

Arbeit zur Erlangung des akademischen Grades  
Bachelor of Science

# **A deep learning based reconstruction of high-energetic muons in IceCube**

Leander Flottau  
geboren in Lünen

2023

Lehrstuhl für Experimentelle Physik V  
Fakultät Physik  
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Dr. Rhode  
Zweitgutachter: Prof. Dr. Westphal  
Abgabedatum: 30. August 2023

---

## Abstract

The atmospheric muon flux results from muons arising from pions and kaons, called conventional, and heavier mesons, called prompt. A significant measurement of the prompt component has never been performed so far, but it expands the knowledge of hadronic interaction models at high energies of PeV and above. The IceCube neutrino Observatory offers the data required to perform this measurement. Due to the distinct energy spectrum compared to conventional muons, a precise measurement of the muon energy is necessary. Since there are only a few high-energetic events expected, large amounts of data have to be reconstructed which requires fast methods. In this thesis, machine learning methods which have recently shown good results in similar applications are used to reconstruct the energy of a muon bundle entering the detector and the energy of the highest energetic muon in a bundle. The reconstruction outperforms a non machine learning based estimator (MuEX) while median runtimes of 0,0092s are achieved. Limitations occur in the reconstruction of the highest energetic muon. The results are promising to measure the prompt muon flux.

## Kurzfassung

Der atmosphärische Myonenfluss resultiert aus Myonen, die aus Pionen und Kaonen entstehen, den so genannten konventionellen Myonen, und Myonen die aus schwereren Mesonen entstehen, den so genannten prompten Myonen. Eine signifikante Messung der prompten Komponente wurde bisher noch nicht durchgeführt, sie kann jedoch das Wissen über hadronische Wechselwirkungsmodelle bei hohen Energien von PeV und darüber erweitern. Der IceCube Detektor bietet die erforderlichen Daten für diese Messung. Aufgrund des veränderten Energiespektrums im Vergleich zu konventionellen Myonen ist eine präzise Messung der Myonenenergie erforderlich. Da nur wenige hochenergetische Ereignisse zu erwarten sind, müssen große Datenmengen rekonstruiert werden, was schnelle Methoden erfordert. In dieser Arbeit werden Methoden des maschinellen Lernens, die kürzlich in ähnlichen Anwendungen gute Ergebnisse erzielt haben, zur Rekonstruktion der Energie eines in den Detektor eintretenden Myonenbündels und der Energie des energiereichsten Myons in einem Bündel verwendet. Die Rekonstruktion übertrifft einen nicht auf maschinellem Lernen basierenden Schätzer (MuEX), während mittlere Laufzeiten von 0,0092s erreicht werden. Einschränkungen ergeben sich bei der Rekonstruktion des höchstenergetischen Myons. Die Ergebnisse sind vielversprechend in Bezug auf die Messung des prompten Myonenflusses.

## Contents

# 1 Introduction

The simulation of atmospheric muons, which are created in air showers by the interaction of cosmic rays and the subsequent decay of produced hadrons [gaisser\_engel\_resconi\_2016], has shown significant differences to the number of muons measured experimentally. The most likely cause of this so-called muon puzzle is the insufficient knowledge about hadronic interactions in the atmosphere [Albrecht\_2022]. A way to improve the knowledge about hadronic interactions is to measure the prompt component of the atmospheric muon flux, which to date has been theoretically predicted but not significantly measured.

Since prompt muons are dominant at high energies, an accurate energy estimator is needed to be sensitive to prompt events. At the same time the estimation has to be runtime efficient due to the high rate of approximately 3000 Hz at which atmospheric muons are detected by IceCube [Aartsen\_2016]. This combination of accuracy and runtime efficiency can be achieved by a convolutional neural network.

Atmospheric muons are mostly expected in bundles of multiple muons generated in the same shower. There are two main labels reconstructed in this analysis: The energy of the muon bundle and the energy of the leading muon, both at detector entry. The leading muon is the most energetic muon within a bundle. Unlike the reconstruction of the energy of a muon bundle, the reconstruction of the leading muon energy is a task not previously attempted. Therefore, it is not clear whether this reconstruction is even possible with the required accuracy using the information provided by the IceCube detector.

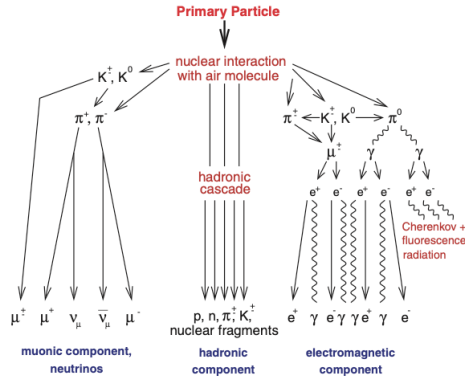
The goal of this thesis is to investigate whether a neural network can provide an energy reconstruction that enables a further analysis of the prompt component by achieving sensitivity for the prompt component at high energies while maintaining sufficient speed to handle the high data rate. Also, the comparison of the resulting reconstruction to existing energy reconstruction methods which are not based on machine-learning is studied. Additionally, this analysis aims to investigate if the reconstruction of the energy of the most energetic muon in a bundle is possible using the information provided by the IceCube detector.

## 2 Theoretical background

At first, the atmospheric muon flux and its components are introduced. Afterwards, the IceCube Neutrino Observatory is described. At last, the functionality of convolutional neural networks is explained.

### 2.1 Atmospheric muons and the prompt component

Cosmic rays are ionized nuclei (mainly protons and alpha particles) hitting the Earth's atmosphere at a rate of approximately 1000 per square meter per second [gaisser\_engel\_resconi\_2016]. When cosmic rays interact with nuclei in the



**Figure 2.1:** Illustration of a hadronic air shower. It consists of an electromagnetic component (right), a hadronic component (middle) and the muonic component (left) which is central for this analysis. [Haungs]

atmosphere they can produce particles who in turn decay or interact with the atmosphere again resulting in a cascade of particles called extensive air-shower. A hadronic shower is illustrated in Figure ???. The development of such events is described by the cascade equations which can only be solved numerically for example using MCEq [gaisser2019precision]. A number of different particles created in air showers can produce muons upon decay and thereby impact the atmospheric muon flux. Atmospheric muons are produced in hadronic cascades,

so a large majority of the corresponding events are bundles of multiple muons instead of single muons further complicating energy reconstructions. The number of muons in those bundles ranges over multiple orders of magnitude. The biggest contribution to the muon flux is made by charged pions and kaons via the channels [gaisser\_engel\_resconi\_2016]:

$$\pi^\pm \rightarrow \mu^\pm + \nu_\mu(\bar{\nu}_\mu)(BR \approx 100\%), \quad (2.1)$$

$$K^\pm \rightarrow \mu^\pm + \nu_\mu(\bar{\nu}_\mu)(BR \approx 63,6\%). \quad (2.2)$$

Pions and kaons are the particles created at the highest rate in primary interactions leading to a high number of atmospheric muons being created. Other leptons like electrons can theoretically also be created in air showers but those decay modes of the  $\pi^\pm$  and  $K^\pm$  can be neglected because of their low branching ratio. Muons stemming from  $\pi^\pm$  and  $K^\pm$  decay are called conventional muons.

Calculating the flux of atmospheric muons is highly complex because decay, interaction and energy loss have to be accounted for. There are also additional uncertainties involved because the production of charged pions and kaons depends on the primary cosmic ray spectrum which is not precisely known.

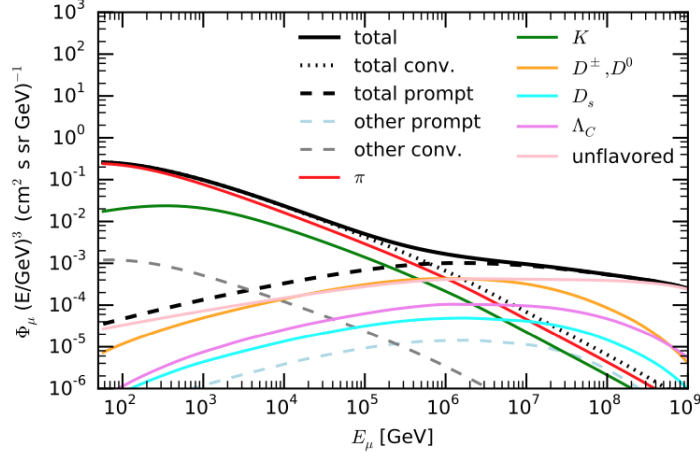
Aside from the conventional channels there are various other modes contributing to the muon flux the most important of those being the decay of charmed hadrons ( $D^0$ ,  $D^+$ ,  $D_s^+$ ,  $\Lambda_c$  as well as their antiparticles) and the decay of unflavored mesons ( $\eta$ ,  $\eta'$ ,  $\rho^0$ ,  $\omega$ ,  $\Phi$ ) into muon-antimuon-pairs [Illana\_2011]. Muons generated this way are called prompt. Even though charmed mesons like the  $D^\pm$  exhibit significant branching ratios in the relevant decay modes for example [pdg]:

$$D^0 \rightarrow K^- \mu^+ \nu_\mu (BR \approx 5,3\%), \quad (2.3)$$

$$D^\pm \rightarrow \bar{K}^0 \mu^\pm \nu_\mu (BR \approx 14,03\%). \quad (2.4)$$

The muons produced that way are the much smaller contribution to the atmospheric muon flux for almost all energies because the corresponding parent particles are produced in the primary hadronic interactions at a much lower rate than pions or kaons. In contrast to that the unflavored mesons have an energy spectrum that is similar to pions or kaons but their branching ratios to the modes including muons are very small.

As previously established prompt and conventional muons are distinguished by definition by their parent particles which in turn differ in their mean lifetime: Pions and kaons have a lifetime of  $\tau_{\pi^\pm, K^\pm} \sim 10^{-8}$ s. In comparison to that charmed particles have a much lower lifetime of  $\tau_{\text{charmed}} \sim 10^{-12} - 10^{-13}$ s [pdg]. Following this pions and kaons usually interact rather than decay in a majority of cases while



**Figure 2.2:** The atmospheric neutrino flux divided into the different contributing factors based on numerical solutions of the Cascade equations. At about  $E_\mu \approx 5 \cdot 10^5 \text{ GeV}$  there is a threshold for the prompt component becoming the dominant. From  $E_\mu = 10^7 \text{ GeV}$  upwards the contribution of charmed particles falls off and the whole flux is dominated by the decay of unflavored mesons. Amongst the charmed particles the  $D^{\pm,0}$  is the most relevant. [fedynitch2015calculation]

the parent particles of other contributions in most cases decay before having interacted hence their name: 'prompt'. Because of time dilatation the decay probability of the prompt component is additionally suppressed at high energies. Resulting from this effect two major differences between prompt and conventional component arise: Conventional muons have a much steeper energy spectrum than their prompt counterpart, therefore the prompt component becomes dominant at high energies as illustrated in Figure ?? [fedynitch2015calculation] and the prompt component has an isotropic angular distribution while the spectrum of conventional muons contains the factor:  $\sec \Theta / E_\mu$  relative to the primary cosmic ray spectrum depending on the zenith angle  $\Theta$  [gaisser\_engel\_resconi\_2016]. Because the parent particles of prompt muons usually don't interact they are isotopically distributed. [Illana\_2011] [gaisser\_engel\_resconi\_2016]

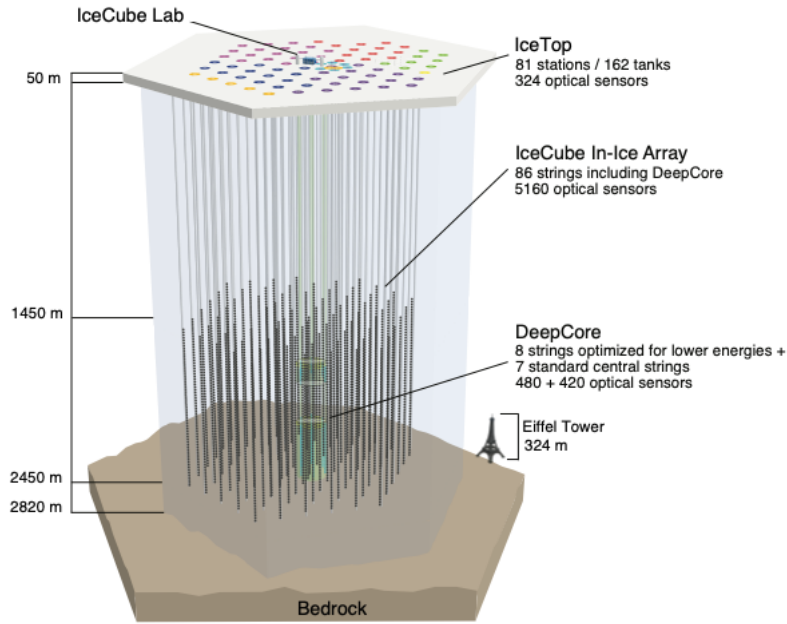
## 2.2 The IceCube Neutrino Observatory

The IceCube Neutrino Observatory is a kilometer-scale detector located in the South Pole ice. When charged particles move through a medium faster than the speed of light in that medium the particle emits light in a cone with a characteristic opening



angle along its track [Mikkelsen\_2016]. This so-called Cherenkov effect is the principle the IceCube detector is based on.

The essential elements of the detector are the digital optical modules (DOMs). They measure charge pulses using Photo-Multiplier-Tubes (PMTs) to detect the Cherenkov light emitted by charged particles (leptons) travelling the ice. The original purpose of the Detector is to measure leptons generated when neutrinos interact with the ice. Atmospheric muons which are the focal point in this analysis are considered background for neutrino measurements.



**Figure 2.3:** Schematic picture of the IceCube Neutrino Observatory. [Aartsen\_2017]

A number of different arrays make up the detector the most important of those being the in-ice array made up of 5160 DOMs deployed along 78 vertical strings between depths of 1450 m and 2450 m covering a total area of  $1 \text{ km}^3$ . The high detection volume is essential to detect neutrinos with a high sensitivity because of their low interaction probability. The strings are positioned in an approximate hexagonal grid (125 m spacing) with each string containing 60 DOMs and function to send the data collected by the DOMs to the surface via copper wires. All cable strings lead to the servers in the IceCube Laboratory (ICL) where the data is analyzed and distributed all over the world. In addition to the main in-ice array the deep-core array composed of 8 strings optimized to register lower energy events. Those strings

are deployed in a tighter non-hexagonal configuration with a mean spacing of (72 m). The DeepCore array is divided into upper and lower DeepCore divided by a dust layer in the ice. The whole IceCube Neutrino Observatory including the different sub-arrays is illustrated in Figure ??.

The south-pole ice is especially suitable for the detection of Cherenkov light because it offers high quantities of interaction material for neutrinos as well as a high optical attenuation length [Aartsen\_2017]. Those properties enable the detector to cover the large volume needed for the detection of neutrinos with a relatively high spacing of DOMs. [Aartsen\_2017]

### 2.3 Convolutional neural networks

The fundamental basics of convolutional neural networks are mentioned at first. Secondly, the specific CNN structure used in IceCube is stated.

#### 2.3.1 General theory on CNNs

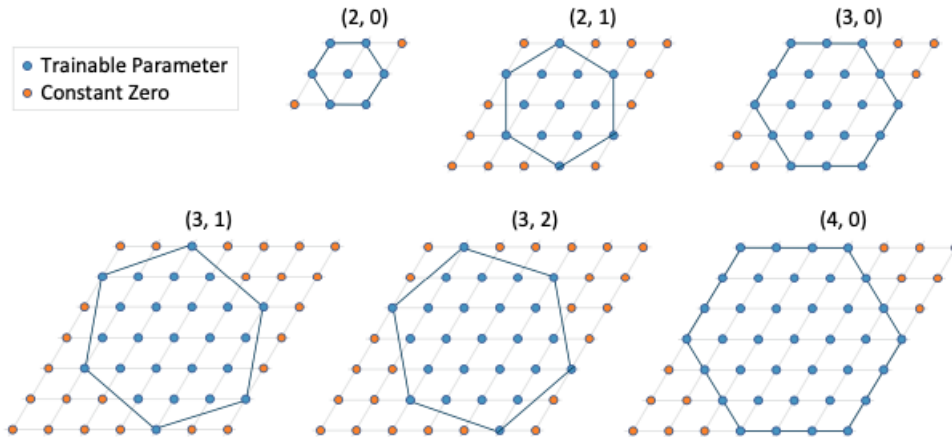
Neural Networks are digital structures designed to mimic the function of the human brain but when broken down they are just a highly complex optimization problem. A Neural Network consists of an input and an output layer as well as one or multiple hidden layers in between whose output is not observed directly. The output of each node is calculated as a linear combination of all outputs by the nodes in the previous layer. The parameters of the individual linear combinations linking the nodes are called weights. Additionally, a so-called activation function is applied to the computed linear combination of inputs to make the model nonlinear. The activation function as well as number and size of the hidden layers have to be chosen based on the application. The fully connected part of the neural network used in this analysis is always identical and consists of one hidden layer with 50 nodes and the 'elu' activation function [clevert2016fast]. The network 'learns' those weights by optimizing a loss function designed to describe how well the regression (or classification) fits the truth. This loss function is then iteratively minimized using complex numerical methods [hastie2009elements]. The exact nature of the loss function and the optimization method are chosen based on the problem, in this case it is the 'Adam' optimizer [kingma2017adam] in combination with a Gaussian-likelihood loss function.

For certain applications like image recognition it is sensible to use a special form of neural network called convolutional neural network. This network architecture is aimed at utilizing translation invariance, the fundamental concept that the same

physical process looks identical no matter where in the detector it happens. To achieve this a filter of a certain shape unique to the dimensionality of the input data is applied locally weighting every node with the ones in its immediate surrounding. This filter is applied equally to every node creating so-called feature maps who are then used by a fully connected network to make a prediction. In addition to that the process of pooling is used to reduce data dimensionality and is especially useful for problems with a high number of input variables like in the case of the IceCube detector. Pooling describes the summarization of multiple data points in a subregion to only one value in most cases by taking the average or the maximum of the values in the designated field. By the application of such a pooling filter the size of the feature maps can be reduced significantly based on the filter size. [dumoulin2018guide]

### 2.3.2 IceCube specific CNN

To suit the unique hexagonal structure and data input format a network was designed specifically to analyze IceCube data <sup>1</sup>. To fit the architecture of IceCube



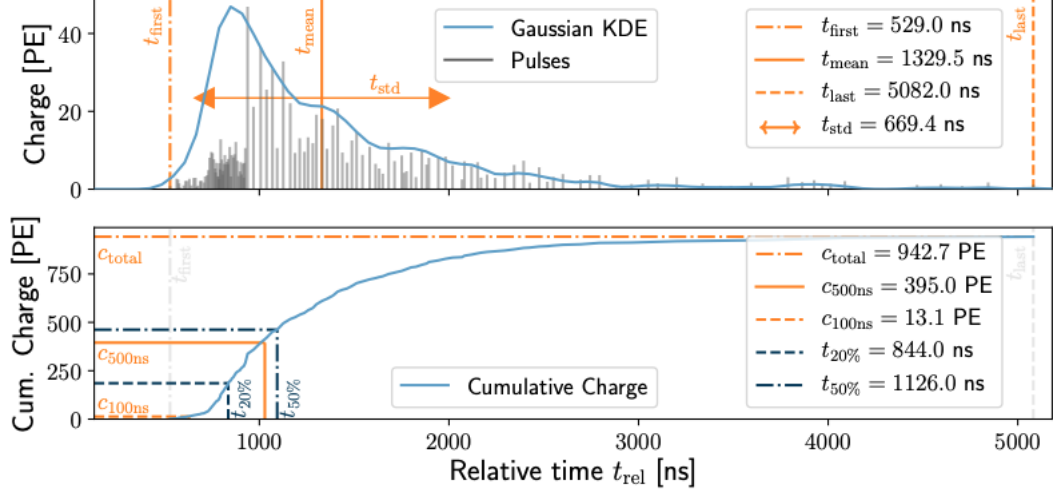
**Figure 2.4:** Hexagonal convolutional Filters. Each Filter can be described by two parameters in the xy-plane: (size, orientation). All networks presented in this analysis have the shape (2,0,3) (top left) with the third number representing the convolution along the z-axis. [Abbasi\_2021]

the convolutional filters of the network are hexagonal shaped as illustrated in Figure ???. A problematic factor when looking at IceCube data is the data input format.

<sup>1</sup>The neural network is available at: [https://github.com/icecube/dnn\\_reco](https://github.com/icecube/dnn_reco)

## 2 Theoretical background

IceCube DOMs measure the deposited charge depending on the time as a pulse series. To achieve an input of constant size for every DOM and every event without



**Figure 2.5:** Construction of the input features based on an IceCube pulse series (top) and the resulting cumulative charge (bottom). [Abbasi\_2021]

making the model to complex each pulse series is broken down to a number of discrete input features. Figure ?? shows a typical IceCube pulse series and the features generated from it. Of the nine input features the network was originally designed on only three are used here to hold runtimes low: The total charge  $c_{\text{total}}$ , the relative time of the first pulse  $t_{\text{first}}$  and the standard deviation of pulse times  $t_{\text{std}}$ .

Having generated the three input features they have to be transformed to ultimately be applicable to a CNN. The hexagonal shaped data of the main array is transformed onto a rectangular grid by aligning the rows and padding the remaining points with zeros. Deep Core is too asymmetric to be put on a rectangular grid, so the corresponding data is aligned next to each other on a 2 dimensional grid. Additionally, all charge and energy related features and labels are put on a logarithmic scale via the transformation:

$$X' = \ln(1 + X) \quad (2.5)$$

to compensate for the big range those values cover. Then all values are centered around zero and normalized to unit variance:

$$X'' = \frac{X' - \bar{X}'}{\sigma_X + \epsilon} \quad (2.6)$$

with a small constant  $\epsilon = 10^{-4}$  to avoid division by zero. [Abbasi\_2021]

## 3 Reconstruction performance

In this chapter the performance of the reconstruction of the bundle and leading muon energy respectively by networks with different architectures is evaluated and compared to an alternative method not based on machine learning.

### 3.1 Bundle energy

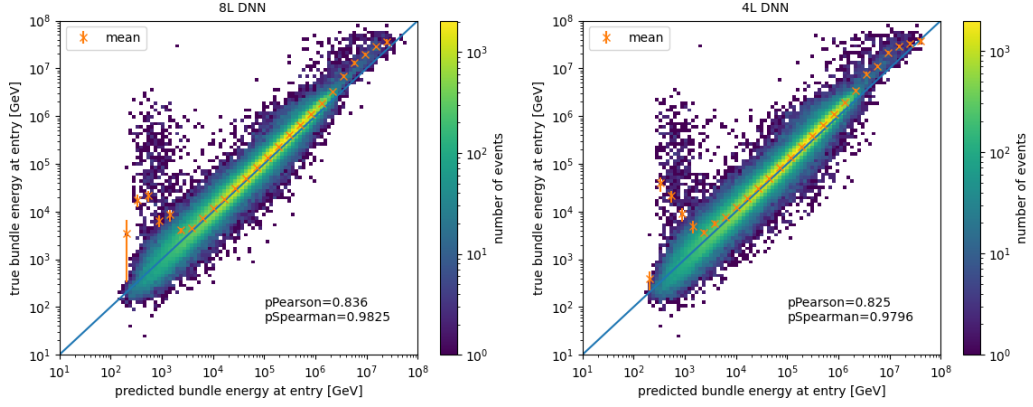
The first reconstructed label is the energy of the muon bundle at detector entry. Because this energy is strongly correlated to the energy deposited in the detector [Aartsen\_2014], which in turn can be inferred from the total charge measured by the DOMs, it is expected that this is a value that can be reconstructed better compared to the energy of the leading muon.

#### 3.1.1 Correlation

A neural network with eight convolutional layers and a kernel shape of  $(2, 0, 3)$  and a network with a reduced number of convolutional layers result in reconstructions as displayed in ??.

Qualitatively both networks perform very similar with only minor improvements in linear correlation for the 8L version. The bias in the high energy region most likely is a result of the detector not being able to resolve all deposited energy leading to systematic underestimation.

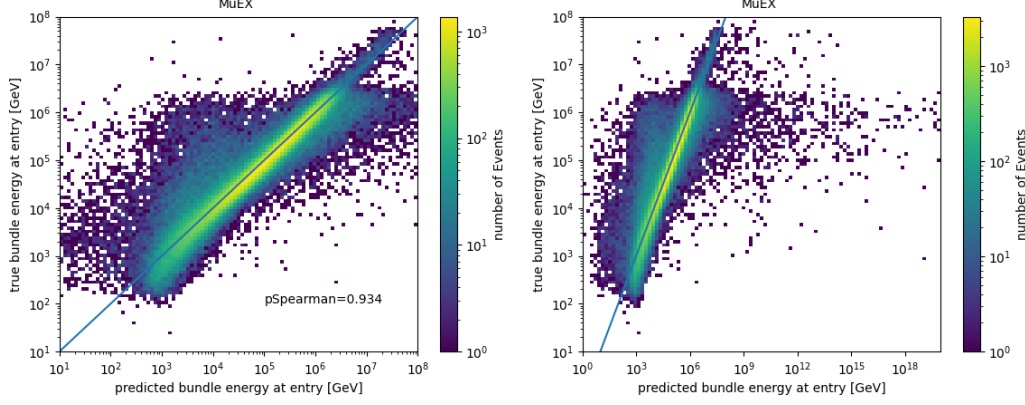
To assess, if the application of a neural network for this reconstruction task makes sense, it is important to also look at standard reconstruction methods and their performance in comparison to the DNN based approach. For this specific task it makes sense to use the MuEX reconstruction [20.500.12030\_2899] as a comparison despite the fact that far more precise methods exist, because those would not be applicable for prompt analysis due to their extremely high runtime. Figure ?? shows that aside from the generally higher spread the MuEX Fit especially shows a high number of outliers in the high energy region, where high energies up to  $10^{28}$  GeV are estimated. As a consequence, in contrast to the DNN, the MuEX Fit is prone



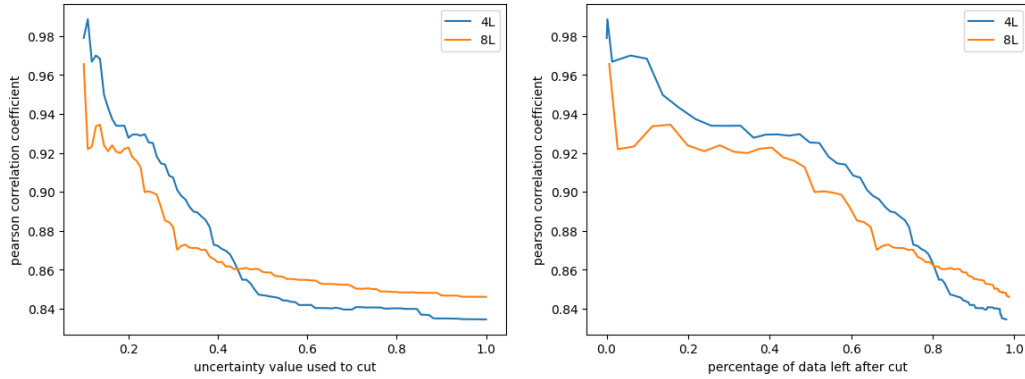
**Figure 3.1:** Correlation of the predicted and true bundle energy as an unweighted two-dimensional histogram for neural networks with four convolutional layers (4L, right) and eight convolutional layers (8L, left) on the main array. The strength of the linear and monotonic correlation between prediction and truth are indicated by Pearson and Spearman correlation coefficients respectively. In the low energy region the mean is dominated by outliers where the energy was significantly underestimated. In the high energy region there are fewer outliers, but the reconstruction is biased towards underestimation.

to overestimation for high energies. This is a particularly problematic effect for the purpose of this analysis, because events with an energy over 100 GeV, which are interesting for prompt analysis, are spread over more than 20 orders of magnitude in the reconstruction likely leading to problems with identifying the prompt component. Furthermore, the MuEX reconstruction does not provide information about the uncertainty of the reconstruction making it harder to cut out the unscientific values and reduce the Data to a subset of well reconstructed events.

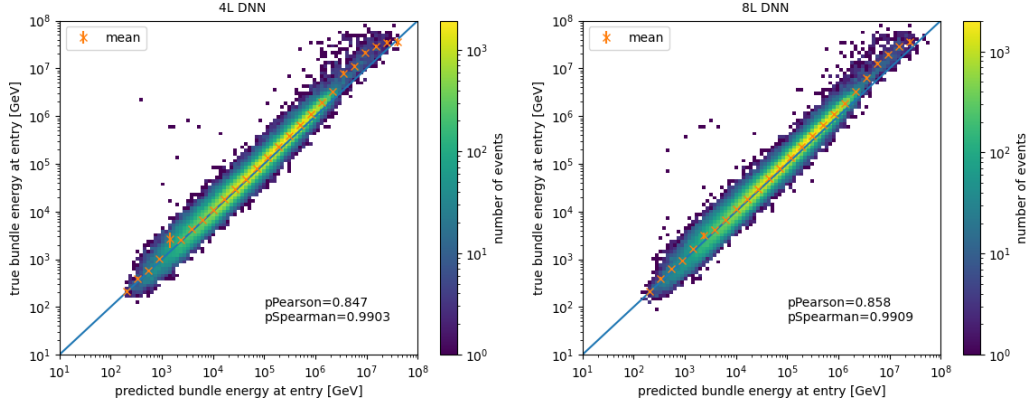
Based on the uncertainty-reconstruction made by a separate subnetwork a cut can be made to filter for events estimated with a lower uncertainty. An uncertainty cut can be made on the absolute value given by the network because it is already made based on the transformed values. The exact value to cut on has to be chosen depending on the application and how much loss of data is affordable. The Dataset used for evaluation only has a low number of events with high energies the cuts are additionally hard to evaluate. One point of reference is the Pearson correlation coefficient depending on the cut. The impact of the different cuts on linear correlation is shown in Figure ?? . As seen in Figure ?? making a quality cut especially eliminates events with small reconstructed energies, where significant smearing effects are seen when not making a cut. The loss of prompt events is especially low, supporting the fact that the cut is actually suitable. After the cut



**Figure 3.2:** Correlation of the true and predicted bundle energy for the MUEX reconstruction. In the right plot, the x-axis is expanded to show the extreme outliers to unrealistically high energies. The monotonic correlation is weaker than that of both DNNs. The Pearson coefficient is not stable against the outliers, so it is less meaningful here.



**Figure 3.3:** Pearson correlation coefficient depending on the chosen uncertainty cut (left) and depending on the percentage of data after the cut (right). In the left plot it is cut for all events smaller than or equal to the displayed logarithmic uncertainty.



**Figure 3.4:** Correlation plot between reconstructed and true bundle energy for events with an estimated log-uncertainty smaller than 0.5. Monotonic and especially linear correlation are significantly enhanced. While this removes the smearing effects seen in Figure ?? for lower predicted energies the bias seen for high energies can not be corrected.

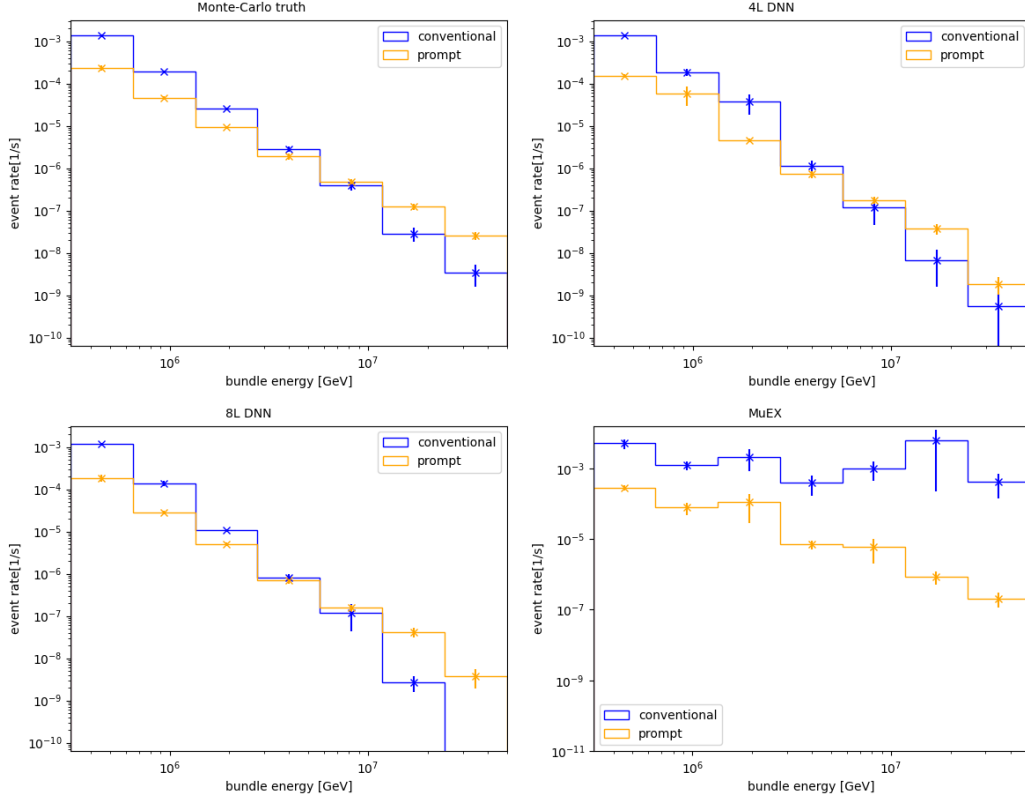
both networks show a linear correlation that is almost equal. The main advantage of the 8L model in this case is the increased amount of Data after the cut (87,1 % compared to 84,9 % for the 4L DNN) because it generally estimates events with a high uncertainty less often.

#### 3.1.2 Energy spectrum

With the energy reconstruction of the neural networks an energy spectrum can be generated, in this case using the weighting for the GaisserH3a-model [gaisser2013cosmic]. In this case only event rates are displayed. To generate an actual energy spectrum one would have to account for systematic effects of the detector such as the detector sensitivity as well as generate the energy on the surface based on the reconstructed energy at detector entry. The resulting spectrum is analyzed in the high energy domain, where the prompt component is expected to be dominant.

The energy spectra displayed in Figure ?? demonstrate that both DNNs predict the prompt component to be dominant for very high energies. Resulting from this, the 3 highest energy bins in both reconstructions provide sensitivity for prompt muons, meaning that they are suitable for prompt analysis. The only slight caveat is the reduced event rate in those bins compared to the Monte-Carlo truth most likely caused by the systematic underestimation mentioned in the previous section. In contrast to that the reconstruction provided by the MuEX fit predicts a very

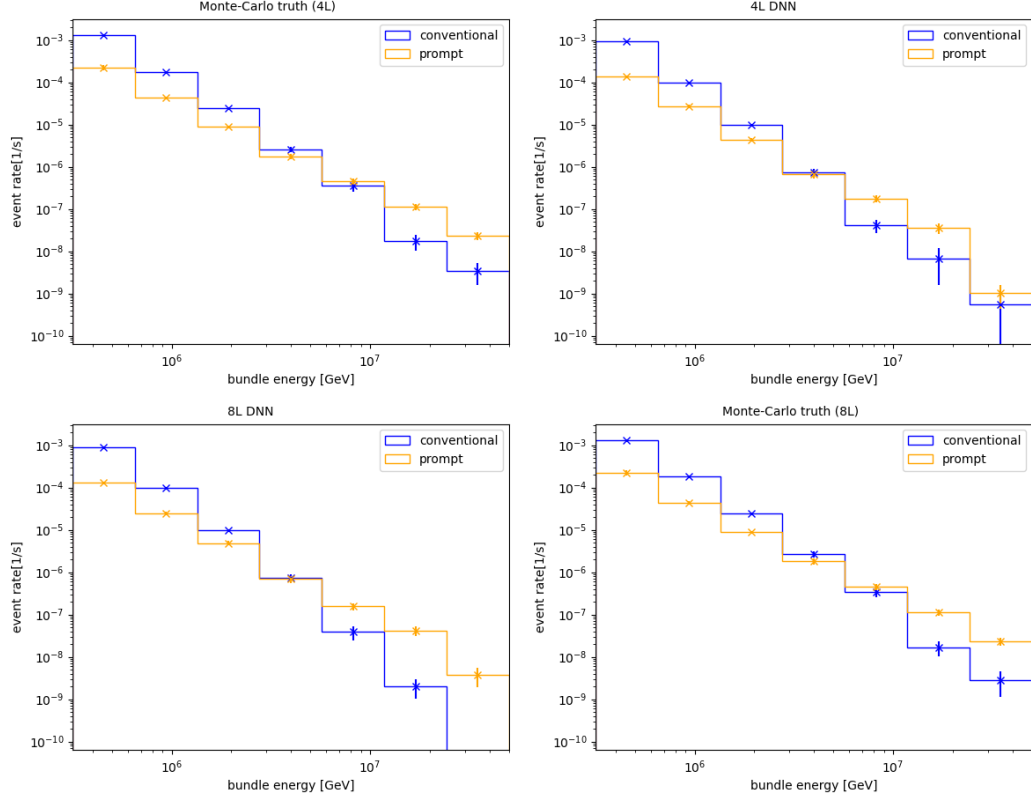




**Figure 3.5:** The energy spectrum for bundle energies between  $3 \cdot 10^5$  GeV and  $5 \cdot 10^7$  GeV according to the Monte-Carlo data, the 4L-DNN, the 8L-DNN and the MuEX Fit. All spectra are divided into a prompt and a conventional component based on the nature of the leading muon. For weighting the GaisserH3a-model is used. The errorbars are calculated according to the Poisson statistic. Both DNNs are capable of displaying a dominant prompt component as expected looking at the Monte-Carlo data and therefore are suitable for prompt analysis. In contrast, the MuEX fit does not reconstruct the bundle energy well enough to see the prompt component as dominant. It is also apparent, that as expected based on the correlation plots shown previously, the MUeX Fit strongly overestimates the event rate in the high energy domain.

### 3 Reconstruction performance

flat energy spectrum that significantly overestimates both prompt and conventional event rate for high energies due to the smearing effects seen in Figure ?? As the MuEX fit is apparently not capable of seeing a dominant prompt component it does not provide sensitivity for prompt muons and is therefore not applicable for an analysis of the prompt component.



**Figure 3.6:** Energy Spectrum of Monte-Carlo simulation and DNNs with the corresponding quality cuts. The cut is network specific so the Monte-Carlo truth has to be calculated for each network individually. Quality cuts on significantly lower uncertainties are hard to evaluate based on this dataset because of the low amount of data in the high energy region.

The spectrum of both networks after the application of the previously shown quality cuts is displayed in Figure ?. While the reconstruction still shows a sensitivity for the prompt component it is not a very significant improvement upon the raw reconstruction because the bias of the network is still present and therefore event rates in the relevant bins are still underestimated compared to the truth. The more noticeable positive effects of the cut are in relevant energies lower than those

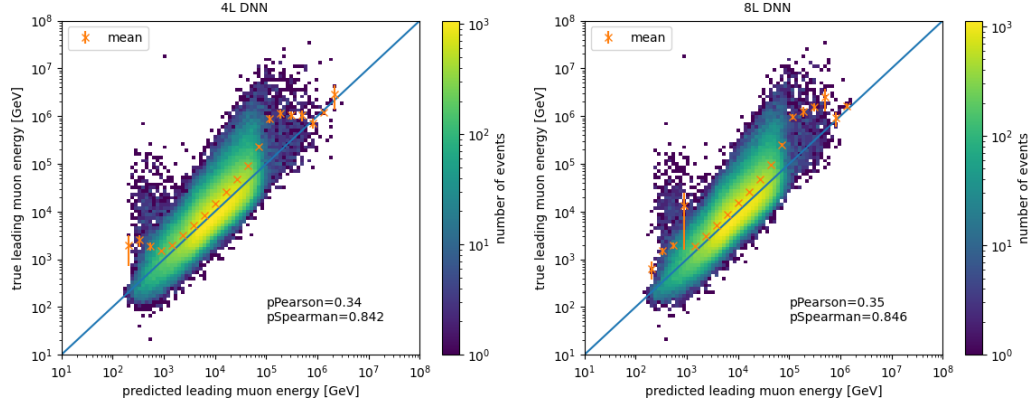
considered in the spectra displayed here. Another important factor to consider when comparing raw and cut Monte-Carlo is that the reduced amount of data results in increased Poisson-uncertainties. This effect is enhanced by the generally small statistic for energies above 1 PeV. This suggests that the size of the data set is not adequate to properly evaluate the neural network with the uncertainty cut applied from a certain threshold of data loss onwards because the Poisson-uncertainty in the prompt sensitive bins becomes too high to make sensible interpretations of the networks' performance. The bigger network allows more events to pass the same uncertainty cut so it offers more statistic.

## 3.2 Energy of the leading muon

Reconstructing the energy of the most energetic muon in the shower is a much more challenging task for a neural network than just reconstructing the energy of the whole muon bundle because it does not suffice to sum up the deposited charge and conclude an energy. Additionally, there has to be an understanding of how dominant the leading muon is within the bundle. In theory, this should be possible by considering the stochasticity of the deposited energy. The basic idea behind this is that bundles where the energy is very equally distributed deposit energy more regularly than bundles with a dominant leading muon, who deposit energy large amounts of energy in singular spots. How well the network can utilize this effect based only on the three input features is a main influence on the resulting reconstruction.

### 3.2.1 Correlation

The neural networks produce correlation plots as seen in Figure ?? . As expected, the correlation in general and the linear correlation in specific are weaker than that of true and reconstructed bundle energy for both networks. In the predicted low energy region between 100 GeV and 10 TeV there are smearing effects similar to those seen for the bundle energy. This suggests, that a bad reconstruction of the bundle energy in general also impacts the reconstruction of the leading muon energy. In addition to that the mean shows strong underestimation from energies of 50 TeV upwards. Especially obvious is, that both networks very rarely predict energies above 100 TeV. The 4L-DNN for example only predicts 8,9 % of the events with a true leading muon energy over 100 TeV to be in that energy region. This results in a large majority of the events expected in this region, where the prompt component is visible in theory, migrating to lower energy bins. This soft threshold in the predicted energy is very problematic as seen in the energy spectrum in Figure



**Figure 3.7:** Correlation of true and reconstructed energy of the leading muon for the 4L and the 8L-DNN. Linear and monotonic correlation are reduced compared to the bundle energy. The mean of the true deposited energy is calculated in every third predicted energy bin. In some bins large outliers or low statistic result in high errors.

?? because it almost completely eliminates the most important part of the spectrum from the prediction.

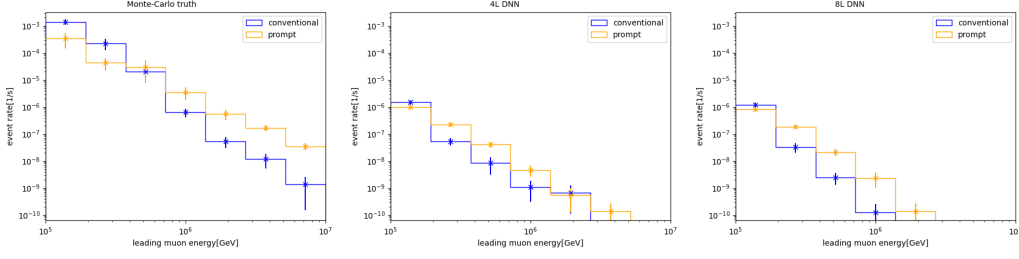
Resulting from this an uncertainty based cut on the data would not be sensible because the amount of data above 100 TeV is already massively reduced. It also appears that the events in that region are generally predicted with a high uncertainty so that they are eliminated at a much higher rate than other events.

#### 3.2.2 Energy spectrum

Looking at the energy spectra of the leading muon in Figure ?? it is apparent, that the systematic underestimation of high energetic muons and subsequent pileup of events in the lower energy-bins leads to reduced event rates in the reconstruction. Resulting from this the reconstruction of the leading muon energy is not sensible to use compared to the bundle energy, although the leading muon energy provide theoretically better information on the prompt component as demonstrated in the Monte-Carlo spectrum in Figure ??.

#### 3.2.3 Leading fraction

The leading fraction is defined as the fraction of the bundle energy carried by the leading muon:  $f_{\text{lead}} = \frac{E_{\text{lead}}}{E_{\text{bundle}}}$ . As it is the essential difference between leading muon

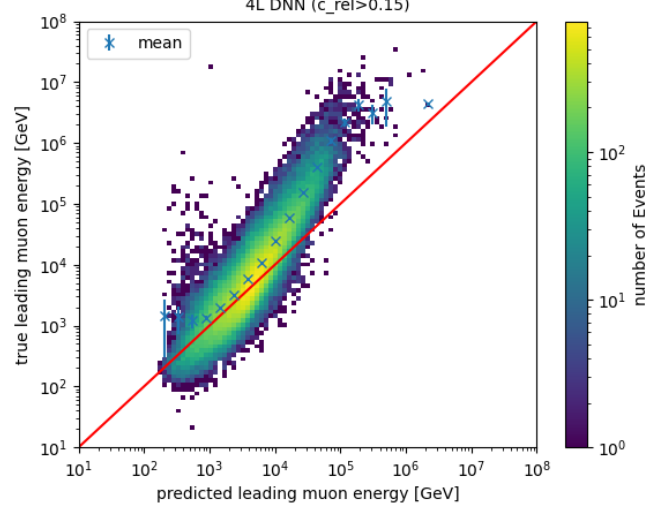


**Figure 3.8:** Simulated and reconstructed spectra for the 4L model (middle) and 8L model (right) of the leading muon weighted with the GaisserH3a-model are shown. The event rate of both components is underestimated by at least three orders of magnitude by both DNNs compared to the simulation even in the lowest energy bin. Moreover, the reconstructed spectra fall off steeper than they theoretically do. Even though both reconstructed spectra offer a sensitivity to the prompt component, the overall quality of the reconstruction does not suffice especially compared to the bundle energy.

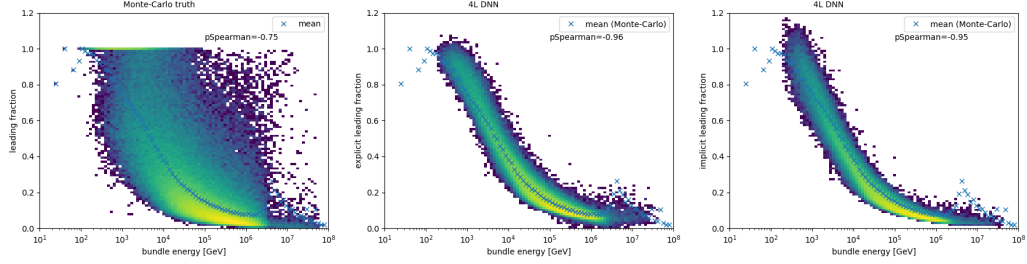
energy and the bundle energy the leading fraction was explicitly estimated by the networks as a separate label.

One way to further evaluate the reconstruction of the leading muon energy is to check for the impact of the leading fraction on the predicted leading muon energy. Based on the assumption, that more dominant leading muons should be easier to identify within the bundle and therefore be reconstructed more precisely, a cut is performed for events in which the leading muon carries a high percentage of the bundle energy, in this case more than 15%. The resulting correlation displayed in Figure ?? demonstrates that the assumption does not hold true. Instead, the cut results in a bias increasing with the predicted energy. Apparently the events with a high leading fraction are exactly those, who are underestimated for high energies by the networks.

To further investigate the reasons for this phenomenon the relationship between bundle energy and leading fraction (??) is studied. The Monte-Carlo simulation shows that while there is a general correlation between those two labels because it is less likely for high bundle energies to be dominated by the leading muon, this correlation is not strong ( $\rho_{\text{Pearson,MC}} = -0.75$ ). In contrast to this the leading fractions estimated by the network are correlated stronger to the estimated bundle energy than they are based on the simulation with Pearson coefficients of  $\rho_{\text{Pearson,im}} = -0.95$  for the implicit and  $\rho_{\text{Pearson,ex}} = -0.96$  for the explicit leading fraction. The maximum of both DNN based deviations is noticeably centered around the mean of the Monte-Carlo. This suggests that the network does not estimate how dominant the leading muon is by using information on the stochasticity of the energy loss



**Figure 3.9:** Predicted and true energy of the leading muons for all events satisfying the condition:  $f_{\text{lead}} > 0.15$ . There is a significant bias to underestimation increasing with the predicted energy. The well reconstructed events in the high energy region are almost completely removed by the cut.

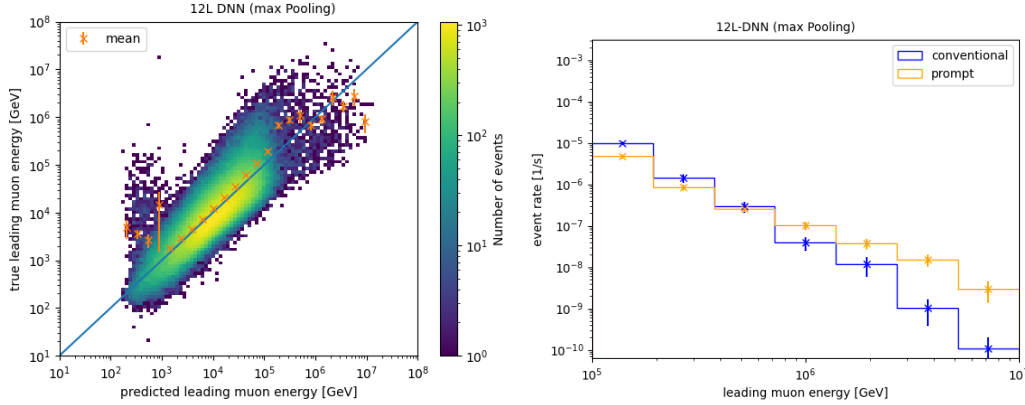


**Figure 3.10:** Correlation between bundle energy and leading fraction based on the Monte-Carlo data and the 4L-DNN. The implicit leading fraction in the right plot is calculated from predicted bundle energy and leading muon energy. The explicit leading fraction in the middle plot is predicted as a separate label by the network. The blue marks are identical in all three plots and represent the mean of the leading fraction in the bundle energy bins based on the simulation.

as was intended but instead learned to build an energy based mean using the reconstructed bundle energy. In the context of prompt analysis the assumption that a high energetic bundle always has a very subdominant leading muon is problematic, because it automatically excludes the network from estimating a high energy above 100 TeV for the leading muon. Instead, it leads to the systematic underestimation of events with a high leading fraction (Figure ??) and the migration of high energetic events to lower energy bins in the prediction (Figure ??). Considering that the explicit leading fraction is slightly less restricted for high energies than the implicit leading fraction there is the possibility that at least small improvements on the results can be made by using the explicit leading fraction to calculate the energy of the leading muon.

### 3.3 Optimization

Based on these results efforts can be made to try and improve the reconstruction of the energy of the leading muon and the bundle energy by changing the network's architecture. By increasing the network size to 12 layers, increasing the number

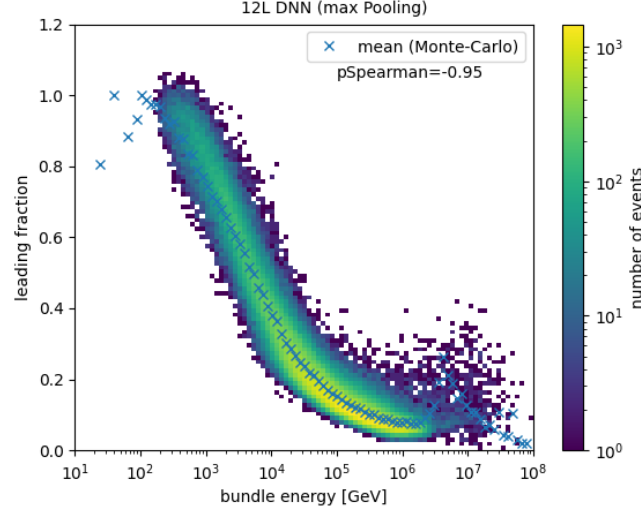


**Figure 3.11:** Correlation plot and energy spectrum for a 12L network. The energy of the leading muon is calculated from the predicted leading fraction and bundle energy instead of being directly estimated. Improvements in linear and monotonic correlation are small, but the resulting spectrum shows significant improvements compared to previous models.

of filters and switching from average to max pooling as well as calculating the leading muon energy from the bundle energy and the explicit leading fraction a significantly improved reconstructed energy spectrum can be reached (Figure ??). Those improvements are not visible in the Pearson or Spearman coefficients and even

### 3 Reconstruction performance

shows a lower Pearson coefficient, but the mean indicates less underestimation for high energies than the other models. The most impactful change inferred from this is the increased amount of statistic above the critical 100 TeV threshold with almost 22,3 % of events actually in that region predicted to be in that region compared to 8,9 % for the 4L version using the direct approach for the leading muon energy. Although this new model offers strong improvements over the other reconstructions



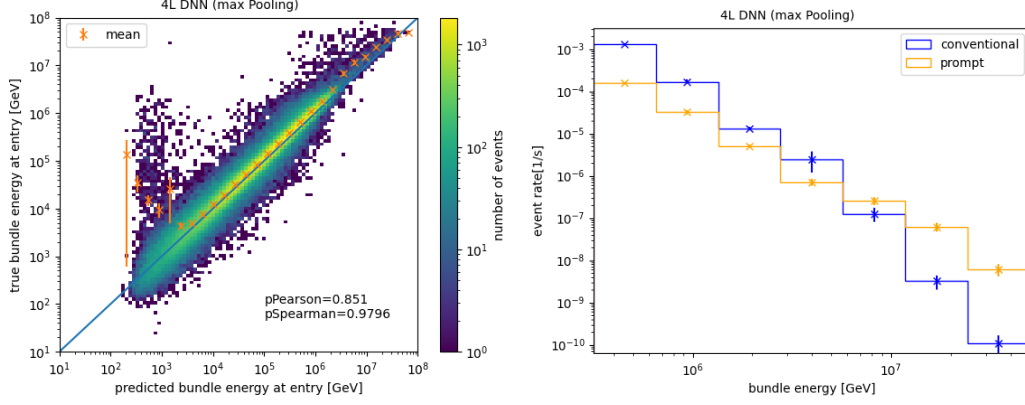
**Figure 3.12:** Correlation between bundle energy and leading fraction for the 12L-DNN with the mean of the corresponding plot done on the Monte-Carlo simulation.

especially in terms of the energy spectrum generally and the prompt sensitivity specifically the Results still deviate from the simulated truth by one or two orders of magnitude and is not competitive compared to the reconstruction of the bundle energy. Another look at the correlation of bundle energy and leading fraction (Figure ??) shows that even when making changes such as increasing the model size, the underlying problem remains the same. As long as the prediction is made solely based on a correlation between bundle energy and leading fraction to an extent made up by the DNN the potential quality of the reconstruction is limited. To find out, if there is a chance to get a higher quality reconstruction the networks' prediction of the leading muon energy first has to be separated from the predicted bundle energy.

In contrast to the prediction of the leading muon energy the bundle energy is already reconstructed well even for small models. The only significant point causing problems in this reconstruction is the bias to underestimation in the high energy region which is most likely a systematic problem that will not be fixed by increasing



network size, so it makes more sense to try improving the smaller network. The easiest way to make an improvement here is to switch from average to max pooling. As seen in Figure ?? the resulting network outperforms the 4L DNN with average



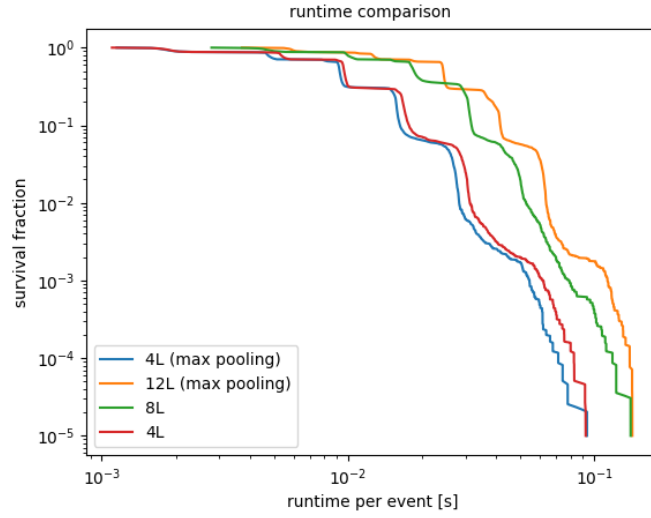
**Figure 3.13:** Correlation plot (left) and energy spectrum (right) of a 4L DNN using max instead of average pooling. Compared to the version of the same network using average pooling (Figures ??, ??) it has a higher Pearson coefficient and better sensitivity for the prompt component especially under the consideration of uncertainties.

pooling and even the 8L DNN when it comes to the linear correlation (Figure ??) and also displays the better high energy spectrum when it comes to prompt sensitivity.

Another logical step is to use a single-label reconstruction for the bundle energy because previous results have shown, that the energy of the leading muon is at least currently not suitable for a prompt analysis. So utilizing the improvements offered by focusing the network on a single label has no significant disadvantages in this scenario.

## 4 Runtime comparison

An important point to consider when evaluating the performance of the networks additionally to the quality of the reconstruction is the runtime. Because atmospheric muons are created and detected at an extremely high rate the reconstructions applied to those events have to have a correspondingly low runtime. The runtime performance



**Figure 4.1:** Displayed on the y-axis is the fraction of events exceeding the runtime on the x-axis. As expected runtime generally increases with model complexity. Per event runtimes are given in reference to the time it takes the network to make a prediction. For the absolute runtime the preprocessing in the form of data transformation has to be considered. When comparing both networks this is unnecessary because the time it takes to preprocess does not depend on the network architecture.

of both networks is shown in the form of survival fractions in Figure ???. The 4L network has a median prediction time of  $\text{median}(t_{\text{pred},4\text{L}}) = 0,0096 \text{ s}$  which is almost twice as fast as the median of the 8L network ( $\text{median}(t_{\text{pred},8\text{L}}) = 0,0185 \text{ s}$ ) which in turn is faster than the median of the 12L network  $\text{median}(t_{\text{pred},12\text{L}}) = 0,0247 \text{ s}$ . A similar result is obtained by looking at the ratio of the runtimes of the DNNs. The median of the ratio  $\text{median}\left(\frac{t_{\text{pred},8\text{L}}}{t_{\text{pred},4\text{L}}}\right) = 1,88$  indicates that the 4L DNN is almost

---

twice as fast at making a prediction. Although there also are instances in which the 8L-DNN is faster, those only make up 0,1 percent of all events.

A more interesting result is the deviating runtime between the 4L model running average pooling and 4L model running max pooling. Both models are otherwise not exactly identical because the model using max pooling has less convolutional layers in the lower DeepCore array, but the effects of this change are very small. Even though both networks share the same architecture aside from the pooling method Figure ?? suggests that the network using max pooling is faster. This is supported by the median runtime of  $\text{median}(t_{\text{pred},4\text{Lmax}}) = 0,0096 \text{ s}$ . The benefit in runtime is only small with a ratio of  $\text{median}\left(\frac{t_{\text{pred},4\text{L}}}{t_{\text{pred},4\text{Lmax}}}\right) = 1,057$  indicating a runtime improvement of over 5 % achieved by using max pooling. But considering this as well as fact that the version using max pooling is faster for 90,9 % of all events it further supports the results obtained in the previous chapter suggesting that at least for a small network max pooling is the better option because it is both faster and delivers better reconstruction results.

## 5 Conclusion and outlook

The energy reconstruction of muons using machine learning methods shows differing results with respect to the application in a potential analysis of the prompt component of the atmospheric muon flux depending on what specifically is reconstructed.

Promising results are obtained for the reconstruction of the bundle energy, which can be estimated well enough even by a relatively small neural network to distinct the prompt component as the dominant contribution in the high energy domain as demonstrated in Figure ???. The study of the non machine learning method MuEX shows worse reconstructions, which makes it impossible to detect the prompt component as shown in Figure ???. Depending on the runtime requirements, a network architecture with an increased number of convolutional layers can be used to achieve slight improvements as seen in Figures ??? and ??? but considering the increase in runtime visible in Figure ??? a smaller alternative is the better option in most cases. With a median runtime of 0,0092 s the 4L DNN with max pooling is twice as fast as the 8L network with a median runtime of 0,0185 s and still shows comparable reconstruction results.

In contrast to that the reconstruction of the energy of the leading muon proves to be more difficult. While it theoretically provides the better sensitivity for a prompt analysis this effect is completely negated by the reconstruction-quality. The energy reconstruction of the leading muon is not competitive compared to the more reliable bundle energy reconstructions. This can be seen in Figure ?? and ???. This discrepancy can be traced back to the strong correlation between bundle energy and leading fraction assumed by the network and displayed in ??? leading it to avoid extreme values. There are roughly three possible reasons for this phenomenon: The input features do not provide enough information to make assumptions about the stochasticity of the energy loss, the stochasticity seen by the DNN does not provide sufficient information about the dominance of the leading muon within the bundle or the stochasticity is not properly used to make a prediction by the network. Whether there is no way to find a better mode of action provided only the three input features used here or whether it was not found by these specific network architectures can not be ultimately concluded.

To investigate the first two possible reasons it is necessary to create a dataset providing a separate label for the stochasticity. Another network can be trained

---

to predict this label, so it can be evaluated, if the stochasticity can be sufficiently estimated. This procedure also enables an evaluation of the assumed correlation between the stochasticity and the leading fraction. If the prediction of the stochasticity performs well, the estimation of the leading muon energy can be improved by changing the network architecture to use the provided information more efficiently. Such optimizations can be performed by a hyperparameter optimization. Possible improvements in performance can be reached by increasing the number of input features to increase the amount of information the network can utilize. Of nine possible features only three are used for this analysis. Additional charge related features like the charge collected in the first 100 ns and 500 ns respectively can improve information about the energy loss. All in all, several networks have been studied and presented to estimate the muon bundle and leading muon energy sufficiently to provide a fast and accurate energy reconstruction to measure the prompt component of the atmospheric muon flux.

## Danksagung

Ich möchte mich zum Schluss in allen Bedanken, die mir den Abschluss dieser Arbeit ermöglicht haben. Besonderer Dank gilt Herrn Prof. Dr. Dr. Wolfgang Rhode sowie Herrn Prof Dr. Carsten Westphal. Weiterhin möchte mich speziell bei Pascal und Mirco für die ausführliche Betreuung meiner Arbeit sowie bei Lucas, Ludwig und Jean-Marco für ihre große Hilfsbereitschaft bedanken. Auch bei den weiteren Mitgliedern der Arbeitsgruppe bedanke ich mich für zahlreiche interessante Gespräche und Diskussionen.

Zuletzt möchte ich mich bei meiner Familie für ihre Unterstützung bedanken, ohne die diese Arbeit nie möglich gewesen wäre.



# Eidesstattliche Versicherung

(Affidavit)

Flottau, Leander  
Name, Vorname  
(surname, first name)

222740  
Matrikelnummer  
(student ID number)

☒ Bachelorarbeit  
(Bachelor's thesis)

☐ Masterarbeit  
(Master's thesis)

Titel  
(Title)

A deep learning based reconstruction of  
high-energy muons in IceCube

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem oben genannten Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present thesis with the above-mentioned title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution before.

Lünen, 29.08.2023  
Ort, Datum  
(place, date)

L. Flottau  
Unterschrift  
(signature)

## Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen.

## Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the Chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, Section 63 (5) North Rhine-Westphalia Higher Education Act (*Hochschulgesetz, HG*).

The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund University will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification.\*

Lünen, 29.08.2023  
Ort, Datum  
(place, date)

L. Flottau  
Unterschrift  
(signature)

\*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.