KIT - Department of Physics
Institute of Experimental Particle Physics (ETP)
and the Institute for Astroparticle Physics (IAP)

Neural Network Analysis of High-Energy Cosmic Ray Air Showers measured with IceCube

Neuronale Netzwerkanalyse von mit IceCube gemessenen hochenergetischen kosmischen Teilchenschauern

Bachelor Thesis
by

Jan Kremer

at the Institute of Experimental Particle Physics (ETP)
and the Institute for Astroparticle Physics (IAP)

Reviewer:	Prof. Dr. R. Engel
Second Reviewer:	Dr. A. Haungs
Advisor:	Dr. D. Kang

Jun 2023 – Sep 2023

Erklärung zur Selbstständigkeit

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der gültigen Fassung vom 24.05.2018 beachtet habe.

Karlsruhe, den , _____

Jan Kremer

Als Prüfungsexemplar genehmigt von

Karlsruhe, den , _____

Prof. Dr. R. Engel

Abstract

Neural Network Analysis of High-Energy Cosmic Ray Air Showers measured with IceCube

The IceCube Neutrino Observatory at the South Pole consists of two main detectors, a grid of ice-Cherenkov tanks at the surface called IceTop and a deep-ice Cherenkov detector called in-ice array. These two detectors are used to measure extensive air showers initiated by cosmic rays in the energy range of PeV to EeV. By measuring various features of these air showers, the characteristics of the primary particle, such as its mass and energy, can be reconstructed. Hereby, the accuracy of the reconstruction can be improved by considering the correlation between the features, e.g. by using neural networks.
Recently, a new method to estimate the density of GeV muons, which is sensitive to the mass of the primary particle, was developed. In order to test the feasibility of the application of this muon density parameter, this study uses this parameter as an additional feature for neural networks. This work is based on the neural network architecture of the previous publication "Cosmic Ray Spectrum and Composition from PeV to EeV Using 3 Years of Data From IceTop and IceCube" [Aar+19].
For this analysis, a Monte-Carlo simulation dataset is used which is equally distributed among four primaries, namely: proton, helium, oxygen, and iron. This work looks at different neural network architectures, with and without the muon density parameter as an additional feature, and shows that the muon density parameter slightly improves the mass but not the energy predictions.

Zusammenfassung

Neuronale Netzwerkanalyse von mit IceCube gemessenen hochenergetischen kosmischen Teilchenschauern

Das Neutrino-Observatorium IceCube am Südpol besteht aus zwei Hauptdetektoren, einem Array von eisgefüllten Cherenkov-Tanks, das IceTop genannt wird und einem Cherenkov-Detektor im tiefen Eis, welcher als In-Ice-Array bezeichnet wird. Mit diesen beiden Detektoren werden die durch hochenergetische kosmische Strahlung verursachte Luftschauer detektiert. Durch die Messung verschiedener Parameter dieser Luftschauer können die Eigenschaften der Primärteilchen, wie Masse und Energie, rekonstruiert werden. Dabei kann die Genauigkeit der Rekonstruktion verbessert werden, indem die Korrelation zwischen den verschiedenen Parametern berücksichtigt wird, z. B. durch den Einsatz neuronaler Netze.
Vor kurzem wurde eine neue Methode zur Bestimmung der Dichte von GeV Myonen entwickelt, die empfindlich auf die Masse des Primärteilchens reagiert. In dieser Studie wird dieser Parameter als zusätzliches Feature für neuronale Netze verwendet, welche auf der Netzwerkarchitektur aus der früheren Veröffentlichung "Cosmic Ray Spectrum and Composition from PeV to EeV Using 3 Years of Data From IceTop and IceCube" basieren, [Aar+19].
Die Analyse verwendet einen simulierten Monte-Carlo-Datensatz, der gleichmäßig auf die vier Primärteilchen Proton, Helium, Sauerstoff und Eisen verteilt ist. In dieser Studie werden verschiedene neuronale Netzwerkarchitekturen mit und ohne den Myonendichteparameter als zusätzlicher Parameter untersucht, und es wird gezeigt, dass der Myonendichteparameter die Massenvorhersagen leicht verbessert, nicht aber die Energievorhersagen.

Contents

1 Introduction

The Earth's atmosphere is constantly bombarded by extraterrestrial particles, which can come from the Sun, the solar system, the Milky Way, or from distant galaxies. These particles are called cosmic rays. Since their discovery about 100 years ago by Victor Hess [Wal12], there have been two main questions in the field of cosmic ray research: what are cosmic rays made of, and where do they come from?
This study attempts to determine the mass and energy of cosmic rays reaching Earth using Neural Networks.
Because high-energy cosmic rays are so rare that they cannot be detected directly, they are observed indirectly through the study of extensive air showers, which are created when a cosmic ray interacts with a nucleus in the atmosphere, leading to a cascading effect that produces a large amount of secondary particles. These particles can then be detected by ground-based observations, one of these being the IceCube Neutrino Observatory at the South Pole.
The IceCube observatory consists of two main parts: IceTop, which is an array of 162 ice-filled Cherenkov detectors on top of a glacier, and IceCube, which is a deep-ice Cherenkov detector, which uses the glacier ice below IceTop as a detector.
This analysis is based on the neural network architecture of the previous publication "Cosmic Ray Spectrum and Composition from PeV to EeV Using 3 Years of Data From IceTop and IceCube" [Aar+19], which takes in five IceCube-specific high-level features of an extensive air shower measured by IceTop and IceCube, and predicts the mass and energy of the cosmic ray initiating that shower.
The goal of this analysis is to include a new feature, the density of GeV muons measured by IceTop, to study the effect of the muon parameter on the quality of the determination of the primary's energy and mass composition. The physics idea behind this is that heavier particles produce more muons for the same energy than lighter particles, and therefore this feature could help predict the mass of the primary particle with better accuracy than without it.

This thesis is structured as follows: Chapter 2 gives a brief overview of cosmic rays, including a section on the history, composition, energy spectrum, and extensive air showers. In Chapter 3 the IceCube Neutrino Observatory is described. Chapter 4 gives a summary of the Monte Carlo simulation used in this thesis. In Chapter 5 an introduction to machine learning is given and in Chapter 6, the data analysis is discussed.

The analysis is structured as follows:
The calculation of the muon density parameter, the training of a benchmark model without the new feature, the comparison of the benchmark model with

a model whose only difference is the addition of the muon density as a new feature, and the search and comparison of a completely new model, which also uses the muon density as a feature, with the benchmark model.

In the last Chapter, a conclusion of the work is given, and an outlook is discussed.

2 Cosmic Rays

2.1 History

After the discovery of spontaneous ionization of gases, the general belief was that this phenomenon is caused by radioactive material in the Earth's crust. However, Charles Wilson was the first to suggest an extraterrestrial origin for this phenomenon [Gin96]. To validate this hypothesis, Victor Hess conducted several balloon flights, during which he measured the atmospheric ionization at different altitudes. His idea being that if the Earth's surface is the source of ionization, then atmospheric ionization would decrease with altitude. On the other hand, if the source is extraterrestrial, ionization would increase. Hess's observations confirmed the latter, as the ionization increases with altitude, effectively ruling out the Earth as the source. Based on this discovery, Hess undertook several additional balloon flights during the night to eliminate the sun as the source of ionization. Finding a consistency of ionization levels independently of the time of day, he concluded that the sun isn't the source of ionization [Wal12][Gin96].
In 1945, Pierre Auger was one of the first to identify the source of this ionization as cosmic ray induced air showers. He reached this conclusion based on his experiments, in which he observed coincident events in Geiger counters placed several meters apart [Aug+39][Com36].

2.2 Composition

There are several types of cosmic rays, but this thesis focuses mainly on charged cosmic rays, which consist mostly of nuclei with charge numbers from $Z = 1$ to around 92 [Mew94]. The distribution of these different nuclei is not uniform, resulting in a higher flux for e.g. proton, helium and iron primaries. Cosmic rays are roughly composed of 89% hydrogen, followed by 9% helium, 2% heavier nuclei, and less than 1% electrons [Feu13]. The odd-even effect, where nuclei with an even number of nucleons are more stable and therefore more abundant, also influences the composition [Gai16].

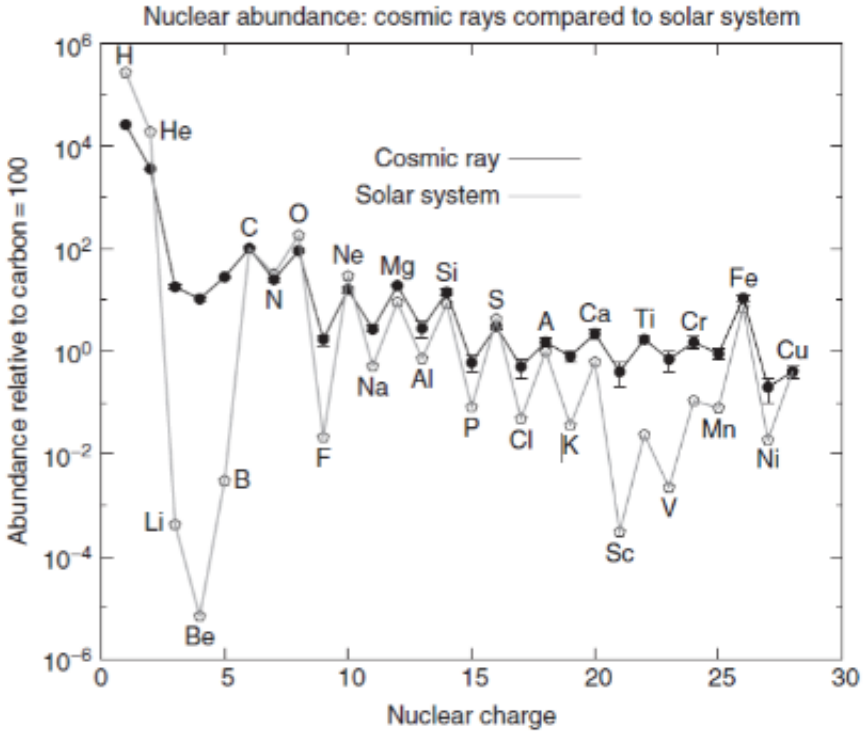


Figure 2.1: The cosmic ray elemental abundances in cosmic radiation measured on Earth compared to the solar system abundances, relative to carbon = 100, for cosmic rays in GeV range [Gai16].

As shown in figure 2.1, the elemental composition of cosmic rays reaching Earth does not match the abundances of the corresponding elements in the Solar System. The fact that elements with $Z > 1$ are much more abundant than protons relative to the Solar System material is not fully understood. Possible reasons for this discrepancy include the relative difficulty of ionizing hydrogen and injecting it into an accelerator process. It is also plausible that the composition at the source differs from that observed on Earth.
The light nuclei Li, Be and B, as well as the heavier nuclei Sc, Ti, V, Cr and Mn, are much more abundant in cosmic rays than in the Solar System. This is a well understood phenomenon. While relatively absent as final products of stellar nucleosynthesis, they are produced as spallation products when higher-mass elements like carbon, oxygen or iron collide with the interstellar medium. For this reason, the ratio of primary particles to spallation products serves as an indicator of the amount of matter that cosmic rays have passed through the interstellar medium [Gai16][Gab+19].

2.3 Energy Spectrum

The energy spectrum of cosmic rays closely resembles the power law:

$$\frac{\mathrm{d}N}{\mathrm{d}E} \propto E^{-\gamma} \tag{2.1}$$

where $\frac{\mathrm{d}N}{\mathrm{d}E}$ represents the flux of cosmic rays given in $\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$, E is the primary energy, and γ is the spectral index, which is about 2.7 for $E < 10^{15}\mathrm{eV}$. To make the important characteristics of the spectrum more visible, it is often multiplied by a power of energy, typically $E^{2.6}$. A plot of the all-particle energy spectrum measured by different experiments can be seen in figure 2.2. The plot shows distinct features, including the knee at around $(3-5)\cdot 10^{15}\mathrm{eV}$, the second knee at around $10^{17}\,\mathrm{eV}$, and the ankle at around $10^{18.5}\mathrm{eV}$.
The origin of these features is not fully understood. The leading theory is that the knee and second knee are caused by cosmic rays from a galactic source reaching their maximum energy. There are two proposed explanations for this. The first suggests that the maximum energy depends on the charge of the cosmic ray. The maximum energy is then given by

$$E_{max} \propto Z \cdot B \cdot R, \tag{2.2}$$

where Z is the charge of the cosmic ray particle, and B and R are the magnetic strength and the radius of the accelerating region, respectively. The second explanation is based on particle physics models and states that the maximum energy depends on the atomic mass A rather than the charge Z.
Thus, it follows that the first and second knees have the same underlying causes, and their position in relation to each other depends on the type of primary, with lighter particles dropping out of the spectrum earlier, than heavier particles. Furthermore, the transition from a galactic to an extra-galactic source is thought to be the reason for the ankle. The suppression at high energies, around $10^{19}\,\mathrm{eV}$, is consistent with the GZK effect. The GZK effect describes how cosmic rays of extra-galactic origin interact with the cosmic microwave background, leading to scattering or decay processes [Alv+19][Uni].

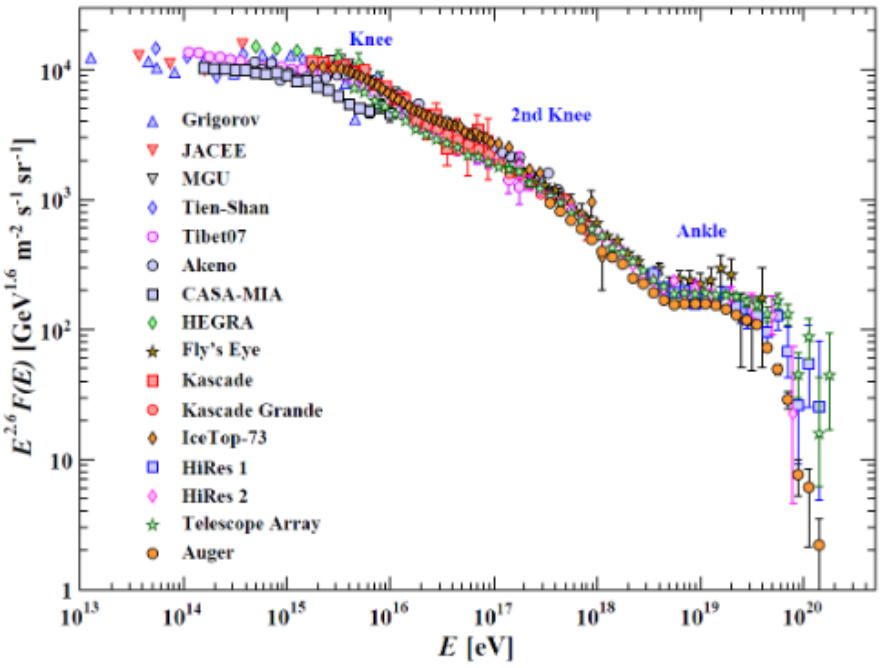


Figure 2.2: All-particle energy spectrum of cosmic rays, multiplied by a factor of $E^{2.6}$, as measured by various observatories [Gro+20].

2.4 Extensive air showers

An extensive air shower is formed when a single high-energy primary cosmic ray enters the Earth's atmosphere and interacts with its air nuclei. This interaction produces secondary particles, which may decay or interact further with the atmosphere or with each other, leading to the production of more particles, resulting in a cascading effect. An air shower can be divided into three components: the hadronic, the electromagnetic, and the mounic components.
The hadronic component consists of nucleons, pions, and kaons, and is produced when the primary particle interacts with an air nucleus, leading to the creation of a region of high-energy hadrons near the shower core. Furthermore, the interaction and decays of the hadronic component lead to the creation of the electromagnetic and muonic components [Mat05].

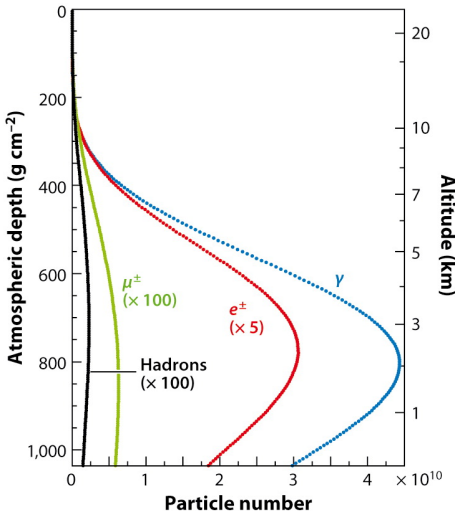


Figure 2.3: Longitudinal shower profile for different particle types for a proton-induced shower with an energy of 10^{19} eV simulated by CORSIKA for the observation level of the Pierre Auger Observatory. The plot shows that the shower size increases until it reaches X_{max}, in this case about 800 g/cm^2, after which it begins to decrease [EHP11].

The electromagnetic component consists of electrons, positrons and photons. The creation of the electromagnetic component begins with the decay of a neutral pion into two photons. These photons can create positron-electron pairs by pair production. At the same time, electrons and positrons produce photons by bremsstrahlung. This interaction causes a positive feedback loop, initiating a particle cascade that continues until the critical energy is reached, at which point the collisional energy losses begin to exceed the radiative losses. Once this energy is reached, the shower reaches its maximum size, after which it begins to shrink. The atmospheric depth at which this maximum number of particles is reached is called X_{max}. An example of this growing and shrinking of the EM component for a proton-initiated shower of energy 10^{19} eV can be seen in figure 2.3.
The majority of muons in an extensive air shower are produced by the decay of charged kaons and pions. Due to the long lifetime of relativistic muons and their small cross section, most of these muons reach the ground without decaying. This also explains the continuous growth of the muonic component shown in the figure 2.3.

3 The IceCube Neutrino Observatory

3.1 Design

3.1.1 Overview

The IceCube Neutrino Observatory is located at the South Pole, near the Amundsen-Scott South Pole Station. The Observatory is involved in the study of neutrinos, dark matter, multi-messenger observations and cosmic ray research. It has four main detector components: IceTop, the IceCube In-Ice Array, the IceCube Laboratory and the DeepCore.
The observatory operates as an ice Cherenkov detector, using the Cherenkov effect to detect the Cherenkov photons emitted by particles traversing the ice at speeds faster than the in-ice speed of light. The main device used to detect these Cherenkov photons is the Digital Optical Module (DOM). Each DOM is encased in a pressure-resistant glass sphere to withstand the pressure of the ice. The DOMs consist of four main parts: a 10-inch PMT (photomultiplier tube) to detect the Cherenkov photons, a high-voltage power supply, an LED flasher board and a main board that processes and digitizes the PMT signals [Aar+17][De 19].

3.1.2 IceCube In-Ice Array

The In-Ice Array consists of 5160 DOMs deployed at depths of approximately 1450 m to 2450 m below the ice. The DOMs are attached to 86 vertical strings, with each string containing 60 DOMs. The primary array consists of 78 of these strings, with a vertical DOM spacing of 17 m on each string. In figure 3.1 it can be seen that the strings are arranged in a triangular grid, with a grid length of about 125 m, while the general outline of the grid forms a hexagonal shape[Aar+17].

3.1.3 DeepCore

The DeepCore Array is a subarray of the In-Ice Array located at its center. It is deployed deeper in the ice, than the In-Ice Array, at approximately 1750 m. The array consists of seven IceCube strings and eight specialised DeepCore strings. The DeepCore strings differ from the IceCube strings in their DOM spacing. Specifically, the DOM spacing on DeepCore strings is smaller. Six of the DeepCore strings use a different PMT than the standard IceCube strings, resulting in a 35% increase in quantum efficiency. This denser arrangement and increased efficiency allows for a lower energy threshold of about 10 GeV compared to the typical IceCube threshold of about 100 GeV, making the DeepCore array better suited for low energy observations and neutrino oscillation studies [Aar+17].

3.1.4 IceTop

IceTop is located on the surface of the ice, directly above IceCube, at an altitude of 2835 m above sea level. It uses roughly the same grid layout as the in-ice array. IceTop consists of 81 stations, each of which contains two ice Cherenkov tanks spaced 10 m apart, for a total of 162 tanks. At the center of IceTop there are eight densely arranged stations corresponding to the DeepCore strings, as mentioned above. The tanks are filled with ice to a height of approximately 90 cm and each tank contains two standard IceCube DOMs, one operating in high gain mode and the other in low gain mode [Aar+17].

3.1.5 Lab

The IceCube Lab is a two-story building located at the center of the array. It serves as the central operational facility for the observatory, where all data is collected and pre-processed [Aar+17].

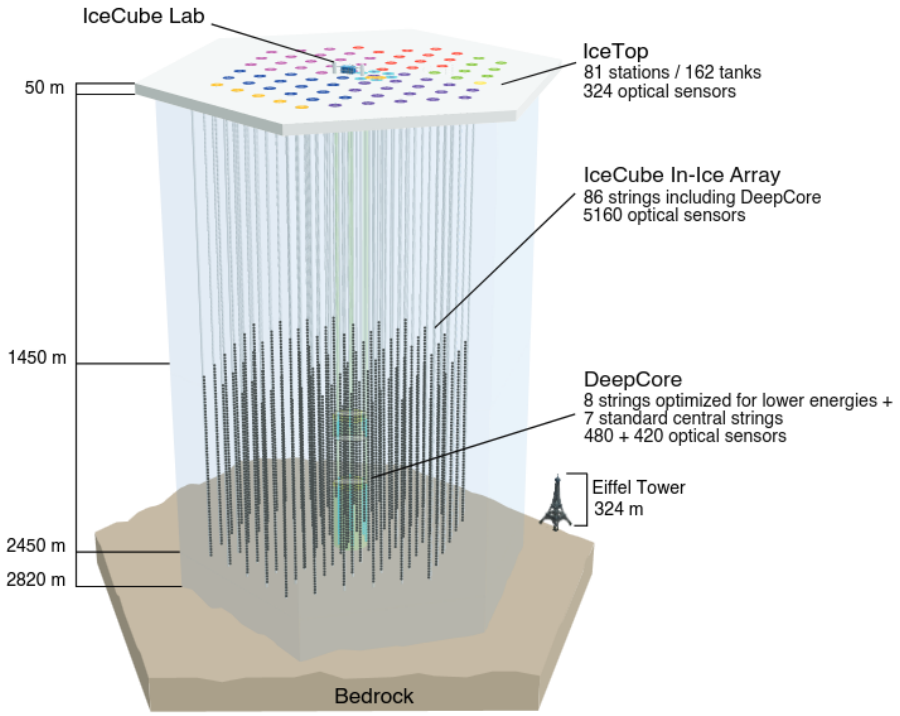


Figure 3.1: Schematic of the IceCube Observatory. The different colors of the stations are for the different years the stations and their accompanying strings were deployed[Aar+17].

3.2 Data Acquisition Systems and Reconstruction

3.2.1 IceTop

Calibration

Single atmospheric muons are used to calibrate the IceTop DOMs. The mean energy of these muons is about 2 GeV[De 19]. Since in this energy range they are MIPs (Minimum Ionizing Particles), their energy loss, in accordance with the Bethe-Bloch formula, is nearly independent of their energy. The unit of calibration is called VEM (Vertical Equivalent Muon), where one VEM represents the photocharge induced by one muon hitting the tank vertically[Aar+17][De 19].

Measurement

Depending on the coincidence level, IceTop has two mutually exclusive types of photocharge measurements: Hard Local Coincidence HLC for high coincidence events and Soft Local Coincidence (SLC) for low coincidence events.
The coincidence level is defined within a station. If both tanks of a station measure an event of which at least one triggers a high gain DOM within a time window of $1\mu s$, it is considered a high coincidence event and the HLC signal is recorded. Conversely, if only one tank of a station detects an event, it is a low coincidence event and the SLC signal is recorded.
For an HLC measurement, the full digitized waveform is stored. For a SLC measurement, only the integrated photocharge and a timestamp is stored [De 19][Aar+19].

Reconstruction

After a shower is measured, it is processed by the so-called Laputop software, which is a maximum-likelihood reconstruction algorithm for IceTop. For each shower, the core position (x_c,y_c,z_c), the direction (θ, ϕ), and the LDF (Lateral Distribution Function) of the photocharge are reconstructed using Laputop. The LDF is described by the parameters κ and β, and is of the form:

$$S(r) = S_{125} \cdot \left(\frac{r}{125m}\right)^{-\beta-\kappa \log_{10}\left(\frac{r}{125m}\right)}. \tag{3.1}$$

In equation 3.1, S_{125} is the charge signal in VEM at the reference distance of 125 m from the shower axis and r is the radial distance from the shower axis. The parameter β describes the steepness and κ characterizes the curvature of the LDF. For hadronic showers, κ remains nearly constant, so by default it is set to a value of 0.303 [Aar+19].

3.2.2 In-Ice

Measurement

Within the in-ice array, signals are also differentiated into HLC and SLC based on the coincidence level. For the in-ice detector, the coincidence level is determined on a string-by-string basis. If an event is registered by a neighboring DOM or a second neighboring DOM in addition to the DOM itself within a time window of 1 μs, then it is classified as an HLC event, else it is an SLC event [De 19].

Reconstruction

The Millipede software is used to reconstruct the in-ice pulses. Millipede reconstructs the in-ice energy loss $\frac{\mathrm{d}E}{\mathrm{d}X}$, in steps of 20 m. From the reconstructed energy loss, three metrics used in this work can be derived, namely $\frac{\mathrm{d}E_{1500}}{\mathrm{d}X}$, as well as a standard and a strong selection of the stochastic energy loss.
The metric $\frac{\mathrm{d}E_{1500}}{\mathrm{d}X}$ is the energy loss at a slant depth of 1500 m, which is approximately the top of the in-ice detector array. The standard and strong selections denote the number of bins in a $X - \frac{\mathrm{d}E}{\mathrm{d}X}$ plot, where the bin width is given by the reconstruction length of 20 m, that exceeds a certain threshold. The following formula defines this threshold

$$\frac{\mathrm{d}E_\mu}{\mathrm{d}X} > a \cdot \left(\frac{\mathrm{d}E_\mu}{\mathrm{d}X}\right)^b_{reco}, \tag{3.2}$$

where $a = 5$ and $b = 0.8$ for the standard selection, and $a = 7$ and $b = 0.9$ for the strong selection [Aar+19]. An example of determining these metrics for one event is shown in figure 3.2.

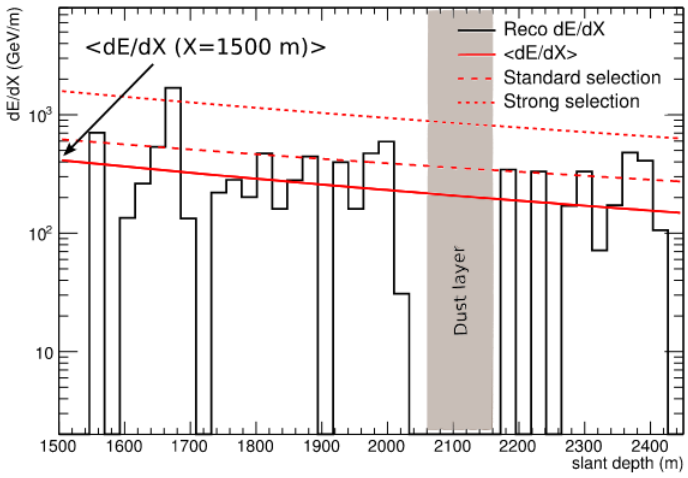


Figure 3.2: An example of reconstructed energy loss profile from a large event [Aar+19]. The solid red line indicates the average energy loss fit, the dashed red line presents the standard stochastics selection, and the dotted red line is the strong stochastics selection. The gray band presents the approximate location of the dust layer.

4 Simulation

4.1 Air Shower Simulation

The Monte Carlo simulation of the extensive air shower is performed using the CORSIKA [Hec+98] (Cosmic Ray Simulation for KASCADE) software. CORSIKA models the development of an air shower following the introduction of a primary particle. To simulate an air shower, the following parameters have to be provided to CORSIKA: the direction and the primary energy of the particle (both can be discrete values or distributions), the particle type, a model of the atmosphere, a hadronic interaction model and an electromagnetic interaction model.
In this thesis, CORSIKA simulations utilizing, the EGS4 [NHR85] model for electromagnetic interactions and the Sibyll 2.1 [Ahn+09] model for high-energy hadronic interactions, are used.
After being given the initial parameters, CORSIKA simulates the shower step by step for each atmospheric altitude slice until the observation level is reached, which for IceCube is about 2837 m [De 19]. During the simulation, the space-time coordinates of each particle are tracked until they fall below a certain energy threshold. When this happens, they are discarded and their energy deposit is recorded at the altitude at which it occurred. For IceCube, only the particles that reach the observation level are of interest, because only these can be detected by the observatory.

4.2 Detector Simulation

The GEANT4 [Ago+03] software is used to simulate the detector response. To ensure comparability with real data, the PMT response, DAQ signal processing, and trigger are simulated. This ensures that the same Laputop / Millipede reconstruction can be applied to both real and simulated data. To achieve higher statistics, the same CORSIKA showers are used multiple times at different locations of the detector array.
At the beginning of the simulation, the shower is placed at a random position in the detector array. Then the particles are read out at the observation level and propagated through the last parts of the atmosphere, the snow and finally the detector, where their response is simulated. In the simulation, only muons with energy greater than 273 GeV [De 19] are propagated into the in-ice array [De 19][Aar+19].

5 Neural Networks

Neural Networks (NNs) are useful when dealing with large amounts of data, especially when the goal is to identify underlying patterns or structures that are not readily apparent using conventional methods. Essentially, a NN is a sophisticated non-linear function f that maps an input vector, X, to an output vector, y, commonly referred to as features and targets, respectively,

$$f(X) = y. \tag{5.1}$$

While there are many types of NNs, this thesis focuses on multi-layer perceptrons (MLPs).

5.1 MLPs

5.1.1 The perceptron

The basic unit of a MLP is called a perceptron. As shown in figure 5.1, a perceptron takes an input vector x, computes the scalar product with its intrinsic vector w, in Machine Learning (ML) the vector w is called a weight vector. In order to account for data that does not intersect the origin, a bias term b is added to the scalar product. It is common to introduce an extra dimension to the input vector, where $x_i = 1$ and $w_i = b$, effectively embedding the bias into the scalar product.
After the scalar product is calculated, a non-linear function σ, called an activation function, is applied to the result. The reason for using the activation function is that since the scalar product is a linear operation, the MLP would not be able to fit non-linear data without using a non-linearity. The equation for a perceptron is then as follows,

$$a = \sigma\left(\sum_i w_i x_i + b\right), \tag{5.2}$$

with a being the output of the perceptron.

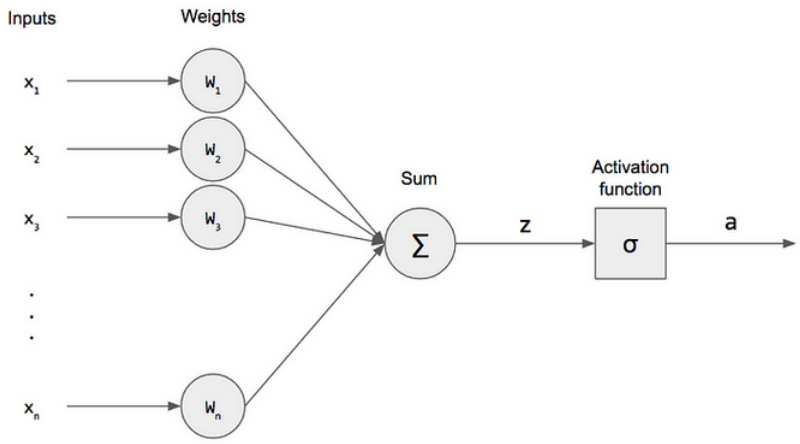


Figure 5.1: Schematic of a perceptron [Duk18].

5.1.2 MLP

A MLP arranges perceptrons into layers, and combines several of them to form a computational graph. In this structure, the perceptrons serve as the nodes of the graph and are therefore commonly referred to simply as nodes. The layers within the MLP are fully connected, which means that each node in a layer is connected to all the nodes in the following layer. Consequently, the input vector x for a node is composed of the outputs of each node in the previous layer. A schematic overview of such a graph/network can be seen in figure 5.2.
There are three types of layers within a MLP: the input layer, the output layer and the hidden layers. As the name suggests, the input layer serves as the entry point of the network and has no preceding layer. It takes the feature vector X as its input. Therefore, its number of nodes is equal to the number of features. The output layer, on the other hand, is the final layer of the network. Its outputs form the target vector y. Thus, the number of nodes in the output layer have to match the dimension of the target vector. The layers between the input and output layers are called hidden layers. Their number and size are not predetermined. Instead, determining their optimal number and size often requires iterative testing. All parameters, except the weights and biases, which are not intrinsically given by the problem have to be optimized by the user, they are called hyperparameters. Algorithms called optimizers are used to optimize the weights and biases.

5.2 Training

As mentioned above, a neural network essentially is a function that maps a feature vector to an output vector. In the context of MLPs, the process of finding

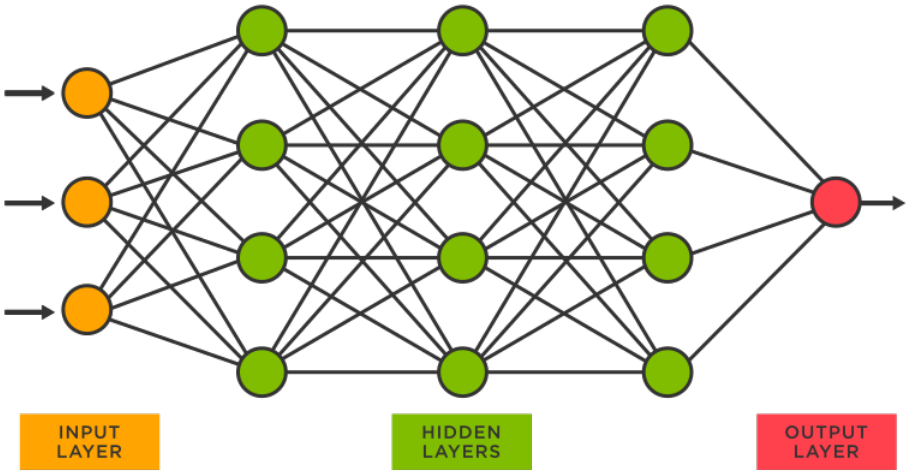


Figure 5.2: Layers of a Neural Network [Inc].

the best parameters for this function to best fit a given dataset, consisting of feature vectors and their corresponding target vectors, is called training. The metric used to quantify how well the MLP fits the data is called the loss function, or simply the loss. There are many loss functions available, one of the most commonly used is the Mean Squared Error (MSE) loss,

$$loss = \frac{1}{N}\sum_{i}^{N}(y_{pred,i} - y_i)^2 = \frac{1}{N}\sum_{i}^{N}(f(X) - y_i)^2, \tag{5.3}$$

which is also the one used in this thesis. In this equation, $y_{pred,i}$ is the prediction of the model for the ith data point, as determined by $f(X_i) = y_{pred,i}$, while y_i is the actual target value for the same data point, and N is the number of data points in the dataset.
This allows us to redefine the task of finding the best parameters for the model as finding the parameters that minimize the loss function.

5.2.1 Optimizer

Finding the minimum of the loss function analytically is difficult. Therefore, iterative algorithms are used to incrementally approach the minimum. These algorithms are called optimizers. There are several optimizers available, in this work the Adam [KB17] and the SGD [Rud16] (Stochastic Gradient Descent) optimizers were tried out.

5.2.2 Training Loop

Since the optimizer converges only step by step, a training loop is used to find the optimal model. In machine learning the number of optimizer steps over the whole dataset is called epochs and the number of epochs is an important hyperparameter because it specifies how long the model is trained for. If it is trained for too short it could be possible not to reach the minimum of the loss function, this is called underfitting, on the other hand if the model is trained for too long, one gets overfitting. At this point, the model is too well optimized for the training data to make good predictions for unseen data (one says the model remembers the data). To prevent the model from overtraining, the so-called validation loss is used. The validation loss is the loss of a subset of the data, called the validation dataset, which is not used for training and therefore can be used to judge the performance of the model on unseen data.
After training, the model's performance is evaluated by calculating the loss on a dataset called the test dataset, which is not used for training or decision-making during the training process.

5.2.3 Frameworks

For the implementation and training of the neural networks in this study, PyTorch [Pas+19] is used as a machine learning framework. Additionally, scikit-learn [Ped+11] is used for data preprocessing.

6 Analysis

6.1 Dataset

For this analysis a dataset of 240,000 showers simulated with CORSIKA for the year 2012, following an E^{-1} distribution, is used. The dataset contains an equal mixture of four primary particles, namely: protons, helium, oxygen, and iron. In CORSIKA, the Sibyll 2.1 model was used as the high-energy hadronic interaction model, and the atmosphere was based on the average atmospheric conditions of April. The simulated showers have an energy ranging from $\log_{10}(E/\mathrm{GeV}) = 5$ to 8 and zenith angles from 0° to 65°.

On this dataset, the following standard quality cuts for IceTop and In-Ice were applied to each event, analog to those in the paper [Aar+19].

- The number of stations after cleaning is required to be ≥ 5.
- The largest snow-corrected charge measured in any tank is required to be at least 6 VEM.
- The station with the highest deposited charge is not at the edge of the detector.
- The neighboring tank in the same station as the tank with the largest signal must have at least 4 VEM.
- The fraction of hit stations within a circle centered on the center of gravity of the shower with outer radius at the furthest hit station must be greater than 0.2.
- The Laputop reconstruction must be successful.
- The LDF slope parameter β is required to be between 1.4 and 9.5.
- A minimum of 8 In-Ice DOMs are required to be hit.
- The Millipede energy loss reconstruction must succeed, with $\log_{10}(rlogl) < 2.0$.
- The total predicted Millipede charge must be at least 90% of the measured charge.
- At least 3 reconstructed cascades remain after all previous selections and after removal of cascades in the dust layer and at the edge of the detector.

Since this work focuses only on vertical showers, an additional cut is applied:

- The cosine of the zenith angle is between 1 and 0.95, which is approximately 0° to 18°.

After the quality cuts, the simulation dataset contains about 56000 events, which will be used for the following analysis. The dataset is nearly equally distributed in the four primary particles and follows the energy distribution shown in figure 6.1.

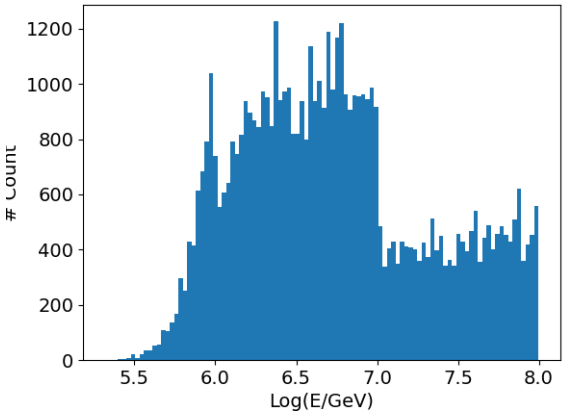


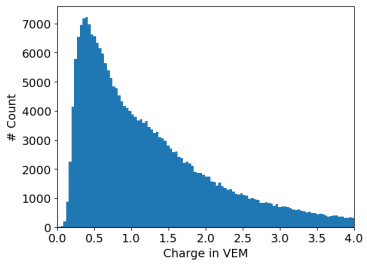
Figure 6.1: True energy distribution of the dataset used in this analysis.

The sudden decrease in the number of events, seen in Fig. 6.1, at $\log_{10}(E/\mathrm{GeV}) = 7$ is due to the change of the resampling radius for the detector simulations at this energy, from 1100 m to 1700 m [De 19]. The steady growth from 5 to 7, on the other hand, is due to the higher detector efficiency at higher energies.

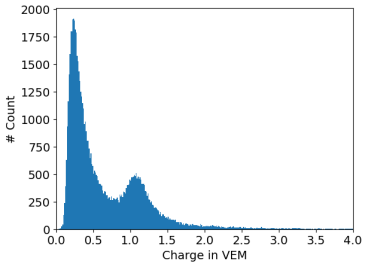
6.2 Muon Density Parameter

6.2.1 Charge Spectrum

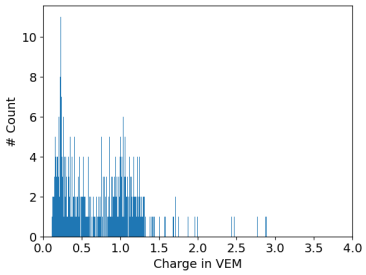
The muon density, as a new feature compared to the analysis described in [Aar+19], is calculated from the combined HLC and SLC charges of the IceTop array. To calculate the muon density, the charge spectrum is examined in radial ranges away from the shower core. This is because near the shower core, the muon peak is overshadowed by the electromagnetic (EM) induced charges. The charge distribution induced by EM components decreases exponentially with distance from the shower core.



(a) Charge spectrum for a radius range of $100\,\mathrm{m} \leq r < 200\,\mathrm{m}$, because this is close to the shower core only the EM induced charges are visible.



(b) Charge spectrum for a radius range of $300\,\mathrm{m} \leq r < 400\,\mathrm{m}$, at this point the EM peak has decreased enough to make the muon peak become visible.



(c) Charge spectrum for a radius range of $800\,\mathrm{m} \leq r < 900\,\mathrm{m}$, at this radius range the EM and muon peaks are well separated, but only a few events induce charges at such great distances, because of this the statistic is not high enough to make good predictions.

Figure 6.2: Charge signal distribution of combined SLC and HLC charges for different radii for all events with $0 \leq \log(S125) < 0.5$.

Furthermore, as a result of the calibration of the IceTop tanks as described in section 3.2.1, the position of the muon peak by definition is fixed at a charge of 1 VEM. Because of these two effects, there exists a radius range away from the shower core, where the EM peak has decreased enough to make the muon peak become visible, this can be seen in figure 6.2b. Here the first peak at about 0.4 VEM is the EM peak, while the peak at 1 VEM is the muon peak. As can be seen in figure 6.2c, at even greater distances the EM peak decreases even further, so that the EM and muon peaks barely overlap. Although this is ideal to calculate the muon density, this radius range should not be used for estimating the muon density, because the sensitive area and density of the detector is too small to get high enough statistics for such large distances. Therefore, a balance

must be struck between the muon peak's prominence and detector's efficiency. Since the height and decay rate of the EM peak depend on the primary energy, the radius range used to calculate the muon density has to be chosen depending on the log($S125$) signal of the shower, which is a proxy for the energy. To find the best range, several radii ranges were considered for each energy range. The best fitting ranges were then selected based on plots of the charge spectrum for different energy and radius ranges, these plots can be found in Appendix A.1. An overview of the chosen ranges for this analysis can be seen in table 6.1.

Table 6.1: $\log_{10}(S125)$ range and the corresponding radius bins.

$\log_{10}(S125)$ Bin	Radius Bin
$0 \leq \log_{10}(S125) < 0.5$	300 m- 400 m
$0.5 \leq \log_{10}(S125) < 1$	350 m- 450 m
$1 \leq \log_{10}(S125) < 1.5$	400 m- 500 m
$1.5 \leq \log_{10}(S125)$	500 m- 600 m

6.2.2 Calculation of the muon density parameter

The estimated muon density is defined as the sum of the muon enhanced charges divided by the effective detector area within a specific radius range. Therefore, the muon density parameter is calculated in the following way,

$$\rho_\mu = \frac{1}{N_{tanks} \cdot A_{eff}} \cdot \sum_{q_i \geq 0.7\,\mathrm{VEM}} q_i. \tag{6.1}$$

Here N_{tanks} is the number of tanks in the considered radius range, A_{eff} is the effective area of one tank, given by $A_{eff} = A_{top} \cos\theta + A_{side} \sin\theta$, and q_i is the combined SLC and HLC charges [KBH21]. The cut at 0.7 VEM is made to select the approximated muon charge. The position of the cut is taken from the valley position of the charge spectrum, between the muon peak and the EM peak. Because the muon peak and the EM peak overlap, this is only an estimate of the muon induced charge. A histogram of the muon density calculated with this method for iron with the radius ranging from 350 m to 450 m and a $\log_{10}(S125)$ range of 0.5 to 1 can be seen in figure 6.3.

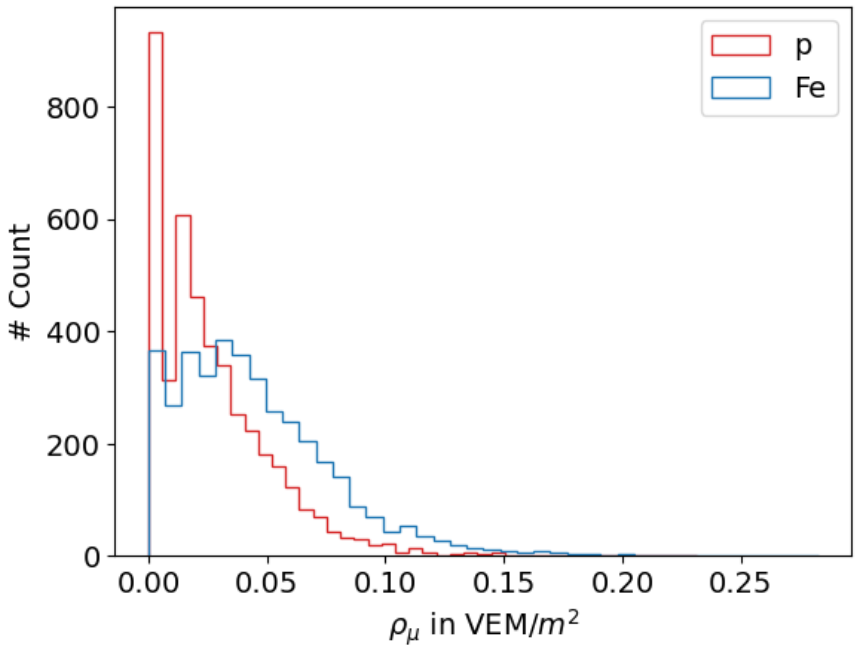


Figure 6.3: An example, histogram of the muon density parameter for proton and iron, with the radius ranging from 350 m to 450 m and a $\log_{10}(S125)$ range of 0.5 to 1. The large number of events with an apparent muon density of zero stems from showers, which hit the edge of the detector and so have a lower probability to induce a charge grater than 0.7 VEM in a tank.

6.3 Targets

The two targets for this analysis, namely $\log(E/\text{GeV})$ and $\ln(A)$, are quite different in nature. While $\log(E/\text{GeV})$ is continuously distributed, $\ln(A)$ has a discrete distribution. Although $\ln(A)$ is discrete, classification was not used to predict $\ln(A)$, this allows for the detection of primary types which are not present in the dataset. Further $\ln(A)$ was chosen over A, because the four primaries: proton, helium, oxygen and iron are nearly equidistant in $\ln(A)$, but not in A.

6.4 Features

6.4.1 Muon Density Parameter

The muon density parameter mainly serves as an indicator for the primary particle of the shower. This is because, for the same energy, the muon density

of an iron induced showers is higher than that of a proton induced showers, which can be explained by the superposition model [KU12].

6.4.2 dE/dX

Since dE/dX is a metric from the in-ice array, it is primarily sensitive to high energy muons. As iron induced showers contain a larger number of muons than proton induced showers at similar energy levels, iron showers result in a larger Cherenkov deposit for the same energy. Therefore, when looking at dE/dX at similar energies, it is primarily an indicator for the primary mass [Aar+19].

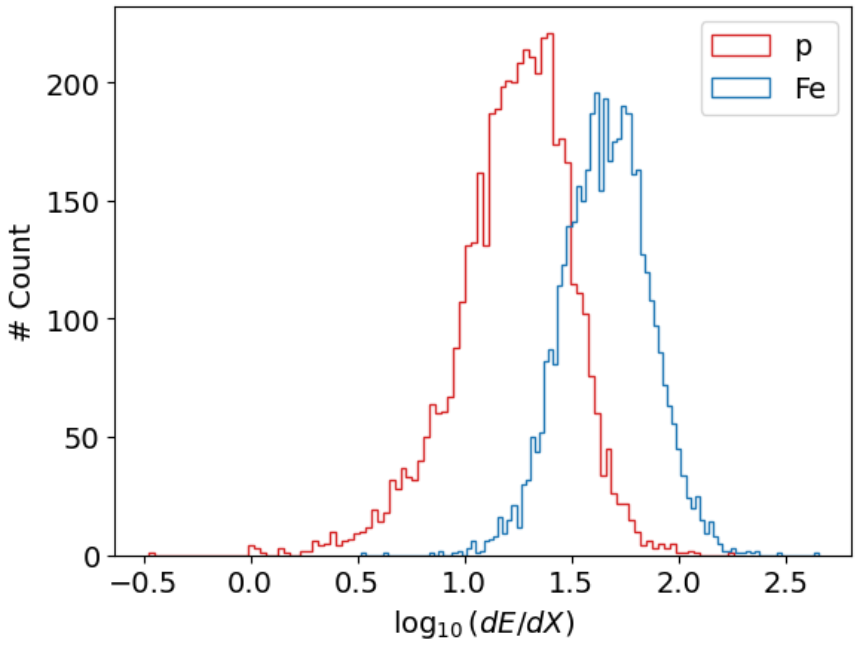


Figure 6.4: An example, histogram for $\log_{10}(\frac{\mathrm{d}E_{1500}}{\mathrm{d}X})$ for iron and proton for an energy range of $6.5 \leq \log_{10}(E/\mathrm{GeV}) < 7$.

6.4.3 Standard + Strong selection

Since the standard and strong selections stochastics are also metrics of the in-ice array, they are also mainly sensitive to high energy muons. As the selections evaluate how many times the energy loss at each slant depth exceeds a certain threshold, which is defined in section 3.2, they serve as a metric for discrete energy losses due to bremsstrahlung. Since, iron produces more but lower energy muons compared to protons, which produce fewer but higher energy muons at the same primary energy. It is expected that at the same primary energy (i.e. $\log_{10}(S125)$) iron showers will have a higher standard selection but a lower

strong selection than proton induced showers [Aar+19]. Thus, the standard and strong selections, when viewed at the same energy, are also mainly indicators of the primary type.

6.4.4 $\log_{10}(S125)$

The reference distance of 125 m for $\log_{10}(S125)$ is chosen in such a way, that the parameter is nearly independent of the primary type and mainly depends on the primary energy. Therefore, $\log_{10}(S125)$ is the main indicator for the primary energy. The figure 6.5 shows the linear relationship between $\log_{10}(S125)$ and $\log(E/\mathrm{GeV})$.

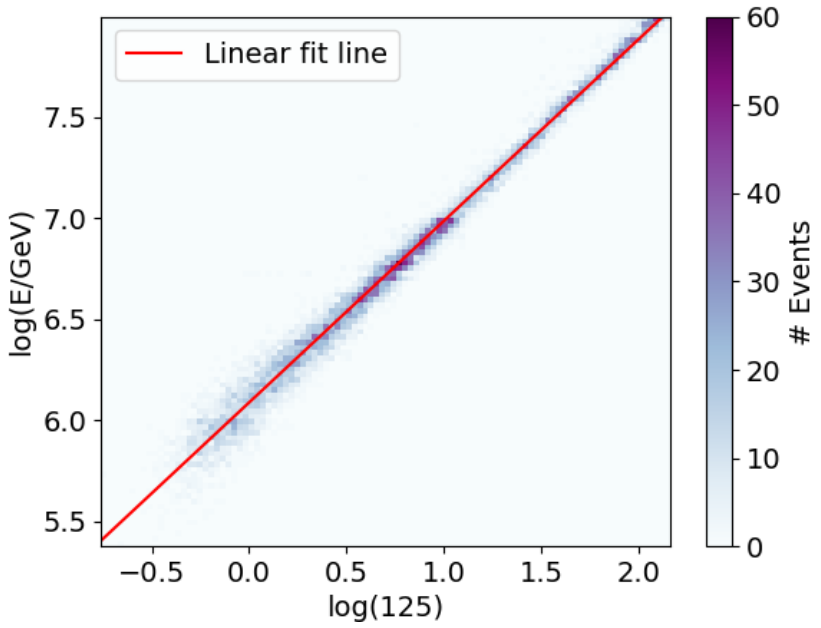


Figure 6.5: Correlation between $\log(E/\mathrm{GeV})$ and $\log_{10}(S125)$ over the dataset used in this study, with a linear fit.

6.4.5 $\cos(\theta)$

Although the zenith angle is not a direct indicator of primary energy or type, it is an important feature, because the other reconstructed features depend on the direction of the shower.

6.5 Training Methodology

To train the Neural Network (NN) for this analysis, a common procedure is used. The procedure is as follows:
Before training begins, the dataset is split into three subsets: the training, the validation, and the test dataset. For this, a split of 60-20-20 is used.
Since the majority of activation functions is most sensitive to changes around zero, the data is downscaled to a range of 0 to 1 using the Scikit-learn Min-Max scaler [Ped+]. For this, downscaling and upscaling layers are added before the input and after the output layer, respectively. This makes it possible to input unscaled data directly into the network, which simplifies operation immensely.
Using the same 5-7-4-2 network structure as in the previous publication [Aar+19], the Adam and SGD optimizers were evaluated on a subset of 10,000 samples, for different learning rates. It can be seen in table 6.2, that their performance is quite close, but SGD performs slightly better, than Adam. Therefore, it is used as the optimizer for the rest of the analysis.
To adjust the learning rate during training, the ReduceLROnPlateau function [Fou] from PyTorch is used. This function reduces the learning rate each time a metric, in this case the validation loss, stops improving.
This is done because using learning rate decay, allows training to start with a higher learning rate, without overshooting the minimum of the validation loss, at the end of training. This speeds up the training of a model, therefore, by using a learning rate decay function, more models can be tried without sacrificing performance, increasing the probability of finding a better performing model.
By testing different parameters for the ReduceLROnPlateau function, the following parameters were found to work best: initial learning rate = 0.01, cooldown = 500, decay factor = 0.5, and patience = 10.
Since it is necessary to train beyond the actual optimum to detect that the validation loss has stopped improving. The model is not in the state with the best validation loss when the decay happens, so to avoid having to continue training with a suboptimal model, each time a decay occurs, the parameters with the lowest validation loss are loaded in.
For this to work, each time the model reaches a new minimum for the validation loss, the model's parameters are saved.
There comes a point where the learning rate has decayed so often that it becomes too small for any significant optimization to take place. Once this point is reached, training can be stopped. For this analysis, the threshold is set to $lr < 0.5$.

Table 6.2: Test and validation loss for the 5-7-4-2 model for different optimizes and learning rates.

learning rate	Optimizer	Trainings Loss	Validation Loss
0.001	Adam	0.7015	0.7409
	SGD	0.7027	0.739
0.005	Adam	0.6981	0.7437
	SGD	0.7021	0.737
0.01	Adam	0.7113	0.7428
	SGD	0.7019	0.7536
0.025	Adam	0.7303	0.7426
	SGD	0.7105	0.7801
0.05	Adam	0.7423	0.7573
	SGD	0.7266	0.7622
0.1	Adam	0.7837	0.8867
	SGD	27.628	28.03

6.6 Metrics

Because it is difficult to evaluate a model's performance on its validation loss alone, four additional metrics are used during training. These are: the R^2 value and the MSE of the energy, for energy predictions, and the separation and MSE of $\ln(A)$, for $\ln(A)$ predictions. Hereby, the separation is defined as follows,

$$separation = \sum_{(i,j)} |\mathrm{KDE}_i - \mathrm{KDE}_j|, \tag{6.2}$$

where i, j are the primary particles and the sum is over all combinations (not permutations).
The separate calculation of the MSE for energy and $\ln(A)$ allows to distinguish between the loss due to the energy prediction and the loss due to the mass prediction. This is done because looking at the combined average of both can be misleading. For example, two models could have the same total loss, but one may have a better energy prediction while the other performs better at mass predictions, or vice versa.
Since in an E_{true}, $E_{predicted}$ plot the relationship between the two should be linear, the R^2 value is used in addition to the MSE as a metric for the energy prediction of a model.
The MSE of $\ln(A)$ serves as a metric for the distance of the predicted mass from the actual target mass, for example, when the MSE of $\ln(A)$ is calculated for a single event with protons as the primary, it shows how much the predicted value for $\ln(A)$ deviates from zero. Therefore, the MSE of $\ln(A)$ is a metric for how well a model performs in predicting the actual mass of a primary.
The separation, on the other hand, is completely independent of the actual mass

of the particles and thus cannot serve as a metric for it. Rather, the separation is a metric of how well the network can distinguish between different primary types.

To evaluate the performance of the model after training, the MSE for energy and $\ln(A)$, and the R^2 value for the energy, are calculated on the test dataset. This is done, because so the model's performance on unseen data can be evaluated, making it a valid indicator of its real world performance.
Instead of using the separation to evaluate a model's performance to differentiate between the different primary types, the normalized overlap areas (NOAs) of the Kernel Density Estimations (KDEs) are used after training. This is done because the NOAs are more physically relevant than the separation. Also, due to its computational cost, the NOAs are not calculated during training.
Furthermore, in order to obtain a single number that reflects the model's performance to differentiate between different primary types over the whole energy range, the mean NOA over the whole energy range and all combination of primary particles types is calculated.

7 Results

7.1 Baseline Model

Since the previous publication [Aar+19] uses a different dataset, with more events and a different energy range, than the one used for this work, and to account for different training methodologies and random fluctuations, a new baseline model analogous to the 5-7-4-2 network from the paper [Aar+19], with the same features, specifically $\log(S125)$, the standard selection, the strong selection, dE/dX, and $\cos(\theta)$, and the same activation function, namely tanh, is trained again for this work.
This model serves as a benchmark against which all other models used in this thesis are compared to. To train the baseline model, the procedure described in the previous section, along with MSELoss as the loss function, was used. The performance of this model will be discussed in the next section.

7.1.1 Performance

Energy Performance

To evaluate the model's performance for energy predictions, both the R^2 and the MSE of the energy are used. For the baseline model, these are: 98.7% and 0.0044, respectively. In addition, to evaluate the performance of the energy prediction in different energy bins, energy resolution and bias are checked. Here, the bias and the resolution of the energy reconstruction are defined as the mean and the standard deviation, of the ratio between the true energy and the reconstructed energy, respectively. A plot of the bias and resolution for different reconstituted energy ranges can be seen in figure 7.1. Since the dataset does not contain enough events with energies lower than $\log_{10}(E/\text{GeV}) < 6$ to make an informed statement about the model's performance for this energy range, it is excluded from the plot.
Furthermore, the sudden drop at the last bin is most likely due to boundary effects, which are caused by incomplete data in the boundary regions when the analysis window gets closer to the edge of the data range [SLL13].

The figure 7.1 shows, that the energy resolution improves with increasing energy. This is expected since higher energy showers contain more particles, and thus the Laputop / Millipede reconstruction for the features gets better at higher energies, leading to better energy predictions.
The plot also demonstrates that for the same energy, particles with heavier masses have better resolution than those with lighter masses, this can be explained by the superposition model, which states that at the same energy, a

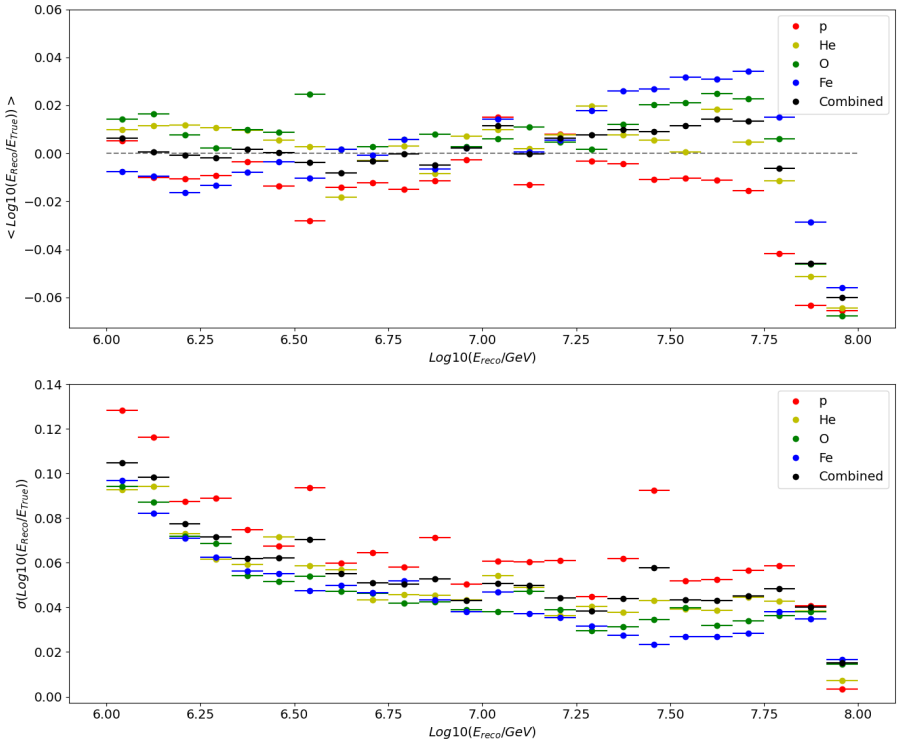


Figure 7.1: Energy reconstruction bias (upper) and resolution (lower) as a function of the reconstructed energy for the different primary types and for an equal mixture of each type.

shower with a heavier mass primary contains more particles than a shower with a lighter mass primary. Therefore, for the same reason that the reconstruction is better at higher energies, the reconstruction for heavier particles is better, than that for lighter ones, at the same energy.

The resolution plot is similar to the one from the [Aar+19] paper, with the two main differences being a consistent offset of about 0.01 over the energy spectrum, which is most likely due to different training methodology, and the absence of the sudden drop of the resolution at $log(E/\text{GeV}) = 8$. The absence of this drop for the [Aar+19] paper is most likely due to the fact, that the [Aar+19] paper's dataset includes higher-energy events, and therefore no boundary effects occur at $log(E/\text{GeV}) > 8$.

The bias plot shows that the bias is close to zero in a range of approximately ± 0.02. From the plot, it can be seen that the bias from $\log(E/\text{GeV}) = 6$ to 7 gets better with increasing energy, which is due to the lower shower fluctuations at higher energies [Aar+19]. Around $\log(E/\text{GeV}) = 7$ there is a change in the

trend of the bias, which coincides with the energy at which the resampling rate for the shower simulations changes, and above which the dataset contains only about half the number of events for each bin, compared to before the change, this can also be seen in figure 6.1. Therefore, the change in trend is most likely, due to the different number of training data in the two ranges.
The sudden drop for the last few bins is again most likely due to boundary effects.

Mass Performance

For the baseline model, the MSE of the mass is about 1.388 and the average NOA is about 0.561.
In figure 7.2 the NOAs for all combinations of primary types and different energy ranges can be seen. To calculate the NOAs the KDEs shown in figure 7.4 are used. The plot 7.2 shows that, for all primary combinations, the NOAs improve with increasing energy. This is due to lower intrinsic shower fluctuations at higher energies. It is also clear that it is easier to distinguish between primaries with large differences in mass than those with closer masses. This explains the relatively high NOAs for elements whose masses are close together.
To evaluate the performance of the baseline model in predicting the actual mass of a primary particle, the MSE of $\ln(A)$ for different energy ranges is plotted, see figure 7.3.

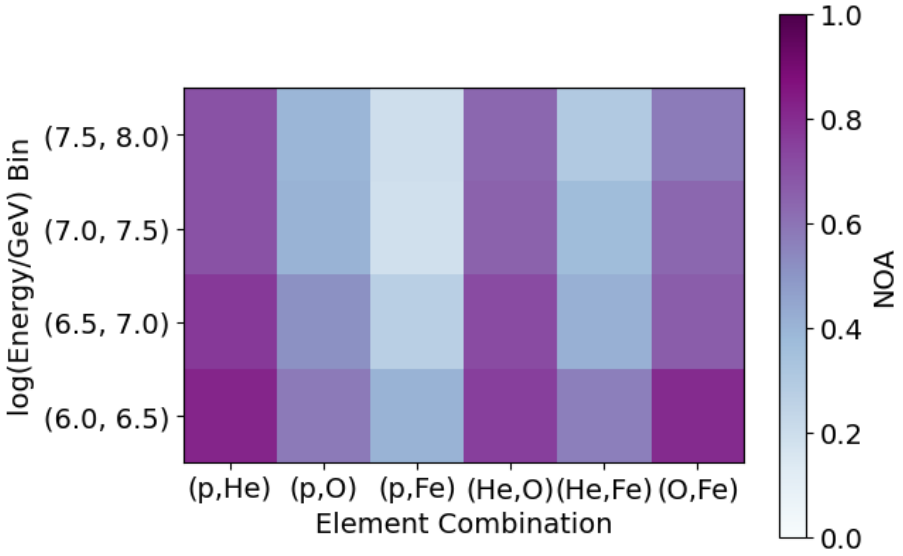


Figure 7.2: Normalized overlap area (NOA) for all combinations for different energy bins. A NOA close to zero means that the two elements are well separated, while a value close to one indicates that they are almost identical.

From the plot 7.3 it can be seen that the mass prediction for oxygen and helium is almost independent of the energy, while the mass prediction for proton and iron, on the other hand, becomes much better at higher energies. This can also be seen in the KDEs in figure 7.4, while the shape of the KDEs for oxygen and helium is similar for all energy ranges, the KDEs of iron and proton are a lot peakier at higher energies.
Comparing the KDEs for the baseline model with the KDEs from the paper [Aar+19], it can be seen that the KDEs for different energy ranges for proton, oxygen, and iron look very similar for both models, while the KDE for helium in the baseline model is peakier, than in the paper [Aar+19]. This is probably due to the fact that [Aar+19] uses A instead of $\ln(A)$ as the target for the mass, which makes it difficult to distinguish between proton and helium, since in A both are much closer to each other compared to the other primaries.

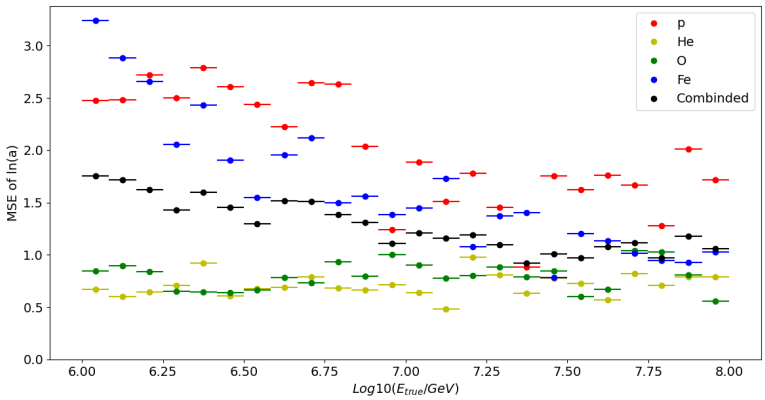
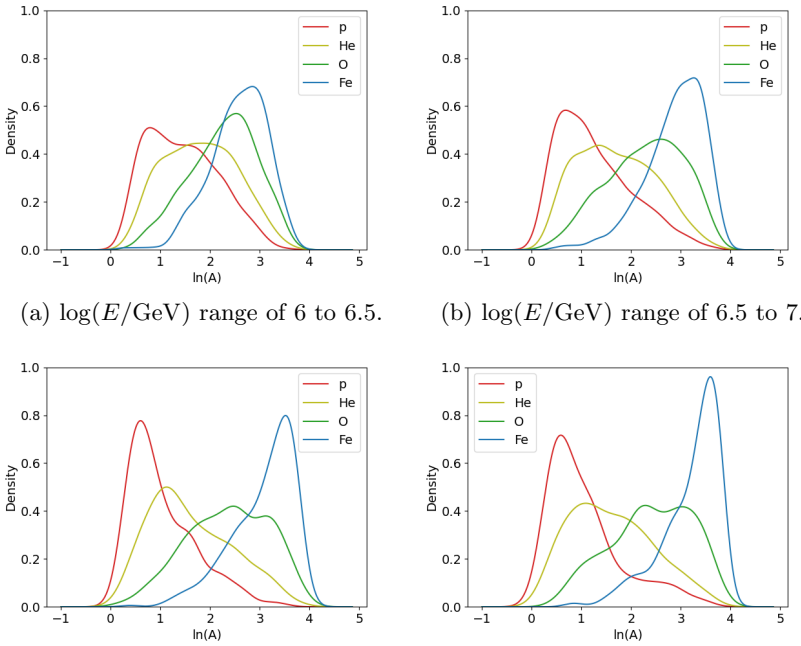


Figure 7.3: MSE of $\ln(A)$ as a function of the true energy for the different primary types and for an equal mixture of each type.



(a) $\log(E/\text{GeV})$ range of 6 to 6.5.

(b) $\log(E/\text{GeV})$ range of 6.5 to 7.

(c) $\log(E/\text{GeV})$ range of 7 to 7.5.

(d) $\log(E/\text{GeV})$ range of 7.5 to 8.

Figure 7.4: KDEs of the baseline model for different energy ranges used to calculate the NOAs.

Conclusion

In conclusion, the baseline model, within the considered energy range, is overall comparable to the model from the paper [Aar+19]. Therefore, it can be used as a benchmark for the models with the added muon density parameter. The effect of the addition of the new parameter can then be determined by comparing the models with the new parameter with this benchmark model. This will be discussed in the next section.

7.2 Model with added muon parameter

The main goal for this study is to determine if the muon parameter has an impact on the performance of the model, and thus a physical merit. A model with the same architecture as the baseline model is trained again, using the same common training method as described above, with the only difference being the addition of the muon density parameter as an extra feature. This requires adding an extra node to the input layer, which makes the model with the added muon parameter a 6-7-4-2 model.

7.2.1 Performance

Energy

This model has an MSE of the energy of about 0.0046 and a R^2 value of about 98.7%. When compared with the model without the muon density parameter, it can be seen that the muon density parameter, as calculated in this thesis, has no significant impact on the energy performance, when looked at over the whole energy range.
To check if the addition of the muon density parameter has an effect in certain energy ranges, bias and resolution distribution are generated, as can be seen if figure 7.5.
Figure 7.6 displays the comparison of bias and resolution between baseline and model with muon parameter for an equal mixture of primaries. In the plot it can be seen, that for the bias, the general tendency is similar, but there is a slight difference at around $\log_{10}(E/\text{GeV}) = 7.25$. The reason might be different amount of simulation dataset, rather than the muon parameter. In case of resolution, both distributions show almost identical.
The fact that the resolution and bias of the baseline model and the model with additional muon density are nearly identical for all energy ranges might suggest that the dependence of the muon density parameter on energy is too small compared to that of the other features to have a significant effect on the energy reconstruction.

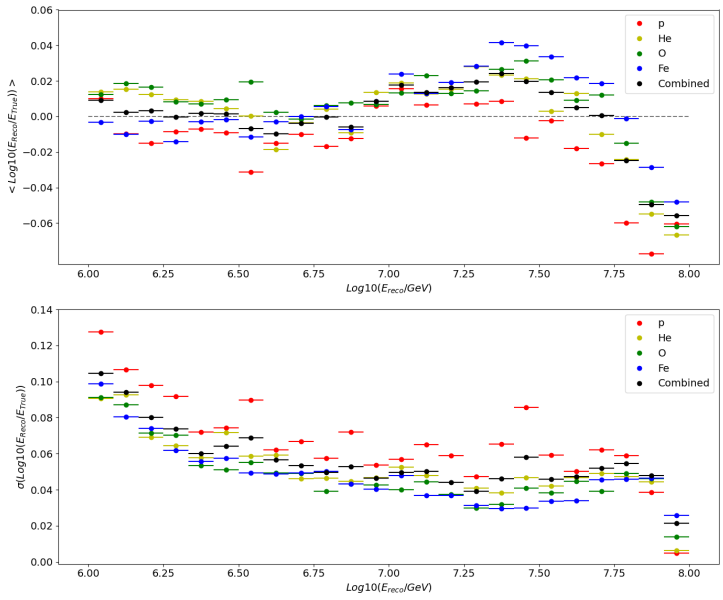


Figure 7.5: Bias (top) and resolution (bottom) of the 6-7-4-2 model's energy reconstruction for different energy bins.

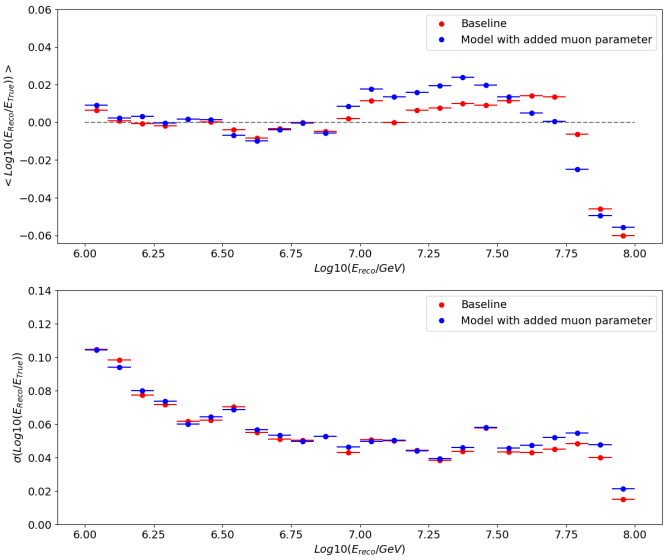


Figure 7.6: Bias (top) and resolution (bottom) for all primaries combined as a function of true energy for a comparison between baseline (red) and model with muon parameter (blue).

Mass

For the model with adding muon density parameter, the MSE of the mass is about 1.36 and the average NOA is about 0.549.
This means that the model with added muon density is better at predicting the true mass of the primary particles and also better at separating the primary particles than the baseline model, when compared over the full energy range. This is the expected behavior since, as described in section 6.4, the muon density is mostly an indicator of mass when looked at the same energy range.
To evaluate the ability of the model to separate between different primary types over different energy ranges, the same NOA plot as for the baseline model, is performed for the model with the added muon density, this plot is shown in figure 7.7. Comparing the two plots, they look almost the same, so to make the differences better visible, a comparison plot is performed, as shown in figure 7.8. From this plot it can be seen that the muon density has only a small overall effect on the NOA, it can also be seen that it has the largest effect at higher energies, with the only outliers being the first two combinations. The improvement at higher energies is probably due to higher muon production at higher energies, which leads to a better reconstruction of the muon density at higher energies.

The KDEs of the 6-7-4-2 model used to calculate the NOAs can be seen in figure 7.9. The figure also shows that the KDEs of the 6-7-4-2 and the baseline model have a similar distribution for all primaries and all energies. This means that the muon density parameter only has a minor effect on the shape of the KDEs.

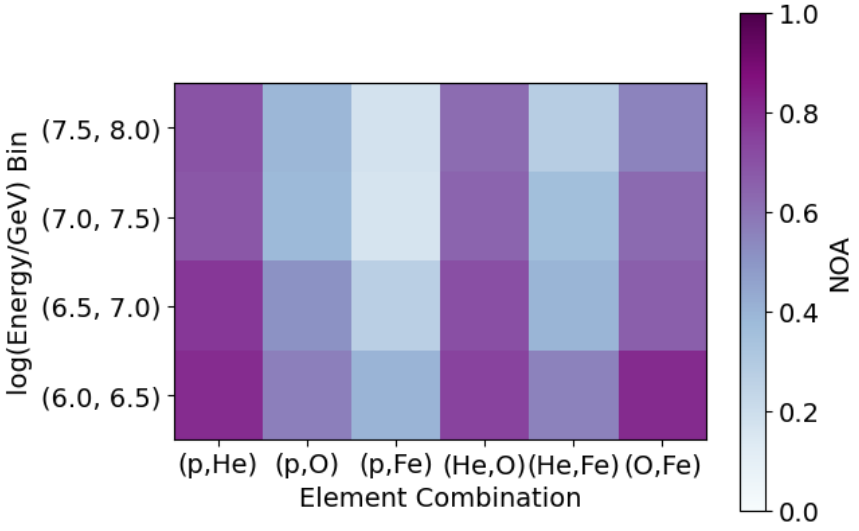


Figure 7.7: NOA for the model 6-7-4-2 model.

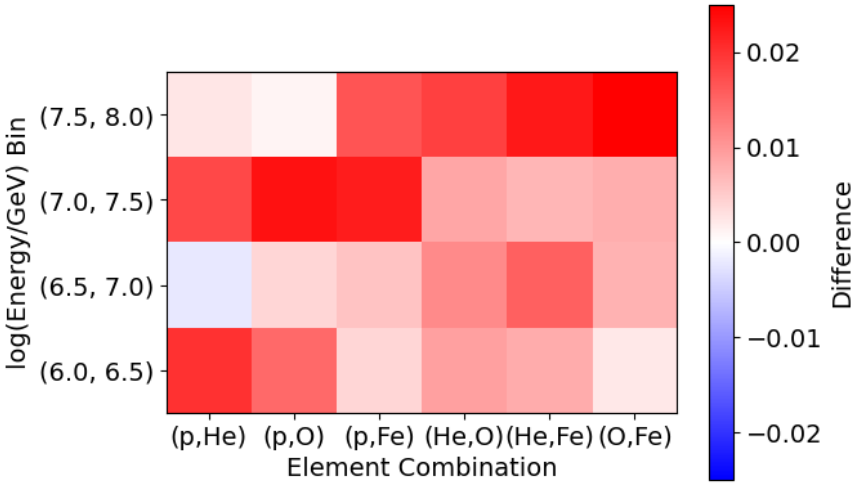
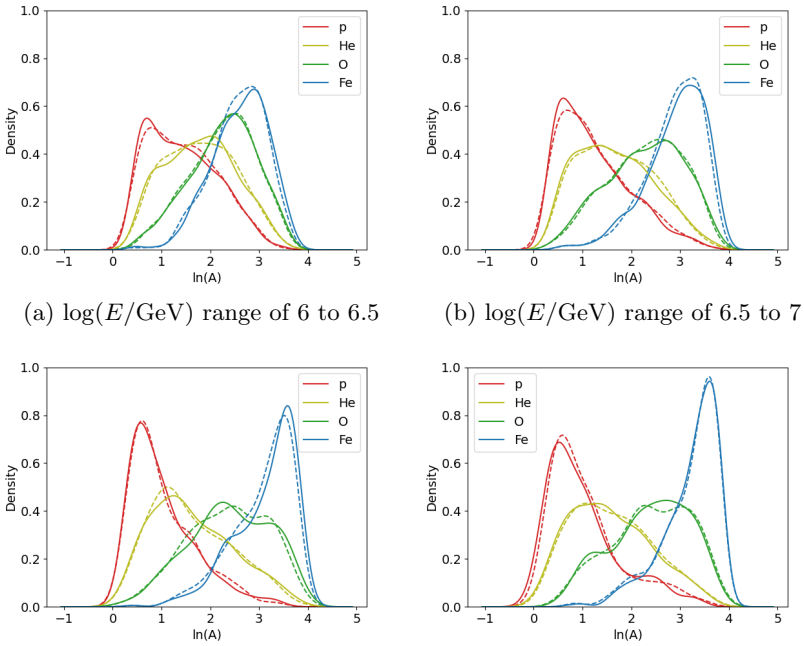


Figure 7.8: Difference between NOA of the baseline model and NOA of the 6-7-4-2 model (higher means 6-7-4-2 has improved).



(a) $\log(E/\mathrm{GeV})$ range of 6 to 6.5

(b) $\log(E/\mathrm{GeV})$ range of 6.5 to 7

(c) $\log(E/\mathrm{GeV})$ range of 7 to 7.5

(d) $\log(E/\mathrm{GeV})$ range of 7.5 to 8

Figure 7.9: KDEs of the 6-7-4-2 (solid) and the baseline (dashed) model for different energy ranges.

To evaluate the performance of the model in predicting the actual mass of the primaries, the same plot as for the baseline model is performed, see figure 7.10. In addition, to better evaluate the differences between the baseline model and the model with added muon density, comparison plots are made, of which the one for the combined masses can be found in figure 7.11, and the ones for the different primaries can be found in the Appendix A.2. From the comparison plot, it can be seen that the MSE over all primaries of the model with added muon density closely resembles that of the baseline model, with the only difference being a slight systematic improvement.

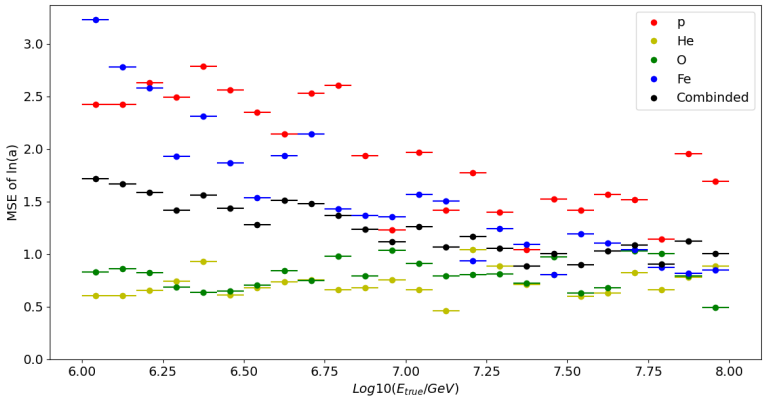


Figure 7.10: Mass prediction of the 6-7-4-2 model for different energy ranges.

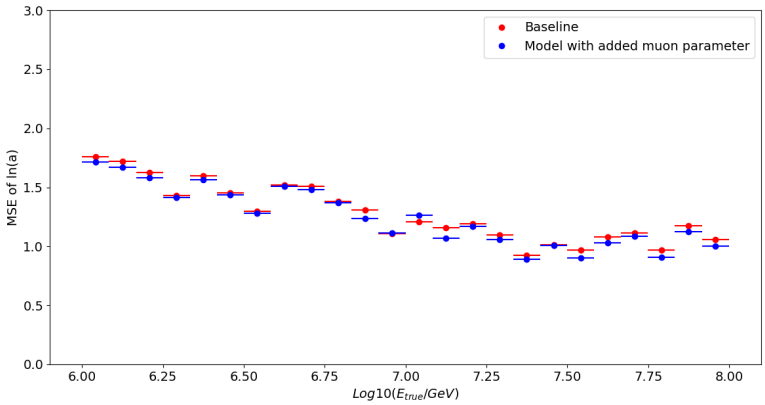


Figure 7.11: Comparison of the mass prediction between the baseline model and the 6-7-4-2 model for different energy ranges.

7.3 Different architectures

Since the 6-7-4-2 architecture was never optimized to use the muon density parameter as a sixth parameter, and to eliminate architectural constraints, several new/different architectures were tried. For this, the following procedure was used:

First, different activation functions were tried out, using the 6-7-4-2 model, hereby it was found that the Exponential Linear Unit (ELU) activation function performs best. Therefore, ELU was used as the activation function for the rest of the trails.
Next, around 400 different combinations of numbers of layers and nodes in those layers were tried on a subset of the dataset with 10,000 events. This was done using the common training method, with the only difference being that the models were trained for a maximum of 5000 epochs. Of the tried models, the 6-18-16-16-2 and the 6-22-14-2 model performed best.
After this, both models were trained on the full dataset. This showed that the 6-18-18-2 model has the best trade-off between energy and mass prediction and separation.
Since all the tried models were much better at predicting the energy than at predicting the mass and separation, different loss functions that favor the mass/separation predictions were tried out next, using the 6-18-16-16-2 model. For this, a weighted MSE between mass and energy, the separation, and the sum of the MSE of the energy and the separation, were tried.
This showed that the loss has a negligible influence on the performance of the model, therefore the final network was trained using the MSE loss.

In conclusion, the best performing model with the added muon density parameter is a 6-18-16-16-2 together with ELU as the activation function and MSE as the loss function, so this model is used for the rest of this analysis, its performance is discussed in the next section.

7.3.1 Performance

Energy

The MSE of the energy of this model is about 0.0042 and the R^2 value is about 98.8%. This means that compared to the other two models in this analysis, the change in architecture provides only a slight improvement in performance over the whole energy range.
In order to see the performance of the model in different energy ranges, bias and resolution distributions are plotted. This can be seen in figure 7.12. This plot shows that the improved model is not effected by the boundary effect in contrast to the other two models in this analysis, suggesting that the improved model is more stable around the boundaries. This can even better be seen in the comparison plot in figure 7.13. From the comparison plot, it can be seen that

the energy performance of the improved model compared to the baseline model is essentially the same up to the boundary, suggesting that the improvement is only due to the improved stability at the edge near $\log(E/\mathrm{GeV}) = 8$.

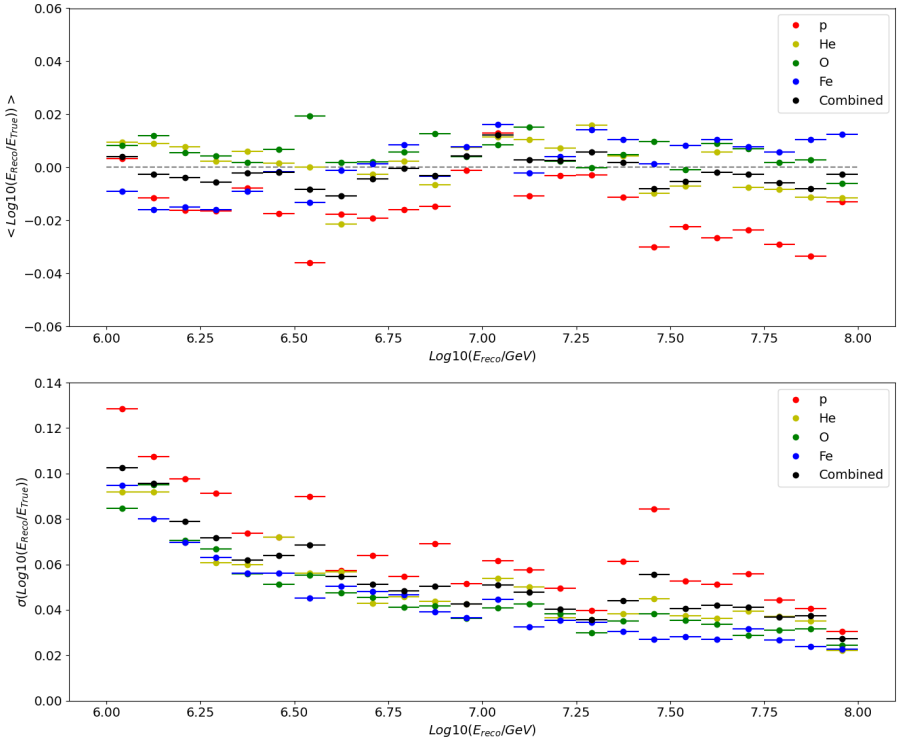


Figure 7.12: Bias (top) and resolution (lower) of the energy reconstruction for the 18-16-16-2 model.

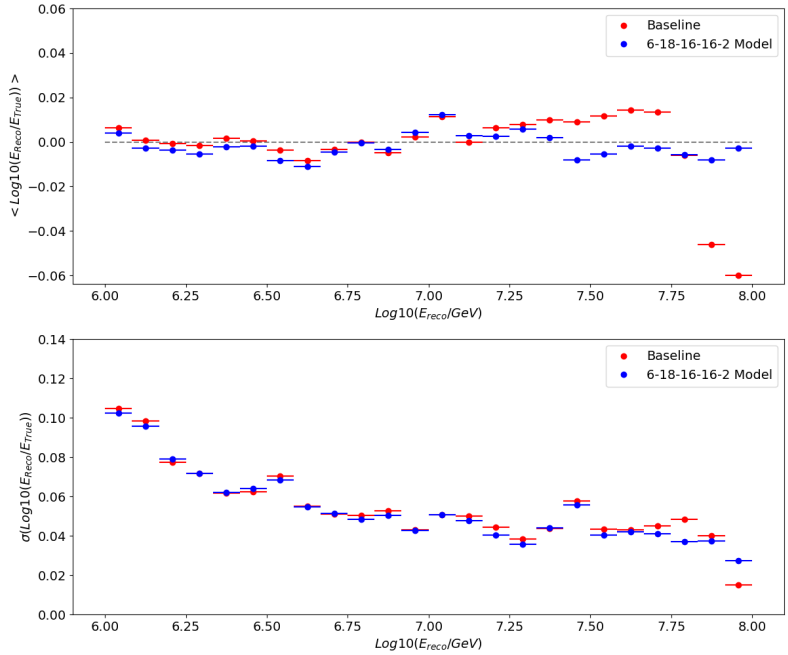


Figure 7.13: Comparison between baseline model and improved model of bias (top) and resolution (bottom) of energy reconstruction for all primaries combined.

Mass

For the improved model, the MSE of $\ln(A)$ is about 1.358 and the average NOA is about 0.55.
As this is very close to the 6-7-4-2 model, this means that the 6-7-4-2 model has no architectural constraints on mass predictions when compared over the whole energy range. To see if there are any differences for different energy ranges, a plot of $\ln(A)$ for different energy ranges, analogous to those for the other two models, is also made for the improved model, this plot can be seen in figure 7.14.
From this it can be seen that the mass prediction of the improved model follows the form of the mass prediction for the 6-7-4-2 model, which means that the architecture change does not affect different energy ranges differently. Furthermore, compared to the energy prediction, the architecture change does not elevate the boundary effect for the mass prediction.
In order to better compare the mass prediction of the improved model with that

of the baseline model, a comparison plot is made, see figure 7.15, from which it can be seen that the muon density, as already seen in the plot 7.11 for the 6-7-4-2 model, does not affect different energy ranges differently, but rather there is a slight systematic improvement over the entire energy range.
To evaluate the ability of the improved model to separate between different primary types over different energy ranges, the same NOA plot as for the other two models, is performed, this can be seen in figure 7.16. From this, it can be seen that the NOA plot of the improved model closely resembles that of the other two models, which means that the architectural changes don't affect the NOA prediction differently for different energy ranges or combinations. To better compare the changes to the NOA of the improved model to the baseline mode, a comparison plot is made, see figure 7.17, which looks very similar to that of the 6-7-4-2, meaning that the improved model also does not change the NOA prediction for different energy ranges or combinations differently compared to the 6-7-4-2 model.
The KDEs used to calculate the NOAs can be seen in figure 7.18. The figure shows that the KDEs of the 6-18-16-16-2 and the baseline model have a similar distribution for all primaries and all energies. This means that the muon density parameter only has a minor effect on the shape of the KDEs.

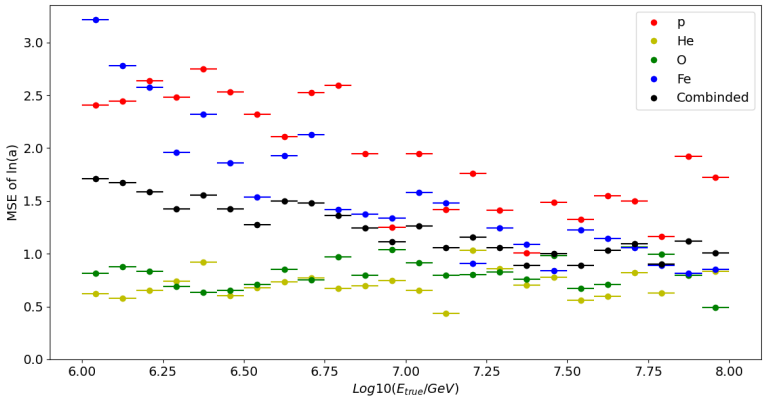


Figure 7.14: Mass prediction of the 6-18-16-16-2 model for different energy ranges.

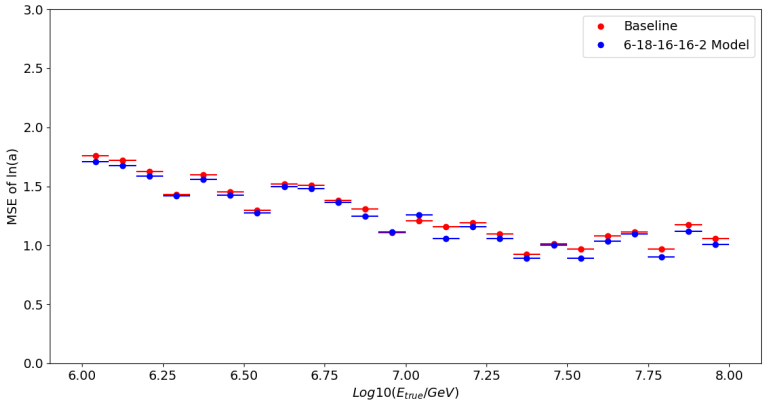


Figure 7.15: Comparison of the mass prediction between the baseline model and the 6-18-16-16-2 model for different energy ranges.

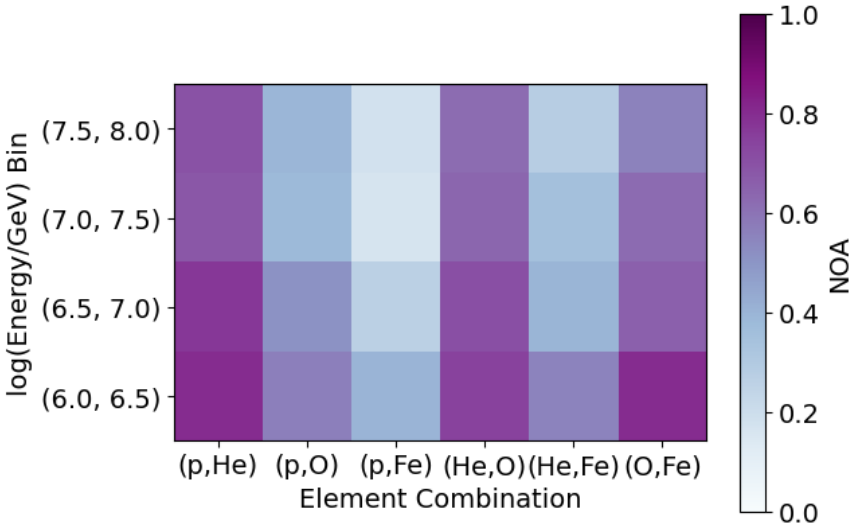


Figure 7.16: NOA for different energy ranges and all combinations of primaries for the improved model.

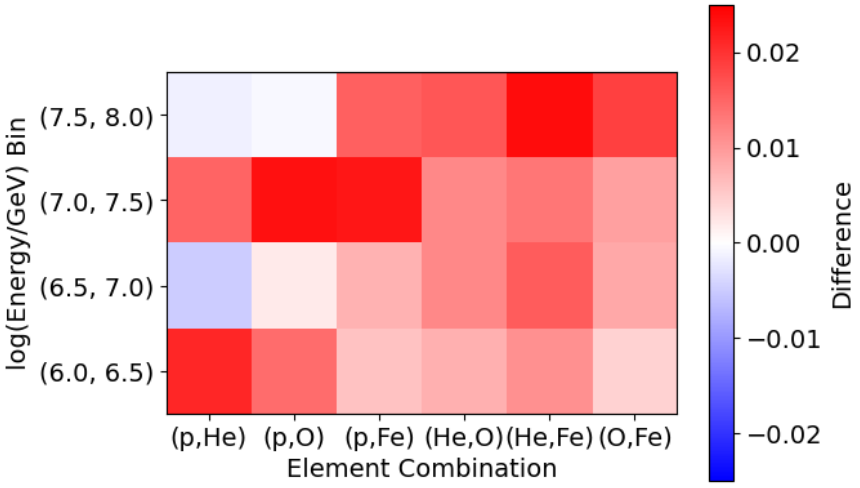
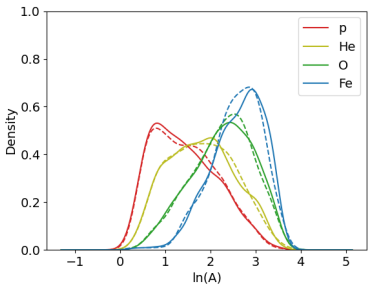
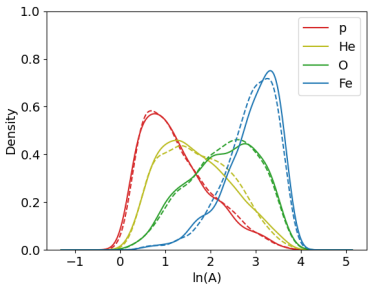


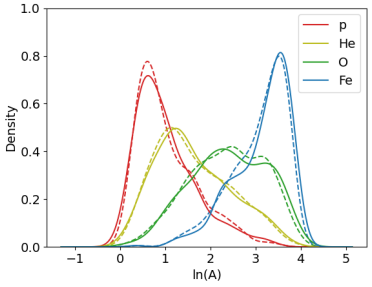
Figure 7.17: Difference between NOA of the baseline model and NOA of the 6-18-16-16-2 model (higher means 6-18-16-16-2 has improved).



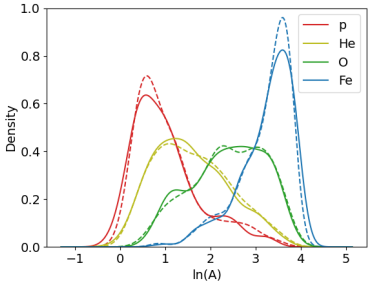
(a) $\log(E/\text{GeV})$ range of 6 to 6.5



(b) $\log(E/\text{GeV})$ range of 6.5 to 7



(c) $\log(E/\text{GeV})$ range of 7 to 7.5



(d) $\log(E/\text{GeV})$ range of 7.5 to 8

Figure 7.18: KDEs of the 6-18-16-16-2 (solid) and the baseline (dashed) model for different energy ranges.

8 Conclusion

In this thesis, the muon density parameter was added as a new feature to a neural network based on a neural network from the previous publication "Cosmic Ray Spectrum and Composition from PeV to EeV Using 3 Years of Data From IceTop and IceCube" [Aar+19], which predicts the mass and energy of cosmic ray particles.

The analysis showed that the muon density as a feature leads to a slight improvement in the mass prediction and in the separation, which is expected since the muon density is mainly a mass sensitive parameter.
It is also shown, that the muon density parameter does not affect the energy prediction significantly. This suggests that the GeV muon density parameter is nearly independent of energy, or its dependence is small compared to that of the other features like $\log(S125)$ to have a significant effect on the energy reconstruction.
The small effect on the mass predictions due to the addition of the muon density parameters indicates that the parameter does not provide much more information about the primary mass than the main mass indicator $\frac{\mathrm{d}E_{1500}}{\mathrm{d}X}$, thus the two are probably highly correlated.

In order to better evaluate the effects of individual features, in the future, a sensitivity analysis could be performed, which describes how the mass and energy predictions respond to changes in the input parameters. This would also show which parameters have the most influence on the prediction of energy and mass, and thus the most physical relevance.

It was also shown that the neural networks used in this thesis perform better at predicting the energy than the mass of the primary. Further, it was shown that the NNs perform better at predicting the correct mass of oxygen and helium than for iron and proton, this could be because iron and proton are at the edge of the dataset, and so boundary effects can influence their predictions, this could be eliminated by using classification. This would have the advantage that it is not effected by the order in the inherent masses and therefore no boundary effects would take place. The disadvantage is that there would be no continuous mass prediction, so such a network could not be used for composition study with template fitting.

Further, it was observed that the models using tanh as the activation function have boundary effects for energy predictions, while the model using the ELU activation function does not have a boundary effect. This indicates that the boundary effect for the first two models is most likely caused by the satu-

ration of the tanh function at the end of the energy range.

It should be noted that the entire analysis in this thesis has been performed with simulations using only one specific hadronic interaction model (Sibyll 2.1). It is known that these models have some deficiencies in describing the entire development of the extensive air shower, in particular the correlations between the different shower components. Here, the additional muon density parameter might be helpful in validating the simulations against data. Both, the use of other hadronic interaction models in the simulations and the application of the NN to a first data set were not possible within the scope of this work.

In conclusion, due to the improvement in mass prediction, the neural network with the density parameter of GeV muons as an additional feature looks quite promising to be used in future analyses, in particular in cross-checking the validation of the hadronic interaction models used for the Monte Carlo simulations.

References

[Aar+17] M.G. Aartsen et al. "The IceCube Neutrino Observatory: instrumentation and online systems". In: *Journal of Instrumentation* 12.03 (Mar. 2017), P03012–P03012. DOI: 10.1088/1748-0221/12/03/p03012. URL: https://doi.org/10.1088%2F1748-0221%2F12%2F03%2Fp03012.

[Aar+19] M. G. Aartsen et al. "Cosmic ray spectrum and composition from PeV to EeV using 3 years of data from IceTop and IceCube". In: *Physical Review D* 100.8 (Oct. 2019). DOI: 10.1103/physrevd.100.082002. URL: https://doi.org/10.1103%2Fphysrevd.100.082002.

[Ago+03] S. Agostinelli et al. "Geant4 a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[Ahn+09] E. J. Ahn et al. "Cosmic ray interaction event generator SIBYLL 2.1". In: *Physical Review D* 80 (2009), pp. 094003/1–17. DOI: 10.1103/PhysRevD.80.094003.

[Alv+19] Rafael Alves Batista et al. "Open Questions in Cosmic-Ray Research at Ultrahigh Energies". In: *Frontiers in Astronomy and Space Sciences* 6 (2019). ISSN: 2296-987X. DOI: 10.3389/fspas.2019.00023. URL: https://www.frontiersin.org/articles/10.3389/fspas.2019.00023.

[Aug+39] Pierre Auger et al. "Extensive Cosmic-Ray Showers". In: *Rev. Mod. Phys.* 11 (3-4 July 1939), pp. 288–291. DOI: 10.1103/RevModPhys.11.288. URL: https://link.aps.org/doi/10.1103/RevModPhys.11.288.

[Com36] Arthur H. Compton. "Cosmic Rays as Electrical Particles". In: *Phys. Rev.* 50 (12 Dec. 1936), pp. 1119–1130. DOI: 10.1103/PhysRev.50.1119. URL: https://link.aps.org/doi/10.1103/PhysRev.50.1119.

[De 19] De Ridder, Sam. "Sensitivity of IceCube cosmic ray measurements to the hadronic interaction models". eng. PhD thesis. Ghent University, 2019. URL: https://biblio.ugent.be/publication/8609789/file/8609790.pdf.

[Duk18] Obumneme Stanley Dukor. *Neural Representation of AND, OR, NOT, XOR and XNOR Logic Gates (Perceptron Algorithm)*. 2018. URL: https://medium.com/@stanleydukor/neural-representation-of-and-or-not-xor-and-xnor-logic-gates-perceptron-algorithm-b0275375fea1.

[EHP11] Ralph Engel, Dieter Heck, and Tanguy Pierog. "Extensive Air Showers and Hadronic Interactions at High Energy". In: *Annual Review of Nuclear and Particle Science* 61.1 (2011), pp. 467–489. DOI: 10.1146/annurev.nucl.012809.104544. eprint: https://doi.org/10.1146/annurev.nucl.012809.104544. URL: https://doi.org/10.1146/annurev.nucl.012809.104544.

[Feu13] Tom Feusels. "Measurement of cosmic ray composition and energy spectrum between 1PeV and 1EeV with IceTop and IceCube". PhD thesis. Ghent University, 2013. URL: https://biblio.ugent.be/publication/4337238/file/8632371.pdf.

[Fou] PyTorch Foundation. *torch.optim.lr_scheduler.ReduceLROnPlateau*. URL: https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html#torch.optim.lr_scheduler.ReduceLROnPlateau.

[Gab+19] Stefano Gabici et al. "The origin of Galactic cosmic rays: Challenges to the standard paradigm". In: *International Journal of Modern Physics D* 28.15 (Nov. 2019), p. 1930022. DOI: 10.1142/s0218271819300222. URL: https://doi.org/10.1142%2Fs0218271819300222.

[Gai16] Thomas K. Gaisser. *Cosmic rays and particle physics*. Ed. by Ralph Engel and Elisa Resconi. Second edition. Cambridge: Cambridge University Press, 2016. ISBN: 9780521016469; 0521016460. URL: http://www.loc.gov/catdir/enhancements/fy1606/2016003557-b.html%20;%20http://www.loc.gov/catdir/enhancements/fy1606/2016003557-d.html%20;%20http://www.loc.gov/catdir/enhancements/fy1606/2016003557-t.html.

[Gin96] Vitalii L Ginzburg. "Cosmic ray astrophysics (history and general review)". In: *Physics-Uspekhi* 39.2 (Feb. 1996), p. 155. DOI: 10.1070/PU1996v039n02ABEH000132. URL: https://dx.doi.org/10.1070/PU1996v039n02ABEH000132.

[Gro+20] Particle Data Group et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf. URL: https://doi.org/10.1093/ptep/ptaa104.

[Hec+98] Dieter Heck et al. "CORSIKA: A Monte Carlo code to simulate extensive air showers". In: *Report FZKA* 6019.11 (1998). URL: https://publikationen.bibliothek.kit.edu/270043064/3813660.

[Inc] TIBCO Software Inc. *What is a Neural Network?* URL: https://www.tibco.com/reference-center/what-is-a-neural-network.

[KB17] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[KBH21] Donghwa Kang, Sally-Ann Browne, and Andreas Haungs. *Studies of a muon-based mass sensitive parameter for the IceTop surface array*. 2021. arXiv: 2109.02506 [astro-ph.HE].

[KU12] Karl-Heinz Kampert and Michael Unger. "Measurements of the cosmic ray composition with air shower experiments". In: *Astroparticle Physics* 35.10 (May 2012), pp. 660–678. DOI: 10.1016/j.astropartphys.2012.02.004. URL: https://doi.org/10.1016%2Fj.astropartphys.2012.02.004.

[Mat05] J. Matthews. "A Heitler model of extensive air showers". In: *Astroparticle Physics* 22.5 (2005), pp. 387–397. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2004.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0927650504001598.

[Mew94] R.A. Mewaldt. "Galactic cosmic ray composition and energy spectra". In: *Advances in Space Research* 14.10 (1994), pp. 737–747. ISSN: 0273-1177. DOI: https://doi.org/10.1016/0273-1177(94)90536-3. URL: https://www.sciencedirect.com/science/article/pii/0273117794905363.

[NHR85] W. R. Nelson, H. Hirayama, and D. W. O. Rogers. *EGS4 code system*. Tech. rep. Stanford Linear Accelerator Center, Menlo Park, CA (USA), Dec. 1985. URL: https://www.osti.gov/biblio/6137659.

[Pas+19] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[Ped+] F. Pedregosa et al. *sklearn.preprocessing.MinMaxScaler*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html.

[Ped+11] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[Rud16] Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: *CoRR* abs/1609.04747 (2016). arXiv: 1609.04747. URL: http://arxiv.org/abs/1609.04747.

[SLL13] Hang Su, Quan Liu, and Jingsong Li. "Adaptive Approach for Boundary Effects Reduction in Rotating Machine Signals Analysis". In: *Advances in Mechanical Engineering* 5 (2013), p. 540172. DOI: 10.1155/2013/540172. eprint: https://doi.org/10.1155/2013/540172. URL: https://doi.org/10.1155/2013/540172.

[Uni] Swinburne University. *Cosmic Ray Energies*. URL: https://astronomy.swin.edu.au/cosmos/c/cosmic+ray+energies.

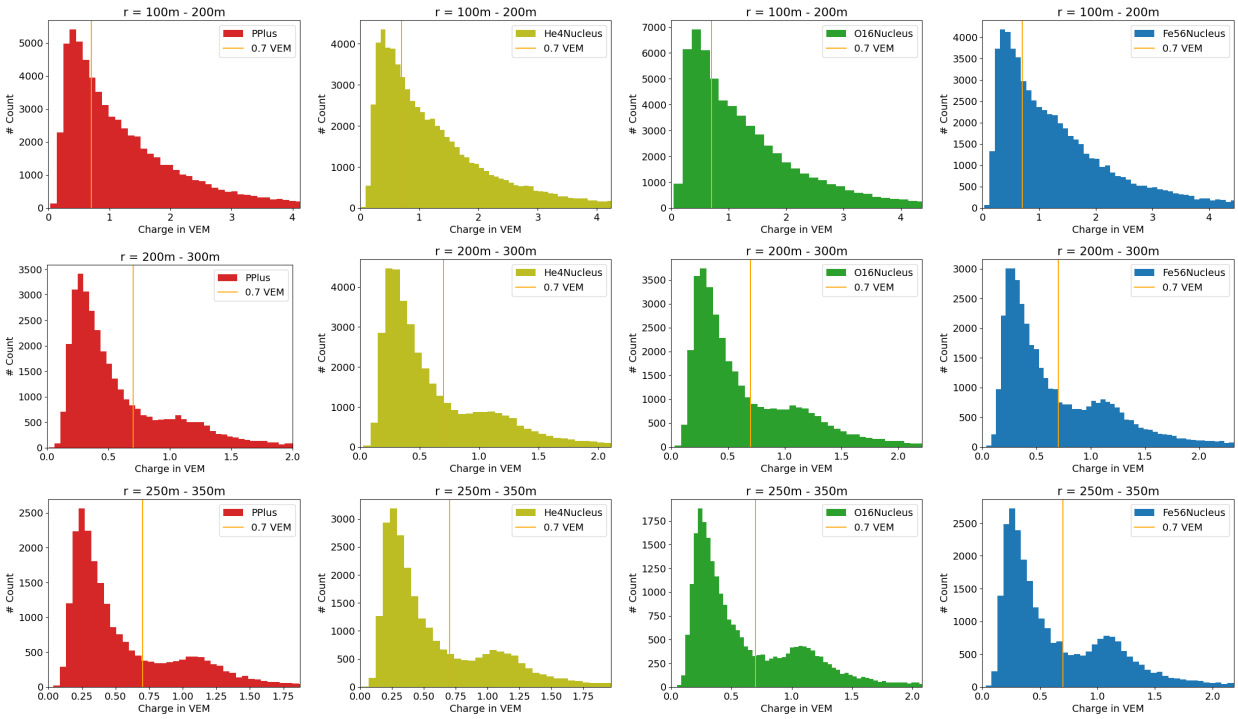
[Wal12] Michael Walter. *100 Years of Cosmic Rays*. 2012. URL: https://www-zeuthen.desy.de/exps/2012_VictorHess/booklet/broschuere_hess_conference_web.pdf.

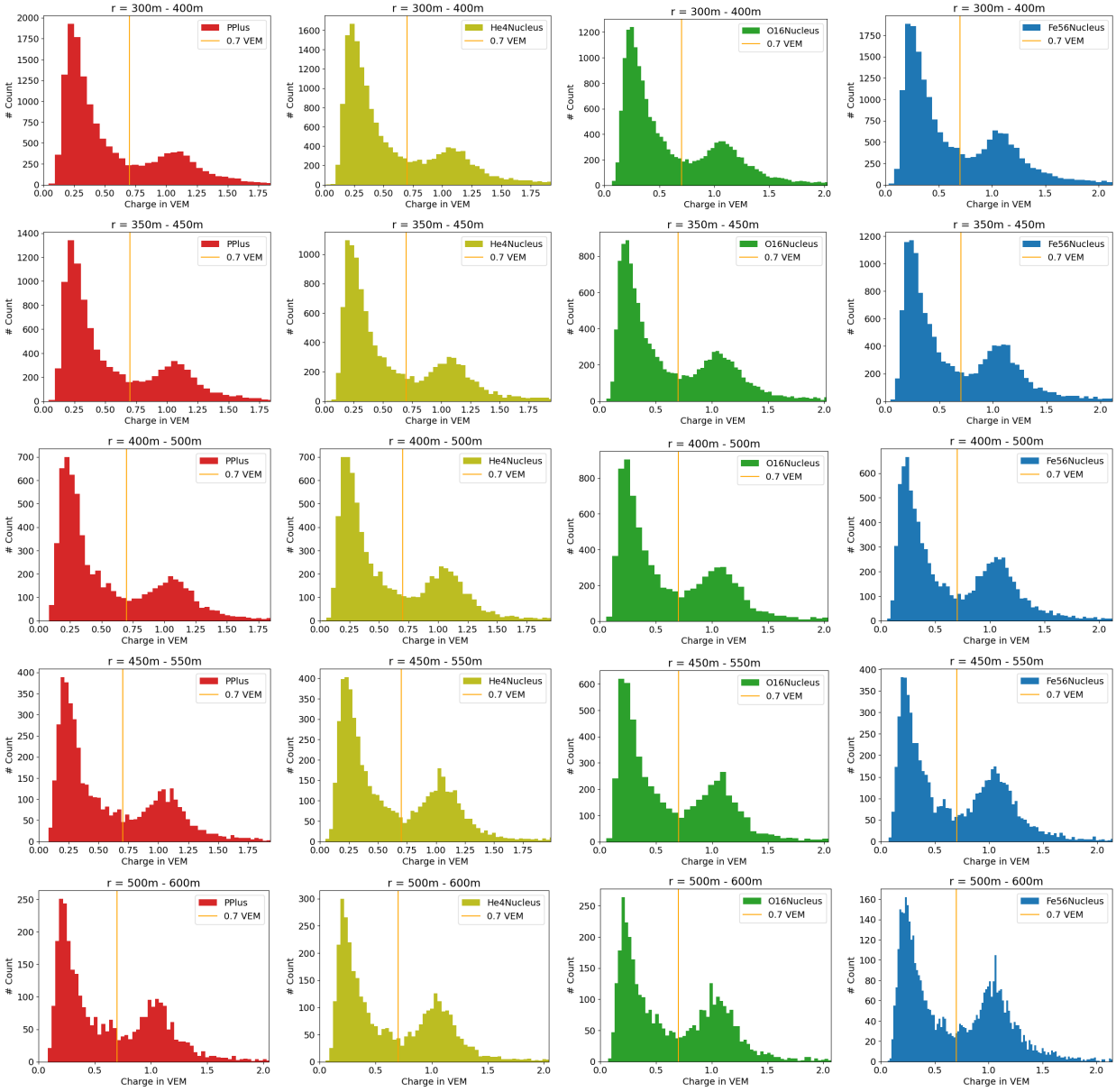
A Appendix

A.1 Combined SLC and HLC charge spectrum for different $\log_{10}(S125)$ and radius ranges

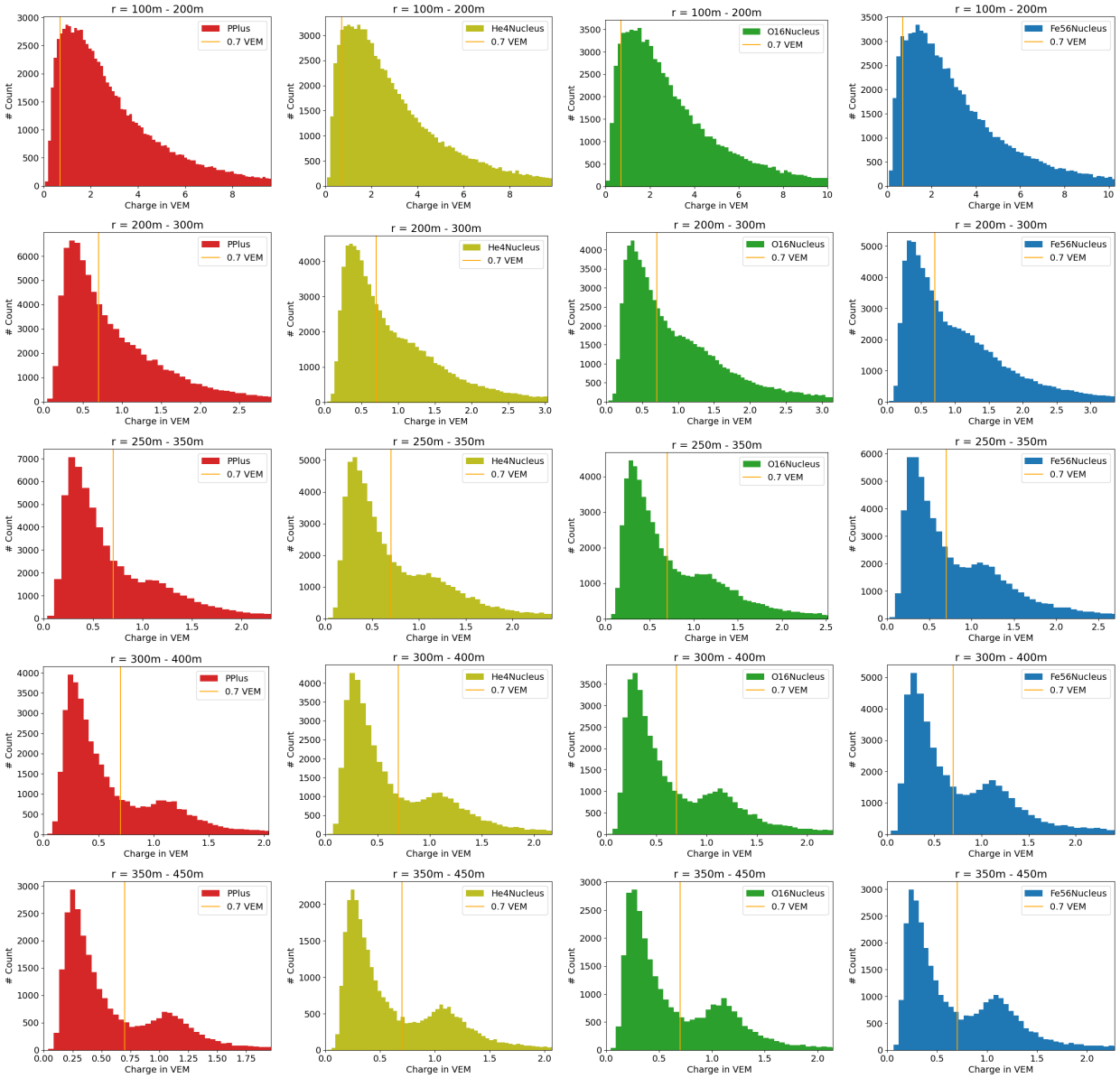
In this section the combined SLC and HLC charge spectrum is shown for different radius bins and $\log_{10}(S125)$ ranges, the orange line at 0.7 VEM is the selection for the muon charge calculation. These plots are used to find the optimal distance for calculating the muon density parameter for different energies.

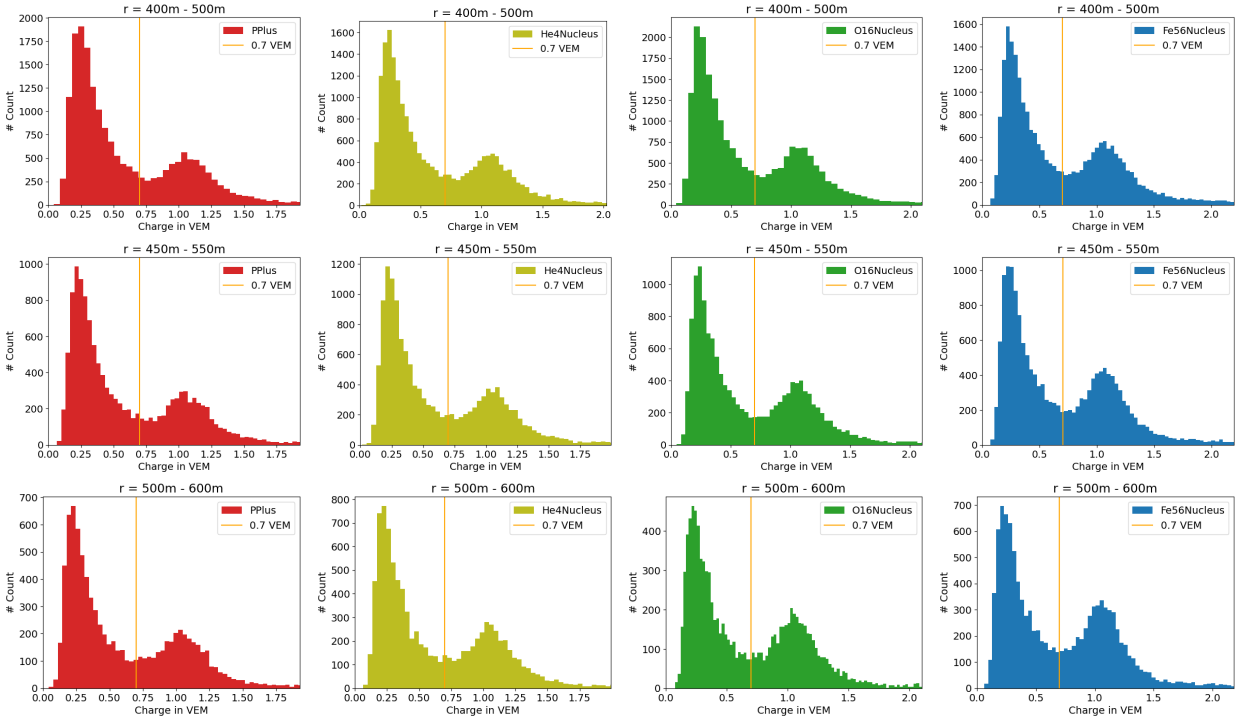
A.1.1 $0 \leq \log_{10}(S125) < 0.5$



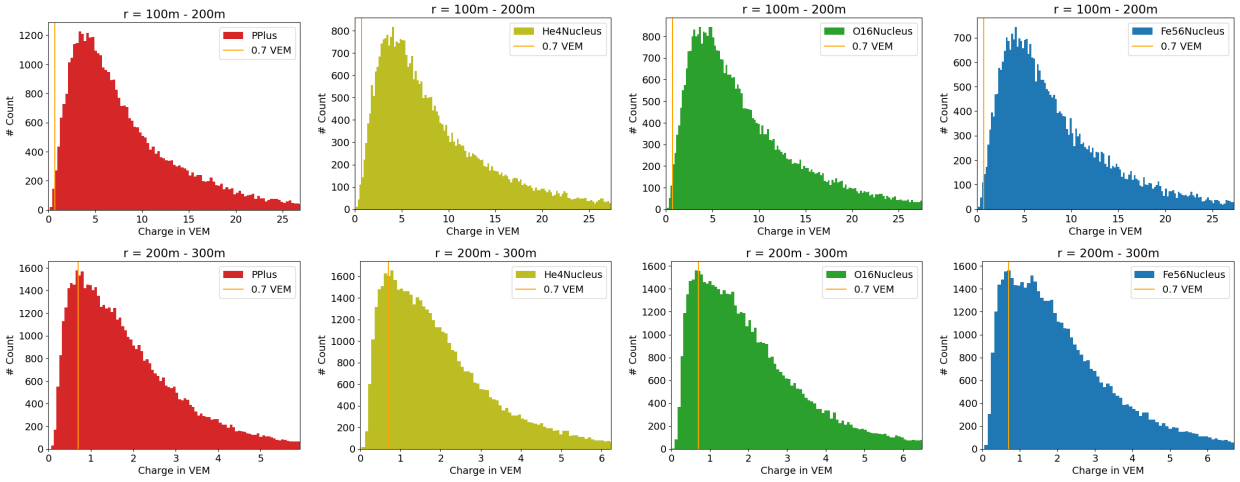

r = 300m - 400m
PPlus
0.7 VEM
Count
Charge in VEM
r = 300m - 400m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 300m - 400m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 300m - 400m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
PPlus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
PPlus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
PPlus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
PPlus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM

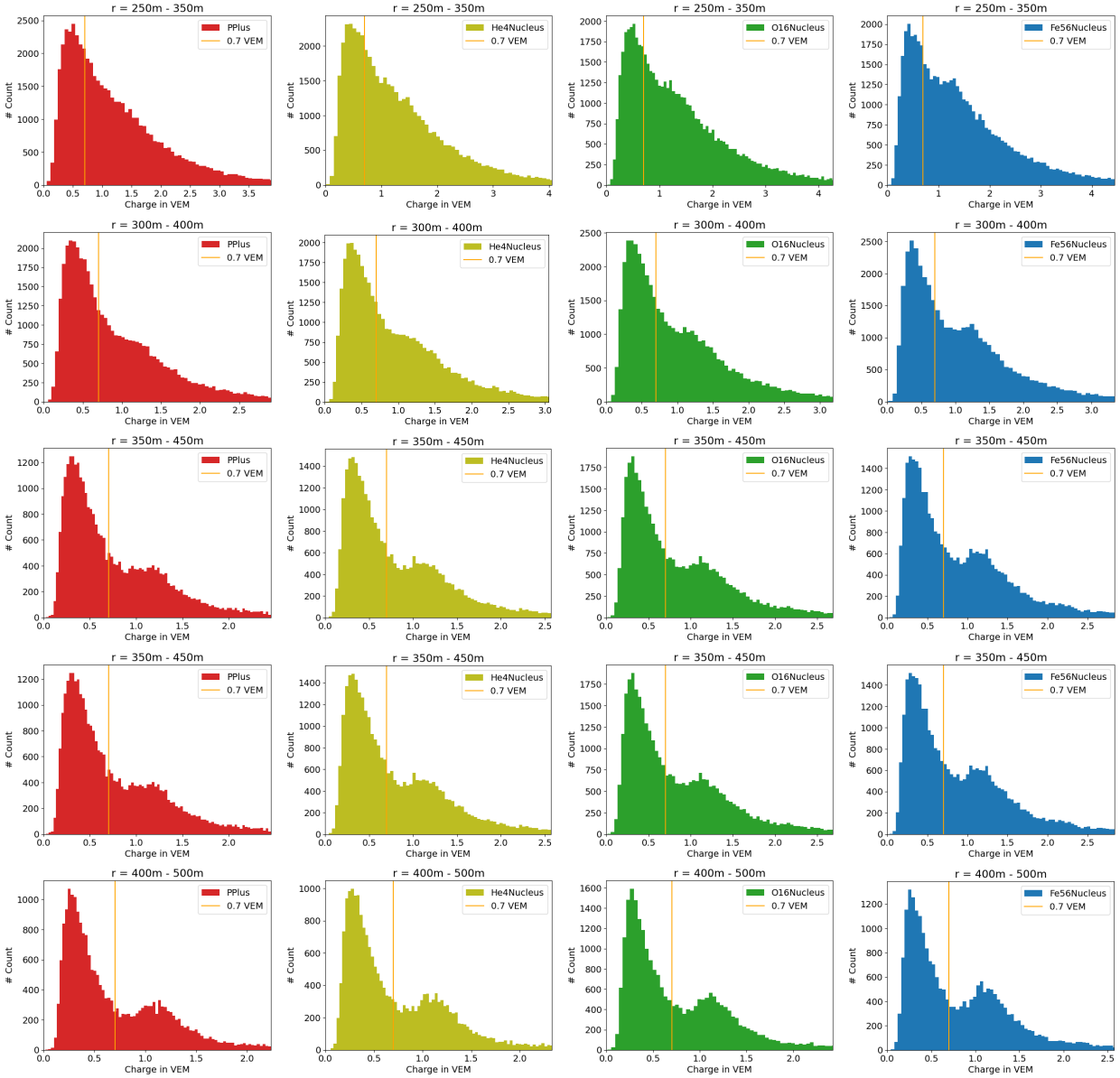
A.1.2 $0.5 \leq \log_{10}(S125) < 1$

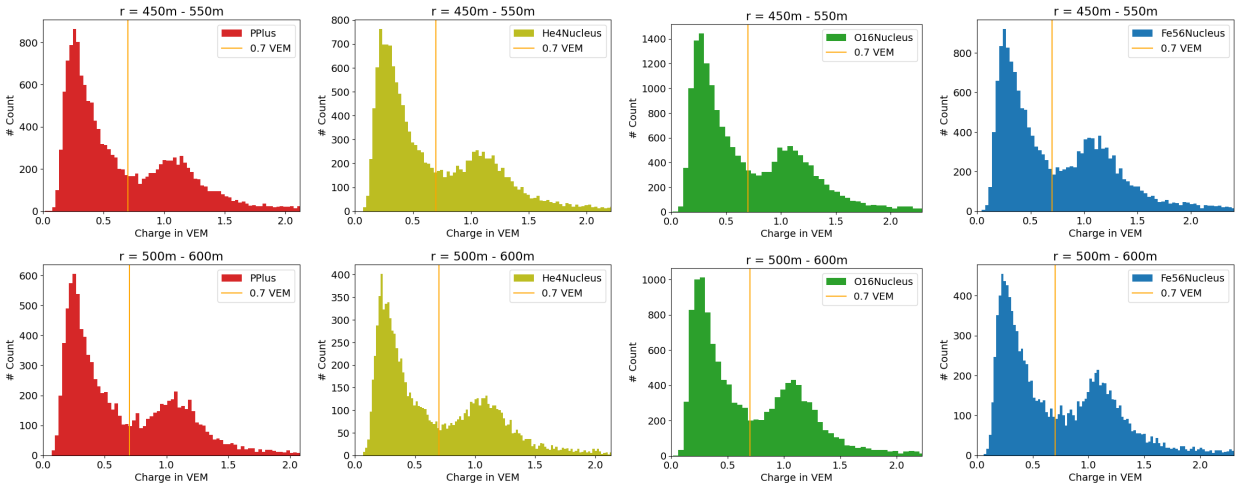




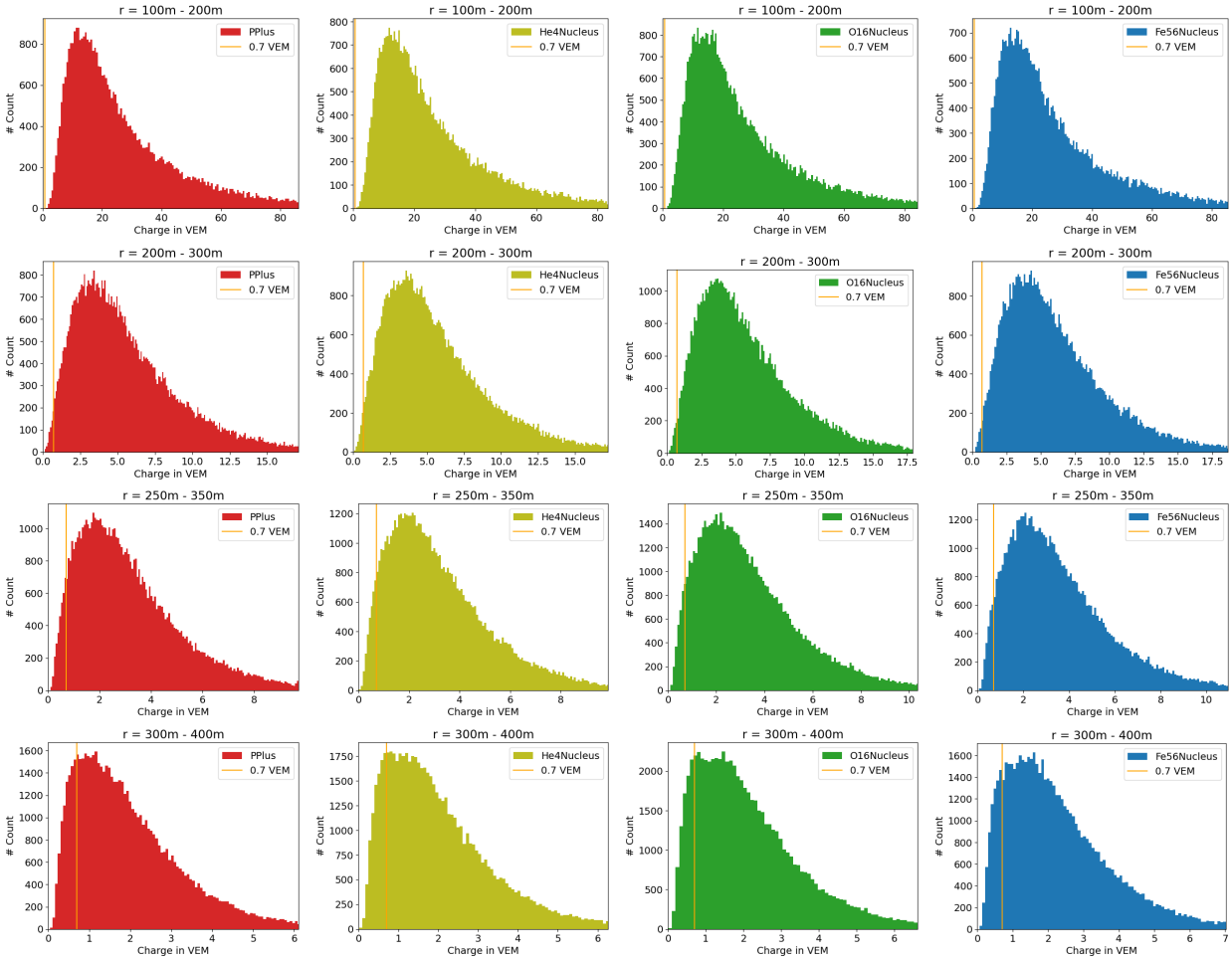
A.1.3 $1 \leq \log_{10}(S125) < 1.5$

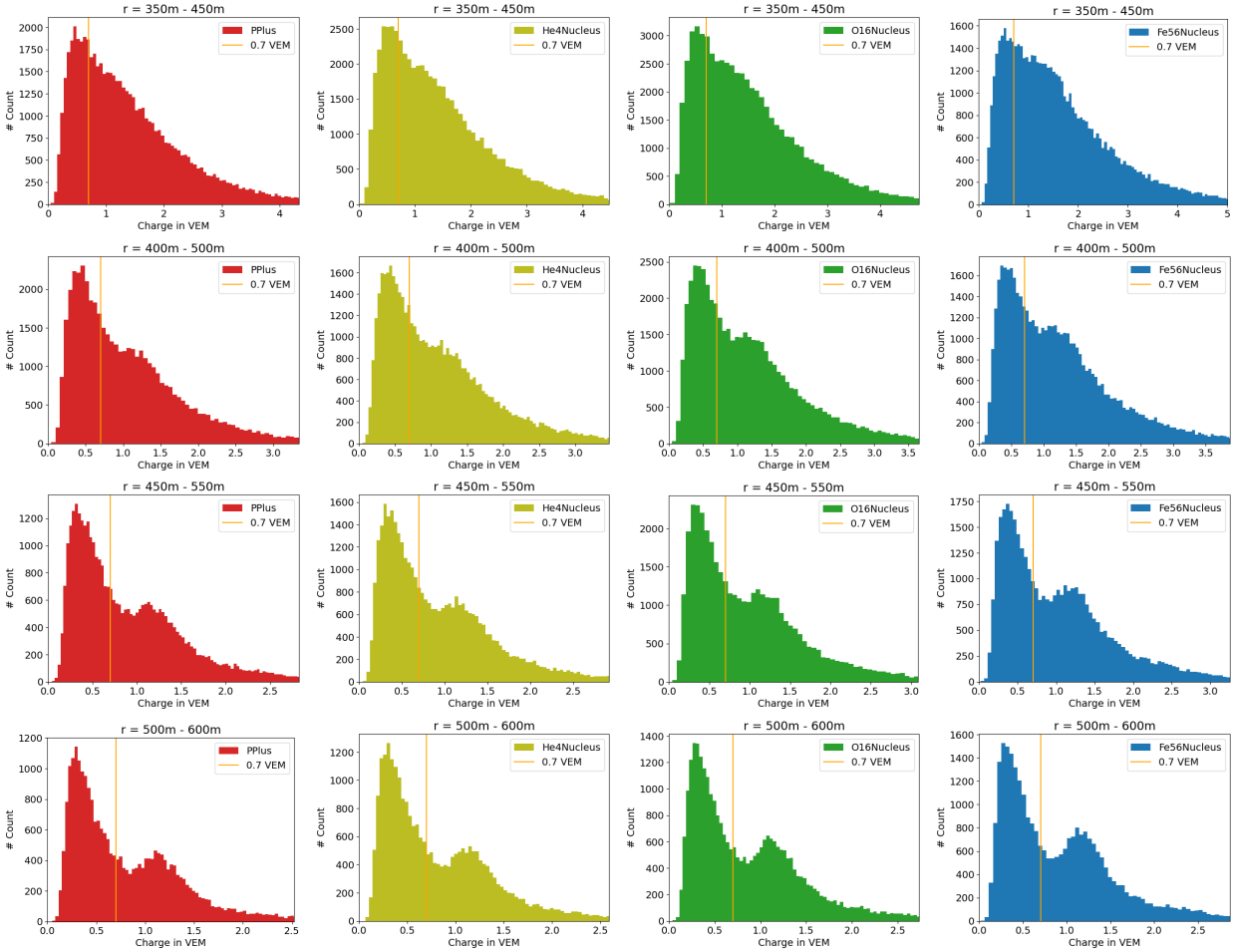



r = 250m - 350m
PPlus
0.7 VEM
Count
Charge in VEM
r = 250m - 350m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 250m - 350m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 250m - 350m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 300m - 400m
PPlus
0.7 VEM
Count
Charge in VEM
r = 300m - 400m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 300m - 400m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 300m - 400m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
PPlus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
PPlus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
PPlus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM


r = 450m - 550m
PPlus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
PPlus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM

A.1.4 $1.5 \leq \log_{10}(S125) < 2$

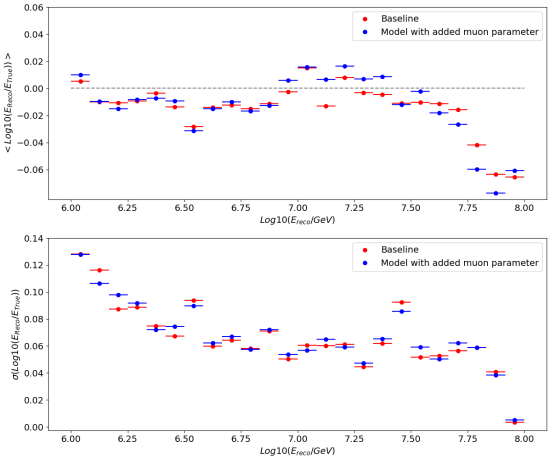



r = 350m - 450m
PPlus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 350m - 450m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
PPlus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 400m - 500m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
PPlus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 450m - 550m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
PPlus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
He4Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
O16Nucleus
0.7 VEM
Count
Charge in VEM
r = 500m - 600m
Fe56Nucleus
0.7 VEM
Count
Charge in VEM

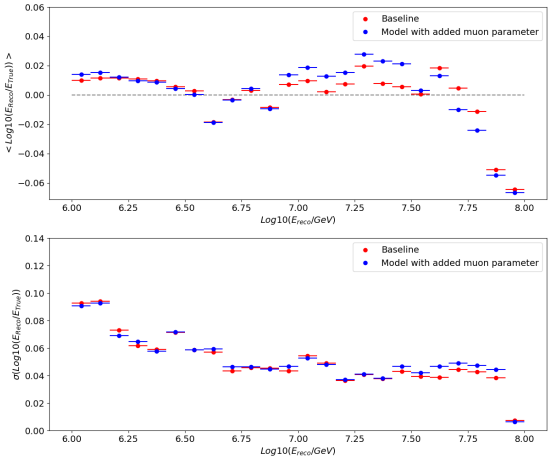
A.2 Comparison to Benchmark Model

A.2.1 6-7-4-2 Model

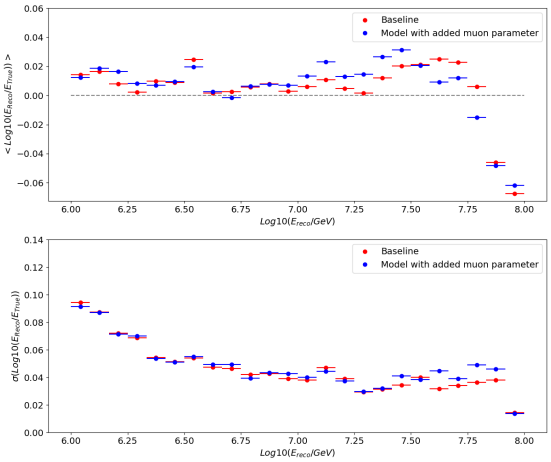
Energy Resolution / Bias comparison between the 6-7-4-2 and the baseline model for different energy bins for all four primaries



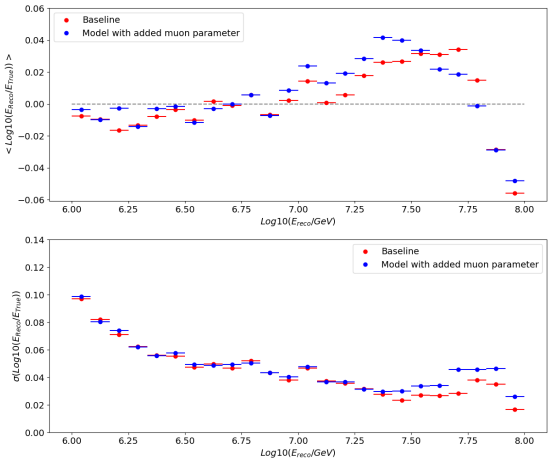
(a) Proton



(b) Helium

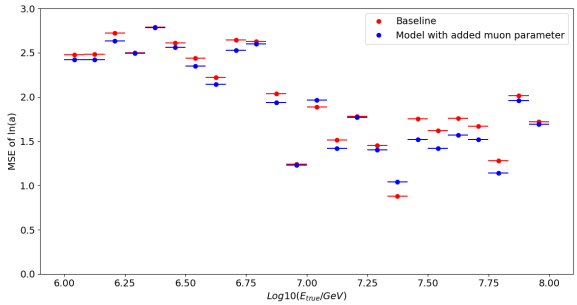


(c) Oxygen

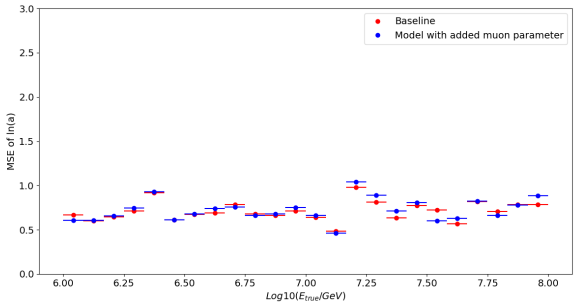


(d) Iron

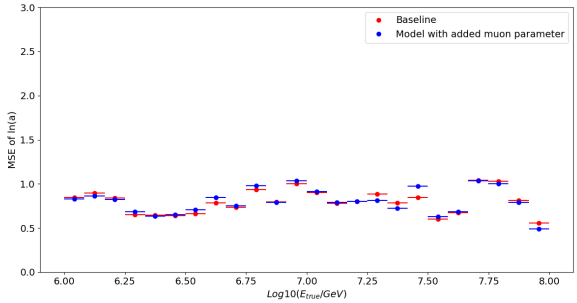
Mean Square Error for of $\ln(A)$ for different energy ranges for all four primaries



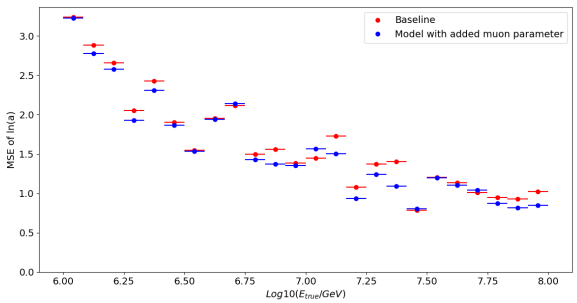
(a) Proton



(b) Helium



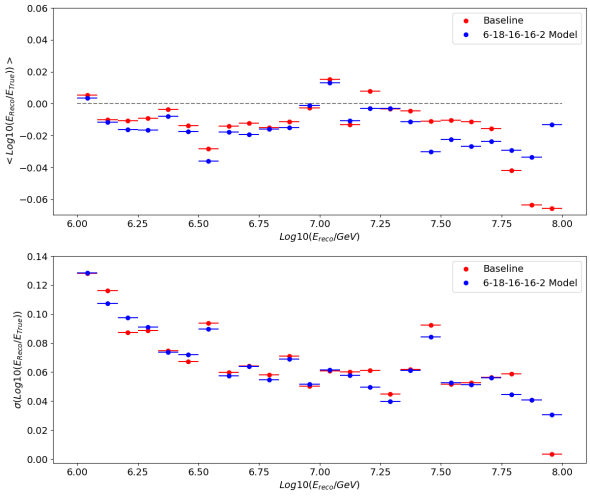
(c) Oxygen



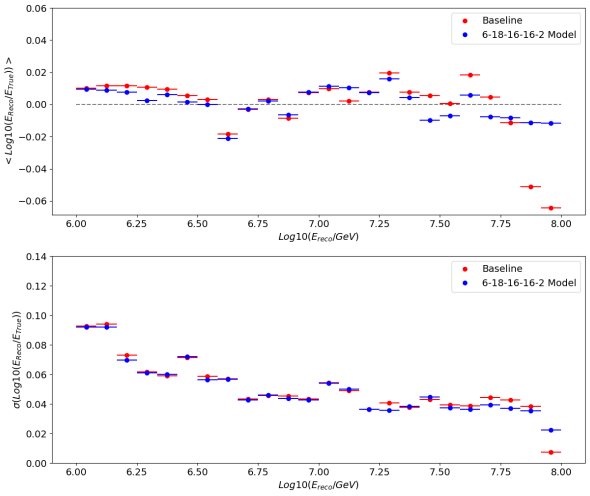
(d) Iron

A.2.2 6-18-16-16-2 Model

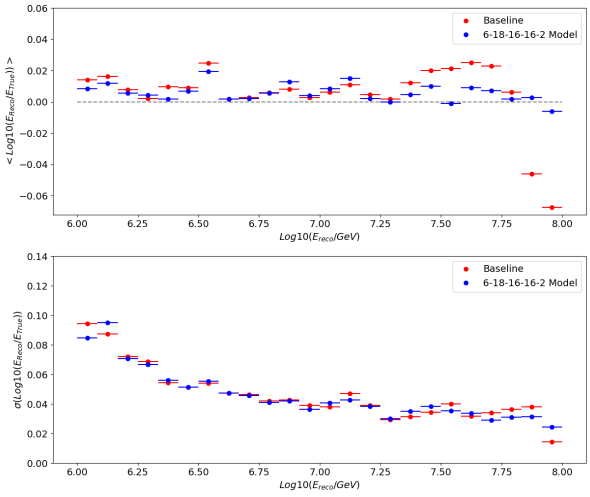
Energy Resolution / Bias comparison between the 6-18-16-16-2 and the baseline model for different energy bins for all four primaries



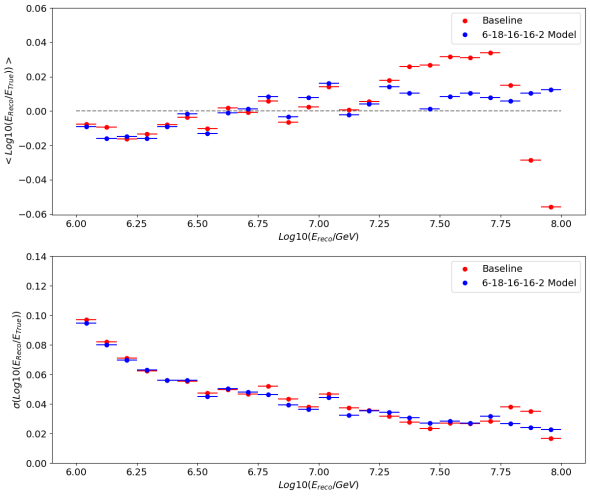
(a) Proton



(b) Helium

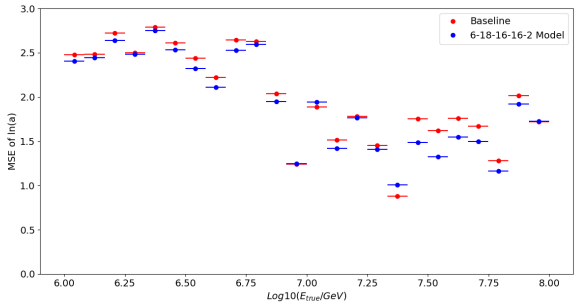


(c) Oxygen

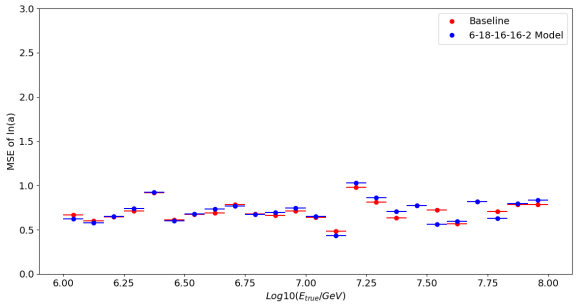


(d) Iron

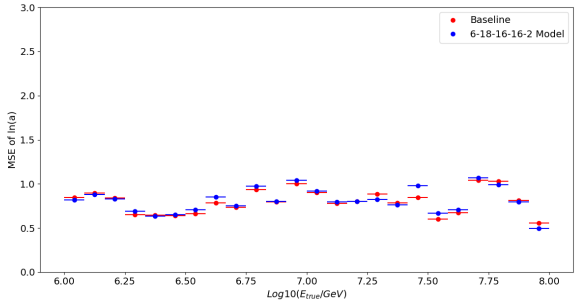
Mean Square Error for of $\ln(A)$ for different energy ranges for all four primaries



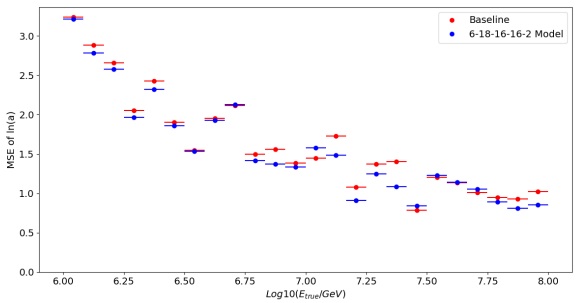
(a) Proton



(b) Helium



(c) Oxygen



(d) Iron

Acknowledgements

I would like to thank everyone at the IAP for their warm welcome and support during this thesis. I would especially like to thank my advisor Donghwa Kang for her amazing support, patience and time she invested in me, without her generous help I would never have been able to complete this thesis.
I would also like to thank Ralph Engel for his help and support.
I also want to thank Andreas Haungs for his great advice, supervision, and encouragement. Furthermore, I would like to thank him for the great feedback and comments, which played a vital role in improving this thesis.
I also wish to thank Julian Saffer and Paras Koundal for their great help, especially in all things related to machine learning. I would especially like to thank Julian for his amazing ideas that helped make this thesis so much better.
I would also like to thank Jelena Köhler for proofreading this thesis and the TexTips she gave me.
Also thanks to langtool and deepl for helping me with spelling, grammar and translation. Furthermore, I would like to thank Doris Wochele for her help in all IT related.

Thank you all for your great support.