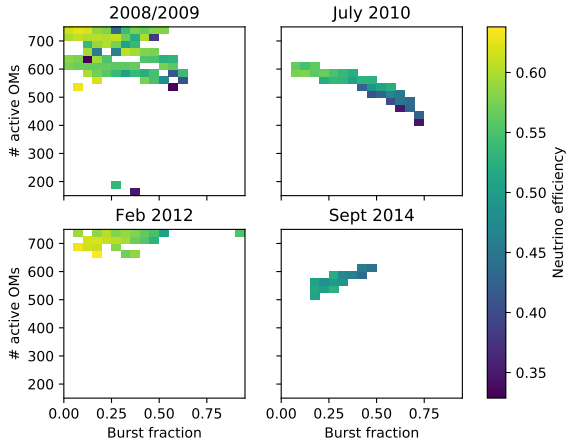
Figure 4.24: Bin-wise neutrino efficiency for the different test episodes for different parameters that characterize the noise.

of the number of active OMs is not that clear and seems to be less strong than for the two already discussed parameters.
The outliers towards small neutrino efficiencies for the test set from 2008/2009 in figure 4.25 could be a result from low statistics in the respective bin. If only a small number of runs contributes to these bins, it is possible that not all event types are included. Hence, the efficiency might not be that meaningful as for the remaining episodes.

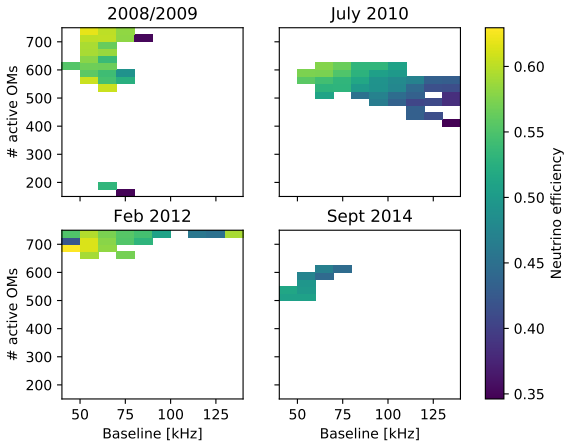
In general, the relation between noise and efficiency are the same for an analysis of cosmic neutrinos (see figures A.10 and A.11 in the appendix). The overall performance is of course worse due to the stricter cut. However, the influence of the single external parameters seems to be less remarkable.
The efficiency reduces only about 1 % for an increased burst fraction of around 0.1



(a) Burst fraction and baseline



(b) Burst fraction and number of active OMs



(c) Baseline and number of active OMs

Figure 4.25: Two dimensional bin-wise neutrino efficiency for the different test episodes for different parameters that characterize the noise.

and around $1 - 2\,\%$ for a higher baseline of about 10 kHz.
Hence, it seems like the influence of noise is less remarkable for cosmic neutrinos but still clearly observable. This indicates again that high energetic neutrino events, as expected for astrophysical neutrinos, are an easier task for the network. In case of lower energies, external parameters such as noise seem to affect the efficiency more severe.

In summary, this analysis shows that some external parameters that characterize the noise definitely influence the performance of the network. This way, these are potential quantities that can be used to separate runs into several categories. Training a network only on events from the respective class could lead to a specialization to the corresponding run conditions. This could improve the performance further.

5 Energy reconstruction

As already mentioned, the distinction between atmospheric and cosmic neutrinos is mainly done due to the different energy spectra. Hence, a good energy reconstruction is important. This chapter will focus on the CNN-based approach. Event selection and network architecture are introduced. Section 5.5 presents the performance results for the full and selected datasets. Additionally, the performance is compared to some standard reconstructions. Finally, a small test on the application to data is done.

5.1 Training set

During this work, some different approaches for the training set have been tested. In principle, the same set as for the background classifier has been used. The training period is set to September 2008 until February 2009. Again, charged and neutral current events have been used for both ν_e and ν_μ as well as the corresponding antineutrinos.
The first approach used the same set as the background classifier (see section 4.1). However, it turned out that only training the network on neutrino events improves the performance significantly. Therefore, the final network was exclusively trained on neutrinos.
As already mentioned in section 1.3.1, the amount of energy deposited inside the detector and hence the energy reconstruction depends on the neutrino signature. Therefore, a 50 : 50 split between track- and shower-like events is used. This reduces the number of available events compared to the background classifier as it is limited by the number of ν_μ CC events. Additionally, the energy deposition is different for the charged and neutral current component of shower-like events. Therefore, these events are also split into 50 % ν_e CC and 50 % NC. The number of events used for training and validation for each flavour and signature is shown in table 5.1. The remaining shower-like events are included into the test set. The energy and direction distribution of the training set can be found in the appendix in figure A.12.
Table 5.1 shows that the split between track-like and shower-like is not completely 50 : 50. However, it turned out that the performance on shower-like events is significantly better, so that a slightly higher amount of track-like events is accepted in order to not increase the performance difference any further.

As it has already been pointed out, the spectrum of neutrino energy is wide. Even though ANTARES is only sensitive to a part of this spectrum, it still covers energies between 10 and 10^8 GeV but is mainly sensitive to TeV − PeV neutrino energies.

Event type	Number of events	Fraction [%]
ν_e CC	1 802 229	22.5
ν_e NC	897 175	11.2
ν_μ CC	4 393 314	54.9
ν_μ NC	908 349	11.4

Table 5.1: Number of events for each event type and percentage with respect to the complete training and validation set of 2008/2009 for the energy reconstruction. The event type includes particles and antiparticles in approximately equal parts.

Hence, the network needs to be trained on energies in this range. Ideally, the distribution should be equally distributed so that the network is not biased to a certain energy range. However, the number of simulated neutrino events is depending significantly on flavour and energy (see figure A.12 for the distribution in the training set). Because of the further limitation of the number of training samples, no additional energy balancing is done. However, it seems likely that this could improve the performance on energy ranges with a smaller simulation statistics.

With respect to the total balanced dataset, 75 % of the events are used for training, 5 % for validation and 20 % for testing. Additionally, about two million muon events are added to the test set in order to evaluate the performance.

5.2 Visible energy

The task of this network is to predict the energy. However, when predicting the true neutrino energy some major performance differences occur. In ν_e CC interactions, all the neutrino energy is deposited inside (or at least close to) the detector and is therefore potentially visible. This does not hold for neutral current reactions as a major amount of the energy is carried away by the neutrino (see figure 1.3). This part of the energy is therefore invisible to the detector and there is no possibility for the network to reconstruct it. Therefore, the network is trained only on the visible shower energy that coincides with the true neutrino energy only for the charged current component. The visible energy of all NC reactions is given by $E_{\text{vis}} = y \cdot E_\nu$ where y is the Bjorken scaling variable. In this way, both shower-like event types are treated equally.
Regarding shower-like events, in this work true energy hence only refers to visible energy. The term true energy coincides with the neutrino energy only in case of charged current interactions.

5.3 Loss function

A very common choice of the loss function for regression is MSE. However, the disadvantage of this loss is that, besides the prediction for the reconstructed quantity, no additional information about its quality is given as this is for example achieved in classification with predicting a probability.
In order to avoid this issue, a loss function was chosen that does not only predict the energy μ_{pred} but also an uncertainty parameter σ_{pred}. This is very useful in order to apply reasonable quality cuts.

In general, the output of the network can be interpreted as a normal distribution:

$$p = \frac{1}{\sqrt{2\pi\sigma_{pred}^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_{pred})^2}{\sigma_{pred}^2}\right) \tag{5.1}$$

with mean μ_{pred} being the actual predicted energy and standard deviation σ_{pred} as uncertainty parameter.
The loss is then defined as the negative log-likelihood of the true energy μ_{true}:

$$L = \frac{1}{2}\left(\log(2\pi) + \log\left(\sigma_{pred}^2 + \epsilon\right) + \frac{(\mu_{pred}-\mu_{true})^2}{\sigma_{pred}^2 + \epsilon}\right) \tag{5.2}$$

The parameter ϵ is added in order to avoid a division by zero. An additional capping of the loss is implemented for stability, which is not shown here.

5.4 Network architecture

The CNN architecture for the energy regression is shown in table 5.2. It is very similar to the one of the background classifier. However, the amount of filters in the convolutional layers is higher for the regression network. A small dropout rate of $\delta = 0.01$ was used in order to avoid overfitting.
The main difference compared to the background classifier is the output layer. As the desired output is two dimensional, there are two distinct layers. Both are calculated by a single dense layer but use different activation functions. The actual energy output μ makes use of a linear function. However, as the uncertainty σ should not be equal to zero, softplus is used as activation function.
As for the background classifier, adam is used for optimization with the standard settings from *OrcaNet*. However, the initial learning rate is set to 0.005 and decayed exponentially.
The training is stopped after 11 epochs which only takes about one and a half day using a Nvidia GTX 1080 (Ti) GPU.

Building block / layer	Output dimension
xyzt input	$4 \times 4 \times 25 \times 100$
Convolutional block (128 filters, $\delta = 0.01$)	$4 \times 4 \times 25 \times 128$
Convolutional block (128 filters, $\delta = 0.01$)	$4 \times 4 \times 25 \times 128$
Convolutional block (128 filters, $\delta = 0.01$)	$4 \times 4 \times 25 \times 128$
Convolutional block (128 filters, pooling, $\delta = 0.01$)	$4 \times 4 \times 12 \times 128$
Convolutional block (256 filters, $\delta = 0.01$)	$4 \times 4 \times 12 \times 256$
Convolutional block (256 filters, $\delta = 0.01$)	$4 \times 4 \times 12 \times 256$
Convolutional block (256 filters, $\delta = 0.01$)	$4 \times 4 \times 12 \times 256$
Convolutional block (256 filters, pooling, $\delta = 0.01$)	$4 \times 4 \times \ 6 \times 256$
Flatten	24576
Dense 256 + ReLu	256
Dense 128 + ReLu	128
Dense 32 + ReLu	32
Dense 1 + Linear (μ) Dense 1 + Softplus (σ)	2

Table 5.2: Network architecture for the energy reconstruction

5.5 Performance

5.5.1 Loss

From equation 5.2, it can be seen that a high loss can not only be the result of a high deviation between true and predicted energy but also from a high uncertainty. Therefore, in addition to the loss, the mean squared error is monitored as an intuitive measure. This ensures that the reconstructed energy really improves. Both quantities are listed in table 5.3 for the different event types.

Event type	Loss	MSE
track	1.88	2.80
shower	0.39	0.31
ν_e CC	0.38	0.31
NC	0.40	0.30
atmospheric μ	2.85	4.55

Table 5.3: Loss and mean squared error of the test set for the different event types

It is obvious that the performance on track-like events is far worse than on shower-like events. However, this is an expected result as the energy reconstruction for ν_μ CC events is much more difficult.

On the one hand, the interaction vertex can lie outside the instrumented volume. In this case, the hadronic cascade might be partially or even completely invisible to the detector. This can also happen for shower-like events but it is far less likely as the longitudinal extension of these events is significantly shorter. In order to detect shower-like events, it is necessary that the interaction takes place inside or at least close to the instrumented volume. For ν_μ CC events, even if the interaction takes place far away from the detector, it can still be possible to detect the resulting muon. In this case, a large amount of energy could be deposited before or after passing the instrumented volume.
Another challenge for the energy reconstruction of track-like events is the different kind of light emission below and above the critical energy. This will be discussed further in the following performance analysis.

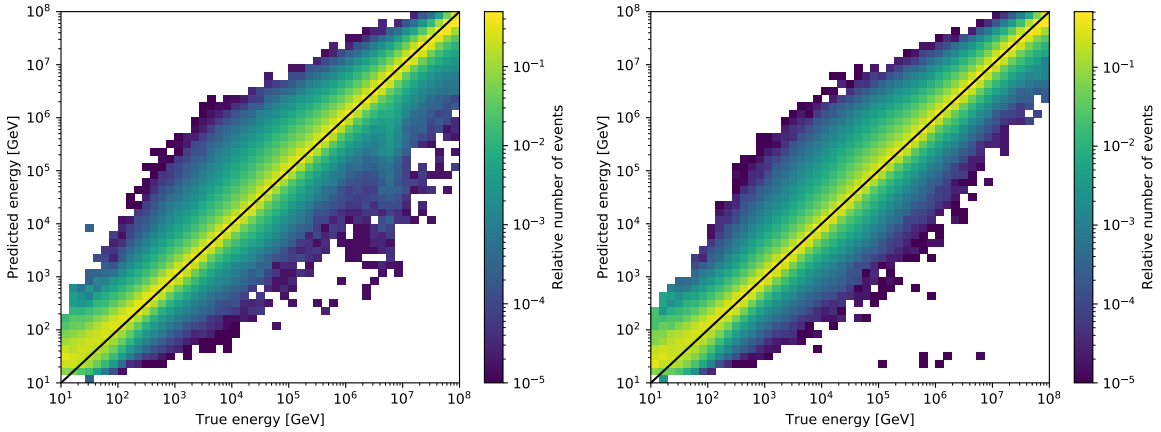
Table 5.3 also shows the distinction between the different components of shower-like events. It can be seen, that there is basically no difference. However, this results from the use of the visible energy.
For a network that is not trained on the visible but the true neutrino energy, this does not hold. In this case, the loss for NC events is significantly worse compared to ν_e CC events. This results from the fact that the network can not predict the invisible energy for NC events other than on a statistical basis.
However, the difference between neutral and charged current component is not the only difference. Even though the input does not change for ν_e CC samples, the loss is much worse if the visible energy is not used. As the network learns that there is some amount of invisible energy from NC events, it probably transfers this assumption to the other event types. Hence, the loss for ν_e CC samples is worse for such a network. In contrast, the loss for the track-like and atmospheric muon events is better, probably because in this case there is also often a certain amount of invisible energy.

It should be mentioned that a network trained without the 50 : 50 split between tracks and cascades leads to an even better performance for shower-like events but a worse performance for track-like neutrino and atmospheric muon events. This is not surprising as the network gets even better adapted to shower-like events. However, this does also indicate that training two distinct networks for the two interaction signatures could lead to an improved performance. Since the training set is already quite small, this approach was not tested here. Besides, this approach requires an additional track-shower classifier in the analysis pipeline.

Finally, the performance on atmospheric muon events is very poor. This can be expected as the network was never trained on such events. Eventually, this can be used to reject muon events additionally to conventional cuts. This is possible if the reconstructed uncertainty σ works as desired, i.e. shows high values for bad reconstructions.

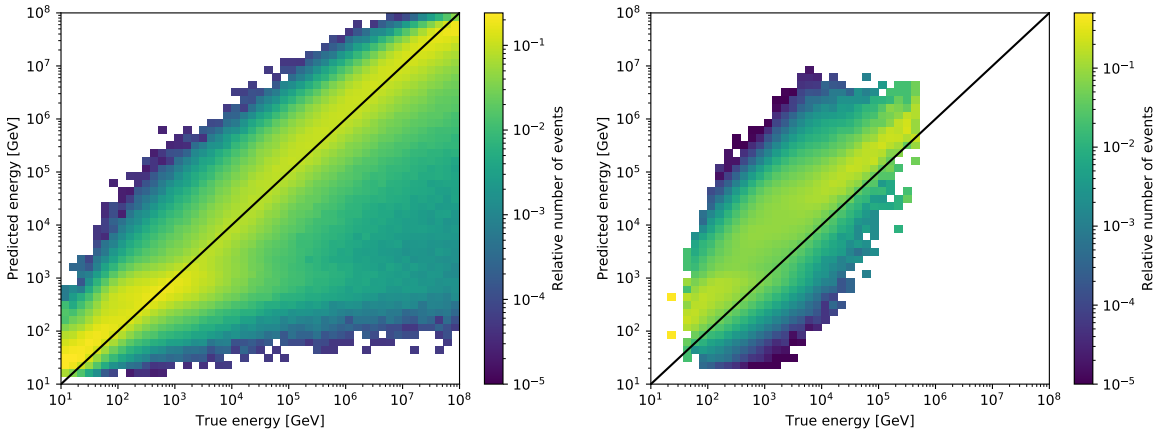
5.5.2 Reconstructed energy

Figure 5.1 shows true and predicted energy for the different event signatures. The term true energy coincides with the neutrino energy only for the charged current interactions. For neutral current interactions, it is the energy that is transferred to the hadronic system.
As it was already indicated by the loss, the performance depends a lot on the event type.



(a) ν_e CC events

(b) NC events



(c) Track-like events

(d) Atmospheric muon events

Figure 5.1: True and predicted energy for all different event types. True energy refers to the label given to the network as explained in section 5.2. The colorbar indicates the normed quantity per true energy bin. Its minimum is restricted to 10^{-5} for all plots.

Considering shower-like events, both ν_e CC and NC events (see figure 5.1 (a) and (b)), show similar results. The only significant difference can be found for a true shower energy of about 6 PeV. For ν_e CC interactions, an outlier of events with

underestimated reconstructed energy can be found. This is clearly a result of the Glashow resonance for $\bar{\nu}_e$ events [47]:

$$\bar{\nu}_e + e^- \quad \rightarrow \quad W^- \tag{5.3}$$

The threshold energy for this process is the energy that is necessary to generate the W^- boson. This corresponds to a neutrino energy of about 6.3 PeV. The W^- boson then decays either into a lepton and the corresponding neutrino or gives rise to a hadronic shower. Thus, the visible energy does not coincide with the original neutrino energy. Therefore, the prediction of the network is too low.
Apart from the Glashow resonance, the performance is basically equal for ν_e CC and NC events. However, this does only hold due to the use of the visible energy as this leads to the same initial situation (see figure A.13 for the results on shower-like events for a network that does not use the visible energy). The energy tends to be overestimated for the lowest energies and the largest variance can be found in the medium energy range.

The result for track-like events (see figure 5.1 (c)) clearly differs from shower-like events. In this case, there is a large amount of samples with underestimated energy, especially for high energetic events. However, considering the lower bound of the energy spectrum, the prediction is slightly too high. This overestimation results in a plateau between 300 GeV and 1 TeV. If the muon is a minimum ionizing particle (MIP), i.e. the energy loss is close to the minimum, its track length is about 4 m/GeV. Hence, for energies above 100 GeV, they can traverse the complete instrumented volume of ANTARES. In case that the interaction vertex is not inside the detector, this leads to a saturation of the deposited energy and explains the extension of the plateau to higher energies. In total, there are more events with a better reconstruction above approximately 10 TeV but at the same time the variance increases dramatically. This underestimation can be expected for ν_μ CC events. As already explained, the muon from such a reaction often only passes the detector. Hence, a large amount of energy is deposited outside the detector and is therefore invisible. Consequently, the reconstructed energy tends to be too small.

The detected overestimation for the lower bound of the energy spectrum can be found for both event signatures. The reason is probably the low statistics in the training set. Hence, the prediction is biased towards higher energies.

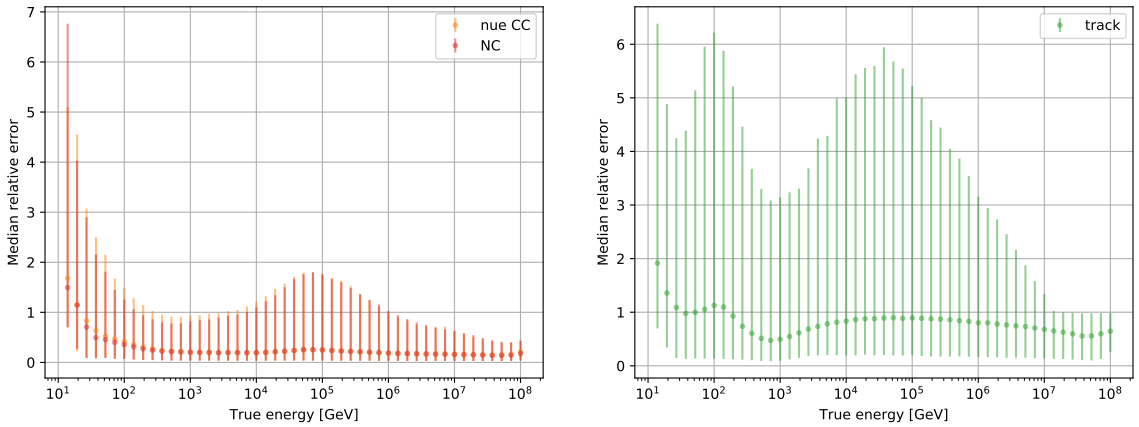
Finally, the performance on atmospheric muon events is evaluated as well (see figure 5.1 (d)). The network's prediction clearly does not match the true energy very well. Different from track-like events caused by neutrino interactions, the CNN tends to overestimate the energy of atmospheric muons. A possible reason is that the network assumes a missing visible hadronic cascade as learned from the ν_μ CC events and hence tries to compensate this with a higher prediction. However, as already pointed out, it is not the goal of this network to reconstruct the energy of atmospheric muon events. Still, the predicted energy range seems to be restricted compared to neutrino events. Hence, especially for high reconstructed energies above 10 PeV, the chance

is high to be a result of a real neutrino event.

Another approach to quantify the performance is given by the median relative error:

$$MRE = \text{median}\left(\left|\frac{E_{true} - E_{pred}}{E_{true}}\right|\right) \tag{5.4}$$

Figure 5.2 presents this quantity for the different energy bins for track- and shower-like events separately. The error bars indicate the 10 % and 90 % quantile in each bin.



(a) Shower-like events (b) Track-like events

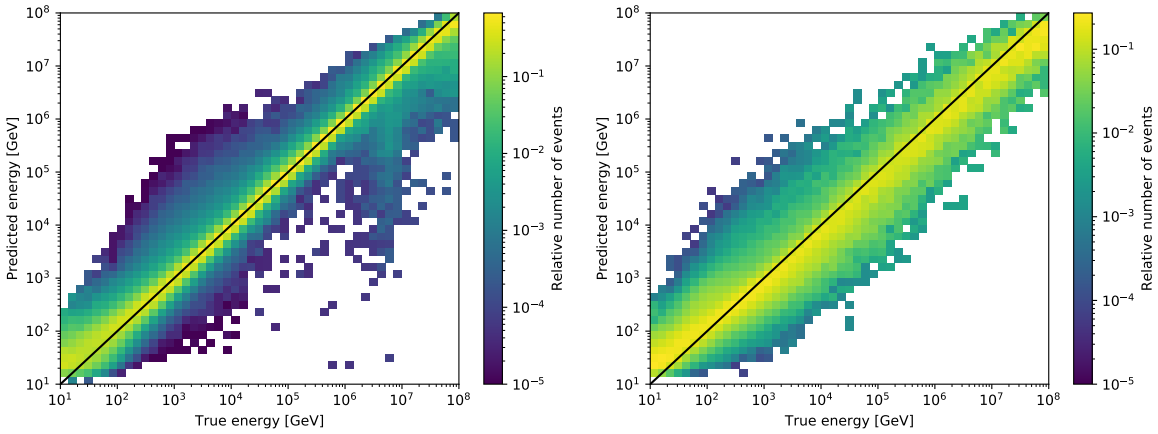
Figure 5.2: Median energy error for all events. For the shower-like events, both ν_eCC (orange) and neutral current NC (red) are demonstrated separately. The error bars indicate the 10 % and 90 % quantile in each bin.

In total, the median energy resolution for shower-like events is very stable, especially above 1 TeV with a value of approximately 0.2. Above 50 TeV it increases slightly to 0.25 until 500 TeV. For higher energies, the MRE drops again down to $0.15 - 0.2$. Below 1 TeV, the deviation between true and predicted energy is more severe but decreases very fast, dropping below 1 at already 20 GeV. Additionally, the fluctuations in this range are large (up to almost 7 but above 50 GeV always below 2). At least for the lowest energies, this might be due to the low statistics and could maybe be improved by adding more low-energetic events to the training set. As there is also a quite hard transition to a lower statistic for visible energies just below 1 TeV, this could also be the reason for the increased error bars above this energy. No significant difference for charged and neutral current component can be noted.
For track-like events, the median energy error is higher compared to shower-like events. The best resolution (MRE=0.5 with 90 % quantile up to 3) is achieved for energies just below 1 TeV. This is approximately the critical energy of muons in water. Below this threshold, the energy loss is dominated by ionization. The light emission in this range is continuous. For these energies, the median resolution is

mostly around 1 and the 90 % quantiles extend up to 4−6. Above the critical energy, more stochastic radiation processes start to take over. Hence, the emission is characterised by a more burst-like shape. The network needs to learn both emission types in order to perform well over the whole spectrum. However, from figure 5.2 (b), it can be seen, that the median resolution is in general better for energies above the critical energy but the 90 % quantile is still very large, partly exceeding the fluctuations below the critical energy. The median error first increases to approximately 0.9 around 100 TeV and is then reduced to 0.6 for the highest energies. The error bars extend up to almost 6 again at $E = 50\,\text{TeV}$. Only for the highest energies above 10 PeV, the 90 % quantile is reduced to approximately 1. The increasing error could be related to the lower statistics in the medium energy range. This also holds for the lowest energies. However, in general the fluctuations are high for almost all true energy bins. This indicates the large variance that is also shown in figure 5.1 (c). As already mentioned before, this is the result of events that are not contained.

Contained events

As already pointed out, a challenge for the energy reconstruction are events that are not contained. In order to verify this, a small performance study on contained events is done. In this analysis, contained refers to the interaction vertex, i.e. the interaction takes place inside the detector. The restrictions are taken from [48], requiring $\rho_{\text{MC}} < 90\,\text{m}$ and $z_{\text{MC}} < 200\,\text{m}$.
From figure 5.3, it can be seen that the variance is getting much better compared to figure 5.1, especially for track-like events.

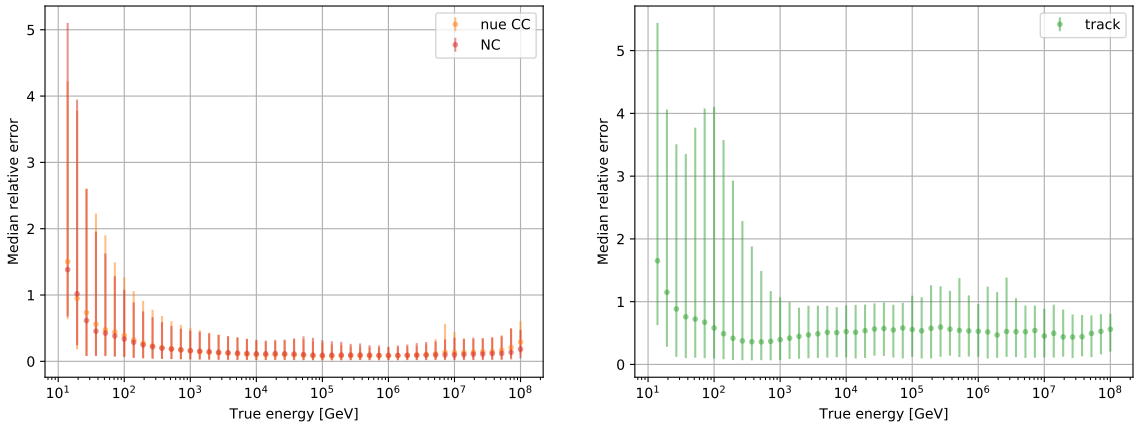


(a) Shower-like events (b) Track-like events

Figure 5.3: True and predicted energy for all contained neutrino events and the different event signatures. True energy refers to the label given to the network. The colorbar indicates the normed quantity per true energy bin. Its minimum is restricted to 10^{-5} for all plots.

As expected, the requirement of containment discards mostly events with an un-

derestimated energy reconstruction. However, the underestimated outliers around the Glashow resonance for shower-like events are not rejected. Even if the vertex of these interactions is inside the detector, still some amount of the energy might be invisible. The median relative energy error therefore decreases, especially for track-like events (see figure 5.4).



(a) Shower-like events (b) Track-like events

Figure 5.4: Median energy error for all contained events. For the shower-like events, both ν_e CC (orange) and neutral current NC (red) are demonstrated separately. The error bars indicate the 10 % and 90 % quantile in each bin.

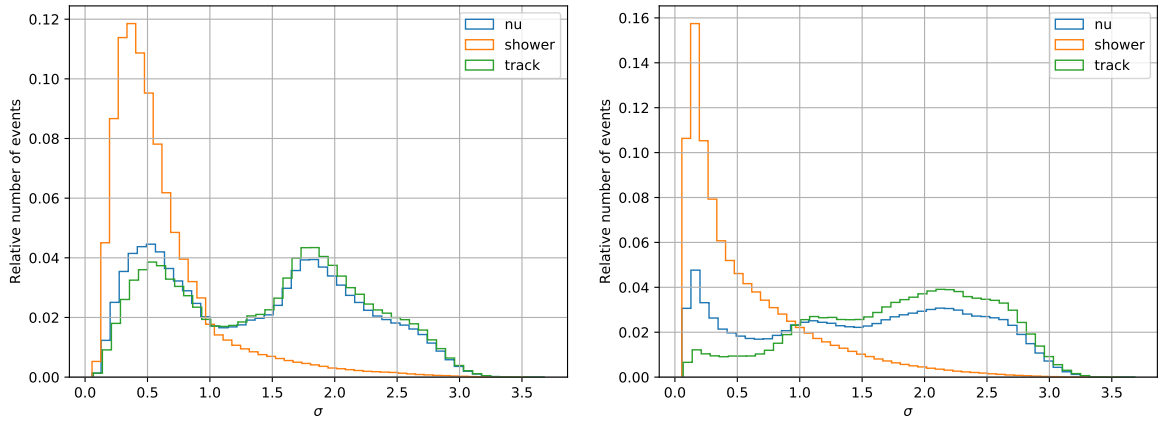
Except for the lowest energies the resolution for track-like events drops constantly below 1. The minimum at the critical energy is reduced about 20 % to 0.35. Except for the highest and lowest energies, the relative improvement is mostly around $30 - 40$ % and reaches its maximum for energies between 100 GeV and 300 GeV with a reduction of approximately 50 %. At the boundaries of the energy spectrum, the improvement is smallest but still about 10%.
Additionally, the extension of the error bars can be greatly reduced. Below the critical energy, the 90 % quantiles are still high around $3 - 4$, corresponding to a decrease of about 10 %. However, above they drop below 2 and mostly even below 1. For energies between a few TeV and about 100 TeV, the improvement is even at 80 %. Since the fluctuations for the highest energies were already comparatively small, the reduction for these bins is only about 10 %.
For shower-like events, the absolute improvement is very small but for energies between approximately 50 TeV and a few PeV, the median error even drops below 0.1. However, the relative improvement for this signature is significant as well. Below 1 TeV, the reduction is only about 10 % but increases to 60 % for energies between 40 TeV and 200 TeV. In case of the highest energies ($E > 30$ PeV), the median error can not be improved or even increases slightly. This holds also for the extension of the error bars. Apart from these energies, the 90 % quantile can be reduced. Between 10 TeV and a few PeV the relative improvement is even around 80 %. In

case of ν_e CC events, the Glashow resonance leads to a higher fluctuation in the respective bin. This is the only significant difference between charged and neutral current component.
In total, the limitation to contained events definitely reduces the median error and the extension of the 90 % quantiles for both event signatures. As expected, the large overestimation for low energetic events and the corresponding large fluctuations in the energy resolution can not be improved.

However, the performance for this pre-selection could probably be improved further if the networks are trained exclusively on contained events. This way, the network would not learn to compensate the underestimation caused by not contained events. Consequently, the limitation to contained events could not only prevent the network from underestimating events from a large distance but also reduce the dominant overestimation in the lower energy regime.
In summary, it can be seen that the performance on events with contained vertex is much better. Hence, a good vertex reconstruction could lead to a major improvement in the energy reconstruction.

5.5.3 Reconstructed uncertainty

The second output of the network σ is intended to be an uncertainty measure for the predicted energy so that it can be used as a quality cut in order to reject badly reconstructed events.
Figure 5.5 presents the distribution of σ for the different event types for an atmospheric weighting as well as a cosmic spectrum $\sim E^{-2}$.

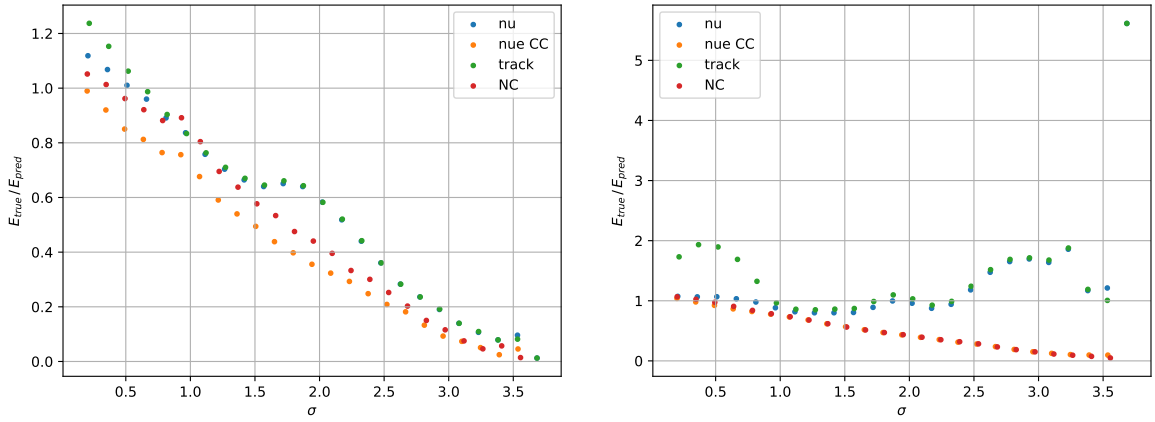


(a) Atmospheric neutrinos (b) Astrophysical neutrinos

Figure 5.5: σ distribution of the different neutrino signatures and different energy spectra. The distribution is shown for all neutrinos *nu* (blue) and separately for shower-like (orange) and track-like (green) events. The number is normed per event type.

Depending on the considered neutrino spectrum, the distribution differs. However, there are definitely more track-like events with a high uncertainty.
Independent of the used flux, the shower distribution peaks below 0.5 and shows a tail of only a few events with large values of σ.
Atmospheric track-like neutrinos also show a peak below 1 but also a second one just below 2. In case of astrophysical neutrinos, the distribution resembles a step function with plateaus between $0-1$ and $1-1.5$ respectively, followed by a broad gaussian peak at 2.25.
Hence, the amount of neutrinos with a high or low uncertainty depends crucially on the selected flux and therefore has to be chosen individually.

However, the distributions in figure 5.5 do not justify a cut on σ. Therefore, the correlation of the median ratio between E_{true} and E_{pred} with σ is analysed. The results are shown in figure 5.6 for atmospheric and cosmic neutrinos.



(a) Atmospheric neutrinos (b) Astrophysical neutrinos

Figure 5.6: Median ratio of true and predicted energy as a function of σ for an atmospheric and an astrophysical neutrino spectrum $\sim E^{-2}$. The distribution is demonstrated for all neutrinos *nu* (blue) as well as separately for the different shower-like events from ν_e CC (orange) and NC (red) interactions and for track-like (green) events.

For atmospheric neutrinos, the correlation is approximately linear for all different event types, i.e. events with high σ show a larger deviation between truth and prediction. As the median value per bin is mostly below 1, a cut on σ should dominantly reject events with overestimated energy. However, the slope of the curve is steeper for shower-like events. Hence, for the same cut, on average events with a high deviation get more effectively rejected than for track-like events.
In case of shower-like events, the same results can be obtained for an E^{-2} spectrum. However, this does not hold for ν_μ CC events. The best median agreement can be found for $\sigma = 1-2.5$ and for $\sigma \approx 3.5$. Apart from these ranges, the median deviation indicates an underestimated reconstructed energy.

The reason for this is probably the higher energy range for astrophysical neutrinos. As shown in figure 5.1, the deviations between truth and prediction are largest for the highest energies. These events are not expected to result from atmospheric neutrinos. Hence, they only affect the performance for a point source spectrum.
However, figure 5.6 only presents the median value per bin. In case of track-like and hence all neutrinos, the fluctuations are large, exceeding ratios of 10. This holds especially for $\sigma > 1.5$, i.e. a range with a good median value. Hence, it seems still reasonable to reject events with such high σ.

In total, these results show that σ is a good quality parameter for atmospheric neutrinos, whereas the performance on track-like events for an E^{-2} spectrum should still be improved. Maybe, increasing the number of training events, especially for high energies, leads to better results. Additionally, the efficiency of σ for track-like events cut be improved by training a network exclusively on this event type.

Application of cuts

Obviously, there is a large difference in the reconstruction and cut performance for shower-like and track-like events. Hence, it would be best to preselect the events and divide them according to their signature. This way, different σ-cuts, i.e. a more strict cut for track-like events, could be chosen. However, at this point there is no CNN-based track-shower classifier. Hence, for this work, a common cut for both signatures needs to be chosen. Therefore, this cut is adapted to the total neutrino set. The here demonstrated cuts are chosen in such a way that the 75 %, 68 %, 50 % and 25 % of events with the lowest σ are kept. The survival rate for each cut is listed for all different event types in table 5.4 for both atmospheric and E^{-2} spectrum.

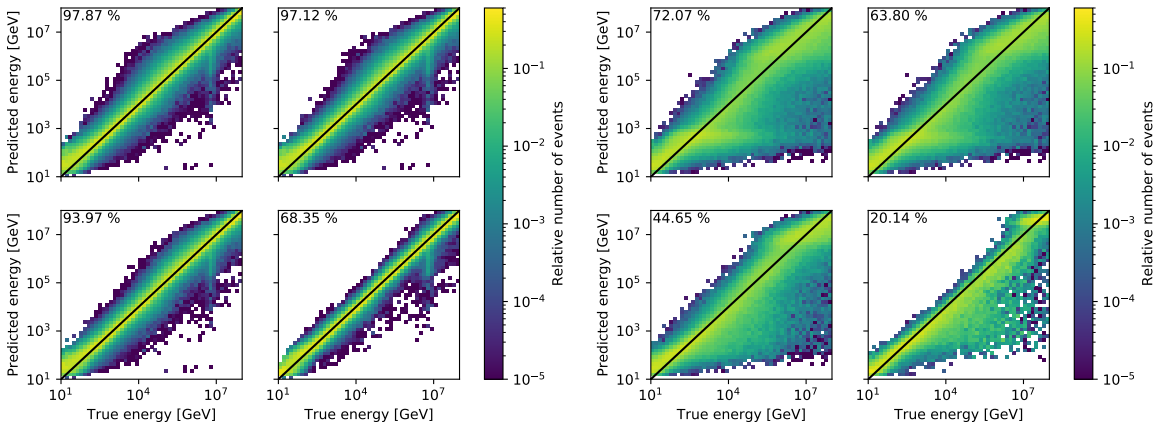
Event type	best 75 %	best 68 %	best 50 %	best 25 %
atmospheric				
all ν	75	68	50	25
track	72	64	45	20
shower	98	97	94	68
mu	97	94	68	1
$\mathbf{E^{-2}}$				
all ν	75	68	50	25
track	69	58	37	10
shower	99	98	94	74
mu	99	98	83	2

Table 5.4: Survival rate of the individual event types for different cuts on σ. Best x % refers to the x % of all neutrino events with the lowest σ.

In both cases, a high cut still leaves the major amount of shower-like events, whereas only a minor fraction of track-like events remains in the set. However, the absolute

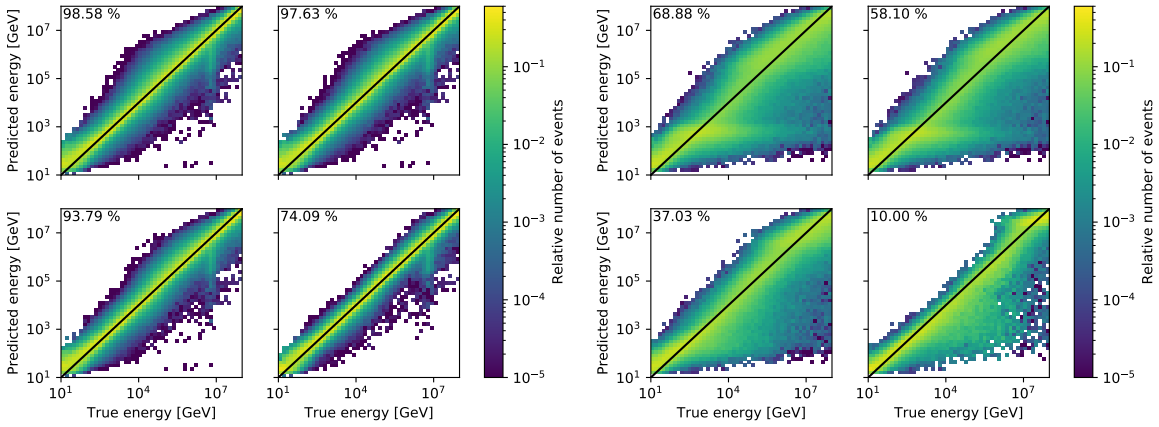
number of expected ν_μ CC events in ANTARES is still larger compared to cascades. It should be noted that a very high cut on σ decreases the amount of atmospheric muon events by two orders of magnitude without any further restrictions like an upwards reconstruction. However, this is of course not sufficient as background rejection. Still, it can maybe contribute to a more pure neutrino set.

Figure 5.7 presents the performance for the different cuts from table 5.4 for the different event signatures and neutrino spectra.



(a) Shower-like events (atmospheric)

(b) Track-like events (atmospheric)



(c) Shower-like events (E^{-2})

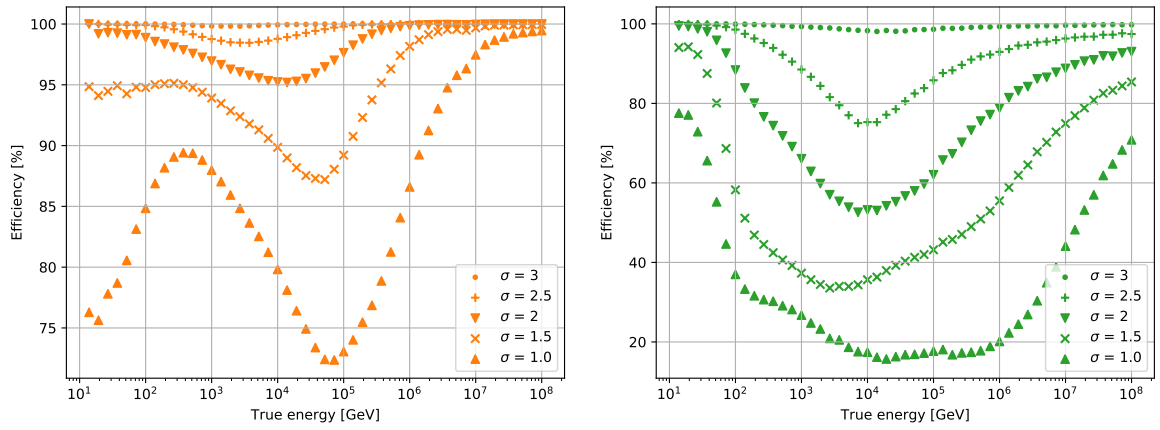
(d) Track-like events (E^{-2})

Figure 5.7: True and predicted energy for track- and shower-like events for an atmospheric and cosmic neutrino spectrum $\sim E^{-2}$ for different cuts on σ. The chosen cuts are the ones presented in table 5.4. The percentage in the upper left corner indicates the efficiency for the signature for the respective cut. All colorbars are limited to 10^{-5} and 0.6 and indicate the relative number of events per true energy bin.

Considering the shower-like events, even when only rejecting 32 % (25 %) for an

atmospheric (cosmic) spectrum, the variance is significantly smaller than for the complete shower set. Especially, the amount of samples with an overestimated energy can be greatly reduced by applying this cut. The underestimation seems to be reduced as well, even if not as successful. Particularly, the dominant excess of underestimation for the Glashow resonance still remains.
A large amount of events with an underestimated prediction remains also for track-like events. In this case, even when rejecting 80 % (90 %) for an atmospheric (cosmic) neutrino spectrum, still the amount of underestimation is very dominant. The overestimation is however nicely reduced. Still, even when applying the cut for maintaining the best 50 %, the variance can be greatly reduced for track-like events in most energy bins. This can be verified by analysing the 90 % quantiles of the median relative energy error for the pre-selected set (see figure A.14). The relative improvement is about $60 - 80$ % for energies above 50 GeV and below 100 TeV. For energies higher than 500 TeV it is only minor or the fluctuations even get larger. However, the median error itself also decreases and is mostly below 0.75. Similar to the results from contained events, the relative improvement is maximum around 100 GeV. For the remaining bins, the MRE is reduced about 20 %.
The improvement for shower-like events is less obvious as only a few percent of the events are discarded. Above 1 PeV, there is no improvement of the performance. However, at least a small reduction of median and variance can be stated for lower energies, with a maximum just above 10 TeV. In general, the relative improvement of the MRE is about $5 - 15$ %, whereas the 90 % quantile can be reduced up to 45 %.

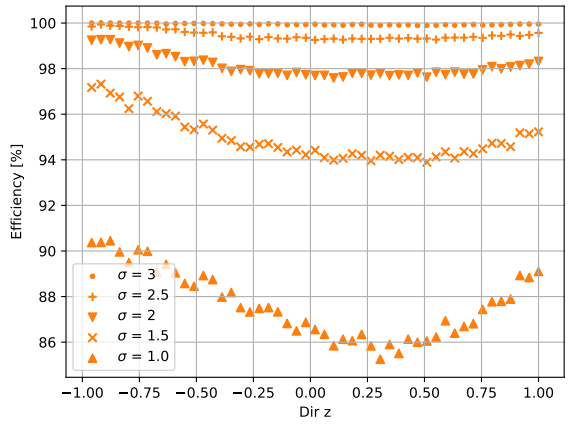
It has been shown, that applying a cut on σ leads to a better performance. However, it is still interesting what kind of neutrino events remain after such a pre-selection. Therefore, the efficiency is analysed in each true energy (see figure 5.8) and direction (see figure 5.9) bin. This analysis is done separately for track and shower-like events.



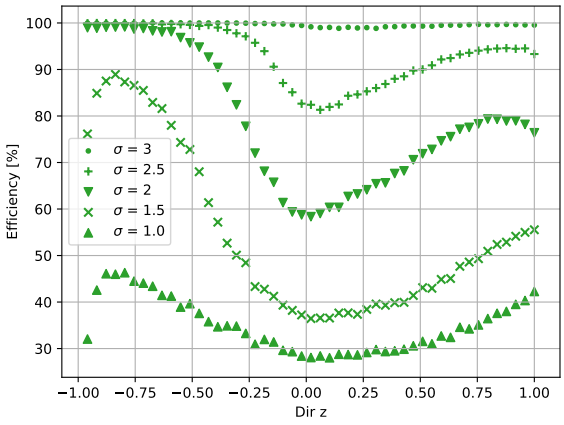
(a) Shower-like events (b) Track-like events

Figure 5.8: Reconstruction efficiency as a function of the true neutrino energy for different cuts on σ.

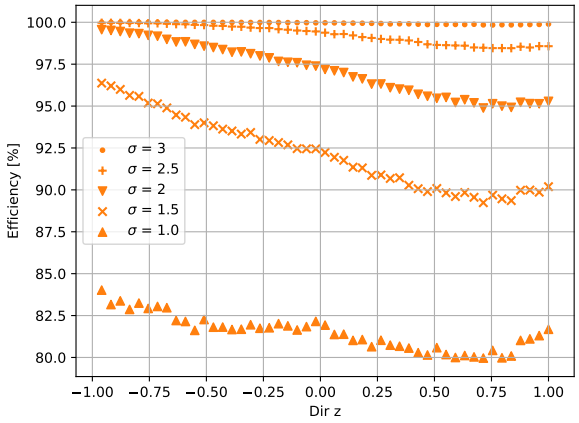
As can be seen from figure 5.8, the lowest efficiency, for track-like events lies in the medium energy range. The stricter the cut, the more the minimum efficiency shifts towards lower energies and the broader it gets. The best efficiency is achieved for the upper and lower boundaries of the energy spectrum, where the rejection is higher for high energy events.
In general, the efficiency is obviously much better for shower-like events. In this case, a stricter cut leads to a shift in the minimum efficiency towards higher energies. High energetic events show the best survival rate. Even for very strict cuts, the efficiency is above 85 % for energies above 1 PeV and even approximately 100 % for energies above 10 PeV.



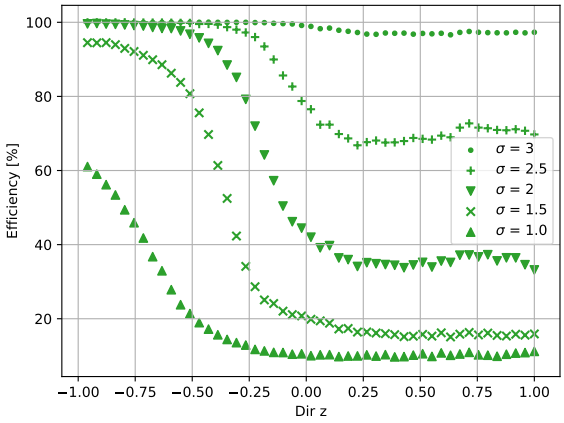
(a) Shower-like events (atmospheric)



(b) Track-like events (atmospheric)



(c) Shower-like events (E^{-2})



(d) Track-like events (E^{-2})

Figure 5.9: Reconstruction efficiency as a function of the true neutrino direction for different cuts on σ for an atmospheric ((a) and (b)) and an E^{-2} ((c) and (d)) spectrum. *dir z* denotes the cosine of the zenith angle, i.e. is equal to 1 (-1) for vertically up-going (down-going) events.

Considering the neutrino direction (see figure 5.9), the efficiency distribution for

track-like events depends on whether an atmospheric or an E^{-2} spectrum is used. However, in both cases, the efficiency is best for down-going events and starts to drop significantly close to the horizon. The efficiency then starts to increase again for atmospheric neutrinos, even though it stays below the efficiency of down-going events. In case of astrophysical neutrinos, the efficiency stays approximately constant for up-going events.
The direction seems to have less influence for shower-like events, where the efficiency is almost constant over the whole spectrum. However, there is also a small preference of down-going events.
This result is quite surprising. ANTARES is optimized for upwards going particles and most analyses exclude down-going particles in order to suppress atmospheric muons. Even the here introduced background classifier rejects most down-going particles. Hence, it is of course unfortunate if the performance on the remaining (mostly) up-going events is poor. However, when restricting the neutrino samples to events with contained interaction vertex, the efficiency becomes more independent of the neutrino direction. Probably, the biggest loss is caused by events that started very far away from the detector. This way, the efficiency depends more on the position of the interaction vertex than on the direction.

5.5.4 Comparison with standard reconstructions

A comparison with standard reconstructions is of course not ideal as most techniques focus on the reconstruction of either track-like or shower-like events and thus reject the not desired signature. This energy reconstruction aims for predicting the energy for all neutrino events. However, suppressing ν_μ CC events could be achieved by applying a rather high cut on σ.

The performance on track-like events is compared with the ANN estimator [45]. Therefore, an E^{-2} spectrum is used. There are two cuts applied for the standard reconstruction. The first one requires $\log(E_{ANN}) > 5$ and the second one aims for a high reconstruction quality $\Lambda \geq -5.3$. With these cuts, 14 % of all track events remain. No further restriction concerning the direction is done. In order to have a fair comparison, the same cuts are used for the deep learning reconstruction.
Figure 5.10 presents the median relative error (see equation 5.4) for both reconstruction algorithms for energies above 100 TeV.
Below a few PeV, the median relative error is smaller for the ANN reconstruction compared to deep learning. Above this energy, the performance is approximately competitive for both reconstructions. However, below 10 PeV, the 90 % quantiles are significantly larger for the CNN-based approach. Hence, even if the median error is slightly better for deep learning, still a quite high amount of worse reconstructed events has to be expected. Finally, above 20 PeV, the median resolution is slightly smaller for deep learning and the fluctuations are mostly also better. Only for the highest energies, i.e. close to 100 PeV, the performance is very similar again. Still, in total, it seems like the deep learning approach is better suited for the highest energetic events.

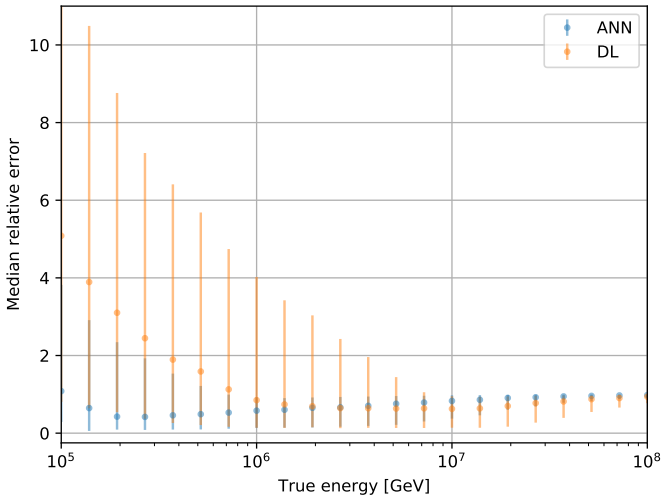


Figure 5.10: Median relative error for deep learning (orange) and the ANN reconstruction (blue). For this analysis, only events with $\log(E_{ANN}) > 5$ and $\Lambda \geq -5.3$ are used. The error bars indicate the 10 % and 90 % quantile.

Additionally, it should be mentioned that the here used cut was chosen such that it is optimized for the ANN estimator. Hence, the performance of the CNN-based approach could be improved with adapted cuts, e.g. on the reconstructed uncertainty σ.

Regarding the energy reconstruction of shower-like events in ANTARES, it is noticeable that classical algorithms tend to underestimate the shower energy [48, 49]. For this work, the results will be compared to the ones from [49]. However, it should be noted that this comparison is not completely fair, as the used dataset is different. Still, it is a good starting point.
Figure 5.11 depicts the ratio E_{MC}/E_{rec} as a function of the shower energy, i.e. the visible energy, for the classical reconstruction algorithm of [49]. The error bars indicate the 10 % and 90 % quantile respectively.
The same quantity is shown for the CNN-based approach in figure 5.12 for the complete shower set as well as for an applied cut, that aims for keeping the best 25 % of all neutrino events. This corresponds to about 75 % of all shower-like events. From this comparison it can be seen that the pre-selection mostly shifts the median ratio upwards. This way, the performance is improved for energies above 100 TeV, which is beneficial in the search of astrophysical neutrinos. Even though the median deviation between true and predicted energy increases slightly for energies between 1 and 100 TeV, the ratio still lies around 1 and the fluctuations are smaller.
As already mentioned, the classical algorithm tends to underestimate the shower energy over the whole energy range. For the deep learning approach this is only true for the highest energies. As for the classical approach, this can be attributed

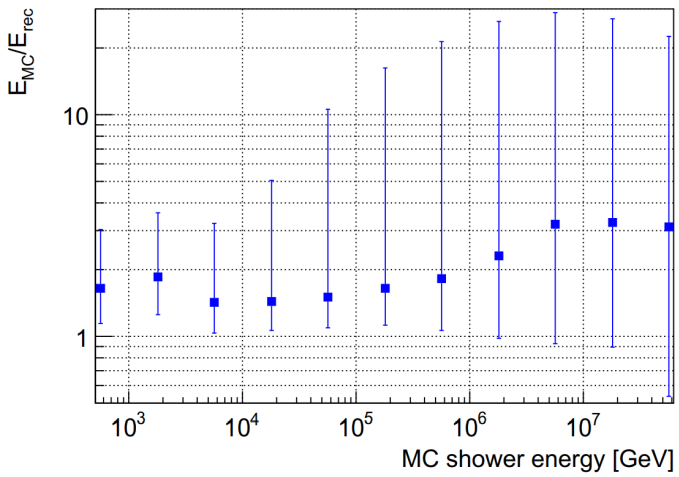


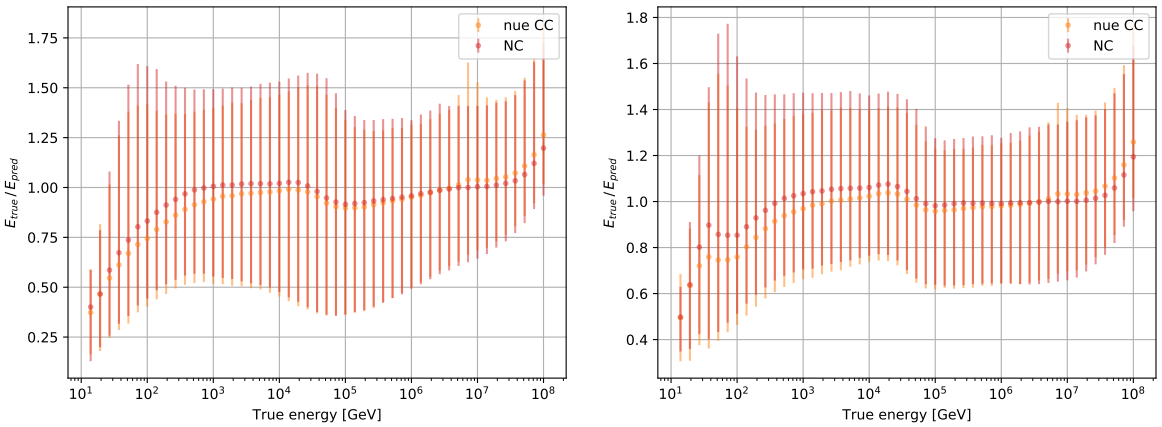
Figure 5.11: Median ratio of true and predicted shower energy for the classical shower reconstruction. The dots represent the median value in each bin. The error bars indicate the 10 % and 90 % quantile. [49]

to the saturation of the detector due to stronger light yield [49].
In general, the results from figures 5.11 and 5.12 suggest that the median ratio is closer to the optimum for the CNN-based method. Additionally, the fluctuations are much less extended and there is less bias towards a too small reconstructed energy.

In summary, the results suggest that following a deep learning based energy reconstruction could lead to an improved performance, especially concerning shower-like events. As already mentioned, it might be worth to test the performance for networks that are exclusively trained on one of the two interaction signatures. However, it should not be forgotten that the network is only trained on a six month period. As the results from the background classifier (see section 4.3.5) suggest, such a network can not inevitably be transferred to other periods which is a disadvantage compared to the classical shower reconstructions. Still, it seems like a deep learning based approach can contribute to an improved reconstruction.

5.5.5 Data

The performance on data is evaluated on the same runs as for the background classifier, i.e. September 2008 to February of 2009. As the energy reconstruction does not aim for the suppression of atmospheric muon events, it is necessary to do a preselection in the first place. Since the classical approaches do not aim for selecting track- and shower-like events, this is done with the background classifier. The number of expected and detected events for the different applied cuts that are discussed in the following are presented in table 5.5.
Since the expected number of cosmic neutrinos is very small for half a year, the first analysis is done for atmospheric neutrinos.



(a) All shower-like events (b) Pre-selected shower-like events

Figure 5.12: Ratio of true and predicted shower energy for the deep learning energy reconstruction of all shower-like events (left) and for a pre-selection after applying a cut on σ (right). Charged current component ν_e CC (orange) and neutral current NC component (red) are demonstrated separately. Markers represent the median value in each bin. The error bars indicate the 10 % and 90 % quantile.

Set	$N_{expected}$	$N_{detected}$
cut on p_ν	2810.02	2012
cut on σ	0.8×10^6	0.6×10^6
combined cuts	766.06	469

Table 5.5: Number of expected and detected atmospheric neutrinos after applying different cuts.

For this analysis, the Monte Carlo simulation set is reduced to events that are included in the test set of the background classifier and the energy reconstruction. This decreases the number of neutrino samples in the test set to about 1.7 million. The measured data runs are the same as for the background classifier. However, sparking and striking runs have been excluded.
The distribution of predicted energy and σ match the expectations quite well and can be found in appendix A.15. However, except for the highest energies, the amount of events in data with energies above 5 TeV is consistently smaller than expected. As the main part of events are atmospheric muons on which the CNN was never trained on, it is possible that the reconstructed energy for these events is not predictable that accurate. On the other hand, there are more events with a lower predicted energy, especially below 100 GeV. This could be a result of pure noise events.
The σ distribution of all events matches the expectation very well. The deviations in each bin are very small.

The first cut that is applied to the dataset is the probability cut for atmospheric neutrinos described in section 4.3.3. Figure 5.13 illustrates the expected and measured distributions of both predicted quantities, i.e. reconstructed energy and σ. The general shape of both distributions is as expected. However, as already pointed out in section 4.3.4, the number of events found in data is smaller than the expected one. Regarding the predicted energy, the tail of high energies seems to be more extended for data. However, this could be a result of the smaller statistics of atmospheric muon events in the simulations. These lead to the peaks in some bins for the Monte Carlo set rather than to a smearing as found for data.
The σ distribution shows the two expected peaks. However, the first peak around $\sigma \approx 0.5$ is significantly smaller for data. The second peak at $\sigma \approx 2$ is also smaller but the deviation is less severe. Hence, the deficit for events with a low uncertainty is larger.

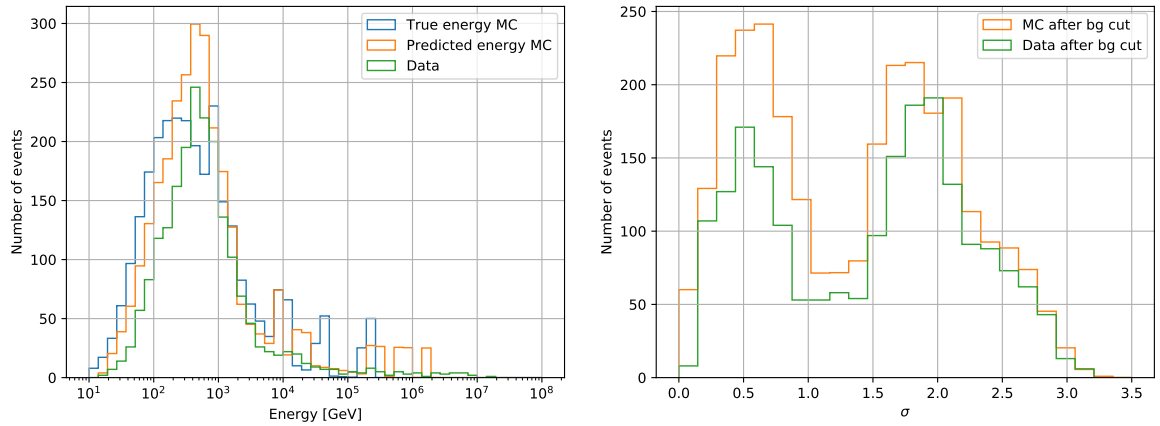


Figure 5.13: Distribution of the predicted quantities for the remaining events after a cut on the neutrino probability from the background classifier. *Left:* True and expected predicted energy for simulations and predicted energy for measured data. *Right:* Expected and predicted uncertainty parameter σ.

Finally, a cut on σ is applied, in order to maintain the best 25 % of all events (see table 5.4). The expected and measured distributions are presented in figure 5.14. The amount of events found in data is almost consistently smaller than expected. The number of expected events is 766.06 (including around 175 atmospheric muon events) but only 489 events are found in data. This is about 36 % less than expected. However, the reason for this seems to be the cut from the background classifier that already showed about 25 % less events and the deficit of events with low σ. The reduction of neutrino events from the selection of the background classifier towards the selection of an additional cut on σ is expected to be reduced by a factor of 0.24

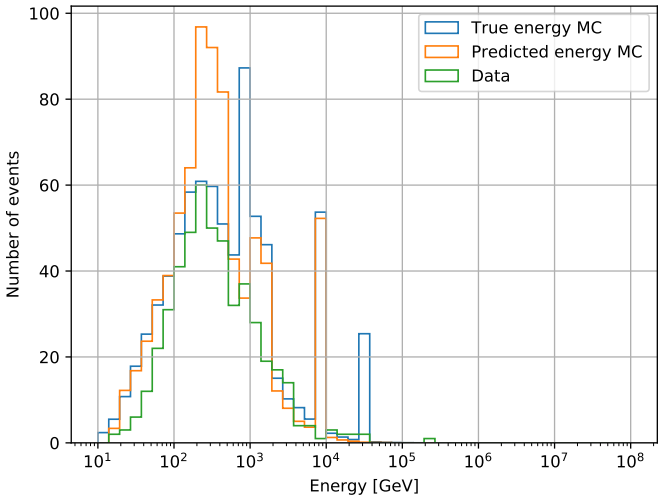


Figure 5.14: Energy distribution for the remaining events after a cut on the neutrino probability from the background classifier and an additional cut on σ.

which is approximately the reduction that is expected with this cut. Unfortunately, the atmospheric muon events are only expected to be reduced by a factor of 0.47, leading to a higher contamination for the final selection. In general, the shape of the distribution for data resembles the expected one.
However, there are more events with a predicted energy above 10 TeV than expected which could be a result of the remaining atmospheric muons. This suspicion is enforced as all of these events are reconstructed as down-going. There is only one event with an energy above 100 TeV. Analysing the zt-projection of the corresponding image does not show a typical down-going track-like signature. Hence, it seems possible that this is indeed a neutrino event.

6 Summary and Outlook

Within the last years, many developments in machine learning and deep learning have been made. Therefore, it is not surprising that today, these techniques play an important role in physics analyses. There is a broad spectrum of applications in particle physics. An overview about modern machine learning applications in this field can be found here [50].
These methods are also very prominent for neutrino detectors, including neutrino telescopes used for neutrino astronomy. Recently, reports of CNN-based reconstructions from KM3NeT as well as IceCube have been published [38, 51].
This work presented the first application of supervised deep learning with convolutional neural networks for the ANTARES detector. Preprocessing and event selection have been demonstrated and the results of a background classifier and an energy reconstruction as examples for classification and regression have been reported.

For the background classifier, a comprehensive analysis on Monte Carlo as well as recorded data has been demonstrated. This network is in general able to extract neutrino events from unseen datasets. However, the performance in the energy and direction space of atmospheric muons is poor. Only a few neutrino events remain in the final selection. The efficiency for up-going events, on the other hand, is significantly better. Hence, the performance on neutrinos from below could be improved by an up-down classifier. However, the advantage of the background classifier is that high energetic down-going events are not rejected.
Additionally, the network was applied to measured data from the period it was trained on, which is about 127 days of lifetime. In order to have an acceptable statistics, a search for atmospheric neutrinos has been performed. Even though about 25 % less events than expected were found, the agreement between Monte Carlo and data is fine. Additionally, it has been shown that the classifier yields the potential to detect irregular runs.
Furthermore, a small detector condition study has been performed. Therefore, the network has been applied to several months of simulations from different episodes than it was trained on. The results confirm that the here trained network can not inevitably be used for other periods. At least, the chosen cuts have to be redefined but even that does not guarantee for a good performance.
Finally, a second background classifier was trained on runs from a different episode in order to investigate if this improves the performance. This analysis showed that training a network on a connected period close to its application time does not necessarily lead to a better efficiency. Rather, it seems that external parameters like baseline and burst fraction, that quantify the noise situation in the detector, determine the performance. These observations suggest that splitting runs into several

categories, defined by their corresponding external parameters, could be a better suited approach.

For regression, the performance of an energy reconstruction for both event signatures, i.e. track- and shower-like, has been investigated. This CNN does not only reconstruct the energy but also an additional uncertainty parameter that can be used to reject badly reconstructed events. Both predicted quantities show promising results for shower-like events. A small comparison with results from a classical reconstruction suggests that this deep learning approach is competitive with standard reconstructions.
However, the performance on track-like events with regard to the search for cosmic neutrinos should be improved. Some of the results from this work indicate that this could already be achieved by training the network separately on the different event signatures. This is of course also beneficial for the reconstruction of shower-like events. Additional improvements could be expected by restricting the training set to contained events and balancing it over the energy range.
The application to data shows in general a good agreement for the final selection of atmospheric neutrinos obtained from the background classifier. However, there seem to be less events with a low reconstructed uncertainty as expected. Only one event with an energy larger than 100 TeV was found.

A nice bonus of using CNNs are the input images as they can be used easily for a by-eye analysis. This way, potential neutrino candidates can be displayed in an intuitive manner. Additionally, this approach has already been used in order to investigate sparking and other remarkable runs.

In general, this work has demonstrated the powerful potential of CNNs for ANTARES. Results along with performance deficits that were encountered could be explained. This way, the frequent argument of artificial neural networks being black boxes, can be mitigated. However, the performance of such networks should always be thoroughly questioned in order to prevent them from learning undesired or wrong features.

In the future, still some systematic investigations have to be done. This includes a more detailed sensitivity analysis that has so far only been done for the background classifier. The influence of different conditions on the energy reconstruction is still completely unclear. In addition, the robustness against changing data taking conditions and other difficulties like poor calibrations and misalignment needs to be analysed.
Furthermore, it should be mentioned that the here shown approach is relatively simple and can easily be developed further. The input information is for example quite sparse, only containing spatial and temporal information of the hits. More details on the exact PMT orientation, missing OMs and charge information of a hit could be included, potentially leading to an improved reconstruction. Of course, this could also be realised by other deep learning approaches.

In summary, this work has shown promising results for a CNN-based classification and regression for the ANTARES detector. These indicate that it is definitely worth to follow the approach and that deep learning can contribute a lot in the future.

7 Bibliography

[1] W. Pauli. "Pauli letter collection: letter to Lise Meitner". Typed copy. URL: http://cds.cern.ch/record/83282.

[2] C. L. Cowan, F. Reines, et al. "Detection of the Free Neutrino: a Confirmation". In: *Science* 124.3212 (1956), pp. 103–104. ISSN: 0036-8075. DOI: 10.1126/science.124.3212.103. eprint: https://science.sciencemag.org/content/124/3212/103.full.pdf. URL: https://science.sciencemag.org/content/124/3212/103.

[3] J. N. Bahcall and R. Davis. "Solar Neutrinos: A Scientific Puzzle". In: *Science* 191.4224 (1976), pp. 264–267. ISSN: 0036-8075. DOI: 10.1126/science.191.4224.264. eprint: https://science.sciencemag.org/content/191/4224/264.full.pdf. URL: https://science.sciencemag.org/content/191/4224/264.

[4] K. Hirata, T. Kajita, et al. "Observation of a neutrino burst from the supernova SN1987A". In: *Phys. Rev. Lett.* 58 (14 1987), pp. 1490–1493. DOI: 10.1103/PhysRevLett.58.1490. URL: https://link.aps.org/doi/10.1103/PhysRevLett.58.1490.

[5] R. M. Bionta, G. Blewitt, et al. "Observation of a neutrino burst in coincidence with supernova 1987A in the Large Magellanic Cloud". In: *Phys. Rev. Lett.* 58 (14 1987), pp. 1494–1496. DOI: 10.1103/PhysRevLett.58.1494. URL: https://link.aps.org/doi/10.1103/PhysRevLett.58.1494.

[6] U. Katz and C. Spiering. "High-energy neutrino astrophysics: Status and perspectives". In: *Progress in Particle and Nuclear Physics* 67.3 (2012), 651–704. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2011.12.001. URL: http://dx.doi.org/10.1016/j.ppnp.2011.12.001.

[7] L. Telescope, M. Aartsen, et al. "Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A". In: *Science* 361.6398 (2018).

[8] M. Aartsen, M. Ackermann, et al. "Neutrino emission from the direction of the blazar TXS 0506+ 056 prior to the IceCube-170922A alert". In: *Science* 361.6398 (2018), pp. 147–151.

[9] R. Stein, A. Franckowiak, et al. "IceCube-191001A: Candidate Counterparts with the Zwicky Transient Facility". In: *GRB Coordinates Network* 25929 (Oct. 2019), p. 1.

[10] R. Stein, S. van Velzen, et al. "A tidal disruption event coincident with a high-energy neutrino". In: *Nature Astronomy* (2021), pp. 1–9.

[11] H. Athar, M. Jeżabek, et al. "Effects of neutrino mixing on high-energy cosmic neutrino flux". In: *Physical Review D* 62.10 (2000). ISSN: 1089-4918. DOI: 10.1103/physrevd.62.103007. URL: http://dx.doi.org/10.1103/PhysRevD.62.103007.

[12] J. K. Becker. "High-energy neutrinos in the context of multimessenger astrophysics". In: *Physics Reports* 458.4-5 (2008), 173–246. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2007.10.006. URL: http://dx.doi.org/10.1016/j.physrep.2007.10.006.

[13] M. Honda, M. S. Athar, et al. "Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model". In: *Phys. Rev. D* 92 (2 2015), p. 023004. DOI: 10.1103/PhysRevD.92.023004. URL: https://link.aps.org/doi/10.1103/PhysRevD.92.023004.

[14] J. Tiffenberg. "UHE Neutrino searches with the Pierre Auger Observatory [Conference Talk]". NUSKY 2011. 2011. URL: http://users.ictp.it/~smr2246/monday/tiffenberg-NUSKY.pdf.

[15] T. K. Gaisser, R. Engel, et al. *Cosmic Rays and Particle Physics*. 2nd ed. Cambridge University Press, 2016. DOI: 10.1017/CBO9781139192194.

[16] R. Gandhi, C. Quigg, et al. "Ultrahigh-energy neutrino interactions". In: *Astroparticle Physics* 5.2 (1996), 81–110. ISSN: 0927-6505. DOI: 10.1016/0927-6505(96)00008-4. URL: http://dx.doi.org/10.1016/0927-6505(96)00008-4.

[17] P. Cherenkov. "Visible luminescence of pure liquids under the influence of γ-radiation". In: *Dokl. Akad. Nauk SSSR* 2.8 (1934), pp. 451–454. DOI: 10.3367/UFNr.0093.196710n.0385.

[18] H. Kolanoski and N. Wermes. *Teilchendetektoren*. Springer, 2016.

[19] M. de Naurois and D. Mazin. "Ground-based detectors in very-high-energy gamma-ray astronomy". In: *Comptes Rendus Physique* 16.6-7 (2015), 610–627. ISSN: 1631-0705. DOI: 10.1016/j.crhy.2015.08.011. URL: http://dx.doi.org/10.1016/j.crhy.2015.08.011.

[20] M. Ageron, J. Aguilar, et al. "ANTARES: the first undersea neutrino telescope". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 656.1 (2011), pp. 11–38.

[21] A. Collaboration and S. Escoffier. *The ANTARES detector: background sources and effects on detector performance*. 2007. arXiv: 0710.0527 [astro-ph].

[22] S Escoffier. "Bioluminescence studies with the ANTARES Prototype Sector Line". In: (2005). URL: https://cds.cern.ch/record/951270.

[23] G Carminati, M Bazzotti, et al. "MUPAGE: a fast atmospheric MUon GEnerator for neutrino telescopes based on PArametric formulas". In: *arXiv preprint arXiv:0907.5563* (2009).

[24] A Albert, M André, et al. “Monte Carlo simulations for the ANTARES underwater neutrino telescope”. In: *Journal of Cosmology and Astroparticle Physics* 2021.01 (2021), p. 064.

[25] *SimpleNeuralNetwork*. URL: https://github.com/snives/SimpleNeuralNetwork (visited on 01/05/2021).

[26] Stanford University. *CS231n Convolutional Neural Networks for Visual Recognition*. URL: https://cs231n.github.io/neural-networks-1/.

[27] G. Cybenko. “Approximation by superpositions of a sigmoidal function”. In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314.

[28] K. Hornik, M. Stinchcombe, et al. “Multilayer feedforward networks are universal approximators”. In: *Neural networks* 2.5 (1989), pp. 359–366.

[29] L. N. Smith. *Cyclical Learning Rates for Training Neural Networks*. 2017. arXiv: 1506.01186 [cs.CV].

[30] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[31] X. Glorot and Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Y. W. Teh and M. Titterington. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: JMLR Workshop and Conference Proceedings, 2010, pp. 249–256. URL: http://proceedings.mlr.press/v9/glorot10a.html.

[32] K. He, X. Zhang, et al. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. 2015. arXiv: 1502.01852 [cs.CV].

[33] N. Srivastava, G. Hinton, et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[34] S. Ioffe and C. Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG].

[35] K. N. S. Chauhan. *Introduction to Convolutional Neural Networks*. URL: https://www.kdnuggets.com/2020/06/introduction-convolutional-neural-networks.html (visited on 02/01/2021).

[36] Stanford University. *CS231n Convolutional Neural Networks for Visual Recognition*. URL: https://cs231n.github.io/convolutional-networks/#norm.

[37] O. Russakovsky, J. Deng, et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252. DOI: 10.1007/s11263-015-0816-y.

[38] M. Moser. “Sensitivity studies on tau neutrino appearance with KM3NeT/ORCA using Deep Learning Techniques”. PhD thesis. Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2020.

[39] Tamás Gál, Moritz Lotze. *km3pipe*. URL: https://git.km3net.de/km3py/km3pipe.

[40] Stefan Reck, Michael Moser. *OrcaSong*. URL: https://git.km3net.de/ml/OrcaSong.

[41] P. Coyle. "The ANTARES Deep-Sea Neutrino Telescope: Status and First Results". In: (Feb. 2010). arXiv: 1002.0754 [astro-ph.HE].

[42] Stefan Reck, Michael Moser. *OrcaNet*. URL: https://git.km3net.de/ml/OrcaNet.

[43] Erlangen National High Performance Computing Center. *NHR FAU*. URL: https://hpc.fau.de/.

[44] A. Heijboer. "Track reconstruction and point source searches with ANTARES". PhD thesis. Amsterdam U., 2004.

[45] J. Schnabel. "Muon energy reconstruction in the ANTARES detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 725 (2013), pp. 106–109. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2012.12.109. URL: https://www.sciencedirect.com/science/article/pii/S0168900212016658.

[46] Philips. *Photomultiplier Tubes: Principles and Applications*. Philips Photonics, 1994.

[47] S. L. Glashow. "Resonant Scattering of Antineutrinos". In: *Phys. Rev.* 118 (1 1960), pp. 316–317. DOI: 10.1103/PhysRev.118.316. URL: https://link.aps.org/doi/10.1103/PhysRev.118.316.

[48] A Albert, M. André, et al. "An algorithm for the reconstruction of neutrino-induced showers in the ANTARES neutrino telescope". In: *The Astronomical Journal* 154.6 (2017), p. 275.

[49] A. Albert, M. André, et al. "An algorithm for the reconstruction of high-energy neutrino-induced particle showers and its application to the ANTARES neutrino telescope". In: *The European Physical Journal C* 77.6 (2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-4979-2. URL: http://dx.doi.org/10.1140/epjc/s10052-017-4979-2.

[50] HEP ML Community. *A Living Review of Machine Learning for Particle Physics*. URL: https://iml-wg.github.io/HEPML-LivingReview/.

[51] R Abbasi, M Ackermann, et al. "A Convolutional Neural Network based Cascade Reconstruction for the IceCube Neutrino Observatory". In: *arXiv preprint arXiv:2101.11589* (2021).

List of Figures

1.1 Measured and expected energy spectra of neutrinos from different sources . . . 8
1.2 Illustration of neutrino-nucleon interactions . . . 11
1.3 Mean value of the Bjorken scaling variable . . . 13
1.4 Illustration of Cherenkov radiation . . . 13
1.5 Detector layout of ANTARES . . . 15

2.1 Illustration of the mathematical operation performed by a single neuron 18
2.2 Deep feed forward neural network with two hidden layers . . . 18
2.3 Schematic illsutration of droput . . . 25
2.4 Illustration of the operation performed by a convolutional layer . . . 27
2.5 Illustration of the operation performed by a max pooling layer . . . 28

3.1 PMT positions and spatial xy-binning . . . 30
3.2 Time distribution of signal hits relative to the first triggered hit and chosen timecuts . . . 31
3.3 Two dimensional projections of an exemplary input image . . . 32
3.4 Internal and external parameters for runs 30000 to 80000 . . . 35

4.1 Loss and accuracy of the background classifier . . . 39
4.2 Neutrino probability distribution . . . 40
4.3 Neutrino probability distribution separately for the different neutrino flavours and interaction types . . . 40
4.4 Neutrino probability distribution dependent on the neutrino energy and neutrino direction . . . 41
4.5 Muon probability distribution dependent on the muon energy and muon direction . . . 42
4.6 Muon contamination and neutrino efficiency for different values of p_ν 43
4.7 Neutrino efficiency for a fixed p_{cut} over energy and direction spectrum for the different event signatures . . . 44
4.8 Neutrino efficiency for a fixed p_{cut} combined over energy and direction spectrum . . . 46
4.9 Energy and direction distribution of all muon and selected muon events 47
4.10 Neutrino efficiency for a fixed p_{cut} over energy and direction spectrum for the different event signatures for a neutrino spectrum $\sim E^{-2}$ . . . 48
4.11 Neutrino probability distribution for Monte Carlo simulations and all data events . . . 51
4.12 Reconstructed energy and direction distribution for all data events and Monte Carlo simulations . . . 52

4.13 Reconstructed energy and direction distribution for data and Monte Carlo simulation events that are selected after applying p_{cut} 53
4.14 Run-wise neutrino probability distribution for October 2008 54
4.15 Ratio R for all runs from October 2008 55
4.16 Reconstructed energy and direction distribution for officially confirmed sparking runs and striking runs 56
4.17 Reconstructed energy and direction distribution for the final run sets 58
4.18 Neutrino probability distribution for different test episodes 60
4.19 Muon contamination and neutrino efficiency for different values of p_ν on the different test sets 61
4.20 Loss and accuracy of the 2010 background classifier 63
4.21 Comparison of muon contamination and neutrino efficiency of both background classifiers 64
4.22 Run-wise neutrino efficiency for both background classifiers 65
4.23 Run-wise neutrino efficiency for both background classifiers for a neutrino spectrum $\sim E^{-2}$ 66
4.24 Bin-wise neutrino efficiency for different parameters 68
4.25 Two dimensional bin-wise neutrino efficiency for different parameters 69

5.1 True and predicted energy for all different event types 75
5.2 Median energy error for all events 77
5.3 True and predicted energy for all contained events 78
5.4 Median energy error for all contained events 79
5.5 σ distribution of the different neutrino signatures 80
5.6 Median ratio of true and predicted energy as a function of σ 81
5.7 True and predicted energy for different cuts on σ 83
5.8 Reconstruction efficiency as a function of the true neutrino energy for different cuts on σ 84
5.9 Reconstruction efficiency as a function of the true neutrino direction for different cuts on σ 85
5.10 Median relative error for two deep learning and the ANN reconstruction 87
5.11 Ratio of true and predicted shower energy for the classical shower reconstruction 88
5.12 Ratio of true and predicted shower energy for the deep learning energy reconstruction 89
5.13 Distribution of the predicted quantities for the remaining events after a cut on the neutrino probability 90
5.14 Energy distribution for the final selection 91

A.1 Neutrino energy and direction distribution of the training set for the background classifier i
A.2 Neutrino probability distribution dependant on the neutrino energy and direction for the different neutrino signatures ii
A.3 Neutrino efficiency for a fixed p_{cut} combined over energy and direction spectrum for the different event signatures for an E^{-2} spectrum . . . ii

A.4 Ratio between data and Monte Carlo for reconstructed energy and direction of all events . . . iii
A.5 Run-wise neutrino probability distribution for the given months . . . iv
A.6 Ratio R for all runs from the given months . . . v
A.7 zt- and xy-projection of exemplary sparking and striking runs . . . vi
A.8 Reconstructed energy and direction distribution for the final selection of neutrino events . . . vii
A.9 Neutrino probability distribution for both networks . . . vii
A.10 Bin-wise neutrino efficiency for different parameters for a neutrino spectrum $\sim E^{-2}$ . . . viii
A.11 Two dimensional bin-wise neutrino efficiency for different parameters for a neutrino spectrum $\sim E^{-2}$ . . . ix
A.12 Visible neutrino energy and direction distribution of the training set for the energy reconstruction . . . x
A.13 True and predicted energy from a network not using the visible energy for shower-like event types . . . xi
A.14 Median energy error for pre-selected events . . . xi
A.15 Distribution of the predicted quantities for all events . . . xii

List of Tables

3.1 General architecture of a convolutional block 36

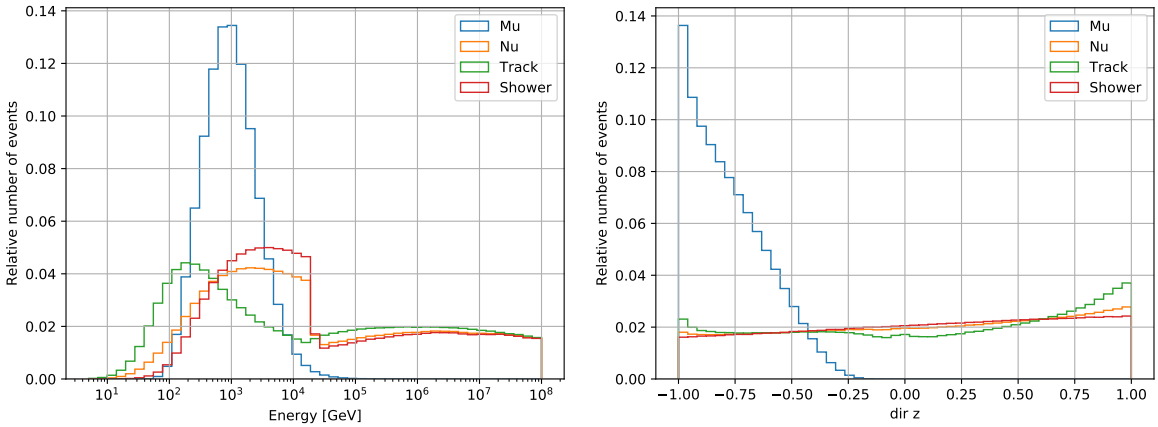
4.1 Number of events per event type in the training set for the background classifier 37
4.2 Network architecture of the background classifier 38
4.3 Number of expected background events and neutrino efficiency for deep learning and a standard reconstruction 49
4.4 Number of expected and detected atmospheric neutrinos 57
4.5 Loss and accuracy for different episodes 59
4.6 Number of atmospheric muons and neutrinos and signal efficiency for all neutrinos of the different episodes 61
4.7 Number of events for each event type in the training set for the background classifier trained on runs from 2010 62

5.1 Number of events per event type in the training set for the energy reconstruction 71
5.2 Network architecture for the energy reconstruction 73
5.3 Loss and mean squared error of the test set for the different event types 73
5.4 Survival rate of the individual event types for different cuts on σ . . . 82
5.5 Number of expected and detected atmospheric neutrinos after applying different cuts 89

A.1 Event type composition for the different test sets iii

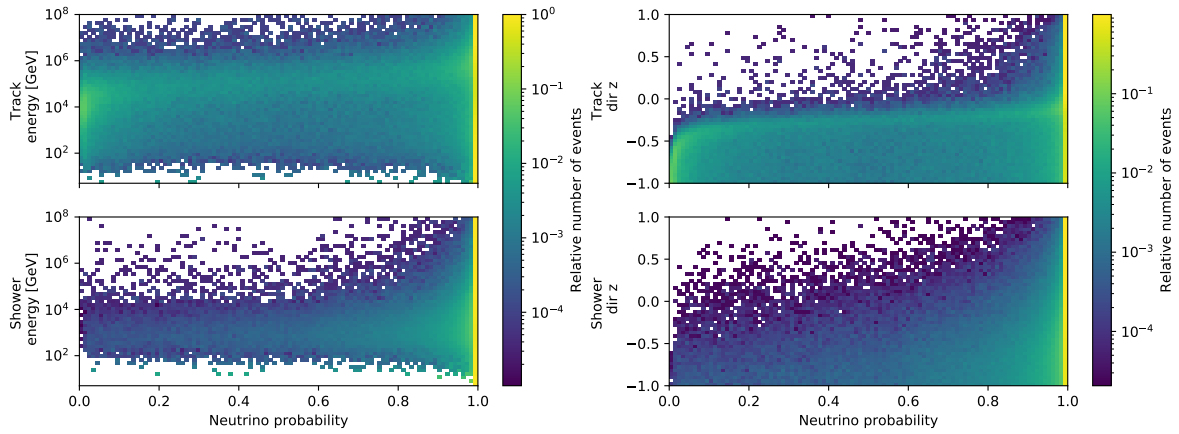
A Appendix

A.1 Background classifier



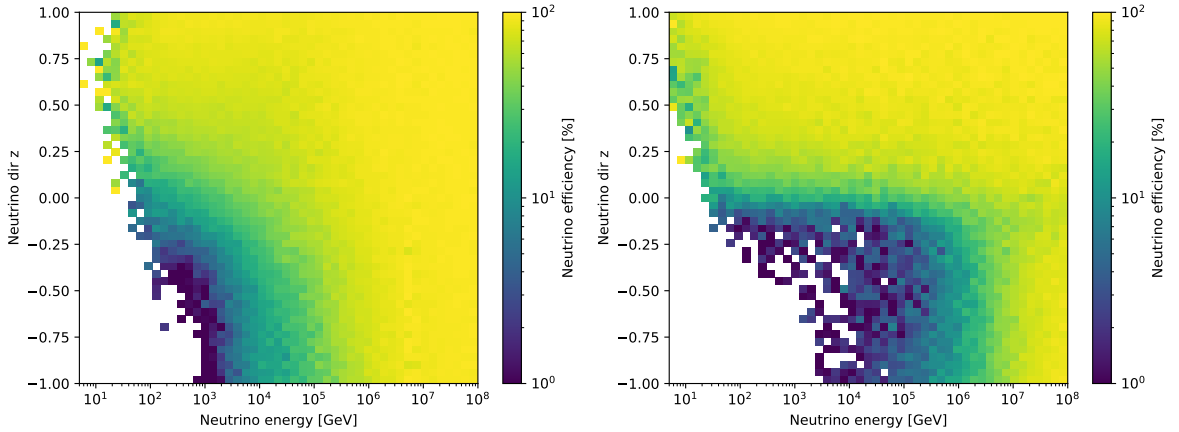
(a) Energy (b) Direction

Figure A.1: Neutrino energy and direction distribution of the training set for the background classifier. The distribution is demonstrated separately for the different events types, i.e. atmospheric muons *Mu* (blue), all neutrinos *Nu* (orange) and separately for track-like (green) and shower-like (red) neutrino events. The number is normalised per event type.



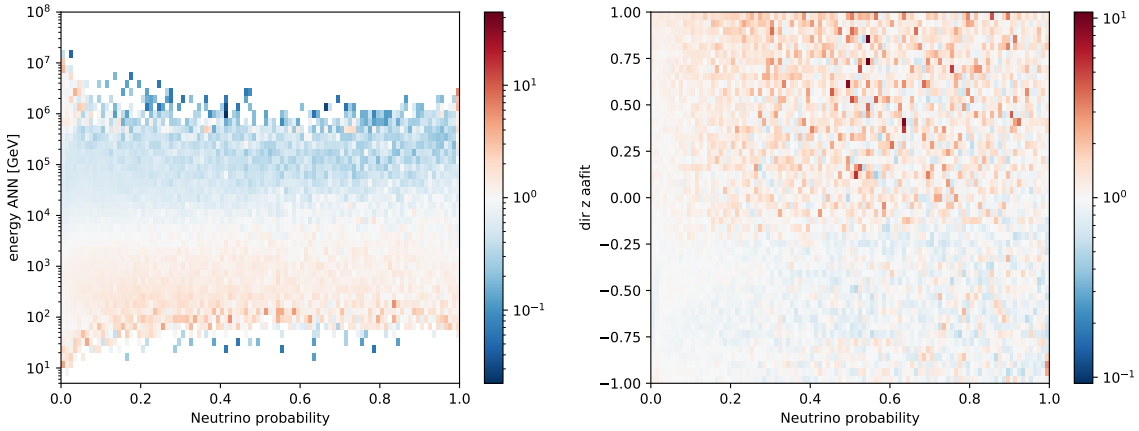
(a) Energy (b) Direction

Figure A.2: Neutrino probability distribution dependant on the neutrino energy (left) and direction (right) for track-like (upper) and shower-like (below) events. The number is normed per energy and direction bin, respectively.



(a) Shower-like events (b) Track-like events

Figure A.3: Neutrino efficiency for a fixed p_{cut} combined over energy and direction spectrum, split for shower-like (left) and track-like (right) events for an E^{-2} spectrum. The probability cut is chosen such that approximately one atmospheric muon per day survives ($p_\nu > 0.999985$).

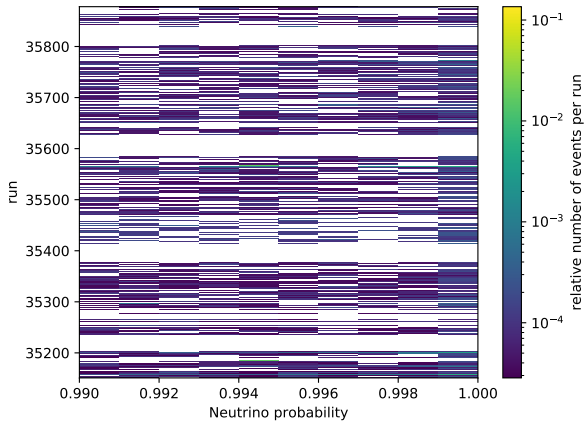


(a) Energy (b) Direction

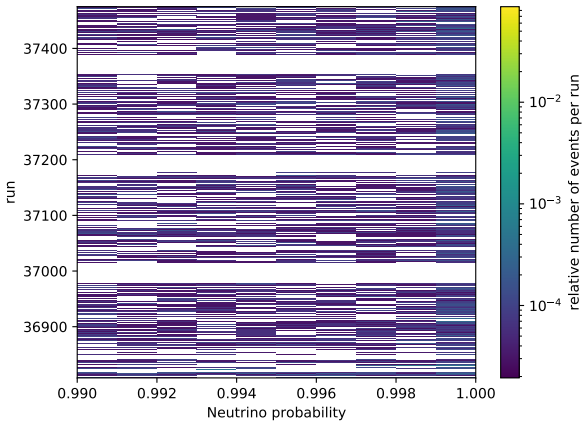
Figure A.4: Ratio $N_{\text{data}}/N_{\text{MC}}$ for reconstructed energy and direction of all events. N indicates the relative amount of events at the given neutrino probability. For MC, only bins for which at least one event is expected are used.

Event type	Number of events	Fraction [%]
July 2010		
Atmospheric μ	2 362 594	50.8
ν_e CC	787 670	17.0
ν_e NC	457 454	9.8
ν_μ CC	585 619	12.6
ν_μ NC	455 482	9.8
February 2012		
Atmospheric μ	2 516 719	47.6
ν_e CC	940 621	17.8
ν_e NC	543 258	10.3
ν_μ CC	737 018	13.9
ν_μ NC	550 026	10.4
September 2014		
Atmospheric μ	1 825 240	73.1
ν_e CC	237 020	9.5
ν_e NC	136 921	6.5
ν_μ CC	160 576	6.4
ν_μ NC	137 570	5.5

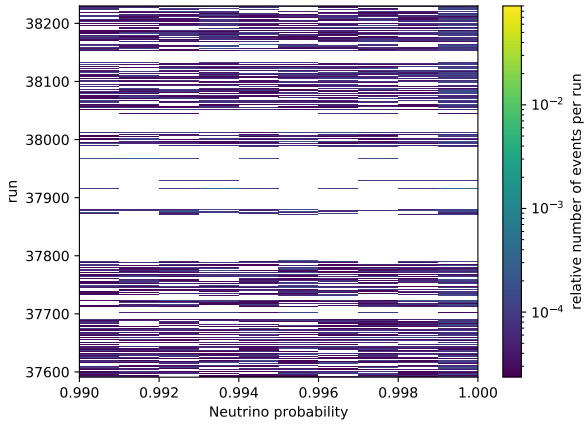
Table A.1: Number of events for each event type and percentage with respect to the respective complete test datasets of July 2010, February 2012 and September 2014 for the background classifier. For neutrinos, the event type includes particles and antiparticles in approximately equal parts.



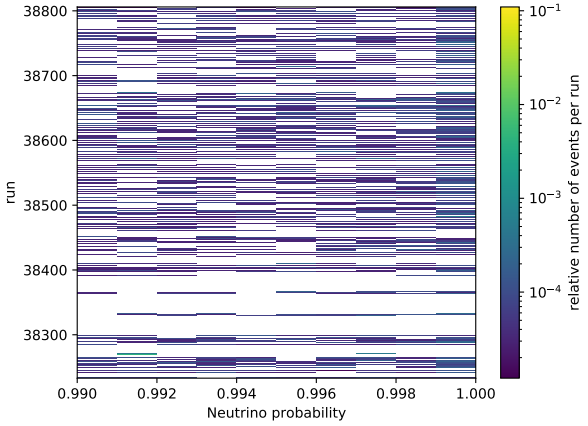
(a) September 2008



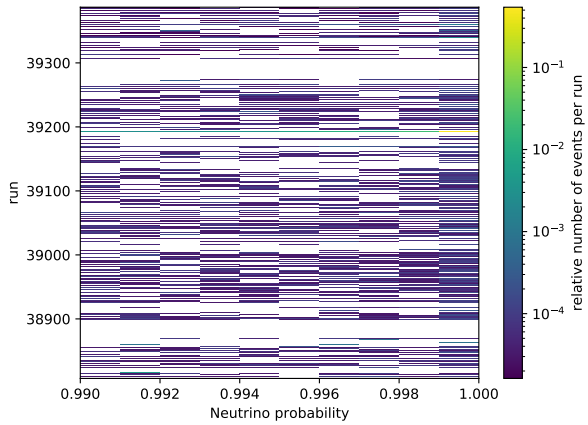
(b) November 2008



(c) December 2008

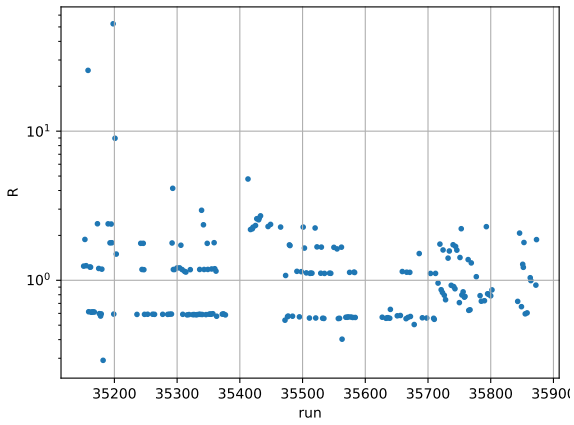


(d) January 2009

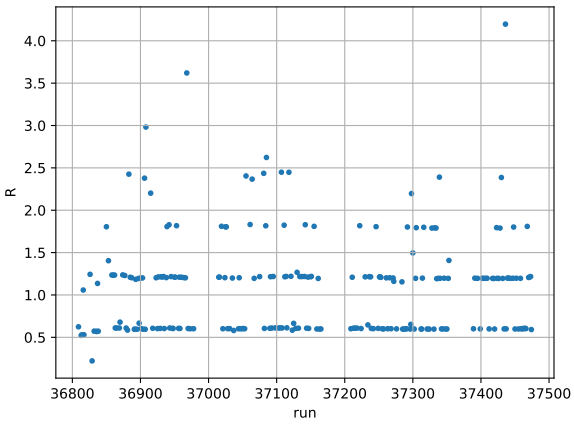


(e) February 2009

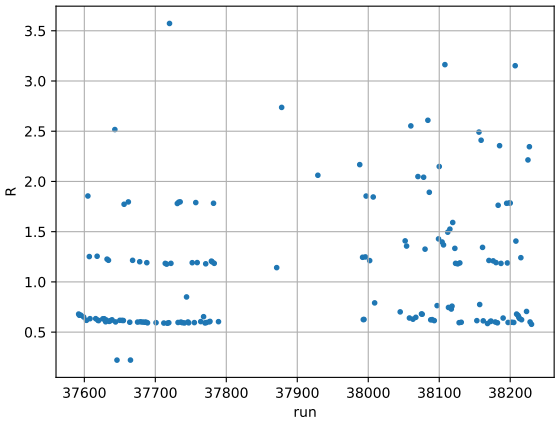
Figure A.5: Neutrino probability distribution for the individual runs from September 2008 to February 2009.



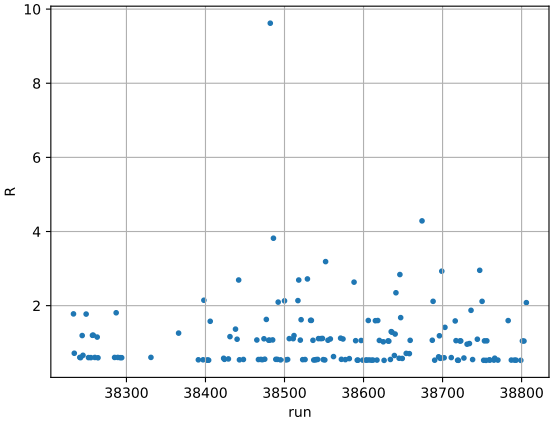
(a) September 2008



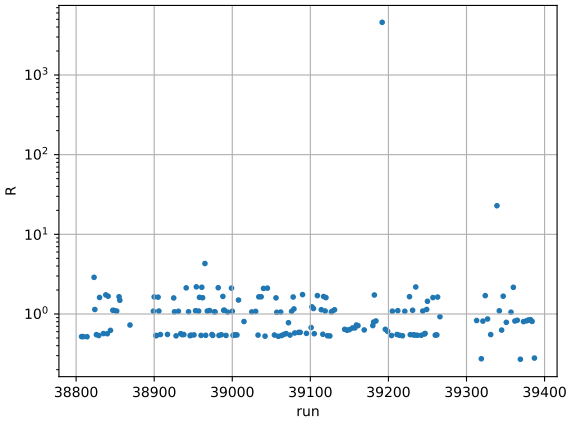
(b) November 2008



(c) December 2008

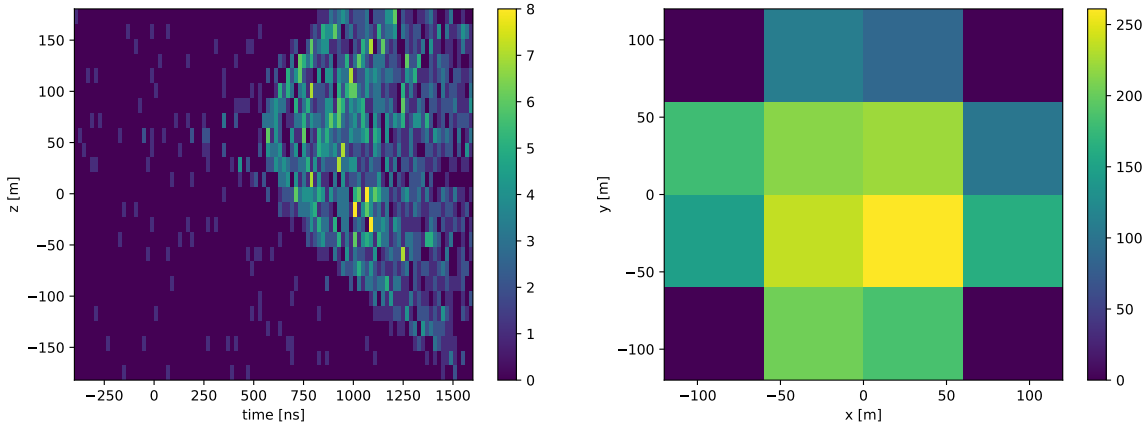


(d) January 2009

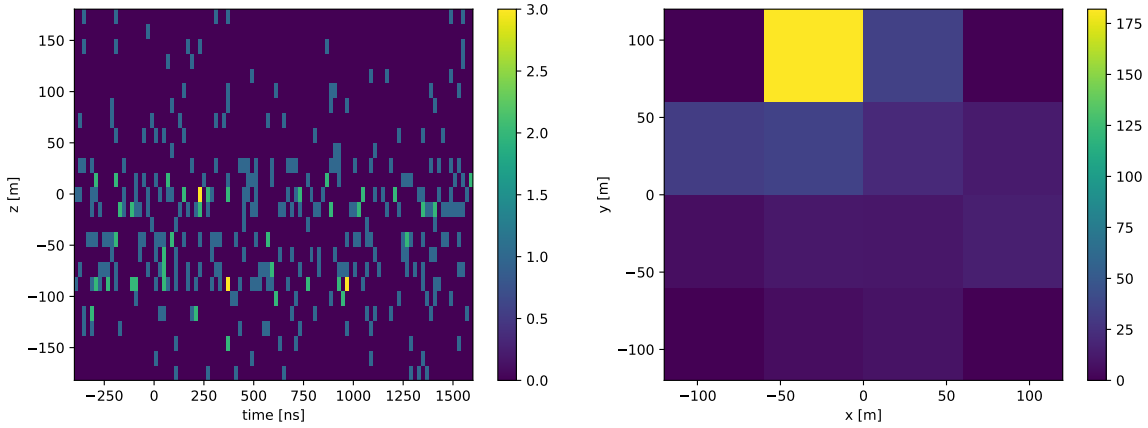


(e) February 2009

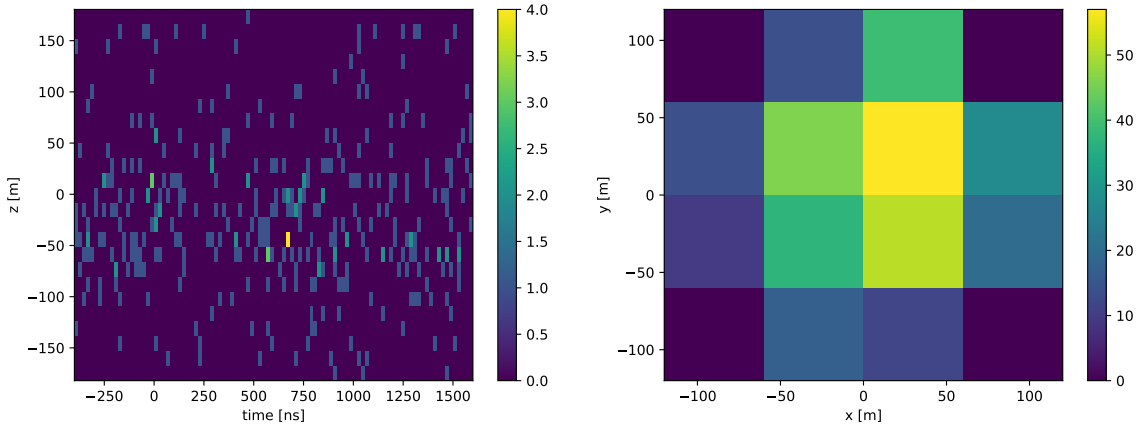
Figure A.6: Ratio R for all runs from the given months. Remarkable high rates can be found for runs 35158, 35198, 35201 (September 2008), 38482 (January 2009) and 39192, 39339 (February 2009).



(a) Run 36666, event 9, triggercounter 2384, $E_{reco} = 6.09\,\mathrm{PeV}$, dir z = -0.84



(b) Run 36600, event 8814, triggercounter 11811, $E_{reco} = 2.2\,\mathrm{TeV}$, dir z = -0.69



(c) Run 35158, event 54128, triggercounter 2209, $E_{reco} = 0.4\,\mathrm{TeV}$, dir z = 0.26

Figure A.7: zt- (left) and xy-projection (right) for exemplary sparking ((a) + (b)) and striking ((c)) runs. The zt-projection of (a) shows the typical cone signature, whereas (b) and (c) illustrate the band signature. The xy-projections of (b) and (c) demonstrate the difference between sparking and striking runs.

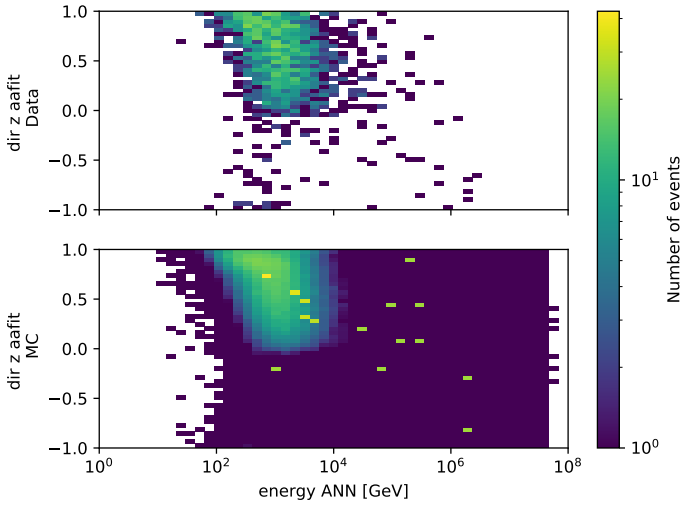


Figure A.8: Reconstructed energy and direction distribution for the final selection of neutrino events for data (above) and Monte Carlo (below). All sparking and striking runs are excluded. The minimum of the colorbar is restricted to 1.

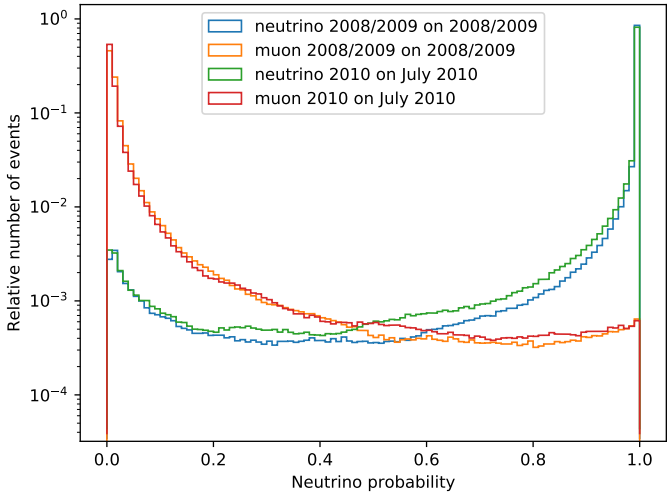
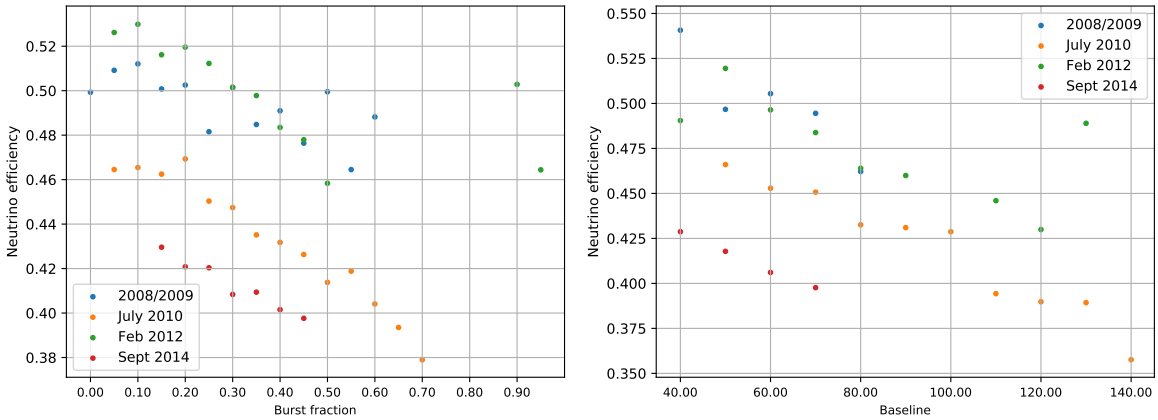
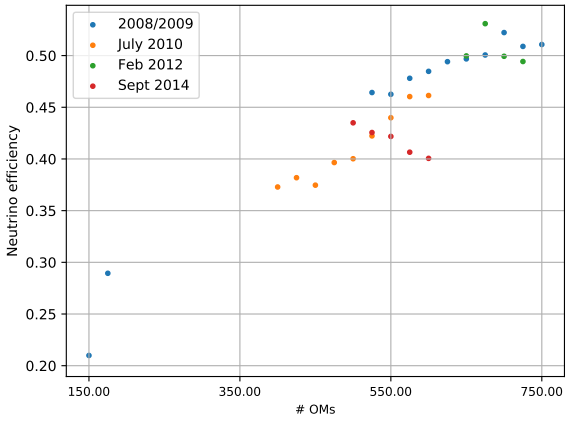


Figure A.9: Neutrino probability distribution of the 2008/2009 and the 2010 network on their corresponding test set.



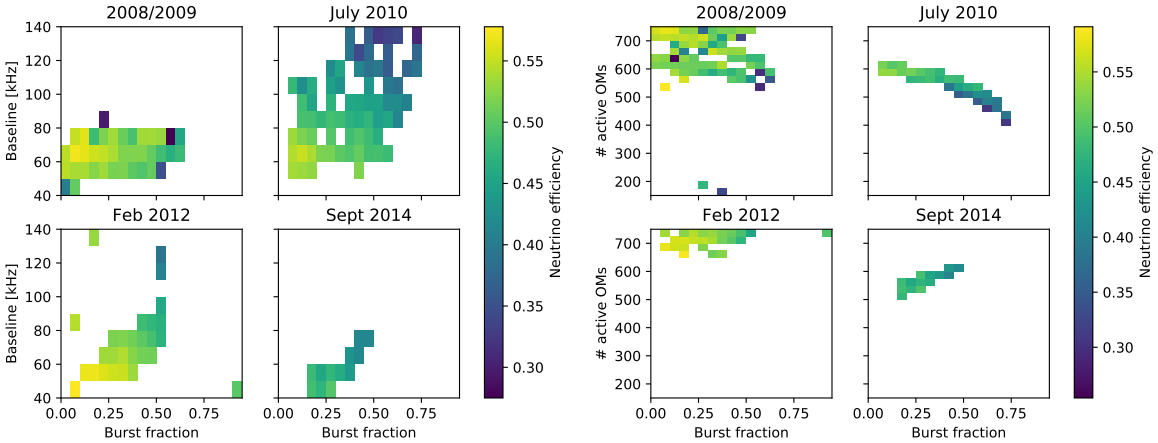
(a) Burst fraction

(b) Baseline



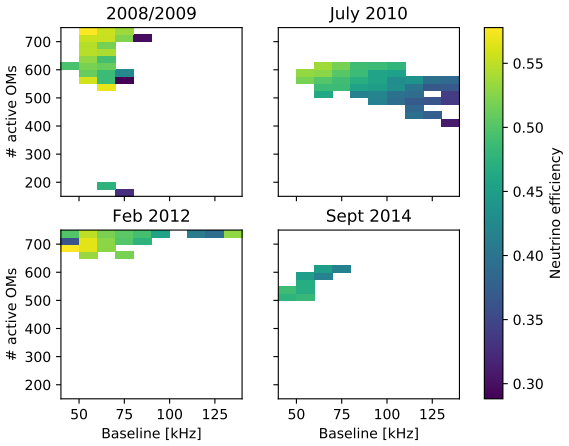
(c) Number of active OMs

Figure A.10: Bin-wise neutrino efficiency for a neutrino spectrum $\sim E^{-2}$ for the different test episodes for different parameters that characterize the noise.



(a) Burst fraction and baseline

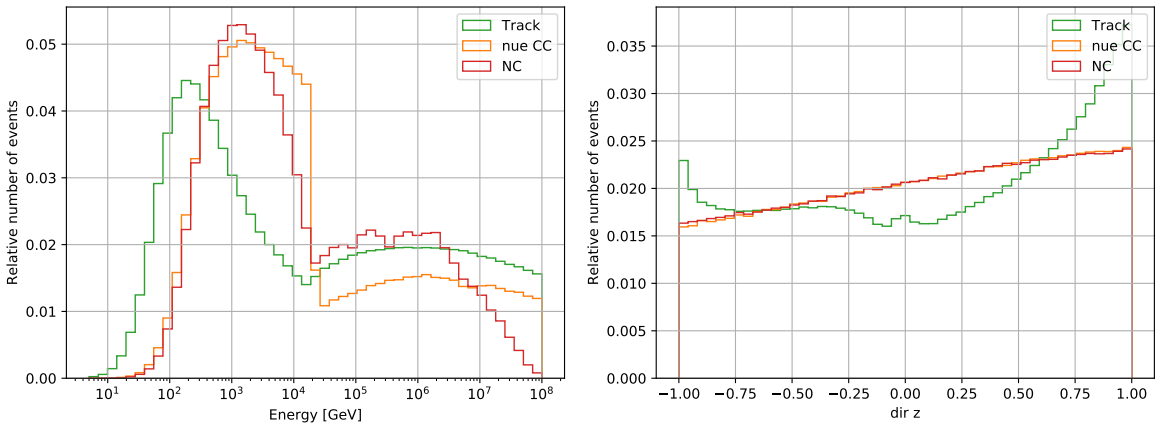
(b) Burst fraction and number of active OMs



(c) Baseline and number of active OMs

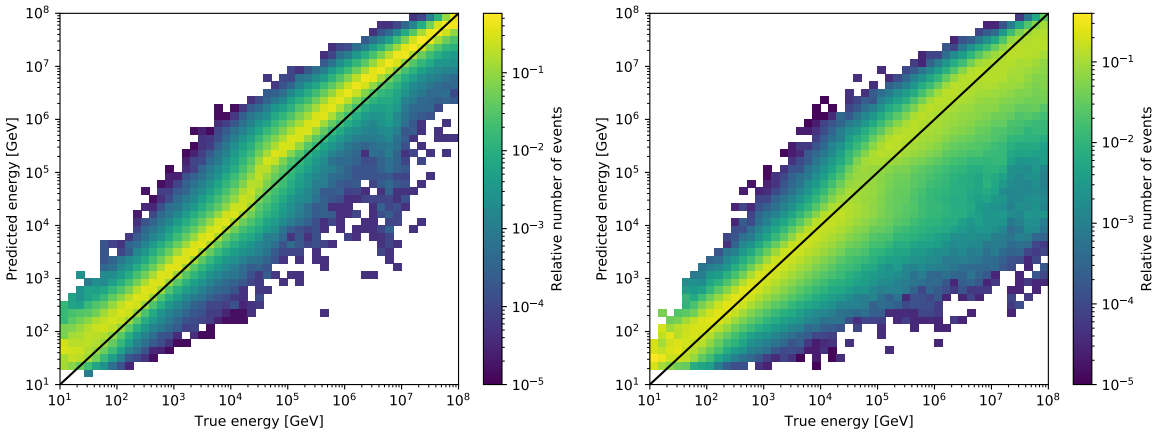
Figure A.11: Two dimensional bin-wise neutrino efficiency for a neutrino spectrum $\sim E^{-2}$ for the different test episodes for different parameters that characterize the noise.

A.2 Energy reconstruction



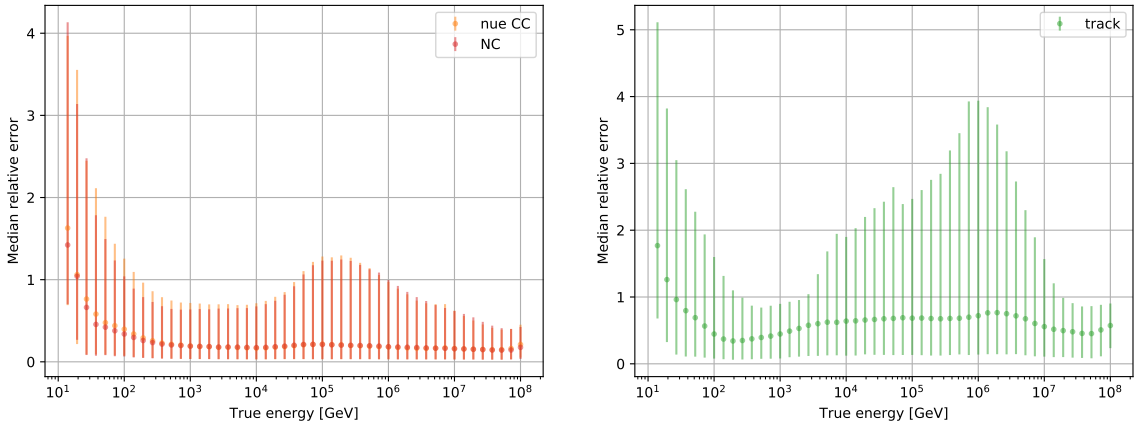
(a) Energy (b) Direction

Figure A.12: Visible neutrino energy and direction distribution of the training set for the energy reconstruction. The distribution is demonstrated separately for the different shower-like events from ν_e CC (orange) and NC (red) interactions and for track-like (green) events. The number is normalised per event type.



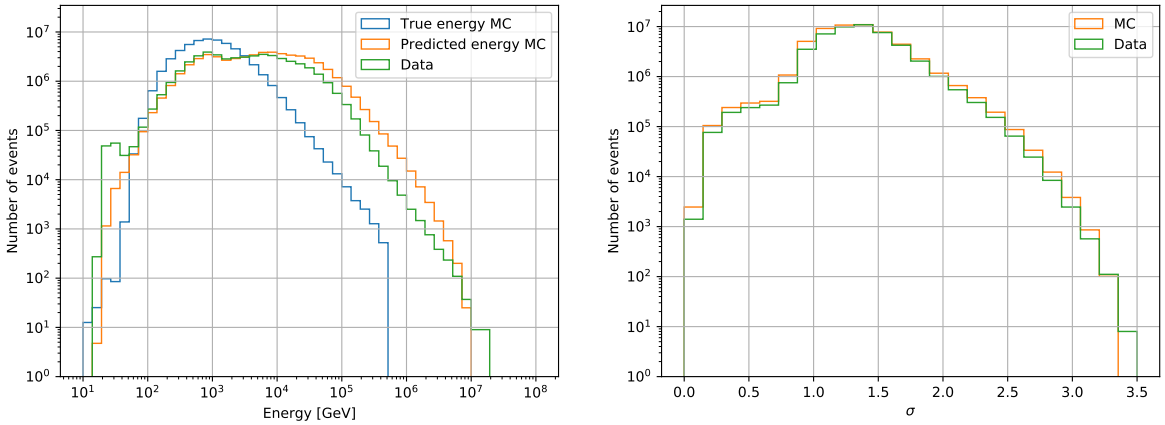
(a) ν_e CC events (b) NC events

Figure A.13: True and predicted energy from a network not using the visible energy for shower-like event types. The colorbar indicates the normed quantity per true energy bin. Its minimum is restricted to 10^{-5}. Clearly the invisible energy leads to the underestimation in case of NC events. The network tries to compensate this and transfers the assumption of the missing energy to the ν_e CC events, leading to the observed significant overestimation over the whole energy range.



(a) Shower-like events (b) Track-like events

Figure A.14: Median energy error for pre-selected events. For the shower-like events, both ν_e CC (orange) and neutral current NC (red) are demonstrated separately. The cut aims for selecting the 50 % of all neutrino events with the lowest σ. In case of shower-like events almost 95 % remain, whereas only about 40 % of all track-like events are included. The error bars indicate the 10 % and 90 % quantile.



(a) Predicted energy (b) σ

Figure A.15: Distribution of the predicted quantities of all events. *Left:* True and expected predicted energy for simulations and predicted energy for measured data. *Right:* Expected and predicted uncertainty parameter σ.

Acknowledgements

At this point, I would like to mention the people who have contributed to the success of this work in so many different ways.
First of all, I would like to thank my supervisor Dr. Thomas Eberl for offering me the opportunity to write this thesis and for encouraging me in the first place, to try out machine learning. Due to your initial support, I was able to gain hands-on experience in this exciting field. I am grateful for all your helpful thoughts, advices and support.
A very special thanks to my office, even if thinned out in the course of my master's thesis and the lack of physical presence due to the Corona pandemic: Michael, for the support at the beginning of my work, guiding me when doing my first steps with deep learning. Steffen, for sharing your great knowledge about the ANTARES detector when I had any questions and of course for the urgently needed variation in home office with our virtual coffee breaks. Stefan, thank you for your help, especially during the time in home office. Whenever I had any questions, your answer came right away. Thanks a lot!
I would also like to thank all ECAP members that have somehow accompanied my way. Thanks for any kind of help and particularly for the nice physics and non-physics related conversations at lunch. I really would have preferred to spend my entire time with you at the institute.
I would also like to thank my friends for their understanding and the distraction during this journey. Special thanks to all the people who shared their coffee breaks with me at ECAP as well as virtually!
Last but definitely not least, I want to thank my family and above all of course my parents. Thank you for your patience and support not only during this thesis but the entire course of my physics studies. Without you none of this would have been possible and I would not have been able to accomplish this. Thank you so much!

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.

Erlangen, April 30, 2021 Nicole Geißelbrecht