Deep Learning Reconstruction of the Muon Signal Fraction for Mass Composition Studies with AugerPrime

Master's Thesis by

Fabian Heizmann

at the Department of Physics
Institute for Astroparticle Physics (IAP)

First Reviewer:	Prof. Dr. Ralph Engel
Second Reviewer:	Prof. Dr. Guido Drexlin
Supervisors:	Dr. Steffen Traugott Hahn
	Dr. Markus Roth
	Dr. Darko Veberič

31.07.2022 – 31.07.2023

Karlsruhe Institute of Technology
Department of Physics
76128 Karlsruhe

Review and Declaration

This thesis has been accepted by the first reviewer of the master thesis.

Karlsruhe, 31.07.2023

Prof. Dr. Ralph Engel

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without properly referencing them in the text. I also declare that the rules of good scientific practice of the Karlsruhe Institute of Technology (KIT) have been followed.

Karlsruhe, 31.07.2023

Fabian Heizmann

Abstract

Ultra-High Energy Cosmic Rays (UHECRs) are the most energetic particles known to mankind. Since the discovery of Cosmic Rays (CRs) a century ago, the questions about their origin and acceleration still remain a mystery. A crucial step towards unraveling the secrets they hold, is to identify their mass, and thus, their charge. A selection of the lightest particles will allow to correlate their arrival directions with candidate sources, as they are minimally deflected by magnetic fields. Furthermore, current hadronic interaction models fail to reproduce the amount of observed muons in Extensive Air Showers (EASs). As the mass of the primary particle is directly linked to the muon content of a shower, we are able to refine our current theory of strong interactions at highest energies.
This work focuses on estimating and relating the muon content of simulated EASs, to the mass of the primary particle utilizing Neural Networks (NNs). To achieve this objective, a recently introduced observable in the Pierre Auger Observatory (Auger), the muon signal fraction in the Water-Cherenkov detector (WCD), is studied using Monte Carlo (MC) simulation data. This work comprises a station-level and an event-level analyses.
On station level, NNs are employed to predict the muon signal fraction by utilizing the information given through the time-dependent signal-traces of the Surface Detector (SD), clearly surpassing the selected analytical methods in performance. Additionally, a direct prediction of the mass on a single-station level with NNs leads to astonishing results in discrimination capability between proton and iron-induced showers.
On an event level, the NN predictions are combined with analytical methods to synthesize a mass-sensitive estimator, leading to encouraging results for future mass-composition studies.

Zusammenfassung

Die ultrahochenergetische kosmische Strahlung besteht aus den energiereichste Teilchen die der Menschheit bekannt sind. Seit der Entdeckung von kosmischer Strahlung vor einem Jahrhundert sind die Fragen nach ihrem Ursprung und ihrer Beschleunigung noch immer ein Rätsel. Ein entscheidender Schritt zur Entschlüsselung der Geheimnisse die sie bergen, ist die Bestimmung ihrer Masse und damit ihrer Ladung. Eine Auswahl der leichtesten Teilchen wird es ermöglichen, ihre Ankunftsrichtung mit möglichen Quellen zu korrelieren, da sie nur minimal von Magnetfeldern abgelenkt werden. Darüber hinaus können die derzeitigen Modelle der hadronischen Wechselwirkung die Menge der beobachteten Myonen in ausgedehnten Luftschauern nicht reproduzieren. Da die Masse des Primärteilchens direkt mit dem Myonengehalt eines Schauer verbunden ist, können wir unsere derzeitige Theorie der starken Wechselwirkung bei höchsten Energien testen.
Diese Arbeit konzentriert sich auf die Bestimmung des Myonengehalts von simulierten extensiven Luftschauern und setzt ihn in Beziehung zur Masse des Primärteilchens mit Hilfe von Neuronalen Netzen (NNs). Um dieses Ziel zu erreichen, wird ein kürzlich eingeführte Beobachtungsgröße des Pierre-Auger-Observatoriums (Auger), der Myon-Signalanteil im Water-Cherenkov-Detektor (WCD), mit Hilfe von Monte-Carlo (MC)-Simulationsdaten untersucht. Diese Arbeit umfasst eine Analyse auf Stationsebene und eine Analyse auf Schauerebene.
Auf Stationsebene werden NNs zur Vorhersage des Myonensignalanteils eingesetzt, indem die Informationen aus den zeitabhängigen Signalspuren des Oberflächendetektors (SD) genutzt werden. Die Leistung übertrifft die der ausgewählten analytischen Methoden bei weitem. Darüber hinaus führt eine direkte Vorhersage der Masse auf Einzelstationsniveau mit NNs zu erstaunlichen Ergebnissen bei der Unterscheidung zwischen Protonen- und Eisen-induzierten Schauern.
Auf Schauerebene werden die NN-Vorhersagen mit analytischen Methoden kombiniert, um einen massensensitiven Schätzer zu synthetisieren, der zu ermutigenden Ergebnissen für künftige Untersuchungen der Massenzusammensetzung führt.

Contents

Appendices

1. Introduction

CRs have become one of the most fascinating fields of fundamental physics to study in the 21st century due to their potential in advancing to new scientific frontiers. These astronomical messengers unveil astonishing insights into the fundamental workings of the cosmos, by bridging the gap between the smallest (particle physics) and the largest (astrophysics) scales. UHECRs are the most energetic particles known to mankind, surpassing the energy scale of the Large Hadron Collider (LHC) by two orders of magnitude. In studying their origin within and beyond our Galaxy, we are able to gain insights into the most violent and fundamental forces in existence.

Since the discovery of CRs over a century ago, their precise origin and acceleration mechanism still remain elusive. To unlock the secrets they hold, it is essential to deduce the mass, and thus, the charge of the CR particles. The identification of sources will be strengthened by composition-enhanced arrival-direction studies. The arrival directions of light particles pin-point back to their origin, being minimally deflected in the omnipresent galactic and extragalactic magnetic fields. Furthermore, measurements of EASs show, that our current hadronic interaction models fail to reproduce the observed amount of muons in EASs. As the mass of the primary particle is directly linked to the muon content of a shower, we are able to refine our current theory of strong interactions, quantum chromodynamics (QCD), by measuring the muon content at the highest energies.

The extreme energies and low flux of CRs only allow for indirect detection techniques. At the time of writing, the Auger is the largest CR observatory in the world, covering an instrumentalized detection area of 3000 km^2 in the argentinian Pampa Amarilla. The observatory uses a hybrid detection technique, where fluorescence telescopes measure the longitudinal shower profile and surface detectors sample the lateral shower footprint of EASs, which are induced by CR particles impinging on Earth's atmosphere.

In recent years, the field of deep learning and NNs has gained a lot of momentum. The ever increasing amount of data, computational power, and the success of NNs in various fields of research, has led to a data-driven approach which are now mature enough to be applied on current problems in physics. The extraction of features from the highly-dimensional spatial and temporal data of an EAS is non-trivial. However, NNs as universal function approximators are perfect candidates to overcome this burden.

The main objective of this work is to link the muon content of EASs, measured by the SD, to the mass of the primary particle. Therefore, a recently introduced variable in Auger is studied on MC simulation data: the muon signal fraction in the WCD. This work can be divided into a station-level and an event-level analysis.

NNs are employed to predict the muon fraction on a single-station level, clearly outperforming classical analytical methods. Even without the need for the second detector of the AugerPrime upgrade, which is crucial for some analytical reconstruction techniques, the NNs provide a promising estimation of the muon signal fraction. Moreover, the NN is able to deduce mass-sensitive features on a single-station level by directly predicting the mass of the primary particle, serving a good discrimination capability between proton and iron-induced showers.
Combining the muon fraction predictions of the NN into an analytical event-level estimator, leads to encouraging results in two developed methods. In the first, a lateral muon fraction function is fitted to the predictions of the NN to derive a muon fraction estimator on the event level. After zenith angle and energy corrections, a suitable discrimination capability between proton and iron is achieved. In the second method, a global parametrization of the expected muon fraction is utilized to construct lateral reference profiles for proton and iron-induced showers on an event-by-event basis. By comparing the NN predictions with the reference profiles, the masses of primary particles can be estimated.

This thesis comprises a condensed summary of CR-physics in Chapter 2, which is followed by the description of the Pierre Auger Observatory in Chapter 3. The fundamentals of Machine Learning (ML) and NNs are introduced in Chapter 4. In Chapter 5, we use air-shower simulations to predict the muon fraction with NNs and validate their superiority over current methods. The predictions of the NNs are combined in analytical event-level estimators in Chapter 6. Finally, the results are summarized and an outlook is given in Chapter 7.

2. Cosmic Ray Physics

CRs are relativistic particles that have an extraterrestrial origin and are omnipresent in the Universe. They reach energies beyond 10^{20} eV, by far exceeding the energies of human-built accelerators, such as the LHC with an fixed target energy equivalent to "only" 10^{17} eV. Thus, they are the most energetic particles known to mankind. Their energy spectrum is well described by a steep power law, which rapidly decreases at the highest energies until it is completely suppressed. At time of writing, it is still unresolved whether the reason for this suppression lies in propagation effects through the galactic and intra-galactic medium or whether the sources reach their maximum achievable energy. In the 21^{st} century, CRs have become one of the most fascinating subjects of fundamental physics to study due to their potential to advance to new scientific frontiers by linking the smallest (particle physics) with the largest (astrophysics) scales known to mankind.
In this chapter, the discovery and the most important properties of CRs are discussed. Possible production mechanisms, propagation effects, and detection techniques are introduced and put into context.

2.1. Historic Overview

Alongside the discovery of natural spontaneous radioactivity by Antoine Henri Becquerel in 1896, the steady discharging of conductors in electrically isolated containers [1] could finally be unraveled. The discharging of electroscopes in dry air [2, 3] had been subsequently attributed to ionized gas atoms by radioactivity of the environment and the Earth's crust. However, the origin of the radiation remained hidden. The scientific consensus around 1909 favored the Earth itself as the primary source containing radioactive elements [4]. In the early 20^{th} century this theory was tested by experiments in various altitudes and environments with Wulf's newly developed electrometer. By measuring the ionization rate above ground level, including the Eiffel Tower [5], and altitudes up to 4500 m [6, 7] using balloon flights, no significant decrease in ionization rate was found. Domenico Pacini took measurements with the Wulf electrometer placed underwater and found a decrease of the ionization rate. He concluded that there exists a sizable amount of radiation independent of the radioactive soil [8].
The next logical step was to conduct measurements at higher altitudes within the atmosphere. It was Victor Hess in 1912 who came up with the correct experiment setup to verify the hypothesis of an extraterrestrial source of radiation. He carried out a successful series of balloon flights throughout day and night, equipped with three hermetically sealed electroscopes. At an altitude of 5350 m, the ionization rate increased up to twice the amount of that at ground level. Hess concluded that those findings can be well described by an omnipresent, highly penetrating radiation entering the atmosphere from out of

space, uncorrelated with the Sun [9]. For the discovery of the cosmic radiation Victor Hess was awarded with the Nobel Prize (NP) in 1936.
The development of new detection techniques in the following years paved the way for all subsequent discoveries in the field of astroparticle physics. They enabled studies of fundamental properties of cosmic rays, which provided the evidence for and against different hypotheses about the CRs particle type. In the 1920s, the scientific community believed that the radiation consisted of high energetic photons which were produced in atom-annihilations in the interstellar space. From this hypothesis, led by Milikan, the name *cosmic rays* was born. Even though cloud chamber (1927) and stereoscopic photograph (1929) experiments from Skobelzyn implied an ionic origin, it was not clear whether the CRs itself or their secondary products are charged particles. The answer was provided by the *east-west effect*, predicted by Rossi in 1934 [10]. Following his hypothesis the East-West asymmetry in intensity originates from predominantly positively charged particles with energies of several GeV. Later, the primary particles could be further decomposed in protons, alpha particles, and heavier nuclei in balloon flights. Extending previous coincidence measurements of CRs by shielding the Geiger-Müller counters [11] with lead and iron absorbers of different thicknesses up to 1 m, Rossi found an increase of coincident signals. He thus draw the conclusion that CRs produce secondary particles which can at least be grouped into a soft and a hard penetrating component [12]. The latter being identified as the muon μ, after its discovery through elementary particle physics in 1936. Due to its high mass ($m_\mu \approx 106\,\mathrm{MeV}/c^2$), resulting in low energy losses through ionization or radiation, and its long lifetime ($2.2{\times}10^{-6}$ s) the excessive penetration power of CRs was explained.
In the year 1939 Pierre Auger provided the first experimental proof of EAS (Sec. 2.5) and shower-cascade theory by increasing the distance between the coincidental measurement apertures up to 300 m. His measurements of the differential CR spectrum do agree with a power law $\mathrm{d}N/\mathrm{d}E \propto E^{-2}$. As a consequence of his initial estimate of the primary energy to exceed 10^{15} eV, he draw the correct conclusion that probably not a single process, but rather enormous and vast electric fields are responsible for the acceleration of CRs [13].

2.2. Acceleration and Sources

Even after 100 years of research, there are still open questions about the exact origin of CRs, like “What are the acceleration mechanisms for CRs?” and “Which sources are capable of providing these extreme energies?”. In general, it is assumed that CRs at the highest energies are accelerated in the most energetic processes of the Universe. Especially, regions with large electromagnetic-field strengths or spatial extents have been identified to provide feasible answers. Among the most prominent sources like Active Galactic Nuclei (AGN), Starburst galaxies, pulsars (highly magnetized spinning neutron stars), Supernovae explosions and accreting black holes, there are probably many more dynamic processes and phenomena which are still not understood in detail.

2.2.1. Hillas Criterion

In the search for possible sources of CRs, the Hillas criterion [14, 15] is indispensable and serves as a starting point. By relating the spatial extend L with the flux density B of the magnetic field in the source region, the particles maximum reachable energy $E_{\max}$ of particles can be examined. The maximum energy is inherently limited by the Larmor radius $R_{\mathrm{L}} = m\,v_{\perp}/(|q|\,B)$ exceeding the spatial extension of the magnetic field provided by the source. Thus, particles with charge $q = Ze$ exceeding the maximum energy $E_{\max}$ will unavoidably escape the acceleration side, leading to a spatial confinement condition, which is further influenced by the age-dependent velocity of the shock front β_{s}

$$E_{\max} \sim \beta_{\mathrm{s}}\, Z\, e\, B\, L. \tag{2.1}$$

The categorization of various known sources in dependency of their magnetic-field strength and spatial size in a diagram has practical implications. Fig. 2.1 shows the directly visible indication of an energy depended source contribution to the CR energy spectrum (Sec. 2.4). Hence, the Hillas criterion already rules out a vast majority of sources for UHECRs. The remaining possibilities are either very compact objects with enormous magnetic fields like magnetars, located in the upper left in Fig. 2.1, or large scale structures with tiny magnetic-field strengths like whole galaxy clusters, located in the lower right. Although the Hillas criterion provides the necessary condition for the energy of CR sources, it is not sufficient to explain the detected rate of CRs because it neglects energy losses in the source environment and during propagation.

2.2.2. Fermi Acceleration

One of the predominant powerhouses for CR production is given by the Fermi acceleration [17]. The theory describes the gradual transfer of kinetic energy from moving magnetized plasma to individual charged particles from the Interstellar Medium (ISM). After [18] the magnetic fields are co-moving with the plasma and thus serving as magnetic mirrors for charged particles. A particle entering the plasma will be eventually diffusely and collision-free reflected by the magnetic field. The particle energy E_n after n encounters is in general modeled as a successive fractional gain α with respect to the injection energy E_0

$$E_n = E_0(1+\alpha)^n. \tag{2.2}$$

Diffusive Shock Acceleration

If the diffusion length inside the plasma is significantly smaller than the spatial expanse of the plasma, the boundary can be modeled as a plane [15] which resembles an elastic reflection against a wall. The most prominent candidate satisfying this condition is the Supernova Remnant (SNR) with its corresponding shock wave. The net energy gain $\Delta E = \alpha E_0$ of the impinging particle with injection energy E_0 after one encounter is proportional to the velocity of the plasma β_{s} [15]

$$\frac{\Delta E}{E_0} = \alpha \propto \frac{4}{3}\beta_{\mathrm{s}}. \tag{2.3}$$

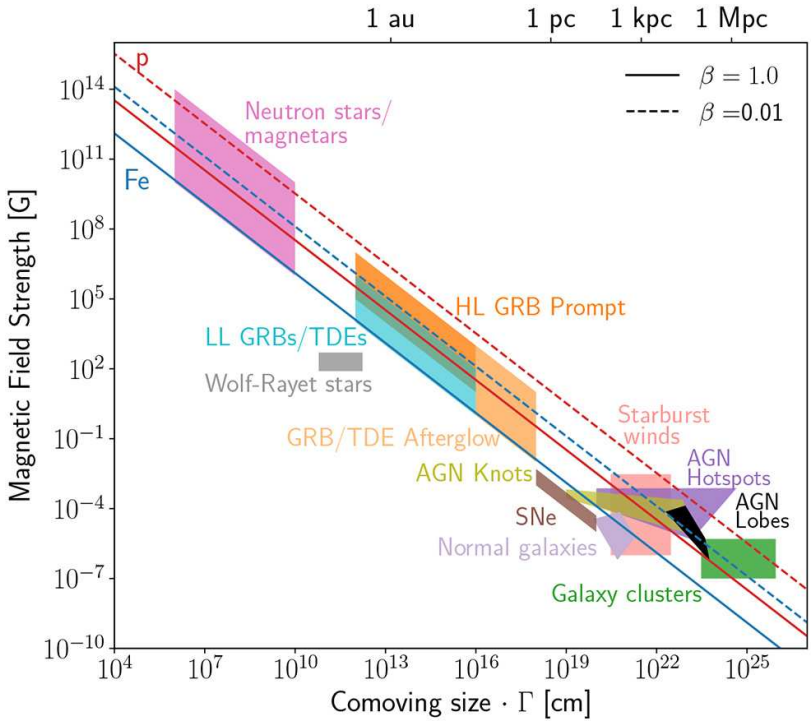


Figure 2.1.: The Hillas diagram locates possible CR sources in dependency of their comoving size and magnetic-field strength. The comoving size is a distance measure that is independent from the expansion of the Universe. The regions beyond the diagonals indicate the possible acceleration of a proton (*red*) or iron (*blue*) nuclei up to $E_{\mathrm{max}} = 10^{20}$ eV. The effect of different shock velocities β is depicted with the solid and dashed lines. Adapted from [16].

With the relative energy gain being linear in the velocity, it is generally referred to as *first-order Fermi acceleration*. Due to typical shock front velocities $\beta_s \leq 0.1$, multiple acceleration cycles are required to explain the observed non-thermal CR energy spectrum. Thus, an additional back-scattering mechanism outside the plasma is necessary, which bends the trajectory back on the shock front for reoccurring cycles. The electromagnetic fields present in the ISM yield this focusing effect.

Second-order Fermi Acceleration

Another energy gain arises if the scattering angle is not constrained as it is the case in moving turbulent plasma clouds. The different underlying geometry of the scattering process leads to a less-efficient energy gain per encounter, being of second order in velocity and is thus usually referred to as the *second-order Fermi acceleration* [15],

$$\frac{\Delta E}{E_0} = \alpha \propto \frac{4}{3}\beta_s^2. \tag{2.4}$$

Predicted Energy Spectrum

In every cycle there exists a statistical probability p inherent to diffusion processes, to escape from the acceleration. Thus, after n cycles the respective remaining number of particles N is given by

$$N(n) = N_0\,(1-p)^n\,. \tag{2.5}$$

Since n cycles are required to reach an energy of $E = E_0(1+\alpha)^n$, the remaining number of particles can be rewritten in terms of their energy as

$$N(E) \overset{\text{Eq. (2.2)}}{=} N_0\,(1-p)^{\frac{\ln(E/E_0)}{\ln(1+\alpha)}} = N_0\left(\frac{E}{E_0}\right)^{\frac{\ln(1-p)}{\ln(1+\alpha)}}. \tag{2.6}$$

By deriving the number of particles per energy interval, a natural differential power law spectrum or flux $\Phi(E)$ arises [19] as

$$\Phi(E) = \frac{\mathrm{d}N(E)}{\mathrm{d}E} \sim E^{\frac{\ln(1-p)}{\ln(1+\alpha)}-1} := E^{-(\gamma+1)}. \tag{2.7}$$

The spectral index $\gamma = -\frac{\ln(1-p)}{\ln(1+\alpha)} > 0$ describes the steepness of the energy spectrum and depends on the age of the SNR (through $\alpha(\beta_S)$) and the injected particle (through p). The Fermi acceleration derives in a simple manner that the expected energy spectrum of CRs should follow a power law which is characterized by the spectral index γ. Results from several studies [16, 20, 21] indicate that the SNRs are the main source of galactic CRs with energies up to a few PeV. However, the maximum reachable energy is limited by the age of the SNR and can thus not account for the contribution up to UHECR.

2.3. Propagation of Cosmic Rays

After various potential CR sources have been introduced in the previous sections, their identification as such is not easily accomplished. If a CR particle is detected on Earth and its arrival direction is successfully reconstructed, backtracking its trajectory to a possible source is only straightforward for neutral particles like neutrinos ν, photons γ or neutrons n. They travel along straight geodesics and possible interactions either lead to the particle not arriving at the detector in the first place (pair production for photons) or are negligible rare (neutrinos) for the most CR observatories.
On the other hand, modeling the propagation of charged CRs is a complex topic and appropriate assumptions and approximations, like boundary conditions, have to be made. The most general form of propagation is described by a transport equation modeling the particle density n_i for particle type i, which is in general untractable to solve due to various coupled differential equations of other particle types j [15],

$$\frac{\partial n_i}{\partial t} + \underbrace{\nabla(v n_i)}_{\text{convection}} - \underbrace{\nabla(D \nabla n_i)}_{\text{diffusion}} \tag{2.8}$$

$$= \underbrace{Q_i}_{\text{source}} - \underbrace{\frac{\partial (b_i(E) n_i)}{\partial E}}_{\text{energy}} + \underbrace{\sum_{j>i} \frac{p_{j\to i} n_j}{\tau_{\text{spal},j}} - \frac{n_i}{\tau_{\text{spal},i}}}_{\text{spallation}} + \underbrace{\sum_{j>i} \frac{p_{j\to i} n_j}{\gamma_j \tau_{\text{rad},j}} - \frac{n_i}{\gamma_i \tau_{\text{rad},i}}}_{\text{weak decay}} . \tag{2.9}$$

To simplify this equation a temporal equilibrium state is assumed in the majority of these models. The particle flow is then caused by diffusion and convection processes which respond to ejected material from the source. For the sake of simplicity, the particle density is assumed and defined to be a quantity of space, time, and energy $n_i := n_i(\vec{x}, t, E)$. The different terms are explained in the following:

- **Convection** $\nabla(\vec{v}(\vec{x}, t) n_i)$
  To account for motion of the containment medium, a space and time dependent velocity field $\vec{v}(\vec{x}, t)$ is introduced. A simple analogy would be the water flow in a river, where the particle density corresponds to a salt concentration.

- **Diffusion** $\nabla(D(\vec{x}) \nabla n_i)$
  The diffusion term describes the random walk of the particles inside the containment volume similar to Brownian motion. In general, the diffusion coefficient $D(\vec{x})$ is a tensor depending on space and accounting for anisotropic diffusion directions.

- **Source** $Q_i(\vec{x}, t, E)$
  The source term injects particles of type i into the containment volume. In general, the contribution is space, time, and energy dependent. Many theoretical models assume a source spectrum of $Q(E) \propto Q_0 E^{-p}$ with $p \in [2, 2.3]$.

- **Energy** $\frac{\partial (b_i(E) n_i)}{\partial E}$
  The energy term accounts for continuos energy losses through interactions with the ISM during propagation and are summarized in $b(E)$. For charged particles radiation and ionization losses dominate.

- **Spallation** $\sum_{j>i} \frac{p_{j\to i}}{\tau_{\text{spal},j}} n_j - \frac{n_i}{\tau_{\text{spal},i}}$
  Spallation of nuclei describes the disintegration of primary nuclei into secondary fragments through highly inelastic collisions. This can destroy particles of type i through collision with the ISM $\left(\frac{1}{\tau_{\text{spal},i}} = \frac{n_{\text{ISM}} v_i}{\lambda_i}\right)$, as well produce them through resulting fragments of heavier particle cascades j. In this transport equation it is assumed that the spallation process occurs with a constant probability $p_{j\to i}$ over a specific time τ_{spal} and is energy independent.

- **Weak decay** $\sum_{j>i} \frac{p_{j\to i}}{\tau_{\text{rad},j}} n_j - \frac{n_i}{\tau_{\text{rad},i}}$
  A weakly unstable particle j has a probability of $p_{j\to i}$ to decay during a Lorentz-dilated lifetime τ_{rad} into lighter particle i. This process happens spontaneously and is energy independent. In this transport equation the decay from heavier particles j, as well as into lighter particles, is accounted for.

2.3.1. Effects of Magnetic Fields

Charged CR particles are deflected by omnipresent galactic and extragalactic magnetic fields on their way towards Earth. While the knowledge of the local galactic field is steadily increasing and sufficient to model its behavior, little is known about the extragalactic magnetic fields where coherence lengths and field strengths vary over several orders of magnitude. The deflection of a charged-particle trajectory originating from its source is described by the Larmor radius R_{L}, which is influenced by particle properties (charge Ze, energy E) that are summarized in the rigidity R (for high energies $E \propto p$) and the magnetic field strength $B_\perp$ perpendicular to the particle trajectory [15],

$$R_{\text{L}} = \frac{E}{ZeB_\perp} = \frac{R}{B_\perp}. \tag{2.10}$$

The rigidity describes the resistance of a particle to deflection by magnetic fields and relates internal particle properties with external source properties $R = \frac{E}{Ze} = R_{\text{L}} B_\perp$. Thus, it is an important quantity to describe the dependency of CR features and to identify source candidates, if the particle properties are known or vice versa to distinguish between different primary particles given the spatial extend and magnetic field strength of a source. In the small-angle approximation where the propagation distance d is in the same order as the Larmor radius, one can derive the angular deflection from the initial trajectory after ejection from the source as

$$\delta\theta = \frac{d}{R_{\text{L}}} = d\frac{ZeB_\perp}{E}. \tag{2.11}$$

Due to the deflection being minimal for highly rigid particles, high-energy light particles, like protons, are suitable candidates to pin-point the trajectory back to a single source.

2.3.2. Galactic Propagation

Leaky Box Model

The Leaky Box Model is one of the simplest models describing the propagation of CRs by approximating Eq. (2.9). It assumes a free propagation in a containment volume with a constant probability of escape per unit time τ_{esc} replacing the convection and diffusion terms. This is equivalent to a small escape probability every time a particle reaches the boarder of a box-like containment volume. This approximation is only valid for mean propagation lengths exceeding the thickness of a galaxy $c\tau_{\text{esc}} \gg h$. The probability of the particle remaining inside the box after time step t is then given by $\exp(-t/\tau_{\text{esc}})$. Thus, τ_{esc} directly reflects the mean time a particle is contained inside the box and thus is directly related to the mean traversed matter $\lambda_{\text{esc}} = \rho\,\beta\,c\,\tau_{\text{esc}}$ [15]. From there, one can derive some interesting implications resulting from the Leaky-Box model in conjunction with measured ratios of elements at given energies.

- **Energy-Dependent Escape Length**
  Looking at the measured ratio between carbon and boron in equilibrium ($\frac{\partial n_i}{\partial t} = 0$), one derives an energy and thus rigidity dependent escape length $\lambda_{\text{esc}} \propto E^{-\delta} \propto R^{-\delta}$.

- **Source Spectrum**
  By only retaining the source and escape term for particles which rather escape than interact $\lambda_{\text{esc}} \ll \lambda_i$, the source spectrum is directly coupled to the observed CR spectrum (Sec. 2.4) and has to be of the form $n_i \propto E^{-(\gamma+1)} \propto Q(E) \propto E^{-\alpha} = E^{-(\gamma+1-\delta)}$

- **Cosmic-Ray Clocks**
  Unstable nuclei with lifetimes on the order of the escape time are suitable as *cosmic-ray clocks*. By measuring the ration of an unstable to stable isotope ($^{10}\text{Be}/^{9}\text{Be}$), the escape time, and hence the average traversed hydrogen density, can be derived. With the resulting density being significantly lower than the density of the ISM, one can conclude that the CRs mainly propagate in the Galactic halo.

2.3.3. Extragalactic Propagation

The energy losses during extragalactic propagation of CRs is dominated by interactions with the Cosmic Microwave Background (CMB) being the only statistically significant interaction medium. As protons are suitable candidates to track back their trajectories to potential sources (Sec. 2.3.1), their energy losses are of a particular interest in order to estimate their track length. A few of the most important interactions are described briefly in the following and are summarized in Fig. 2.2.

GZK Cutoff

During propagation through the interstellar medium, UHECR protons can undergo an inelastic scattering with the omnipresent CMB photons, creating an unstable Δ^+ resonance. By the prompt decay of this resonance the resulting proton (or neutron) quickly loses its

energy and falls below the reaction threshold,

$$\mathrm{p} + \gamma_{\mathrm{CMB}} \rightarrow (\Delta^{+})^{*} \rightarrow \begin{cases} \mathrm{p} + \pi^{0} \\ \mathrm{n} + \pi^{+} \end{cases} \tag{2.12}$$

This energy threshold is called Greisen-Zatsepin-Kuzmin (GZK) cutoff, named after its discoverers Greisen [22] and independently Zatsepin and Kuzmin [23], and describes the starting point of the photo-pion production at about $E_{\mathrm{p}} \approx 6{\times}10^{19}$ eV energy. This value is determined by the kinematics of the interaction, namely the four momenta of proton q_{p}, photon $q_{\gamma_{\mathrm{CMB}}}$, the mean energy of the CMB Planck distribution $E_{\gamma_{\mathrm{CMB}}} \approx 1.1\,\mathrm{eV}$ and the rest masses of proton m_{p} and pion m_{π},

$$(q_{\mathrm{p}}^{\mu} + q_{\gamma_{\mathrm{CMB}}}^{\nu})^{2} = (m_{\mathrm{p}} + m_{\pi})^{2} \tag{2.13}$$

$$\Leftrightarrow E_{\mathrm{p}} = \frac{m_{\pi}^{2} + 2m_{\mathrm{p}}m_{\pi}}{4E_{\gamma_{\mathrm{CMB}}}} \approx 6{\times}10^{19}\,\mathrm{eV}, \tag{2.14}$$

$$\text{and} \qquad \frac{\Delta E_{\mathrm{p}}}{E_{\mathrm{p}}} = \frac{m_{\pi}}{m_{\mathrm{p}} + m_{\pi}} \approx 0.125. \tag{2.15}$$

This energy loss indicates a natural limit to the mean free path of UHECR protons with an energy around 10^{20} eV to less than 100 Mpc as depicted in Fig. 2.2.

Pair Production and Photo-Disintegration

Another important energy-loss process is the Bethe-Heitler pair production through the CMB [15],

$$\mathrm{p} + \gamma_{\mathrm{CMB}} \rightarrow \mathrm{p} + \mathrm{e}^{+} + \mathrm{e}^{-}. \tag{2.16}$$

The energy threshold and the relative energy loss is given by the mass of the electron m_{e} and the rest mass of the proton m_{p}

$$s_{\mathrm{th}} = (m_{\mathrm{p}} + 2m_{\mathrm{e}})^{2} \approx 6{\times}10^{17}\,\mathrm{eV}, \tag{2.17}$$

$$\text{and} \qquad \frac{\Delta E_{\mathrm{p}}}{E_{\mathrm{p}}} \approx \frac{2m_{\mathrm{e}}}{m_{\mathrm{p}}} \approx 10^{-3}. \tag{2.18}$$

Due to the relative energy loss being only of order 10^{-3} for protons and decreases even further for heavier primaries, it is small in comparison to the GZK energy loss at energies beyond 100 EeV.
In the case the CMB or infrared photon gets absorbed by the nuclei, photo-disintegration occurs and the nucleus is split into two lighter nuclei. Typical fragments are neutrons, protons or alpha particles, thus the reactions are denoted by (γ, n), (γ, p), (γ, α)

$${}^{A}\mathrm{X} + \gamma \rightarrow {}^{A\text{-k}}\mathrm{Y} + k \qquad\qquad k = 1, 4, \ldots. \tag{2.19}$$

The photo-disintegration scenario predicts a composition-dependent change of the spectrum proportional to the atomic mass number A at the highest energies.

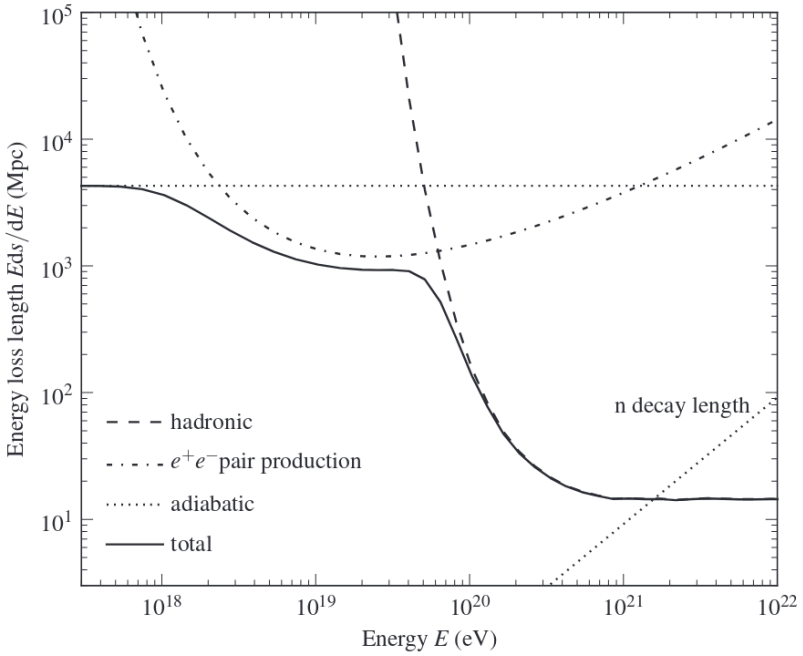


Figure 2.2.: Dominant energy losses of protons during extragalactic propagation through the CMB. The GZK cutoff (*hadronic dashed line*) limits the mean propagation path drastically to 100 Mpc around 10^{20} eV. Taken from [15].

Adiabatic Energy Loss

Due to the expansion of the Universe, the energy of CR particles is reduced by the adiabatic energy loss. The energy loss is given by the expansion rate of the Universe given by the Hubble constant $H_0 \approx 70\,\mathrm{km\,s^{-1}}$ Mpc,

$$-\frac{1}{E}\frac{\mathrm{d}E}{\mathrm{d}t} = H_0. \tag{2.20}$$

2.4. Energy Spectrum of Cosmic Rays

The all-particle spectrum of CRs is given in Fig. 2.3 and spans over a huge range of eight orders of magnitudes in energy, from PeV to more than 100 EeV and is thus plotted in log-log scale. Measured data from various experiments have to be put together to account for this enormous range in energy, as typically one experiment focuses on a certain energy range. The detected rate of particles ranges from frequently observed events $10^6\,\mathrm{km^{-2}\,s^{-1}}$ in the GeV range, down to events occurring barely once a year per $\mathrm{km^2}$ in the EeV energy regime. As the flux follows a steep power law, which was already indicated for energies up to PeV in Secs. 2.2.2 and 2.3.2, it is scaled by $E^{-2.5}$ to counteract the rapid fall-off and highlight certain features. These characteristic features correspond to a change in the spectral index γ, and are consequently reflected as kinks in the spectrum. Resulting

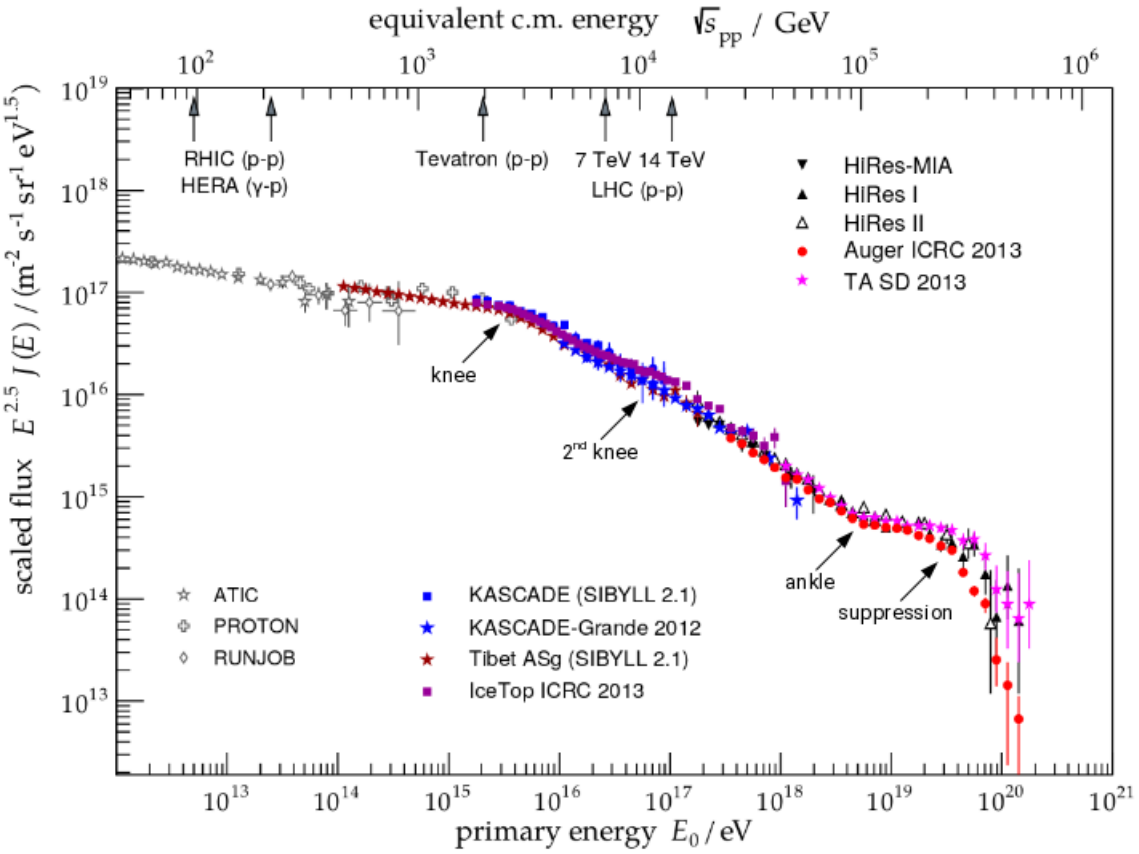


Figure 2.3.: Combined energy spectrum of the CRs as measured by various experiments. The flux is scaled by a factor $E^{2.5}$ to highlight characteristic features. The achievable energy references of human-made accelerators are indicated on the top, lead by the LHC with an fixed-target energy equivalent of 10^{17} eV, barely reaching the second knee. Modified from [15].

implications and interpretations of those features and their corresponding energy intervals reveal interesting insights into properties and the origin of CRs. Upon writing of this thesis, four distinct spectral breaks above an energy of ~ 1 PeV have been identified, which are discussed in the following.

The Knee ($E \approx 3{\times}10^{15}$ eV)

The first break of the all-particle spectrum occurs around $3{\times}10^{15}$ eV and is generally referred to as the *knee*. Composition measurements in the form of individual single particle spectra, as it were done by KASKADE-Grande [24], observed an energy dependent steepening, proportional to the charge $E \approx Z\,E_{\mathrm{p,max}}$ of the primary. Thus, the break in the knee is assigned to a break in the spectrum of the lightest particles (protons). Whether this rigidity-dependent break is caused by Galactic sources, like SNR reaching their maximum energy (Secs. 2.2 and 2.2.1), or highly rigid particles start escaping the Galaxy due to limited confinement by the Galactic fields (Sec. 2.3), is still an open question.

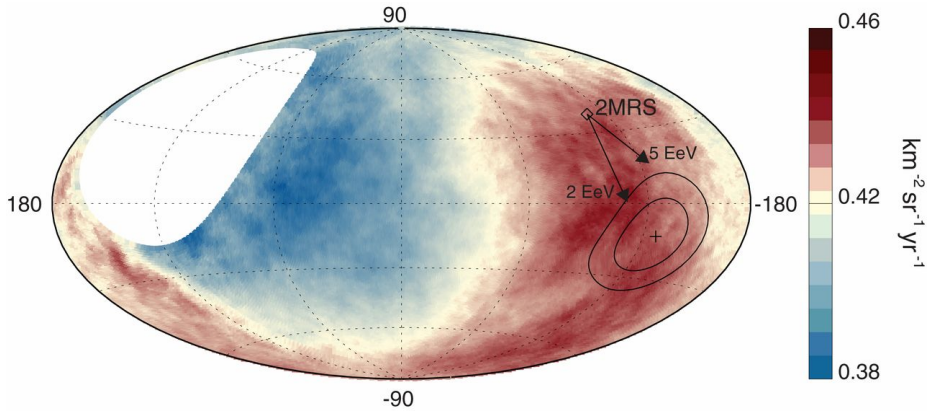


Figure 2.4.: Sky map of the smoothed CR flux above 8 EeV in Galactic coordinates with the Galactic center at the origin as reported by Auger. The large-scale anisotropy is described by a dipole with a 5.2σ significance. The cross indicates the direction of the dipole together with the 1σ and 2σ confidence contours. The arrows indicate the rigidity dependent deflection of particles $R = E/Z = 2.5\,\text{EeV}$ due to the Galactic magnetic field. Because the dipole has no correlation with the Galactic plane, this is a strong indication for an extragalactic origin and supports the transition towards extragalactic UHECR at the *ankle*. Taken from [26].

The Second Knee ($E \approx 10^{17}$ eV)
Following the rigidity reasoning as done in the explanation of the knee, the *second knee* naturally emerges at $26\,E_{\text{p,max}} \approx 10^{17}$ eV, when the rigidity conditions for heavier particles like iron are met. Hence, it is sometimes also referred to as the *iron knee*.

The Ankle ($E \approx 4{\times}10^{18}$ eV)
The *ankle* describes the hardening of the energy spectrum around $4{\times}10^{18}$ eV. According to the current scientific agreement, it depicts the transition from Galactic towards extragalactic CRs. This scenario is heavily supported by arrival-direction studies and the reasoning is two fold. Firstly, if sources within our Galaxy would be capable of accelerating particles up to this energies, their deflection over a distance on the order of our Galaxy size is negligible. Hence, the reconstructed arrival direction would equal the trajectory and straightly point towards the source, which would result in an observation of anisotropies and a high correlation of the arrival directions with the Galactic plane. However, studies show no significant deviation from isotropy below 10^{18} eV [25]. Secondly, the Auger reports a 5.2σ significant large-scale anisotropy in form of a dipole (Fig. 2.4), pointing towards 100° in right ascension and -24° in declination away from the Galactic center. The dipole correlates with source catalogues of extragalactic surveys [26, 27] within a few 100 Mpc.

The suppression ($E \sim 6{\times}10^{19}$ eV)
The end of the spectrum is characterized by a rapid fall-off usually referred to as the *suppression*. The transition point where the spectral index changes again around $E \approx 6{\times}10^{19}$ eV, is in coincidence with the expected maximum energy which arises from losses during propagation, dominated by the GZK cutoff (Sec. 2.3.3). But due to the lack of

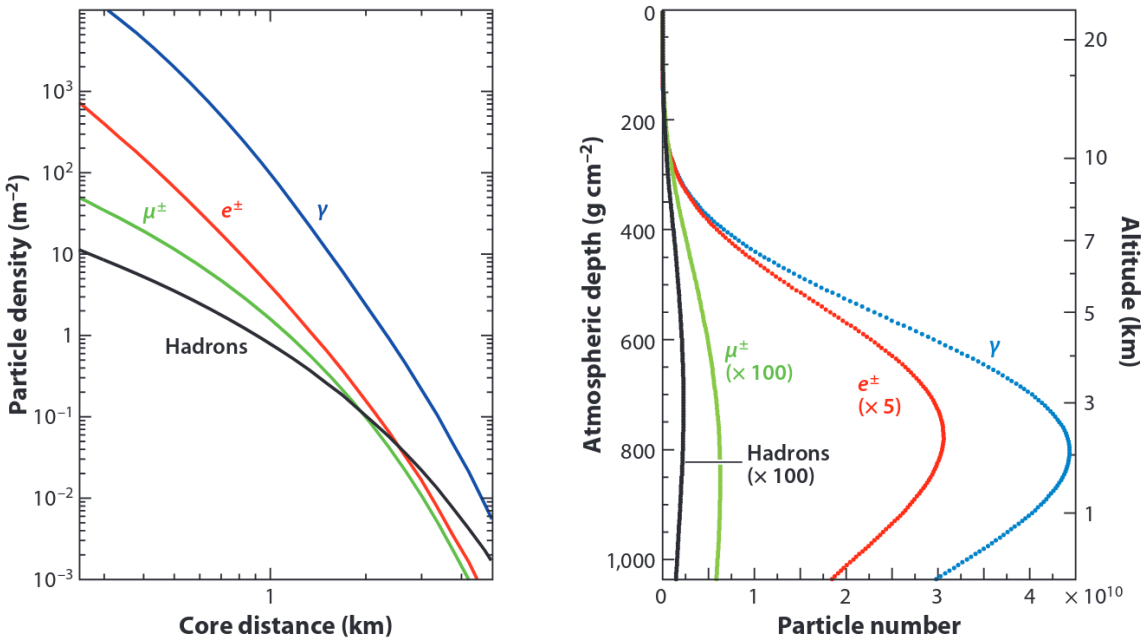


Figure 2.5.: The lateral (*left*) and longitudinal profile (*right*) of different components in an air shower, induced by a vertical primary proton with 10^{19} eV energy. The number of particles in the shower maximum, which can be seen in the longitudinal profile, is dominated by the electromagnetic component, which dies out eventually due to attenuation. The muons make up the *hard* shower component, because they experience less attenuation through the atmosphere, which results in the particle number staying almost constant after reaching its maximum. Taken from [29].

statistics in composition measurements at the highest energies, it is in general hard to distinguish this scenario from extragalactic sources reaching their highest energies. Currently, the GZK is disfavored by results from Auger [28].

2.5. Extensive Air Showers

As a primary CR impinges on Earth's atmosphere, the interactions with the atomic nuclei in the air produce a zoo of new particles. By subsequent interactions of the secondary particles, a cascade forms among the extended trajectory of the primary, until the EAS eventually hits the ground, covering a footprint of up to a few 10 km^2. At an altitude of 15 km to 35 km, the first interaction initiates the growth of the shower until it reaches the shower maximum of several billion particles, before the creation rate of particles drops below their decay rate (Fig. 2.5). The usage of the atmosphere as a calorimeter, where the primary particle deposits all of its energy and the detection of only secondary particles, is referred to as an *indirect* detection technique for CRs.
In the following sections a semi-empirical description of air shower and cascade physics is provided to derive relations of the most important shower features in absence of detailed computationally-intensive MC simulations.

2.5.1. Heitler Model

First insights into elementary air-shower physics have been provided by the simplistic electromagnetic cascade model of Heitler [30]. It sufficiently describes the basic characteristics of electromagnetic cascades in the first few generations, by relating the longitudinal profile (Fig. 2.5) with the energy of the primary particle. The cascade is initiated by a single photon with energy E_0 impinging on the atmosphere and creating an electron-positron pair ($\gamma \rightarrow e^+ + e^-$) after a depth $d = \lambda_r \ln 2$ (Fig. 2.6). It is assumed that the energy is split equally between the outgoing particles, thus each one carries an energy of $E_0/2$ after the first interaction. Consecutive interactions take place every time the depth d is traversed again, doubling the number of particles every time. After n generations the number of particles is $N_n = 2^n = e^{X/\lambda_r}$, and the traversed longitudinal slant depth is $X = n\,\lambda_r \ln 2$. The cascade comes to a halt when the ionization-energy losses dominate over the radiative losses, which is at the critical energy of $E_c = 85\,\text{MeV}$. The number of particles in the cascade then reaches its maximum N_{max} with each particle carrying an energy of $E_c = E_0/2^{n_c}$ after n_c generations. This leads to the following two central relations in electromagnetic cascade theory within the Heitler model:

$$N_{max} = 2^{n_c} = \frac{E_0}{E_c}, \quad \text{and} \quad X^{em}_{max} = n_c\,\lambda_r \ln(2) = \lambda_r \ln\left(\frac{E_0}{E_c}\right), \tag{2.21}$$

The number of particles in the shower is proportional to the primary energy E_0 and the depth of the shower maximum X^{em}_{max} increases logarithmically with energy. This approximation is valid for electromagnetic showers at high energies and validated with extensive numerical MC simulations, which for instance take into account the energy dependency of the cross sections and the varying atmospheric profile.

2.5.2. Heitler-Matthews Model

By extension of the electromagnetic cascade model of Heitler to a more general hadronically-induced air shower, Matthews derived further detailed shower characteristics [32]. The key implications relate the average atomic mass A of the primary to the number of muons N_μ in a shower, as well to the average depth of the shower maximum X_{max}. The cascade is initiated by a single proton with initial energy E_0 that produces all pion types (π^0, π^+, π^-) at an equal Branching Ratio (BR) (1:1:1) in a hadronic interaction (Fig. 2.6). While the neutral pions promptly decay into photons ($\pi^0 \rightarrow 2\gamma$ or $\pi^0 \rightarrow e^+ + e^- + \gamma$ BR (99:1)) and feed the electromagnetic component of the shower, the charged pions successively interact after a fixed hadronic interaction length $\lambda_h \approx 120\,\text{g cm}^{-2}$ in air, producing $2N_{ch}\,\pi^\pm$ and $N_{ch}\,\pi^0$ per interaction. The hadronic-component cascade ceases when the energy of $\pi^\pm$ falls below the critical energy $E^\pi_c \approx 20\,\text{GeV}$ in analogy to the Heitler model (Sec. 2.5.1). In case of the charged pions, this critical energy is determined by the mean interaction length becoming shorter than the mean decay length and thus the charged pions exclusively decay into muons ($\pi^+ \rightarrow \mu^+ + \nu_\mu$ and $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$). After n_c generations, the energy

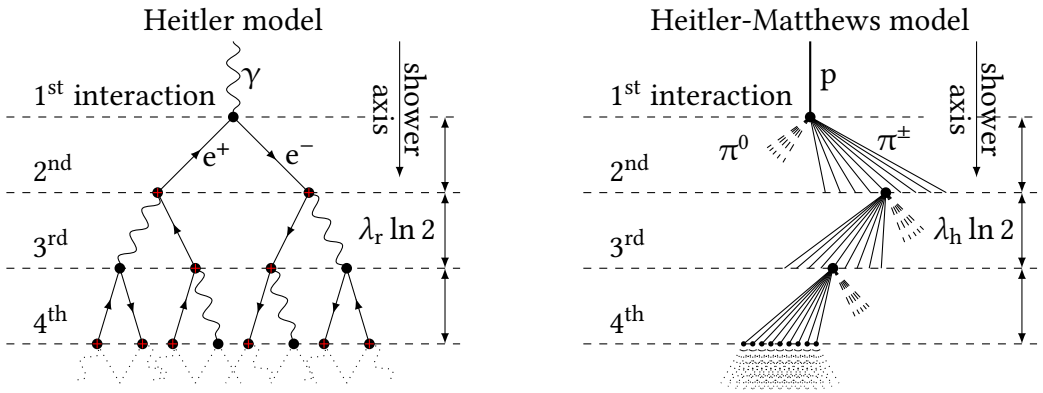


Figure 2.6.: Schematic representation of the shower development according to the Heitler model (*left*) and the Heitler-Matthews model (*right*). In the Heitler model a purely electromagnetic cascade with characteristic radiation length λ_r is initiated by a single photon undergoing pairproduction. The resulting electron-positron pair produces further photons which can themselves undergo pair production, leading to an exponential growth of the cascade until a critical energy is reached. The Heitler-Matthews model extends the electromagnetic shower development to hadronicly induced air showers. An impinging proton produces charged and neutral pions. The neutral pions promptly decay into photons, feeding the electromagnetic shower component. The charged pions undergo similar hadronic interactions as the proton, producing further charged and neutral pions after the characteristic interaction length λ_h. The cascade comes to a halt when a critical energy is reached where the charged pions preferably decay into muons instead of interacting again. Reproduced from [31].

per charged pion is given by $E_{\pi^\pm} = E_c^\pi = \frac{E_0}{\left(\frac{3}{2}N_{ch}\right)^{n_c}}$, leading to

$$n_c = \frac{\ln\left(E_0/E_c^\pi\right)}{\ln\left(\frac{3}{2}N_{ch}\right)} \overset{(N_{ch}=10)}{=} 0.85 \lg\left(\frac{E_0}{E_c^\pi}\right). \tag{2.22}$$

The number of muons N_μ are then given by the number of charged pions $N_{\pi^\pm} = (N_{ch})^{n_c}$ at the hadronic shower maximum,

$$\ln N_\mu = \ln N_\pi = n_c \ln N_{ch} = \frac{\ln N_{ch}}{\ln\left(\frac{3}{2}N_{ch}\right)} \ln\left(E_0/E_c^\pi\right) = \beta \ln\left(E_0/E_c^\pi\right) \tag{2.23}$$

$$\Leftrightarrow N_\mu = \left(\frac{E_0}{E_c^\pi}\right)^\beta \qquad \text{with } \beta \in [0.85, 0.95]. \tag{2.24}$$

The depth of the shower maximum is obtained in analogy to Eq. (2.21) from the electromagnetic shower starting at depth $X_0 = \lambda_h \ln 2$ with an energy of $E_0/(3N_{ch})$ per photon,

$$X_{max}^p = X_0 + \lambda_r \ln\left(\frac{E_0}{3N_{ch}E_c^{em}}\right) = X_{max}^{em} + X_0 - \lambda_r \ln(3N_{ch}). \tag{2.25}$$

Its rate of change per decade of logarithmic energy E_0, is given by the elongation rate theorem $\Lambda = \mathrm{d}X_{max}/\mathrm{d}\lg E_0$ [32]

$$\Lambda^p = \Lambda^{em} + \frac{\mathrm{d}}{\mathrm{d}\lg E_0}\left[X_0 - \lambda_r \ln(3N_{ch})\right]. \tag{2.26}$$

Superposition Principle

The extension of the proton induced hadronic shower to a heavier primary nucleus with atomic mass number A, is modeled by the superposition of A independent proton showers at first order. This approximation is valid, if the interaction energy is much greater than typical binding energies E_b per nuclear $E_0 \gg E_b \approx 8\,\mathrm{MeV}$ and neutrons produce a similar cascade as protons. The primary energy E_0 is assumed to be distributed uniformly over A individual independent protonic showers initiated at same depth with energy E_0/A. The shower characteristics for any primary nuclei are then deduced by inserting the reduced energy per proton shower E_0/A into the previous derived equations, and summing over all showers, thus yielding

$$N_\mu^A = A^{1-\beta} N_\mu^p, \quad \text{and} \tag{2.27}$$

$$X_{max}^A = X_{max}^p - \lambda_r \ln A. \tag{2.28}$$

The first key feature, which originates from this model, describes the increase in number of muons with the mass number of the primary nuclei in comparison to a pure proton shower at same primary energy. It is expected that an iron-induced shower will have up to $56^{0.15} = 1.8$ as many muons as a proton shower at the same energy on average. The second important feature states that, on average, heavier nuclei will reach their shower maximum earlier and not penetrate as deeply as proton showers would at the same primary energy. Hence, the number of muons as well the depth of the shower maximum are important quantities giving insights into the mass of the primary particle.

3. The Pierre Auger Observatory

The flux of UHECRs is extremely low. Less than one particle per century arrives on an area of 1 km^2 at $E \approx 10^{19}$ eV (Fig. 2.3). Therefore, a huge detection area is required to obtain sufficient event statistics at the highest energies. This is only feasible using indirect detection methods, which utilize the atmosphere as a calorimeter, measuring secondary particles that originate from shower cascades (Sec. 2.5).
Named after the French physicist Pierre Auger, the Pierre Auger Observatory is the largest CR detector in the world, specifically designed to study UHECRs above 10^{17} eV. It is located in the Pampa Amarilla in the western part of Argentina, at an average altitude of 1400 m above sea level. The observatory covers a huge instrumentalized area of 3000 km^2 (Fig. 3.1). The data taking started in 2004 and is still ongoing [33]. Its exposure until 2015 exceeded 40 000 km^2 sr yr at energies above $3{\times}10^{18}$ eV and is higher than in any other CR experiment. The observatory features a hybrid detector design consisting of the SD and the Fluorescence Detector (FD), enabling important cross-checks and calibration [34, 35]. The FD observes the longitudinal profile of the shower cascade in the atmosphere by detecting the emitted fluorescence light. The SD samples the lateral density profile of the shower footprint by measuring the light yield from Cherenkov radiation (WCD) or scintillation (Surface Scintillator Detector (SSD)), which is emitted by charged particles traversing the detectors. The detection techniques of the FD and SD have inherently different systematics [33], which allow for high-quality measurements in so called *Golden Hybrid events*.
Auger aims to probe the origin and characteristics of CRs, as well as to study hadronic interactions at energies beyond the reach of terrestrial accelerators.

3.1. Surface Detector

The main SD array is composed of 1660 cylindrical WCDs set up in a triangular grid with 1500 m spacing (Fig. 3.1). Each of the SD stations includes a 3.6 m diameter tank which contains 12 t of ultra-pure water. Charged relativistic particles traversing through the water produce Cherenkov radiation, which is diffusely reflected on the inner surface and collected by three 9 inch Photomultiplier Tubes (PMTs). The recorded signals are digitized via a Flash Analog-to-Digital Converter (FADC) and calibrated in Vertical Equivalent Muon (VEM) units, which is the most likely equivalent signal of one muon traversing vertically through the center of the detector [36]. Every 61 s a charge and a pulse-height histograms are filled with the signals from atmospheric background muons passing a single tank at a rate of 3 kHz [37], producing a peak in both of them. The peak of the charge histogram represents the most likely integrated signal of the muon buffers as a measure of deposited energy and is thus used to calibrate the time traces which are used for high-level reconstruction. The corresponding unit is given by VEM ($\mathrm{VEM_{charge}}$) or MIP

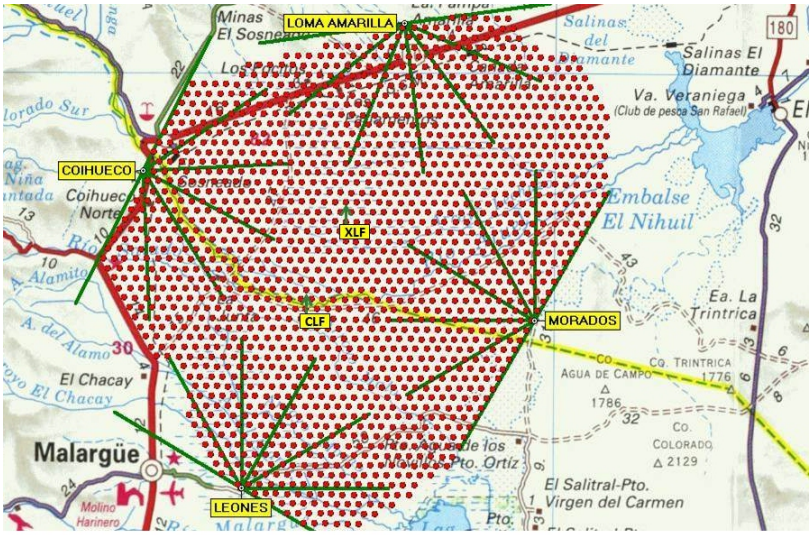


Figure 3.1.: Overview of the Auger Observatory Array in the Pampa Amarilla, Western Argentina. The four FDs (*green*) are overlooking the array of SD stations (*red dots*). The total instrumentalized area covered by the triangular grid amounts to 3000 km^2. Taken from [33]

.

($\mathrm{MIP}_{\mathrm{charge}}$). The pulse-height histogram on the other hand is exclusively used to calibrate the signal time traces and local trigger thresholds of the SD stations in units of $\mathrm{VEM}_{\mathrm{peak}}$ or $\mathrm{MIP}_{\mathrm{peak}}$ on site. By relating both quantities the area-over-peak ratio a_{p} is derived

$$a_{\mathrm{p}} = \frac{Q_{\mathrm{VEM}}}{I_{\mathrm{VEM}}}. \tag{3.1}$$

In simulations, the area over peak a_{p} is fixed to 7.3586 (WCD) and 3.5339 (SSD) for the Upgraded Unified Board (UUB). Thus, if one is using measured data with non-constant area-over-peak values, due to detector aging, they have to be taken into account for every station. Each SD station is self-powered by a battery and solar panel and communicates with the Central Data Aquisition System (CDAS). An example upgraded SD station is illustrated in Fig. 3.2. The main array was designed to measure UHECR induced air showers above 10^{19} eV, with a duty cycle of 100%. To extend the measurements to lower energies the *Infill* sub-array was commissioned in 2011, containing 61 stations with a triangular grid spacing of 750 m. It is located near the Coihueco FD side and spans a total area of 23.5 km^2.

Triggers

To identify an UHECR air shower as such, the Auger has implemented various trigger levels to optimize the efficiency of recorded events while suppressing as much background as possible. They range from simple threshold triggers on a station level up to coincidence triggers between stations to form an event. The SD triggers exceed 99% efficiency for energies above $3{\times}10^{18}$ eV [33]. The trigger levels can be divided into three stages.

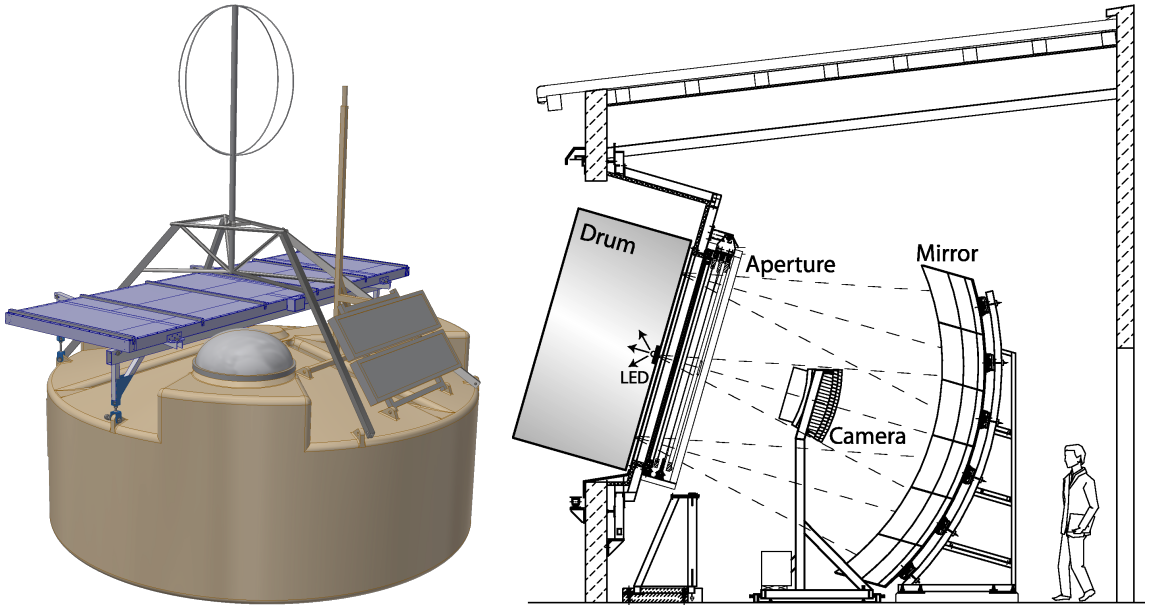


Figure 3.2.: Illustration of an SD (*left*) and an FD (*right*) detector of Auger. The self-contained, upgraded SD station is comprised of the WCD, the SSD, the radio antenna, the solar panel and the Data AcQuisition (DAQ) communication antenna. The FD telescope consists of a segmented mirror and a camera made out of PMTs. Taken from [38, 39].

- **T1 (Single Station Level)**
  The T1 triggers feature the most straightforward way to decide if a signal should be recorded or not. This is done by imposing a fixed threshold requirement on the signal amplitude. This threshold is calibrated in such a way that the trigger rate is reduced from 3 kHz due to atmospheric muons to only roughly 100 Hz [37].
- **T2 (Single Station Level)**
  The T2 triggers are based on similar principles as the T1 triggers, but in a more sophisticated way. It implements additional temporal and spatial coincidence stages on a station level in two ways. One is a temporal-coincident threshold trigger for all three PMTs in the tank. The other one requires the signal amplitude of two PMTs exceeding a certain threshold within a temporal window.
- **T3 (Event Level)**
  On top of the two station level triggers, the CDAS implements a third trigger level, which is based on the coincidence of at least three T2 triggers in a triangular grid (T3) of neighboring stations (Fig. 3.3). The T3 trigger, as an array trigger, is used to form an event and initiates data acquisition and storage.

3.2. Fluorescence Detector

The FD consists of four observation sites (Los Leones, Los Morados, Loma Amarilla, and Coihueco) built atop small elevations on the perimeter of the SD array, facing towards the

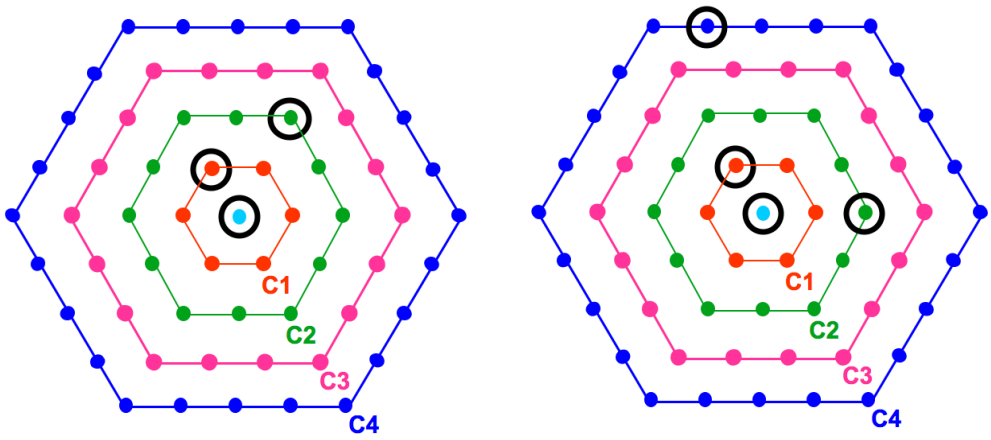


Figure 3.3.: Schematic trigger configurations of the SD T3 trigger level that form an event. At least three T2 triggers have to coincide in a hexagonal grid of neighboring stations. Taken from [37].

interior of the array (Fig. 3.1). Each FD site contains six independent telescopes, with an individual Field of View (FoV) of 30° in azimuth and 30° in zenith, summing up to a total of 180° coverage in azimuth. Additionally, three high-elevation fluorescence telescopes (*High-Elevation Auger Telescopes (HEAT)*) are located at the Coihueco side, overlooking the *Infill* array. The inclination of about 60° opens the observation of low-energy showers, extending the FD energy range down to 10^{17} eV. Due to the very high sensitivity of FD telescopes, it is only possible to operate them during dark nights, resulting in a low duty cycle of 10%.
The telescope main component is the 13 m^2 spherical segmented Schmidt , which focuses the collected UV fluorescence light onto a camera of 440 pixels with photomultiplier light sensors arranged in 22 rows and 20 columns. To improve the signal-to-noise ratio, a UV filter is mounted on the aperture system behind the shutter, filtering out the background light flux.
Charged particles traversing the atmosphere excite nitrogen molecules which emit fluorescence light in the 300 nm to 430 nm wavelength during deexcitation [40]. The electromagnetic energy losses by the charged particles are proportional to the number of emitted fluorescence and thus, measuring their rate as a function of slant depth X, results in the development of the longitudinal profile $\mathrm{d}E/\mathrm{d}X$ of an air shower. Integrating this profile gives the total electromagnetic calorimetric energy E_{cal}, which corresponds to approximately 90% ($\epsilon = 0.9$) of the primary energy E_0. Due to a limited FoV of the FD, the longitudinal profile is not always completely observable, which requires an appropriate function to extrapolate into unobserved depths. The Gaisser-Hillas function [41] $f_{\mathrm{GH}}(X)$ provides a good description of measured longitudinal profiles and is used as a parameterization of

the shower development,

$$E_0 = \epsilon^{-1} \times E_{\mathrm{cal}} = \epsilon^{-1} \int \frac{\mathrm{d}E}{\mathrm{d}X}\,\mathrm{d}X\,, \tag{3.2}$$

$$\text{where} \quad \frac{\mathrm{d}E}{\mathrm{d}X} = f_{\mathrm{GH}}(X) = \left(\frac{\mathrm{d}E}{\mathrm{d}X}\right)_{\mathrm{max}} \left(\frac{X - X_0}{X_{\mathrm{max}} - X_0}\right)^{\frac{X_{\mathrm{max}} - X_0}{\lambda}} \exp\left(\frac{X_{\mathrm{max}} - X}{\lambda}\right). \tag{3.3}$$

Detecting UHECRs through nitrogen fluorescence light is a well established technique, used various times before, for example in in Fly's Eye [42], HiRes [43], and today in Telescope Array [44] experiments. Major advantages of the FD technique are the provision of a mass-composition estimate by direct observations of X_{max}, and an hadronic-model-independent cross calibration of the SD array.

Atmospheric Monitoring

The fluorescence yield of nitrogen molecules depends on the atmospheric conditions, such as pressure, temperature, and humidity. To correct for a change in the environment, the FD array is equipped with a network of weather stations, measuring the atmospheric conditions at each FD site. The data is used to correct the measured fluorescence light flux for the atmospheric conditions at the time of the observation [33].

3.3. Air-Shower Event Reconstruction

The main goal of the air-shower reconstruction is to infer the direction of the impinging primary particle, and its properties like the energy and mass, by interpreting the recorded data.

3.3.1. Fluorescence Detector

In a first step, the shower axis is reconstructed from the timing information of triggered pixels that form a track on the camera [40]. After the shower-detector plane and the shower axis are known, the collected signal per pixel amounts to the differential calorimetric energy deposit as a function of slant depth in the atmosphere. Before fitting and integrating the Gaisser-Hillas profile (Eq. (3.3)) to obtain the total calorimetric energy, corrections for the light attenuation through the atmosphere, Cherenkov radiation, and multiple scattered light have to be made. Because neutrinos, muons, and other particles carry away an *invisible energy* of 10% to 20% [45], the resulting energy estimate has to be upscaled, to get an unbiased energy of the shower [40].

3.3.2. Surface Detector

For an accurate reconstruction of the shower parameters with the SD array, two additional event-selection conditions T4 and 6T5 are implemented in $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ Analysis Framework ($\overline{\mathrm{Off}}\underline{\mathrm{line}}$) to ensure high data quality. The first one checks the time compatibility with the estimated shower front and the latter one imposes that the station with the largest signal is surrounded by six working stations.

Geometric Reconstruction

The arrival direction of the approximated shower front is obtained by fitting a speed-of-light inflating sphere with origin $\vec{x}_0$ to the start times of the signals t_i of individual SD stations. Their positions on the ground are denoted by $\vec{x}_i$, i.e.

$$c(t_i - t_0) = |\vec{x}_i - \vec{x}_0|. \tag{3.4}$$

From this four-parameter fit, the shower age t_0 and its virtual origin $\vec{x}_0$ in the atmosphere are inferred. After obtaining the virtual shower origin $\vec{x}_0$ and the impact point on the ground $\vec{x}_c$, which follows from the Lateral Distribution Function (LDF) fit and is initialized as the signal-weighted barycenter $\vec{b}$, the normalized shower axis $\hat{a}$ is calculated by

$$\hat{a} = \frac{\vec{x}_0 - \vec{x}_c}{|\vec{x}_0 - \vec{x}_c|}. \tag{3.5}$$

If only three stations trigger, the shower-front is approximated to be planar, and the normalized shower axis $\hat{a}_{pf}$ is derived by fitting the start times t_i to an arbitrary station $(\vec{x}_j, t_j)$ using the plain-front model

$$c(t_i - t_j) = (\vec{x}_i - \vec{x}_j) \cdot \hat{a}_{pf}. \tag{3.6}$$

The geometrical reconstruction has an angular resolution of 0.4° to 1.14° (68% to 96% Confidence Level (CL), respectively) [46].

Shower Size

To derive an SD estimator for the energy of the shower, an average reference signal is constructed. It is assumed that the signal strength is rotationally symmetric in the Shower Plane Distance (SPD), perpendicular to the shower axis $\hat{a}$. As an empirical approximation of the lateral distribution function LDF, a modified Nishimura-Kamata-Greisen (NKG)-function [47] (Eq. (3.7)) is used. The fit is done by maximizing the negative log-likelihood (Eq. (3.8)) of the shower impact point on the ground $\vec{x}_c$, and the shower size $S(r_{opt})$, given the station signals S_i and their locations $\vec{x}_i$,

$$S(r) = S(r_{opt})\, f_{LDF}(r) = S(r_{opt}) \left(\frac{r}{r_{opt}}\right)^{\beta} \left(\frac{r + r_1}{r_{opt} + r_1}\right)^{\beta+\gamma} \tag{3.7}$$

$$-\ln \mathcal{L} = -\sum_i \ln p(S(r_{opt}), \vec{x}_c | S_i, \vec{x}_i). \tag{3.8}$$

For the SD array, with 1500 m spacing, the optimal distance [48] is $r_{opt} = 1000$ m, chosen to minimize fluctuations from the array geometry, and $r_1 = 700$ m [46]. Due to an often insufficient station multiplicity per event, it is impossible to determine the shower size $S(r_{opt}) \coloneqq S_{1000}$ and the shape (β, γ) of the LDF simultaneously. As a workaround, the parameters β and γ are fit on a sub set of data with high multiplicity and parametrized in terms of θ and S_{1000} [46].

The uncertainties of the derived S_{1000} estimator are driven by statistical uncertainties originating from the finite number of particles sampled per detector and shower-to-shower fluctuations, as well as by systematic uncertainties introduced by the empirical

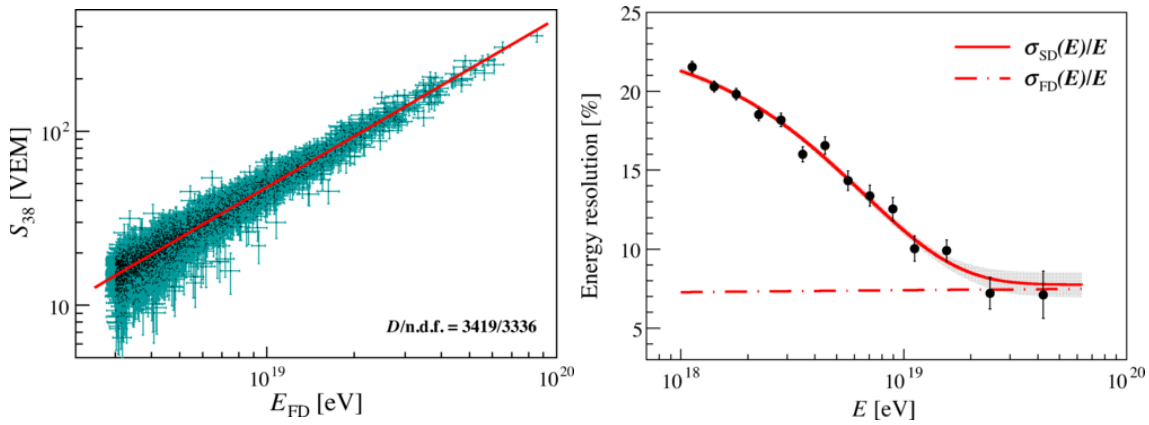


Figure 3.4.: Cross-calibration of the S_{38} SD shower-size estimator using the reconstructed FD energy E_{FD} from 3338 high quality hybrid events (*left*). The red line depicts the fitted power-law $E_{FD} = A\,S_{38}^{B}$. The SD energy resolution (*right*) converges towards the energy resolution of the FD (dashed red line) around $2{\times}10^{19}$ eV. Taken from [52].

parametrization of the LDF. To further reduce the systematic biases, one has to correct for variations in the atmosphere [49] and the influence of the geomagnetic field [50]. On top of that, more inclined showers will be attenuated by an increasing slant depth $X(\theta) \approx X(0)/\cos\theta$ at first order, which has to be accounted for as performed using the *Constant Intensity Cut (CIC)* method.

Constant Intensity Cut

The CIC is a method to correct the zenith dependency of the S_{1000} estimator, which is caused by attenuation effects through the atmosphere. Under the assumption of an isotropic flux, the S_{1000} estimator is transformed to a reference signal $S_{38} = S_{1000}/f_{CIC}(\theta)$ that would have been produced by the same shower with an inclination of $\theta_{ref} = 38°$ [51]. This reference originates from the median of the zenith angle distribution of reconstructed Auger events [46]. The parametrization of the attenuation curve is given by a polynomial

$$f_{CIC}(x) = 1 + ax + bx^2 + cx^3, \tag{3.9}$$

with $x = \cos^2\theta - \cos^2 38° = \sin^2 38° - \sin^2\theta$, reflecting the assumption of an isotropic flux impinging on a flat detector $\Phi \propto \sin^2\theta$.
The relation between S_{38} and the energy of the primary particle can be inferred through the measured FD energy in high-quality hybrid events [52] to calibrate the SD energy estimator. It is well described by a single power-law

$$E_{FD} = A\left(\frac{S_{38}}{\text{VEM}}\right)^{B}. \tag{3.10}$$

The hybrid calibration is depicted on the *left* in Fig. 3.4, together with the SD energy resolution on the *right*, which converges from roughly 20% at EeV down to the energy resolution of the FD with approximately 7% above $2{\times}10^{19}$ eV [52].

3.4. Ongoing Upgrade: AugerPrime

As discussed in the preluding chapters and sections (Sec. 2.3, Sec. 2.4), the separation from light and heavy particles is crucial to unravel the persisting puzzles in cosmic ray physics and beyond. Currently, most of the composition studies depend on the measurements of the first and second moment of X_{max} through the FD with a limited duty cycle of 15%. Consequently, the statistics at highest energies are not sufficient for a detailed composition study. The ongoing *AugerPrime* update approaches this problem by adding new detector types. In essence, the primary mass can be acquired by the disentanglement of the shower subcomponents at a given energy, as derived in the Heitler-Matthews superposition model (Sec. 2.5.1). A layered detector system in which each detector has a distinct response to each subcomponent of the shower is particularly suitable. In the following, a brief overview of the additional components and their objective is given.

Surface Scintillator Detector (SSD)

Extending the existing WCD by mounting a plastic scintillator on top, allows for a complementary measurement of shower particles. The measured signal is calibrated in units of Minimum-Ionizing Particle (MIP) which corresponds to one muon traversing the detector vertically. Due to both detectors having different responses to the muonic and electromagnetic sub components of the shower [38, 39], composition analysis at high statistics with only the SD are possible.

Small PMT

A fourth Small PMT (SPMT) will be added to the WCD, extending the dynamic range of the SD even further by the ability to measure non-saturated signals as close as 300 m to the shower core [38, 39].

Upgraded Unified Board (UUB)

Upgrading the electronics of the SSD stations is inevitable to facilitate the processing of both WCD and SSD signals. The data quality will increase by three-times faster sampling (40 kHz $\longrightarrow$ 120 kHz) of the ADC traces with a 12 bit FADC, leading to an increased timing accuracy and a higher dynamic range [38, 39].

Underground Muon Detector

Directly measuring the shower muon content and its time structure with the Underground Muon Detector (UMD) will provide important cross checks to muon reconstruction models based on the layered detector system of the SD. It is part of the Auger Muons and Infill for the Ground Array (AMIGA) upgrade and will also provide a measure of muons which is independent of hadronic models. Therefore, an array containing 61 muon counters, 30 m^2 each, at a spacing of 750 m, covering 24 km^2 buried under a layer of 2.3 m ($\approx$280 $\mathrm{g\,cm}^{-2}$) soil is deployed [38, 39].

Radio Detector (R&D)

Radio antennas sensitive to the electromagnetic shower component are mounted on top

of the WCD, based on the results from Auger Engineering Radio Array (AERA) [38]. They will extend the composition sensitivity of the SD to highly inclined showers, where the SSD suffers from a low effective cross section.

Increase of FDs Duty Cycle

Lengthening the operation mode of the FD into nights with high sky background will lead to an increase of about 50% of the current duty cycle. This is achieved by reducing the PMTs gain through lowering the supplied high voltage [38, 39].

3.5. Scientific Milestones and Open Questions

Auger has provided essential contributions to the understanding of the origin of CRs and particle physics at the highest energies. In the following section a few scientific milestones are highlighted and open questions that are related to this thesis are outlined.

3.5.1. Muon Puzzle

The investigation of particle physics at the highest energies through Auger has revealed that current air-shower simulations fail to accurately describe the longitudinal and lateral shower profiles [53, 54, 55, 56]. The number of muons from measured shower footprints is significantly higher than the number of muons obtained in simulations. This muon deficit in simulations ranges from 30% up to 80%, depending on the hadronic interaction model at 10^{19} eV [54].
These results are obtained through a preselection of highly-inclined hybrid events, where the electromagnetic component is mostly absorbed in the atmosphere. Hence, the signal at the ground is dominated by muons and provides a direct measurement of the muon number, which is in turn related to the mass of the primary (Eq. (2.27)). The number of muons is derived by scaling simulated reference profiles of the expected lateral muon density profile until they fit the measured data [54]. The results imply a contribution of elements heavier than iron (Fig. 3.5, *left*), which is unphysical and not compatible with studies based on X_{max} [57]. The discrepancy is attributed to hadronic interaction models failing to reproduce the observed amount of muons.

3.5.2. Spectrum

One of the original design objectives of Auger, was to measure the CR energy spectrum at the highest energies with significant statistics. After a few years of operation, it undoubtedly confirmed the suppression of the flux beyond the ankle in 2008 [58]. Even though the region coincides with the theoretical proposed GZK cutoff, we are not able to distinguish it from the scenario where nearby sources reach their acceleration limit. Improving the currently limited knowledge about the mass composition, and hence the fraction of protons, at the highest energies is crucial to disentangle those two scenarios.

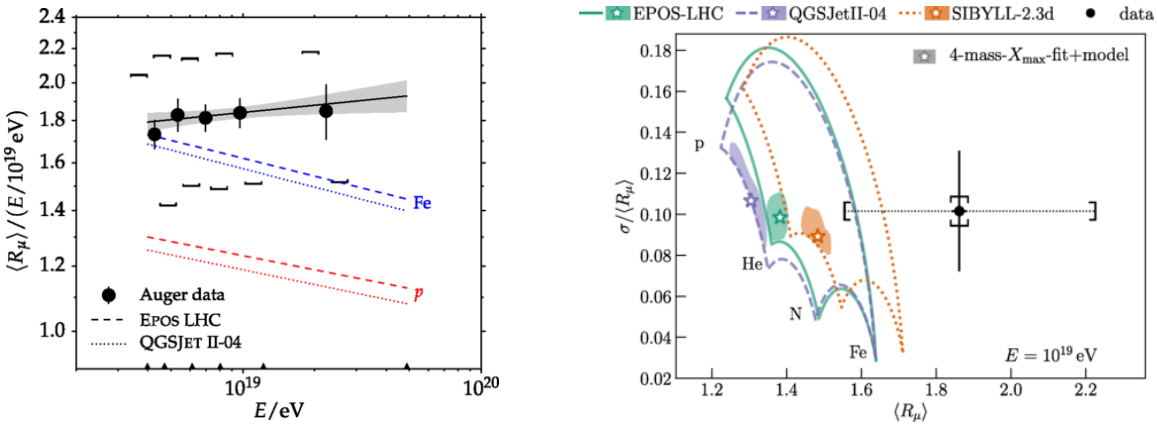


Figure 3.5.: The muon puzzle between simulation and measured data by Auger. The average relative muon content $\langle R_\mu \rangle$ per shower as a function of energy E (*left*) [54] would imply a contribution of elements heavier than iron, which contradicts studies performed with $X_{\max}$. The divergence between simulation and data is even more prevalent in the first and second moment of the relative muon content, $\langle R_\mu \rangle$ and $\sigma(R_\mu)$, respectively (*right*) [56]. The expectations of any mixture between p, He, N, and Fe (*colored contours*) for three hadronic interaction models, have no significant overlap with the measured data (*black*) within its systematic uncertainties. The star symbols indicate the preferred mixture which is derived from the $X_{\max}$ measurements. The discrepancy is attributed to hadronic interaction models failing to reproduce the observed amount of muons [54]. Taken from [54].

3.5.3. Mass composition of UHECRs

The key ingredient to solve many different open questions in CR physics lies in the determination of the primary particle on a shower-to-shower basis. Currently composition studies are only possible on a statistical level by measuring the first and second moment of the shower-depth distribution via the FD. Showers of lighter particles penetrate deeper and have larger fluctuations in shower depth than showers of heavier particles, which translates into higher $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ respectively. The measurement of these two moments, binned in energy, is depicted in Fig. 3.6 and shows a gradual change from lighter to heavier particles with increasing energy.
Another very prominent mass-sensitive variable is the muon content in air showers. In theory the number of muons N_μ in an iron shower could be up to 1.8 times higher than in a proton shower (Sec. 2.5.2), providing a promising discrimination. Currently it is only possible to estimate the muon content for very inclined showers, see Sec. 3.5.1. In the context of the AugerPrime upgrade (Sec. 3.4), it should be possible to measure the muon content on a shower-to-shower basis. In the scope of this thesis the benefit of adding an additional detector on station level shall be explored. The results of the layered two-detector system will in the future be cross-checked by the direct measurements of the muon content through the UMD.

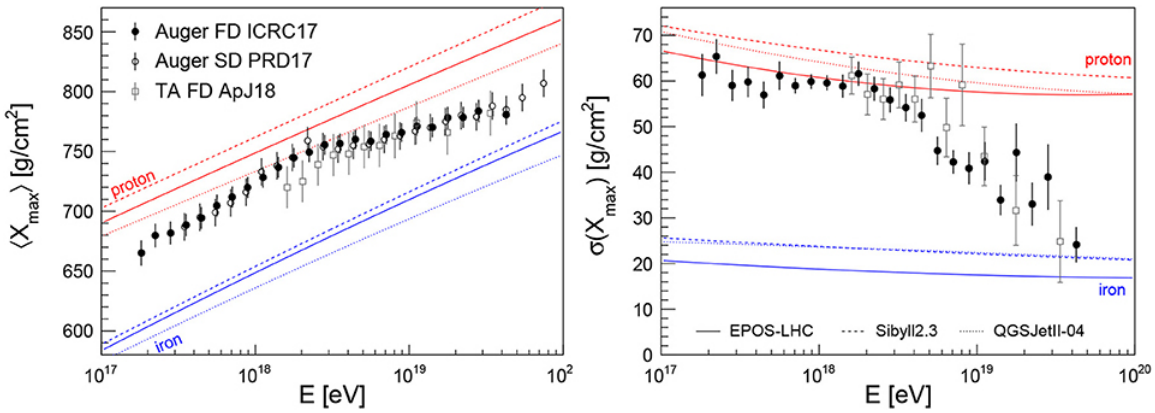


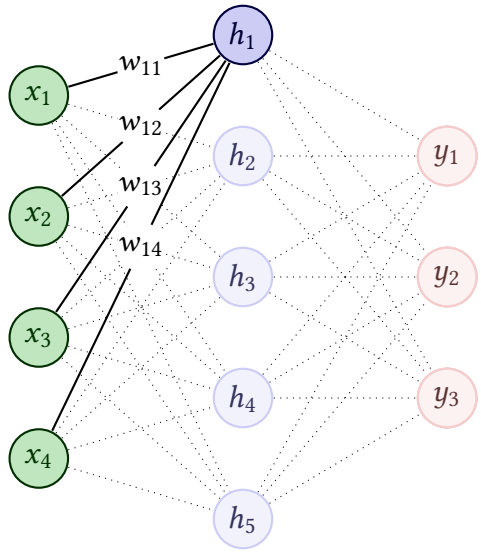
Figure 3.6.: First (*left*) and second (*right*) moment of the shower depth, $\langle X_{\mathrm{max}} \rangle$ and $\sigma(X_{\mathrm{max}})$, respectively, as a function of energy, measured by Auger and Telescope Array project (TA). Both experiments indicate a gradual change from lighter to a heavier composition above $10^{18.3}$ eV. The dashed, dotted, and solid lines are the expected values for $\langle X_{\mathrm{max}} \rangle$ and $\sigma(X_{\mathrm{max}})$ for air shower simulations using different hadronic interaction models. Taken from [16].

4. Machine Learning and Neural Networks

The field of *ML* studies the capability of systems to *learn* by extracting patterns from large amounts of data [59]. It is part of the field of general *Artificial Intelligence (AI)* which tries to reproduce the intelligence of humans by machines. The methodology of ML is based on optimizing an imposed metric with respect to a given data set, referred to as training. During the training process a ML model learns the mapping from inputs, called features, to outputs, denoted as targets. A major problem of ML is that the performance heavily relies on the choice of hand-crafted feature representations [59].
The main training paradigms, which describe the feedback type during training, are divided into supervised, unsupervised, and reinforcement learning. In the context of supervised learning, the model is given a data set, which consists of inputs and their corresponding outputs. During training, the model learns a generalized mapping between inputs and outputs to accurately perform predictions. Example applications are classification and regression tasks with recent major breakthroughs in Natural Language Processing (NLP) through ChatGPT [60]. On the contrary, unsupervised learning does not provide any target information and the model is taught to learn the underlying structure inherent to the data. Use cases are anomaly detection and clustering. Reinforcement learning is a special case of supervised learning where the model takes the role of an agent, taught to maximize a reward function by interacting with the environment. Most prominent applications exist in robotics and in games like Go (AlphaGo [61]).

Deep Learning

Deep Learning addresses the aforementioned burden existing in ML by loosening the imposed fixed feature representations to be instead learnable by the model itself, usually denoted as *Representation Learning* [59]. This is done by introducing simple trainable representations that hierarchically merge together to form more complex high level representations. In the context of image detection this would correspond to the representation learning of corners and contours that merge together to form complex features closer related to the object itself [59], like nose, eyes, and lips to identify a face.
The recent success of Deep Neural Network (DNN) models is mainly attributed to the enormous growth in computational power, the advance of algorithms, and the availability of big amounts of data. The former enables the training of much larger models which are capable of complexer representations and the latter is especially important as the models need to be trained on a huge amount of data to generalize well.



$$h_1 = \sigma\left(w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4 + b_1\right) = \sigma\left(\sum_{i=1}^{n=4} w_{1i}x_i + b_1\right)$$

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{pmatrix} = \sigma\left[\begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}\right]$$

$$\mathbf{h} = \sigma\left(\mathbf{W}\mathbf{x} + \mathbf{b}\right)$$

Figure 4.1.: Schematic representation of a simple feed-forward neural network with one hidden layer. Example calculation of the activation of the first hidden node is shown on the *top right* as well as the activation of the complete hidden layer on the *bottom right*. The activation function is denoted as σ and not explicitly specified.

4.1. Artificial Neural Networks

Artificial Neural Networks (ANNs) have emerged as a powerful technique for solving complex problems in a wide range of fields. They are inspired by the biological neural network of the human brain. The human brain consists of billions of neurons which are connected to each other by synapses. In analogy the nodes in an artificial network represent the neurons and the edges the synapses connecting the neurons. The activation resemblance of a node is computed by applying an activation function on the weighted sum of the directly connected nodes. Increasing the complexity by stacking multiple layers of nodes together and connecting them via weights w_{ij} and biases b_i yields an artificial neural network as depicted in Fig. 4.1. In the context of this thesis only feedforward neural networks are considered. Thus the terms neural network and feedforward neural network are used interchangeably except when explicitly stated otherwise. The information flow is directional and starts from the input $\mathbf{x}$, migrates through the chain of hidden layers until it reaches the output layer where it tries to match the label $y = f(\mathbf{x})$. The length of the chain corresponds to the depth of the model. The terminology of DNNs stems from a long and hence deep chain of layers.

Supervised training

In a process called training, the parameters of the model are adjusted with respect to a performance measure. The performance measure is given by a *loss* function $\mathcal{L}$, which qualitatively compares the predicted output $\hat{y}$ to the ground truth y. Common choices of loss functions for regression tasks are the Mean Squared Error (MSE) or the categorical cross-entropy for multi-class classifications. The loss function is then optimized by ad-

justing the trainable parameters θ of the network. In essence this is a minimization or maximization problem with the loss as an objective function

$$\hat{\theta} = \arg\min_{\theta} \left[\mathcal{L}(y|\theta, x)\right] \quad \text{or} \quad \hat{\theta} = \arg\max_{\theta} \left[\mathcal{L}(y|\theta, x)\right]. \tag{4.1}$$

In practice the optimization is accomplished by the *backpropagation* algorithm which computes the gradient of the loss function with respect to the parameters. The gradient is then used to update the weights and biases in the direction of the negative gradient. In the most simplest form a gradient descent update reads as

$$\theta_{i+1} = \theta_i - \eta \frac{\partial \mathcal{L}}{\partial \theta_i}, \tag{4.2}$$

with the learning rate η and the iteration index i. There exist several more sophisticated variants of stochastic optimization algorithms like *Stochastic Gradient Descent (SGD)* [62] or *Adaptive Moment Estimation (Adam)* (adaptive moment estimation) [63] which implement a decay of the learning rate, are able to avoid local minima through momentum, and thus have an overall better convergence behavior.
To ensure that the model is really capturing the underlying representations of the data and not simply memorizing it, the available data is split into a *training*, *validation*, and *test* dataset. The training data is used to train the model while simultaneously validate, tune, and design the model hyperparameters[1] on the validation data. The test data is used in a final step after the model finished training to evaluate the model performance on unseen data. In this way a measure of the generalization capability can be obtained.

4.2. Convolutional Neural Networks

In the context of computer vision, deep neural networks rapidly explode in parameter size. For instance, a one-megapixel gray-scale image has an input dimensionality of 10^6. If the first layer of a network contains 1000 nodes, the first layer already contributes $10^6 \times 10^3 + 10^3 \approx 10^9$ parameters. This makes deep neural networks infeasable to train and prone to overfitting [64].
By exploiting symmetries inherent to the natural data representation, one can impose a model inductive bias [65], which ultimately leads to a reduction of parameters and a decrease in computational complexity. *Convolutional Neural Networks (CNNs)* [66] are a special class of neural network architectures especially designed to integrate translational invariance and locality in the spatial structure of images [66]. The convolutional layer as the core of the CNN is responsible to extract feature maps from adjacent nodes in the input. It utilizes a trainable kernel window which is slid over the input, computing an element-wise sum to reduce the input dimensionality (Fig. 4.2, *left*). They are usually followed by pooling layers to downsample the input dimensionality even further (Fig. 4.2, *right*). The output part of the network consists of fully connected layers that are used to make the

[1] Hyperparameters are parameters that are not optimized during training. They specify the model architecture (number of nodes, number of layers, used activation functions etc.) or parameters related to the training process itself (learning rate, batch size, epochs etc.)

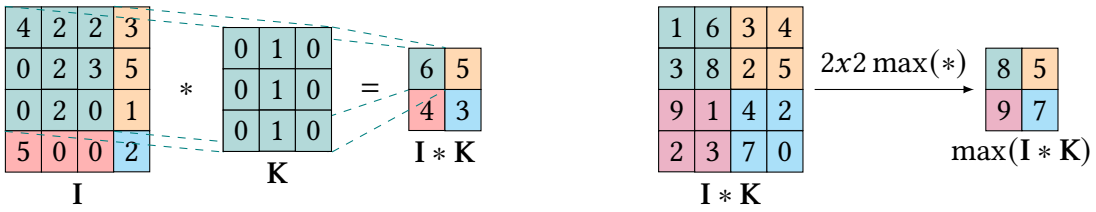


Figure 4.2.: Basic kernel operation of a convolutional layer (*left*). In this case the kernel $\mathbf{K}$ is a 3×3 matrix filtering vertical edges with a stride of 1 and a padding of 0. The kernel is moved step-wise (stride = 1) over the input image $\mathbf{I}$. At each point, the sum over all elements of the element-wise product is computed, producing a feature map $\mathbf{I} * \mathbf{K}$. The MaxPooling operation (*right*) is used to reduce the spatial dimensionality of the feature map. It is a non-linear operation that computes the maximum value of a kernel window of size (2×2) over the input. The kernel window is moved over the input with a stride of 2 and a padding of 0. The output is a spatially-reduced feature map.

prediction, i.e. to classify the image. Through successive stacking of convolutional and pooling layers the effective area connected to a single node grows. This effective area is called the *receptive field* in analogy to the discoveries in neurophysiology [67, 68, 69]. It enables the network to learn more high-level features and thus, to generalize better.

4.3. Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are powerful models for processing sequential data and are especially designed to capture temporal dependencies. They are widely used in the fields of NLP and speech recognition [70]. The core idea of RNNs is to process sequential data by maintaining an internal state vector $\mathbf{h}^{(t)}$ that is updated at each time step t. The update is accomplished by a non-linear function f that takes the current input $\mathbf{x}^{(t)}$ and the previous state vector $\mathbf{h}^{(t-1)}$ as input, to produce the new state vector $\mathbf{h}^{(t)}$,

$$\mathbf{h}^{(t)} = f(\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}). \tag{4.3}$$

This recursive relation led to the name of RNNs. As the calculation of the state at any given time step t depends on all previous applications of the recursive relation, the training is computationally expensive. For long sequences, traditional RNNs suffer from catching long-term dependencies and vanishing gradients, as the backpropagation is done over multiple time steps.
A more state-of-the-art architecture that overcomes this burden is given by a Long Short-Term Memory (LSTM) network [71]. It incorporates the memory cell (Fig. 4.3) as its basic building block. The key idea is that it can selectively forget or restore information over time. This basic components can be divided into the *input gate*, the *forget gate*, and the *output gate*, which all have the ability to add or remove information to and from the cell state $\mathbf{c}^{(t)}$. The amount of information change is regularized by the sigmoid functions σ that output a scalar between 0 and 1. In case of the forget gate, the sigmoid takes the last hidden state $\mathbf{h}^{(t-1)}$ and the current input $\mathbf{x}^{(t)}$ to modify the previous cell state $\mathbf{c}^{(t-1)}$

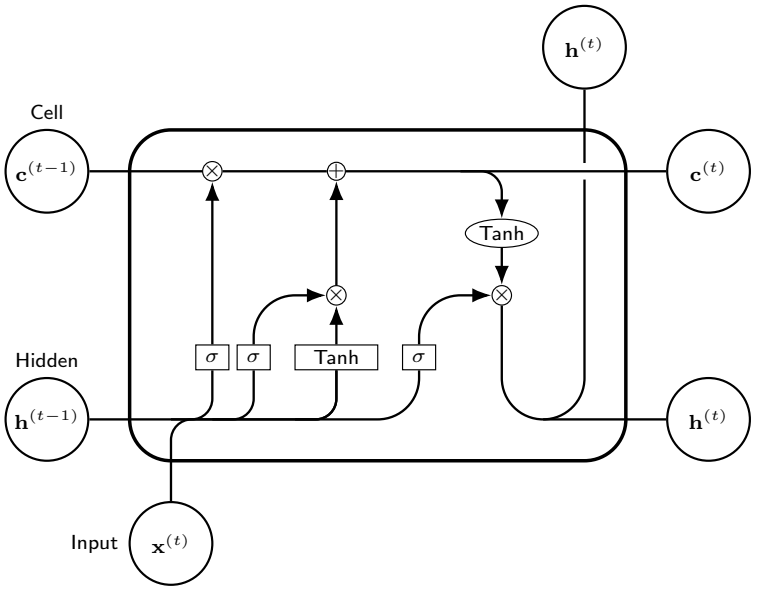


Figure 4.3.: Illustration of a LSTM cell. The cell state $\mathbf{c}^{(t)}$ is updated by the forget gate $\mathbf{f}^{(t)}$, the input gate $\mathbf{i}^{(t)}$, and the output gate $\mathbf{o}^{(t)}$. The forget gate controls the amount of information that is removed from the cell state. The input gate controls the amount of information that is added to the cell state. The output gate controls the amount of information that is passed to the output. The cell state is updated by the element-wise multiplication of the forget gate and the cell state and the element-wise multiplication of the input gate and the input. The output is computed by the element-wise multiplication of the output gate and the cell state.

by multiplication of the sigmoid scalar. A value of 0 completely removes the previous information while a value of 1 keeps the internal state. In a next step, the amount of newly added information to the cell state is controlled via the input gate. Its main purpose is to extract features and modify the previous cell state accordingly. In a final step, the updated previous cell state is combined with the last hidden state in the output gate to form the next hidden state $\mathbf{h}^{(t)}$ of the cell.

4.4. Mixture Density Networks

In the context of classical regression and least-squares optimization, the approximated function gives the most probable mean function conditioned on the input data [72]. Thus it contains no information about any uncertainty in general. The uncertainty can be divided into *aleatoric* and *epistemic* uncertainty. The former is known as stochastic uncertainty and describes randomness inherent to the data itself. The latter is commonly known as the systematic uncertainty and describes uncertainty attributed to imprecise models. One approach to qualify the uncertainty of the neural network prediction is through *mixture density networks* initially proposed by Bishop [72]. They are a special class of neural

networks that are able to model the *probability density* of the target, via conditioning on the input data. By building a normalized linear combination of multiple Probability Density Functions (PDFs) p_k, it is in principle possible to approximate arbitrary conditional probability densities in the same manner as neural networks represent arbitrary functions [72]. For properly normalized PDFs, the mixture coefficients α_k have to sum to one. The network is trained to predict the characteristic parameters $\mathbf{\Theta}_k$ of the mixture PDFs together with the mixture coefficients. The resulting joint probability density function characterized by $\mathbf{\Theta}_k, \alpha_k$ is used to calculate the negative log-likelihood of the targets $\mathbf{y_i}$ with respect to the inputs $\mathbf{x}_i$ as

$$-\ln \mathcal{L} = -\sum_{i=1}^{N} \ln \left(\sum_{k}^{m} \alpha_k(\mathbf{x}) p_k(\mathbf{y_i}|\mathbf{x}_i, \mathbf{\Theta}_k) \right) \qquad \text{with} \quad \sum_{k=1}^{m} \alpha_k = 1. \tag{4.4}$$

In practice the normalization of the α_k is achieved through a softmax activation function. In case of a Gaussian mixture model $\mathbf{\Theta}_k = \{\vec{\mu}, \Sigma\}$ it is recommended to ensure positivity of the standard deviation by applying a corresponding activation function. The *softplus* function, $f(x) = \ln(1 + \exp x)$ is a smooth continuous version of a Rectified Linear Unit (ReLU) activation function and commonly used.

4.5. Explainable Artificial Intelligence

The increasing complexity in ML models and their rising applications in fields like medicine, healthcare, and autonomous systems has led to the need to understand and interpret the decisions made by these models. Explainable Artificial Intelligence (XAI) is an emerging field of research that developes methods and techniques to explain how ML models derive their predictions to ultimately increase trust, reliability, and causality for humans [73]. In general, there exists a tradeoff between the interpretability of a model and its performance [74]. Well performing models like ANNs are usually so called *black-box* models because they are systems with hidden internal mechanisms, which can not be deduced by simply looking at their parameters. More simpler models like linear regression, on the other hand, are *white-box* models because their internal mechanisms can be easily understood by looking at their parameters.
The field of XAI can be divided into *model-specific* and *model-agnostic* methods. Model-specific methods are only applicable to a specific class of models like decision trees or linear regression. Model-agnostic methods, on the other hand, are applicable to any supervised ML model after they have been trained [75]. They usually analyze how changes in the input get reflected in the model output. Depending on the specific use case the model-agnostic methods can be divided further into *global* and *local* methods. Global methods describe the average behavior of the model while local methods explain individual predictions. In the context of this thesis the focus lies on the global method to derive measurements of feature importances with respect to a regression target. In the following section some elementary XAI methods are introduced.

4.5.1. Model-Specific Methods

The easiest way to achieve model interpretability is to use a model that is inherently interpretable as the aforementioned linear regression or decision trees.

Mean Decrease in Impurity (MDI)

In case of a non-linear relationship between input features and target variable, linear regression fails to capture the underlying relation. In this case decision trees are a good alternative. They are a non-parametric model that can be used for both classification and regression tasks and are easy to interpret as they can be visualized in a tree structure. The prediction of a decision tree is based on a set of rules that are build from the training data. Through recursive splits of the data into subsets, depending on the input features, the tree is built. The splitting is generally done in a greedy manner by maximizing the *information gain or impurity (I)* (Gini Impurity, Entropy, Variance or MSE) at each split [76]. The process is repeated until a stopping criterion is met. The prediction of a decision tree is the average of the target values of the leaf node the input sample falls into.
By averaging multiple noisy, but unbiased estimators, a reduction of the variance is accomplished. This is the key principle of bagging [76]. Decision trees are optimal candidates that fullfil this requirements and are usually collectively used in ensemble methods like *Random Forest* [77].
The importance measure I of a specific feature x_j can be derived by adding up the weighted impurity decrease ΔI at each node t where the feature is used inside a tree T, and average it over all N_T trees in the forest [78],

$$I(x_j) = \frac{1}{N_T} \sum_{T} \sum_{t \in T : v(s_t) = x_j} \frac{N_t}{N} \Delta I(s_t, t) \tag{4.5}$$

$$\text{with:} \quad \Delta I(s_t, t) = I(t) - \sum_{k=1} \frac{N_k}{N_t} I(k) \tag{4.6}$$

Here N_t/N denotes the fraction of samples at node t and $v(s_t)$ the variable used in the splitting condition s_t. The impurity decrease ΔI is the difference between the impurity of the parent node t and the weighted impurity of the children nodes k. Due to multiple possible impurity measures I, this method is generally referred to as *Mean Decrease in Impurity (MDI)* [78]. One drawback of this method is that it does not reflect the importance of a feature for a model to generalize well. This comes from the fact that the impurity measure is computed on the training set statistics. Due to the regression scope of this thesis, the variance or MSE is used as impurity measure.

4.5.2. Model-Agnostic Methods

The separation of model interpretability from the underlying model is a key concept of model-agnostic methods. They provide great flexibility because they can be applied to any ML model, including complex black-box models like ANNs. They provide insights into the model behavior and can improve the understanding of the data.

Permutation-Feature Importance

Permutation-feature importance (PFI) is a model-agnostic inspection technique that can be used on any fitted model. It is defined via the decrease in the model score if a single feature is randomly shuffled in the data set. This shuffling breaks the relationship between the feature and the target and the resulting drop of the model score indicates how much a model depends on that feature. Thus, this technique does not reflect the intrinsic predictive value of a feature itself but rather the dependency of the model on that feature. A feature is considered as unimportant if shuffling its values leaves the model error unchanged, while a feature is considered important if shuffling its values increases the model error.
The plain algorithm after Fisher, Rudin, and Dominici [79], with the inputs of the trained model $f(x) \rightarrow \hat{y}$, the dataset X, the target y, and an error measure $L(y, \hat{y})$ roughly reads as

1. Estimate original model error $e_{\text{orig}} = L(y, \hat{y}(X))$
2. For each feature $j \in \{1, \ldots, p\}$ do:
   a) permute feature j in the data X
   b) estimate error $e_{\text{perm}} = L(y, \hat{y}(X_{j=\text{perm}}))$
   c) calculate permutation feature importance as difference $I_j = e_{\text{perm}} - e_{\text{orig}}$ or quotient $I_j = e_{\text{perm}}/e_{\text{orig}}$
3. Sort features by descending I

The advantages of this method include that it can be computed on the test data and therefore provides better insights into feature importance for generalization of the model. Additionally, the comparison of permutation-feature importances between train and test set can reveal overfitting of a model. One drawback is the breakdown for highly correlated features, which can result in a decrease of the importance of both features. This can be overcome by clustering correlated features and only keeping one feature per cluster.

SHAP (SHapley Additive exPlanation) Values

SHAP values are proposed as a unified measure of feature importance and are a conditional expectation function of the original model [74]. They originate from cooperative game theory and aim to explain the contribution of each feature to the final prediction.
The classical Shapley value is given by the average marginal contribution of a feature values across all possible coalitions,

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S)], \tag{4.7}$$

where F denotes the set of all features, S denotes the set of features that are included in the coalition, x_S represents the input features values, f_S is the model trained with the features included and $f_{S \cup \{j\}}$ the model with the feature withheld. Due to the exponential growth of computation time with the number of features, this method is not feasible for multiple input features. A few model-agnostic approximation methods like *Kernel SHAP*

or *Shapley Sampling Values* exist [74]. The *Shapley Sampling Values*, approximate the effect of removing a variable by integrating over samples from the training dataset and to additionally apply sampling techniques to Eq. (4.7) [74].
The global importance of a feature j over all samples N can be derived by the mean of the absolute Shapley values ϕ_j,

$$I_j = \langle |\phi_j| \rangle = \frac{1}{N} \sum_{i=1}^{N} |\phi_j^{(i)}|. \tag{4.8}$$

One of the big differences to other feature importance measures is that SHAP is based on the magnitude of feature attributions [75].

5. Reconstruction of the WCD Muon Signal Fraction on Station Level

In this chapter, a brief overview of current reconstruction methods for the muon content in Auger are given in Sec. 5.1. Afterwards, the muon signal fraction is introduced and motivated as a new composition-dependent quantity in Sec. 5.2, which is central to this work. The huge simulation data sets, which are used for the training and evaluation of the machine learning models, are described in Sec. 5.3, together with various preprocessing steps and normalization techniques. In addition, the reference performance metrics are defined in Sec. 5.4. Insights into important features for the prediction of the muon content are provided by Sec. 5.5, followed by introducing the baseline and NN models in Secs. 5.6 to 5.7 and their training setup. The CNN architecture is then extensively studied, by first validating it against other architectures and then investigating possible improvements through different architecture configurations. Throughout the analyses, the importance of the SSD is investigated in Sec. 5.8.3 and the best performing models are chosen to evaluate their composition dependency in Sec. 5.8.4. Finally, a crosscheck of the CNN performance on another hadronic model is done in Sec. 5.9. The chapter completes with a direct estimation of the primary particle on a single station level in Sec. 5.10.

5.1. Current Methods for Muon-Content Estimation

The estimation of the muon content on an event-by-event level is one of the prime objectives of high-level analyses in Auger. The muon content is an important quantity because its close relation to the mass of the primary. It will advance our current understanding of CRs and air shower physics in many aspects. For instance, the fine-tuning of hadronic interaction models at highest energies will be improved as well as the study of composition-enhanced arrival directions that might pin-point to CR sources.

5.1.1. Matrix Inversion Method

With the ongoing AugerPrime upgrade (Sec. 3.4), two complementary detectors in the SD allow for the separation between the muon and the electromagnetic component. The local flux of muons $\mathcal{F}_\mu$ and electromagnetic particles $\mathcal{F}_{\mathrm{em}}$ can be related to the total signals S^{WCD} and S^{SSD} in each detector of a station. In a first-order approximation, the relation between the fluxes and the total signals is linear and can be expressed using a matrix $\mathbf{M}$ [80, 81, 82, 83, 84],

$$\begin{pmatrix} S^{\mathrm{WCD}} \\ S^{\mathrm{SSD}} \end{pmatrix} = \begin{pmatrix} S_\mu^{\mathrm{WCD}} + S_{\mathrm{em}}^{\mathrm{WCD}} \\ S_\mu^{\mathrm{SSD}} + S_{\mathrm{em}}^{\mathrm{SSD}} \end{pmatrix} = \mathbf{M} \begin{pmatrix} \mathcal{F}_\mu \\ \mathcal{F}_{\mathrm{em}} \end{pmatrix}, \qquad \text{where} \qquad \mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{5.1}$$

By inverting **M**, the fluxes can be calculated from the total signals. Taking into account the relation between the fluxes and the corresponding sub-component signals, the relations can be reduced to quantities that are accessible in MC data. Using the coefficients of the matrix we arrive at

$$\begin{pmatrix} S_\mu^{\mathrm{WCD}} \\ S_\mu^{\mathrm{SSD}} \end{pmatrix} = \frac{1}{ad-bc} \begin{pmatrix} ad & -ab \\ cd & -bc \end{pmatrix} \begin{pmatrix} S^{\mathrm{WCD}} \\ S^{\mathrm{SSD}} \end{pmatrix}. \tag{5.2}$$

The coefficients show only a weak dependence on the hadronic interaction model, the primary mass, and the shower geometry [38]. The relative resolution of the muon signal is between 20% to 30%. Utilizing the lateral distribution function and $S_\mu(800)$ of the muon signal, merit factors for proton and iron induced showers of up to $\mathrm{MF}(S_\mu(800)_\mathrm{p}, S_\mu(800)_\mathrm{Fe}) = 1.5$ can be obtained [38].

5.1.2. Air-Shower Universality

Some characteristic properties of EAS exhibit universal behavior, irrespective of the primary particle or its energy. For instance, the electron-positron distributions only depend on the depth of the shower maximum and the number of particles at the corresponding depth [85]. This is called *air-shower universality*.
It can be exploited by using a multivariate analysis of EAS data, which parametrizes the main sub-components of a shower by a few characteristic shower variables [39]. The leading components inside an EAS in terms of particle numbers are

- the electromagnetic component: $e\gamma$,
- the muon component: μ,
- the hadronic component and successive decay products: $e\gamma(\pi)$,
- the electromagnetic component from muon decays: $e\gamma(\mu)$.

For each component a model relates the time-dependent expected signal, induced by a specific sub-component inside a detector, with the global shower parameters. Commonly used global shower parameters are the energy of the shower E, the relative muon content $R_\mu = N_\mu / \langle N_\mu^\mathrm{p} \rangle$ describing the number of muons N_μ relative to an average proton shower, the depth of shower maximum X_max, the zenith angle θ, the relative plane-front time $\Delta t_\mathrm{sp} = t_i - t_\mathrm{sp}$ and the shower-plane distance r. The Likelihood $\mathcal{L}_\mathrm{uni}$ is thus given by

$$\mathcal{L}_\mathrm{uni} = \mathcal{L}(E, R_\mu, X_\mathrm{max}, \theta, \Delta t_\mathrm{sp}, r). \tag{5.3}$$

The main object of universality is to simultaneously reconstruct R_μ and X_max to obtain the primary mass A. Both quantities are related to $\ln A$ via

$$\ln A \overset{\text{Eq. (2.27)}}{=} \ln(56) \frac{\ln R_\mu - \ln\langle R_\mu^\mathrm{p} \rangle}{\ln\langle R_\mu^\mathrm{Fe} \rangle - \ln\langle R_\mu^\mathrm{p} \rangle}, \quad \text{and} \tag{5.4}$$

$$\ln A \overset{\text{Eq. (2.28)}}{=} \ln(56) \frac{X_\mathrm{max} - \langle X_\mathrm{max}^\mathrm{p} \rangle}{\langle X_\mathrm{max}^\mathrm{Fe} \rangle - \langle X_\mathrm{max}^\mathrm{p} \rangle}. \tag{5.5}$$

5.2. Muon Signal Fraction

In this work, the muon signal fraction $f_\mu = S_\mu^{\mathrm{WCD}}/S^{\mathrm{WCD}}$ is proposed as a new and better suited composition estimator than the muon signal S_μ^{WCD} itself. The historical common reasoning for the reconstruction of the mass of the primary through the SD is the following. Since the mass is proportional to the number of muons in a shower (Sec. 2.5.2), it would be straight-forward to derive them by estimating the muon signal by disentangling the sub-components using different detectors (Sec. 5.1.1). Even if the MC value of the muon signal is used, this is not easily accomplished. The issue at hand has its roots in the *energy-migration effect*, which is caused by the uncertainty and the systematic bias of the energy reconstruction of the SD.

5.2.1. Energy-Migration Effect

The energy estimator of the SD (S_{1000}), is derived by fitting the LDF to the signals S^{WCD} of a radial shower footprint. It is systematically driven by muon signals. As the electromagnetic component exhibits universality for all primaries and is far more attenuated than the muon component, showers with a higher number of muons will deposit a larger signal in the detectors (Fig. A.1). Showers initiated by heavier primaries produce more muons and thus, get assigned to higher reconstructed energies. This results in a migration effect of lower-energetic heavier particles towards higher reconstructed energies. In case of measured data, this *energy-migration effect* is increasingly asymmetric, as the CR spectrum is steeply falling and much more heavier particles migrate towards higher energies, than lighter towards lower energies. The migrated heavier particles produce a lower total signal and muon signal than they would produce at the same MC energy. This results in protons of a given reconstructed-energy range being indistinguishable from lower energetic irons by using the muon signal to identify the different primaries. On the contrary, the relative muon signal fraction is almost independent of this *energy-migration effect* because it also takes into account the reduced total signal, reducing the energy dependence induced by the SD reconstruction dramatically. On a statistical level, the muon fraction allows for the separation between showers induced by protons and irons (Fig. 5.1). Note that Fig. 5.1 represents uniformly distributed MC data. The muon signal LDFs of proton and iron are indistinguishable, while the muon fraction stays separated when binned in reconstructed energy instead of MC energy. The relative energy bias of the SD amounts to -18% and is slightly shifted towards iron (Fig. 5.1, *bottom right*).

5.2.2. Calculation of the Muon Signal Fraction

In simulations, $\overline{\mathrm{Off}}\underline{\mathrm{line}}$ provides signal traces originating from different sub-components $c \in \{\mathrm{e}, \gamma, \mu\}$ for each PMT of the SD stations. By averaging the three PMT traces in the WCD for each component, the signal trace of each component $s_c[i]$ is computed by

$$s_c[i] = \frac{1}{N_j} \sum_{j=1}^{3} s_c[i, j], \tag{5.6}$$

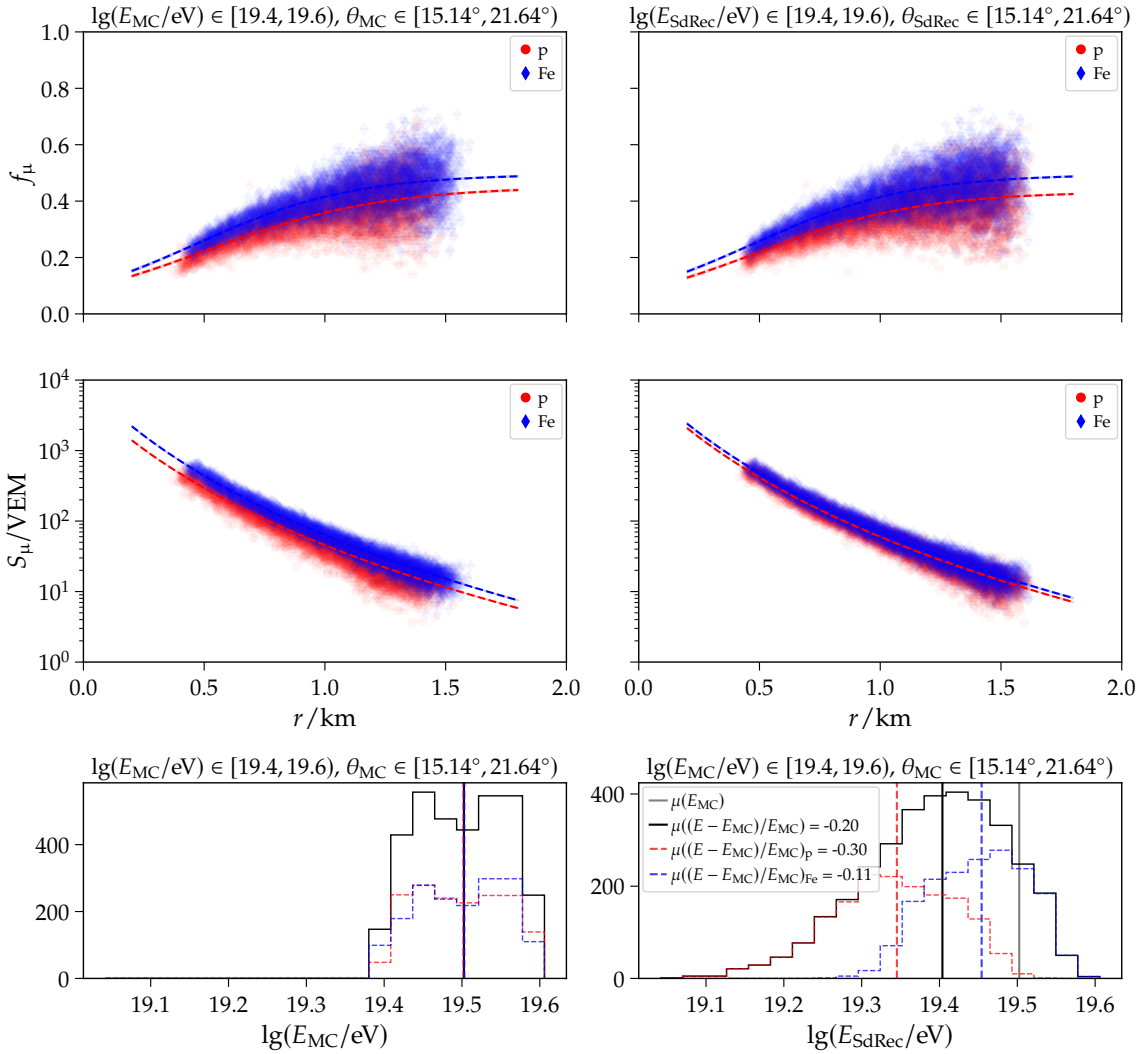


Figure 5.1.: Comparison of the discrimination power of the muon signal fraction f_μ (*first row*) and the muon signal S_μ (*second row*) in stations, triggered by air showers induced by proton (*red*) and iron (*blue*). The dotted lines represent fitted LDFs and Lateral Fraction Functions (LFFs) per primary. Due to the *energy-migration effect* (Sec. 5.2.1) the discrimination power of the muon signal is lost when reconstructed quantities E_{SdRec} and θ_{SdRec} are used (*right*). This is not the case for the muon signal fraction. Since it has a very weak energy dependence, the *energy-migration effect* has a minimal impact. The *energy-migration effect* is illustrated in the *bottom right* panel, which shows the widening of the distribution of the reconstructed energy in a chosen MC energy bin. Iron showers get a higher reconstructed energy assigned than proton showers even tough they have a similar MC energy (*bottom left*).

where i is the time bin of the trace and $s_c[i,j]$ is the component trace of the PMT j. N_j denotes the number of PMT traces that exceed an integrated signal of $\sum_i s_c[i,j] \geq 10^{-3}\,\mathrm{VEM_{peak}}$. Since the individual component traces are historically simulated independently from the total signal traces in Offline, to avoid additional statistical fluctuations the relative fraction is computed with respect to the sum of all components

$$f_\mu = \frac{S_\mu}{\sum_c S_c} = \frac{\sum_{i=b_s}^{b_s+L_{tr}} s_\mu[i]}{\sum_c \sum_{i=b_s}^{b_s+L_{tr}} s_c[i]} \qquad \text{where } c \in \{e,\gamma,\mu\}. \tag{5.7}$$

The trace length $L_{tr} = 400$ is given in number of bins (1 bin = $8.\bar{3}$ns for UUB) after the signal start bin b_s, found by the trigger algorithm of Offline. In the following f_μ always denotes the relative muon signal in the WCD, except when explicitly stated otherwise. Using Eq. (5.7), the muon signal is $S_\mu = f_\mu S_{tot}$. The hadronic component is completely neglected throughout this work.

5.3. Data Sets and Preprocessing

Large data sets are needed for the training, validation, and testing of machine learning models. In UHECR physics they are provided by complex MC simulations of air showers and their corresponding detector responses. The main advantage of simulation data is the access to quantities that are otherwise not available in measurements, such as the muon fraction which suits as a target for the supervised training of machine learning models. To improve the training process, the data is preprocessed to improve the quality of the data and to reduce the computational complexity. The air-shower simulations, the data set, and the preprocessing steps are described in the Secs. 5.3.1 to 5.3.6.

5.3.1. Air-Shower Simulations

In Auger, the detector simulations are provided by two successive MC simulations. In the first step, the air-shower cascade is simulated with the COsmic Ray SImulations for KAscade (CORSIKA) simulation package, which can use different hadronic interaction models. In this work the QGSJetII-04 Hadronic Interaction Model (QGSJetII-04) is predominantly used, while the EPOS-LHC Hadronic Interaction Model (EPOS-LHC) suits as a reference. In the second step, the detector responses to the simulated shower is simulated with the Offline framework. Since the simulation of UHECR CORSIKA showers is computationally expensive, every shower is re-used ten times on the array to increase statistics. After the detector simulation, the standard reconstruction of Offline is used to extract characteristic shower properties, such as the energy of the primary particle. The detector response, as well some reconstructed station and shower properties are used for the detailed station-level study.
To generate the detector responses, the Praha/Napoli library [86] is used. The showers are continuously sampled in all shower variables, in contrast to the discrete Karlsruhe air

Table 5.1.: Number of triggered stations that passed the quality cuts for the MC study in different logarithmic energy intervals.

$\lg(E_{\mathrm{MC}}/\mathrm{eV})$			18 to 18.5	18.5 to 19	19 to 19.5	19.5 to 20	20 to 20.2
QGSJetII-04	UUB	p	34 488	64 392	126 417	221 622	130 590
		He	32 270	68 783	127 966	224 830	133 306
		O	34 465	73 680	131 367	234 063	136 093
		Fe	36 065	70 460	139 585	242 221	140 552
EPOS-LHC	UUB	p	20 588	80 740	129 972	232 594	130 602
		He	25 130	70 026	134 584	240 721	136 682
		O	27 917	72 954	139 433	246 674	142 177
		Fe	29 685	75 830	145 649	242 116	142 529

shower Library (KaLib). This helps the network to rely less on extrapolation in between the sampling points, instead of working only with discrete values. The air showers are distributed uniformly in the energy range from 10^{18} eV to $10^{20.2}$ eV and uniformly in $\sin^2\theta$ (isotropic flux), with θ ranging from 0° to 60°. For the primaries a mixed, uniform composition in proton, helium, oxygen, and iron is used. The number of stations per energy interval and per primary is depicted in Table 5.1. In total 721 450 (691 139) showers with 2 403 215 (2 466 603) stations remain for QGSJetII-04 (EPOS-LHC) after the quality cuts (Sec. 5.3.4). Even though the showers are sampled uniformly, the number of stations are non-uniform and increases with increasing energy. This is driven by the larger size of the shower footprint and better trigger efficiencies.

5.3.2. Preprocessing of the Traces

For the UUB time traces the first $L_{\mathrm{tr}} = 400$ bins after the signal start bin b_{s} are used as an input for the network. Using a predetermined trace length allows for a reduced network size and removes the need for zero padding or cutting off of the traces during training or prediction. Furthermore, the first 400 bins contain the relevant signal information. The muon signal fraction is barely altered by cutting the traces (Fig. 5.3). To further reduce the size of the input, the traces of the three PMTs of the WCD are averaged, taking into account only traces with an integrated PMT signal greater than $\geq 10^{-3}\,\mathrm{VEM}_{\mathrm{peak}}$ (Eq. (5.6)).

5.3.3. Custom Input Features

The feature representation of the input data is crucial for the performance of a machine-learning model. Because the available information on station level can be highly correlated, additional high-level features are created with the aim to disentangle those cases and increase predictive power.

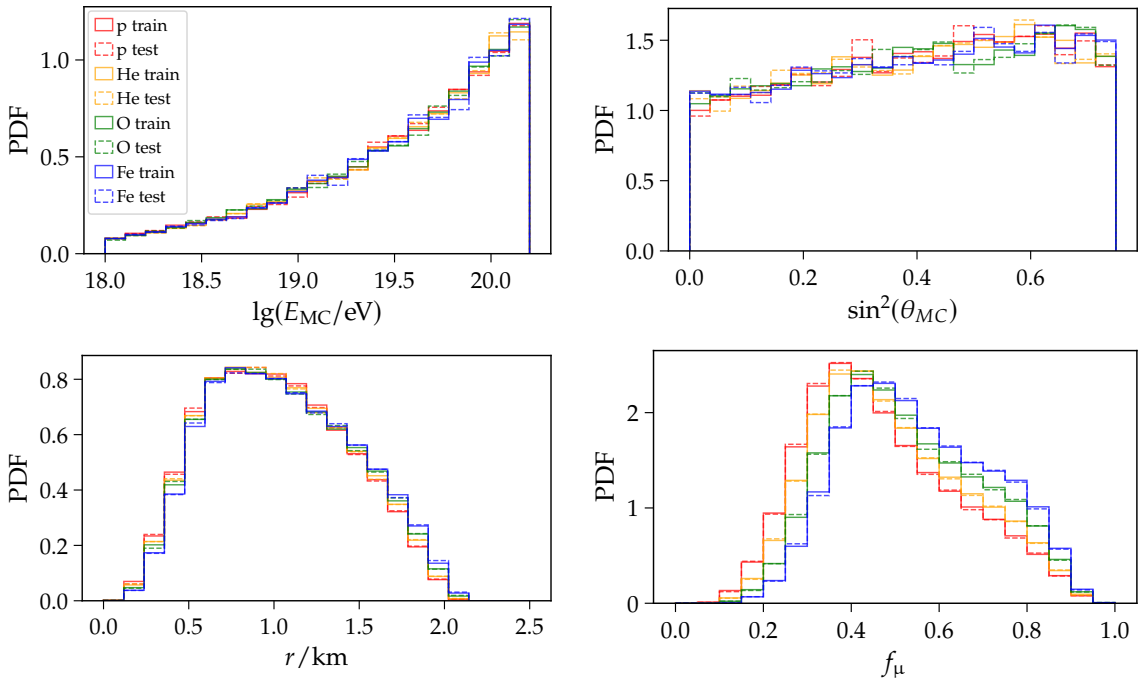


Figure 5.2.: PDFs of train and test data sets for each primary, as a function of logarithmic MC energy E_{MC} (*upper left*), zenith angle $\sin^2 \theta_{\mathrm{MC}}$ (*upper right*), shower-plane distance r (*lower left*) and muon signal fraction f_μ (*lower right*). Since more energetic primaries trigger more stations and stations with low signals are cut away predominantly in low energy showers, the distribution in log-energy is not uniform anymore on station level. This overweights high energy showers. The train:test split is chosen to be 9:1.

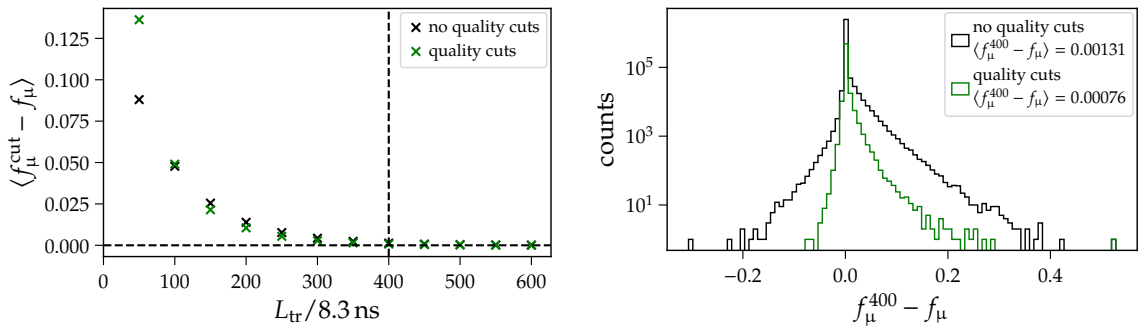


Figure 5.3.: Mean difference between the muon fraction f_μ, computed from complete traces, and the trimmed muon fraction f_μ^{cut} (Eq. (5.7)), as a function of cut length L_{tr} (*left*), and distributions of differences between the muon fraction f_μ^{400} from the $L_{\mathrm{tr}} = 400$ bin cut and f_μ (*right*). The differences are computed with (*green*) and without quality cuts (*black*). The cut length is given in number of bins after the signal start bin b_{s}. The reference muon fraction f_μ results from the standard Offline cuts, which are defined by *SdRecStation.SignalStartSlot* b_{s} and *SdRecStation.SignalEndSlot* b_{e}.

- $S^{\mathrm{WCD}} = a_{\mathrm{p}} \sum_{c\in\{\mathrm{e},\gamma,\mu\}} \sum_{i=b_{\mathrm{s}}}^{b_{\mathrm{s}}+L_{\mathrm{tr}}} s_c^{\mathrm{WCD}}[i]$,
  denotes the total signal in the WCD detector derived from the cut traces of the sub-components and is converted from $\mathrm{VEM}_{\mathrm{peak}}$ to VEM by a_{p}, see Eq. (3.1).

- $S^{\mathrm{SSD}} = a_{\mathrm{p}} \sum_{c\in\{\mathrm{e},\gamma,\mu\}} \sum_{i=b_{\mathrm{s}}}^{b_{\mathrm{s}}+L_{\mathrm{tr}}} s_c^{\mathrm{SSD}}[i]$,
  denotes the total signal in the SSD detector derived from the cut traces of the sub components and is converted from $\mathrm{MIP}_{\mathrm{peak}}$ to MIP by a_{p}, see Eq. (3.1).

- $S^{\mathrm{r}} = S^{\mathrm{SSD}}/S^{\mathrm{WCD}}$,
  denotes the total signal ratio between the SSD and the WCD detector.

- $\Delta h = r \sin\theta \cos\psi_{\mathrm{sp}}$,
  denotes the height of the station along the shower axis, relative to the impact point of the shower core.

- $\Delta t_{\mathrm{c}} = t_i - t_{\mathrm{c}}$,
  denotes the time difference between the impact time of the shower core on the ground t_{c} and the start time of the individual stations t_i.

- $\Delta t_{\mathrm{hs}} = t_i - t_{\mathrm{hs}}$,
  denotes the time difference between the start time of the hottest station t_{hs} in an event and the start time of individual stations t_i. The hottest station is defined as the station with the largest signal in the WCD for each shower.

- $\Delta t_{\mathrm{sp}} = t_i - t_{\mathrm{sp}} = t_i - \left(\frac{1}{c}(\vec{x}_i - \vec{x}_{\mathrm{c}})\hat{a} + t_{\mathrm{c}}\right)$,
  denotes the time difference between the shower-plane front t_{sp} and the station start times t_i. It encompasses information about the curvature of the shower-front, which is related to the shower depth, and thus to the mass of the primary.

5.3.4. Selected Quality Cuts

Removing noisy or irrelevant data from a data set to improve its overall quality is a vital step in machine learning. In general, it helps to reduce the computational complexity and improves the model performance. The cuts need to be carefully chosen to ensure that no valuable data is removed or any bias is introduced.

Exclusion of low-gain saturated stations

Low-gain (LG) saturated stations are excluded from the analysis, due to an unrecoverable loss of signal. Consequently, no reliable muon fraction can be derived.

Zenith cut $\theta \leq 60^{\circ}$

Very inclined showers are excluded from the analysis to stay in the regime, in which the SD has the full trigger efficiency. In future studies similar to this thesis, the extension to more inclined showers could provide important cross checks to already existing studies on very inclined events.

Distance cut $r \leq r_{\text{cut}}$

The maximum distance r_{cut} is the distance where a signal of $S(r) = S_{\text{th}} = 30\,\text{VEM}$ is expected. To determine r_{cut}, the following equation has to be solved numerically for r, with given S_{1000}, β and γ from the standard reconstruction per event

$$S(r) = S_{1000} \left(\frac{r}{r_{\text{opt}}}\right)^{\beta} \left(\frac{r + r_1}{r_{\text{opt}} + r_1}\right)^{\beta+\gamma} = S_{\text{th}}, \tag{5.8}$$

where $r_{\text{opt}} = 1000\,\text{m}$ and $r_1 = 700\,\text{m}$ for the main array. The requirement of the total expected signal $S(r)$ being above the threshold S_{th} prevents the LDF to be biased towards higher signals due to upward fluctuations close to the trigger threshold, as it would be the case if one cuts on the total signal per station. Cutting away the low-signal stations would shift the average signal for larger distances upwards and thus modify the scale of the LDF, as well the distribution in the expected signal. The distance cut further improves the overall uncertainty of the muon fraction $\sigma(f_\mu)$. Currently no sophisticated error model exists for the muon fraction. However, first insights can be derived under the assumption of a Gaussian error propagation and Poissonian statistics ($\sigma(S) \approx \sqrt{S}$) in the detector response for low signal strengths,

$$\begin{aligned}\sigma(f_\mu) &\approx \sqrt{\left(\frac{\partial f_\mu}{\partial S}\right)^2 \sigma^2(S) + \left(\frac{\partial f_\mu}{\partial S_\mu}\right)^2 \sigma^2(S_\mu) + 2\frac{\partial f_\mu}{\partial S}\frac{\partial f_\mu}{\partial S_\mu}\rho(S_\mu, S)\sigma(S)\sigma(S_\mu)} \\ &= \sqrt{\frac{S_\mu}{S^2} + \frac{S_\mu^2}{S^3} - 2\frac{S_\mu^{3/2}}{S^{5/2}}\rho(S_\mu, S)} = f_\mu \sqrt{\frac{1}{S_\mu} + \frac{1}{S} - 2\frac{\rho(S_\mu, S)}{\sqrt{S_\mu S}}}\end{aligned} \tag{5.9}$$

A dependence plot of the approximated $\sigma(f_\mu)$ on the total Signal S and muon signal S_μ is shown in Fig. 5.4 for the cut simulation data with $\rho(S_\mu, S) = 0.91$. It qualitatively shows the decrease in uncertainty of the muon fraction with increasing signal strength. Around $S \approx 30\,\text{VEM}$ the absolute error of f_μ is below 0.08 for $\rho(S_\mu, S) = 0.91$.

Physics cuts

Due to baseline fluctuations, a very tiny fraction of stations have a muon fraction outside of the physical limits $f_\mu \notin [0, 1]$ and thus need to be excluded from the analysis. Therefore, $S_\mu \geq 0\,\text{VEM}$ and $S \geq S_\mu$ are imposed to also capture cases in which both signals are negative and still would map to a physically valid $f_\mu \in [0, 1]$.

5.3.5. Standardization

The standardization of scalar features is a commonly used preprocessing technique that aims to normalize the scale and distribution of features

$$x \to \frac{x - \langle x \rangle}{\sigma(x)}. \tag{5.10}$$

It transforms the data by subtracting the sample mean $\langle x \rangle$ and dividing by the standard deviation $\sigma(x)$, which results in a zero mean and unit variance. Standardization reduces

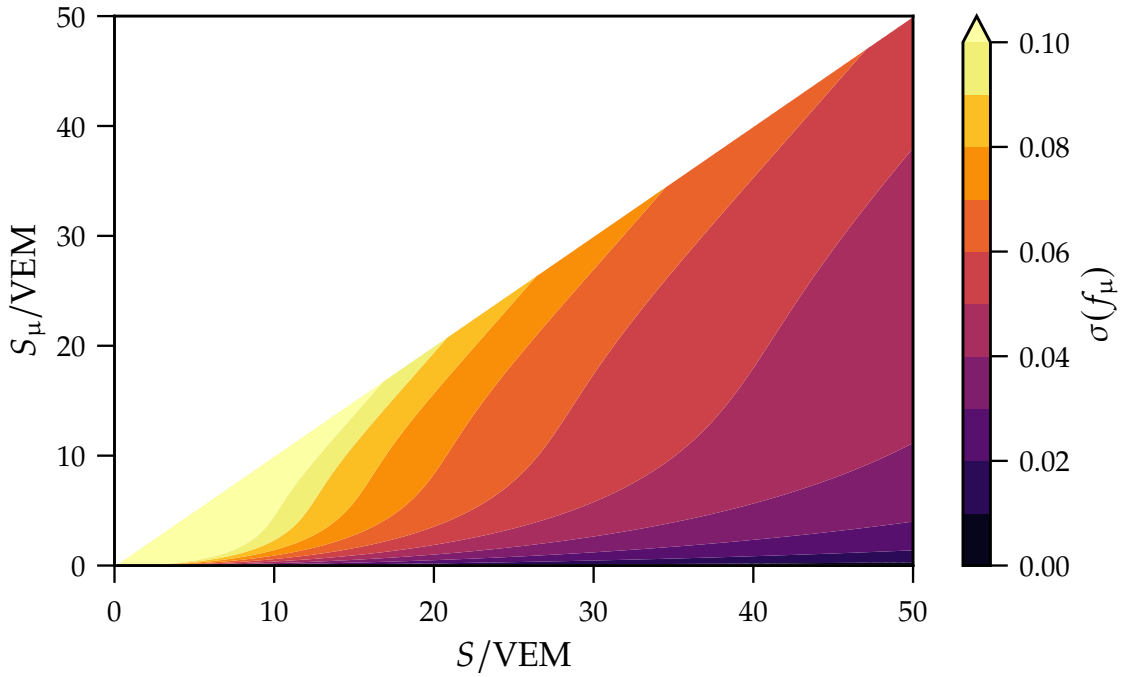


Figure 5.4.: Approximated muon fraction error model $\sigma(f_\mu)$ (Eq. (5.9)) as a function of the total Signal S and muon signal S_μ. It is evident that the error on f_μ decreases with increasing signal strength. The error is below $\sigma(f_\mu) \leq 0.08$ for $S \geq 30\,\mathrm{VEM}$.

the bias introduced by varying scales of the input data. Hence, it is beneficial to transofrm the input data to a common scale and unit-less quantities. Furthermore, it increases the assessment of interpretability and feature importance. The feature-wise weights are common to the same scale and thus directly comparable to evaluate the contribution of each feature to the model. The mean and standard deviation of the standardization are calculated on the training data and applied to the test data as well. It ensures that no bias is introduced into the data that the model receives as an input, which would be the case if the test data set has a different distribution. The standardization is applied to all timing features.

5.3.6. Custom Normalizations and Polynomial Features

In cases where the distribution of an input feature is highly asymmetric, peaked or skewed, a specific non-linear transformation is more beneficial thana a linear one. For instance, such transformations can be used to linearize a non-linear relationship between an input feature and the target. In this way, complex relationships can be modeled with linear regression. Other use-cases include flattening of peaky distributions, compressing the range of values or reducing the influence of outliers.
In the context of this thesis, the energy E exhibits some of these characteristics, as it is distributed as $1/E$. A log-transformation is applied, resulting in an uniform distribution on an event-basis $E \longrightarrow \lg E$. A similar transformation to achieve a uniform distribution is

applied to the zenith angle. Under the assumption of an isotropic CR flux, the transformation uniformizes $\theta \rightarrow \sin^2 \theta$.
Another method to capture non-linear relationships or interactions between features, is the generation of polynomial features. This is done by creating new features as the product of existing features up to a certain polynomial degree P. Given a feature set $\{x_1, x_2, \ldots, x_j\}$, all unique products $x_1^{p_1} x_2^{p_2} \ldots x_j^{p_j}$ that satisfy $p_1 + p_2 + \cdots + p_j \leq P$ are generated, which creates a the new feature set

$$\{x_1, x_2, \ldots, x_j\} \rightarrow \{x_1, x_2, \ldots, x_j, x_1 x_2, \ldots, x_1 x_j, x_1^2, \ldots, x_j^2, \ldots, x_1^P, \ldots, x_j^P\}. \tag{5.11}$$

5.4. Evaluation Metrics and Statistical Quantities

To make different models comparable and unify the metrics that are used to evaluate their performance in the subsequent sections, the following statistical quantities are introduced. The mean $\langle x \rangle$ of a sample of size N of a random variable x is defined as

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{5.12}$$

The sample variance σ^2 and the standard error on the sample mean $\sigma_{\langle x \rangle}$ are defined as

$$\sigma^2(x) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \langle x \rangle)^2 \tag{5.13}$$

$$\sigma_{\langle x \rangle} = \frac{\sigma(x)}{\sqrt{N}}. \tag{5.14}$$

The evaluation metrics are used to describe the quality of any model prediction $\hat{y}$ with respect to the target value y. These metrics can be computed globally, or in bins of a certain quantity, like the shower-plane distance r. The bias of a model $\langle \Delta y \rangle$, is defined as the mean difference between the prediction and the target values, which are sometimes also referred to as the residuals $\Delta y_i = \hat{y}_i - y_i$

$$\langle \Delta y \rangle = \langle \hat{y} - y \rangle = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i). \tag{5.15}$$

A bias around zero $\langle \Delta y \rangle \approx 0$ indicates that the model is not systematically over- or underestimating the target value and ensures an accurate prediction on average. To further quantify the precision of the model, which is given by the spread of the residuals, the standard deviation is used

$$\sigma(\Delta y) = \sigma(\hat{y} - y) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\Delta y_i - \langle \Delta y \rangle)^2}. \tag{5.16}$$

The bias and the resolution usually depend on different primaries, and thus a primary-dependent bias $\langle \Delta y \rangle_{\mathrm{p,Fe}}$, which tracks the absolute difference between proton and iron

bias, is introduced as well. Additionally, the resolution is also calculated separately for each primary $\sigma(\Delta y)_{\mathrm{p,He,O,Fe}}$

$$\langle \Delta y \rangle_{\mathrm{p,Fe}} = |\langle \Delta y \rangle_{\mathrm{Fe}} - \langle \Delta y \rangle_{\mathrm{p}}| \tag{5.17}$$

$$\sigma(\Delta y)_z = \sqrt{\frac{1}{N_z} \sum_{i=1}^{N_z} (\Delta y_{i,z} - \langle \Delta y \rangle_z)^2} \quad \text{for: } z \in \{\mathrm{p,He,O,Fe}\}. \tag{5.18}$$

In the following sections the y and $\hat{y}$ are replaced by the target at hand, which usually is the muon signal fraction f_μ and $\hat{f}_\mu$ respectively. The bias, bias standard error and resolution are denoted by $\langle \Delta f_\mu \rangle$, $\sigma_{\langle \Delta f_\mu \rangle}$ and $\sigma(\Delta f_\mu)$.
The regression metric to evaluate the performance of the model for the prediciton of the muon fraction is chosen to be the MAE. This results from the muon fraction describing a relative quantity that should be optimized to have an accurate prediction, irrespective of the total signal size. The MAE is defined as

$$\mathrm{MAE} = \langle |\Delta y| \rangle = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i|. \tag{5.19}$$

Merit Factor

The *merit factor (MF)* is a metric that describes the separation of two standard normal distributions, given by the subsets y_1 and y_2 of a random variable y. It is a quantity often used in Auger and is defined as

$$\mathrm{MF}(y_1, y_2) = \frac{|\langle y_1 \rangle - \langle y_2 \rangle|}{\sqrt{\sigma^2(y_1) + \sigma^2(y_2)}}. \tag{5.20}$$

For composition analyses, usually the sample mean and the sample variance of a variable y with respect to iron (1 = Fe) and proton (2 = p) subsets are used. Higher merit factors indicate a better separation of the two distributions, as long as the distributions are Gaussian.

5.5. Importance of Features

To gauge the predictive power of models, a study on which features are important for the prediction is often helpful and can provide insights in the underlying relations between features. For all future analysis of Auger using the SSD, it is very important to quantify the importance of the second detector for the prediction of the muon fraction f_μ in the SD (Sec. 3.4). First, linear correlations are explored. Then, various XAI importance measures are evaluated on ensemble models and compared qualitatively.

Pearson Correlation

To measure linear correlations between features themselves and between features and the target f_μ, the Pearson correlation coefficient ρ_p is used. As possible input features all

previously introduced handcrafted variables are taken into account, together with a selected subset of shower and station variables available from the standard reconstruction. The correlation matrix is depicted in Fig. 5.5. With the focus on the target f_μ, the zenith angle and the relative signal between SSD and WCD tend to have the highest absolute correlation, which is intuitively expected. The attenuation of the electromagnetic component is proportional to the zenith angle and the fraction of the signal of two different detectors encompasses the different responses to the muon and electromagnetic component of the shower. The absolute correlations with f_μ are followed by the shower-plane distance and the total signals in the detectors. Again, this can be interpreted in terms of electromagnetic attenuation increasing with the shower-plane distance and in the first order the total signal being composed of a mean constant fraction between electromagnetic and muon signals. The highly correlated feature cluster of Δh, t_i, Δt_{hs} and Δt_{c} has a light linear correlation with f_μ of the same absolute magnitude for each feature. They all encode information about the curvature of the shower front and thus the shower age. A second highly correlated feature cluster is given by Δt_{sp}, $t_{1/2}$ and t_{f}, which describes the temporal structure of the total signal trace and is related to the time dependent mixture of sub-component signal traces. Even though this cluster has almost no linear correlation with f_μ, non-linear transformations of these features will probably have an higher predictive power. To exploit the feature importance outside the linear regime, tree models are used in the following.

Ensemble Importance Study

To gain further insights into features that are important for a model to predict the muon fraction on a station level, the *Mean Decrease in Impurity (MDI)* (Sec. 4.5.1), *Permutation Feature Importance (PFI)* (Sec. 4.5.1) and *SHapley Additive exPlanations (SHAP)* values (Sec. 4.5.2) of ensemble methods are studied. The python library *scikit-learn* [87] provides various ensemble methods[1] for regression tasks. As *XGBoost* [88] and *LightGBM* [89] come with faster training, better evaluation metrics, and a scikit-learn Application Programming Interface (API), they are easily included and the preferred ensemble models. All methods are configured to use decision trees as base learners, the MSE as splitting criterion or impurity measure, and the total gain as feature importance. To additionally align all methods, the number of estimators is chosen to be 500 for each ensemble, because increasing it further does not improve the overall performance, but rather drives overfitting. The advantage of decision trees is the ability to model non-linear relationships between the input and the target, and thus to exploit further dependencies. The ensemble models are trained on the UUB QGSJetII-04 data set in Table 5.1. For numerical stability some of the input features are transformed to map to the order of unity, even tough tree methods are inherently independent of feature transformations. Since the timing information is redundantly encoded in multiple timing features, some of the low level features, namely t_{c}, t_i, $\Delta t_{\mathrm{hs}} \approx \Delta t_{\mathrm{c}}$ and ψ are dropped in this study. It will be shown in Sec. 5.6, that they hold similar information content.

[1] AdaBoostRegressor, ExtraTreesRegressor, GradientBoostingRegressor, RandomForestRegressor, HistGradientBoostingRegressor

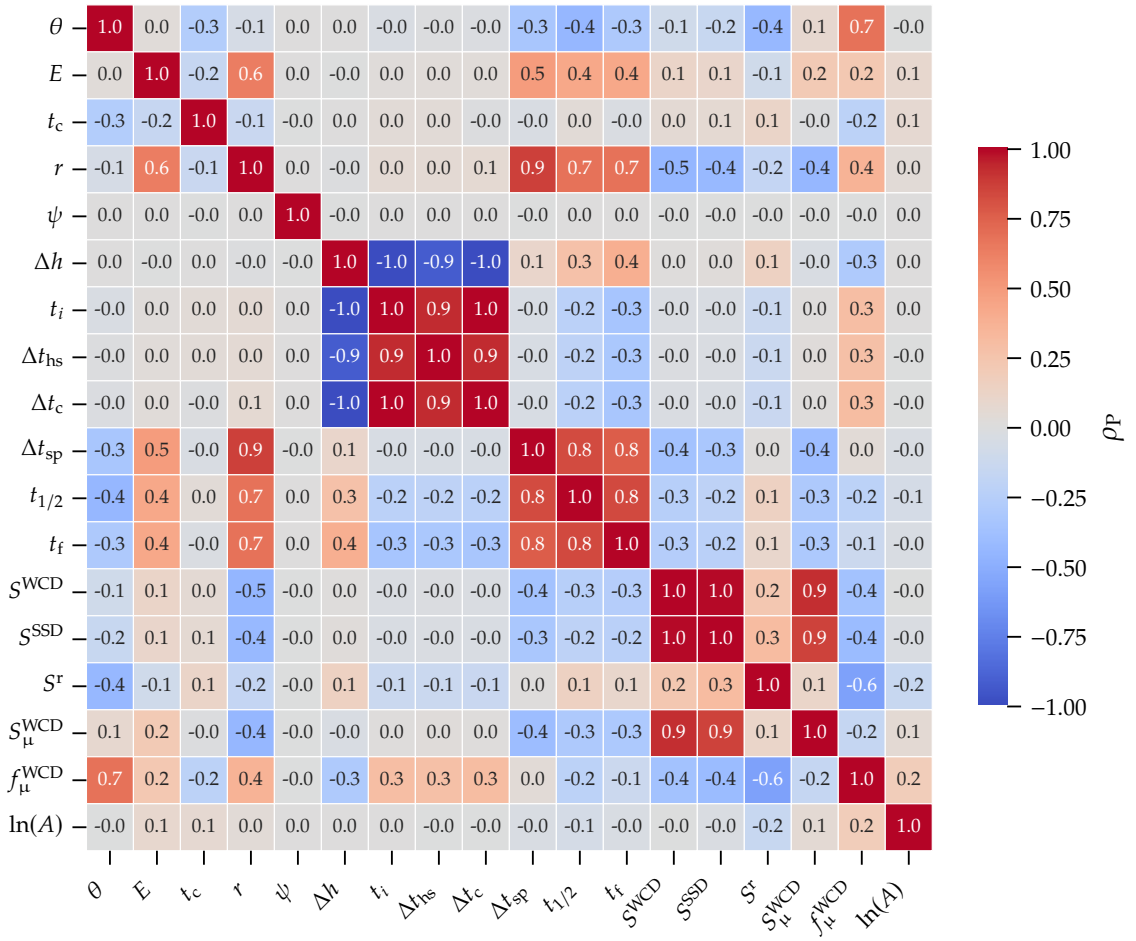


Figure 5.5.: Pearson correlation coefficient ρ_p of selected shower- and station-level features. The shower variables are given by the energy E, zenith angle θ, the impact time of the core on the ground t_c and the mass of the primary as $\ln A$. The station variables are given by the shower-plane distance r, the azimuth angle of a station inside the shower plane ψ, the height Δh, the signal start time t_i, the time difference between signal start time in each station and the hottest station $\Delta t_\mathrm{hs} = t_i - t_\mathrm{hs}$, as well to the shower core $\Delta t_\mathrm{c} = t_i - t_\mathrm{c}$, the time delay with respect to the plane-shower front $\Delta t_\mathrm{sp} = t_i - t_\mathrm{sp}$, the signal-rise time $t_{1/2}$, the signal fall time t_f, the signal of the detectors from the integration of the cut traces S^WCD and S^SSD, the signal ratio of both detectors $S^\mathrm{r} = S^\mathrm{WCD}/S^\mathrm{SSD}$, the muon signal S_μ^WCD and the muon signal fraction f_μ^WCD in the WCD derived by integration of the cut trace.

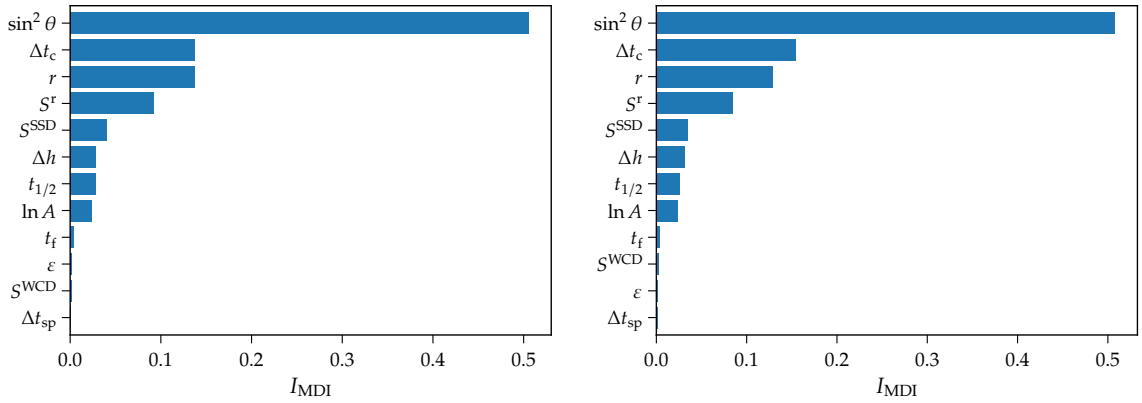


Figure 5.6.: *Mean decrease in Impurity (MDI)* for LGBMRegressor (*left*) and XGBRegressor (*right*). The MDI feature importance I_{MDI} is normalized to the sum of all feature importances and sorted by their value. Both ensembles provide very similar importance rankings. The zenith angle θ is by far the most important criterion for the reduction of the MSE at each node in the trees. The high importance of the signal ratio $S^r = S^{SSD}/S^{WCD}$ supports the benefit of adding a second detector to determine the muon fraction f_μ. The small energy $\varepsilon = \lg E/10^{19}$ eV dependence of f_μ is reflected in its negligible MDI value.

Mean Decrease in Impurity

The MDI feature importances for the *LightGBM* and *XGBoost* models are depicted in Fig. 5.6. The importance ranking of both models is almost identical and confirms their validity. The zenith angle θ is by far the most important feature for the reduction of the MSE at each node in the trees, which is in line with the findings of Sec. 5.5. It is followed by the time difference between the station start time and the impact time of the shower core on the ground Δt_c and the shower-plane distance r. The fourth most important feature is the total signal ratio S^r between the SSD and the WCD, which is much more important than the absolute total signal (S^{WCD}, S^{SSD}) of each detector and motivates the matrix formalism. Together with the total signal of the SSD, the information content provided by the total detector signals is exhausted. This results in only the signal of the WCD to be seemingly not important at all. Surprisingly, this still holds if the ensembles are trained without the signal fraction, as done in Fig. A.2. As pointed out in Sec. 5.2 already, the muon fraction is little dependent on the energy, which is reflected in the low importance of the logarithmic energy $\varepsilon = \lg E/10^{19}$ eV in the MDI feature importance. The logarithm of the mass of the primary $\ln A$ is also introduced to explore to what extent the relative muon signal fraction is related to the primary mass.

Permutation Feature Importance

For the study of the *Permutation Feature Importance (PFI)* (Sec. 4.5.1) the same data and ensemble models are used as in the MDI study, and the number of times a feature is permuted is fixed to 10 to provide appropriate statistics. To have access to feature importances describing the generalization capability of the model, the PFI studies are done on the test data set and the results are depicted in Fig. 5.7. A crosscheck to reveal potential overfitting

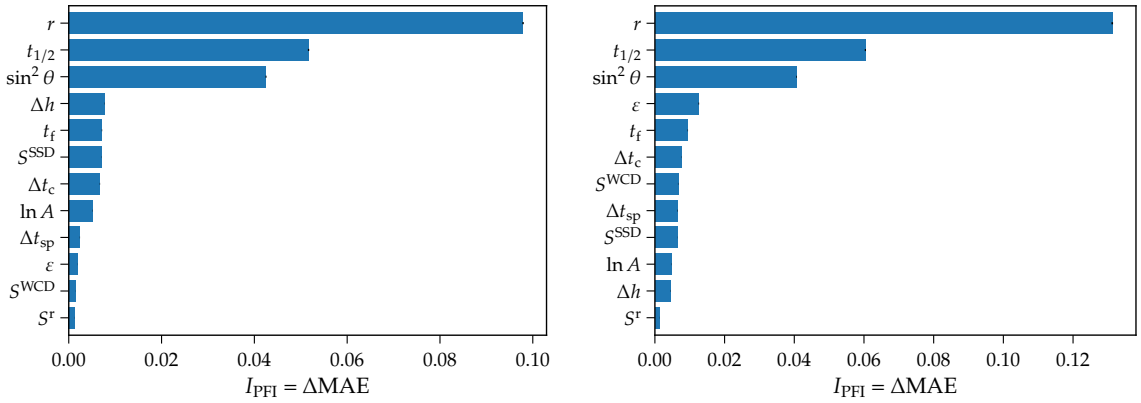


Figure 5.7.: *Permutation Feature Importance (PFI)* I_{PFI} = ΔMAE for LGBMRegressor (*left*) and XGBRegressor (*right*). The MAE is chosen as the PFI metric because the muon fraction denotes a relative quantity. The shower-plane distance r and the zenith angle θ are one of the most important features. For generalization capability, the signal risetime $t_{1/2}$ seems to be an important quantity as well for both ensemble models.

is done in Fig. A.3, where the PFI is also calculated on the train data set. As evident, the models are not overfitted because the PFIs are similar for both train and test data sets.
By comparing the most important features of the PFI with the MDI study, two major differences occur. Firstly, the signal rise time is of great importance for the generalization capability of the ensemble models in the PFI study, whereas it is less important in the MDI study. Secondly, the order of the other important features has changed slightly, but are comparable in their scale. For instance r is undoubtedly leading, while θ has become less important. Other differences are given by the reduced importance of the signal ratio of both detectors and each model having a slightly different feature ranking. This may be attributed to the randomness of the feature wise shuffling and the break down for highly correlated features. Despite the mentioned differences, the less important features are in good agreement for both studies.

SHapley Additive exPlanations Values

For the calculation of the *SHapley Additive exPlanations (SHAP)* values [74], the SHAP python library[2] is utilized. The framework provides a fast and exact algorithm for tree ensemble methods [90]. The same data and ensemble models as in the MDI and PFI studies are used. The SHAP values are computed using a data sample of size 20 000, which is randomly drawn from the test data set without replacement.
The global importance $\langle|\phi_j|\rangle$ of feature j is calculated by averaging the absolute SHAP values over the test data set (Eq. (4.8)). The results are depicted in Fig. 5.8. The order of the features with respect to their importance is quiet similar to the results of the MDI and PFI studies. The two features contributing the most to the model prediction are given by the shower-plane distance and the zenith angle. The following features of medium importance are of quiet similar magnitude, but there are differences in the ordering. More

[2] https://github.com/slundberg/shap

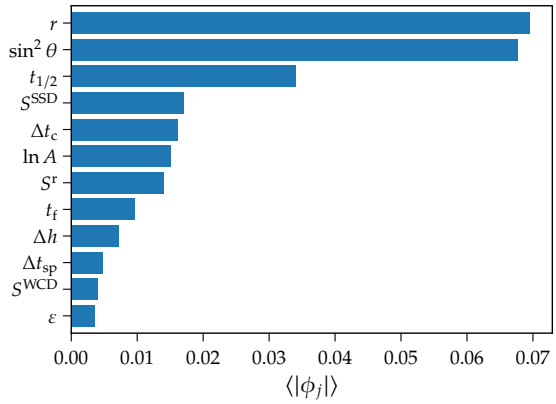


Figure 5.8.: Global importance $\langle|\phi_j|\rangle$, given by the mean absolute *SHapley Additive exPlanations (SHAP)* values for the LGBMRegressor. The zenith angle θ and the shower-plane distance r attribute the most for the prediction of the muon fraction f_μ.

detailed insights about the attribution of a feature with respect to a model's prediction can be obtained by the feature-wise distribution of the SHAP values (Fig. 5.9). In combination with the value of the feature, specified by the color code, the direction of impact on the model output can be investigated. In case of θ and r, high feature values lead to a positive additive attribution ϕ_j to the model predictions and vice versa. Meaning that stations far from the shower core or showers with high inclination contribute to a higher prediction of the muon fraction. This perfectly resembles the physical expectation due to increased attention of the electromagnetic component.

Feature Importance Summary

The findings of the MDI, PFI, and SHAP feature importance studies for the ensemble models are qualitatively summarized in the following. The zenith angle θ and the shower-plane distance r are consistently the most important features for the prediction of the muon fraction f_μ. The signal rise time $t_{1/2}$, the time difference between station start time and the impact time of the shower core on the ground Δt_c, the signal ratio between both SD detectors, the SSD total signal, and the primary mass are of medium importance for the ensemble models. The features with the least importance are the signal in the WCD, the time difference relative to the plane shower front Δt_{sp} and the energy of the shower E. The high importance of the timing features motivate the utilization of further time dependent trace features.

5.6. Baseline Models

At time of writing, no analytical method to estimate the muon fraction f_μ on station level is implemented in the standard reconstruction of $\overline{\text{Off}}\underline{\text{line}}$. In the field of machine learning, especially in training complex models, it is essential to have an easily interpretable benchmark for comparison. If this is not the case, one could end up in an *echo chamber*,

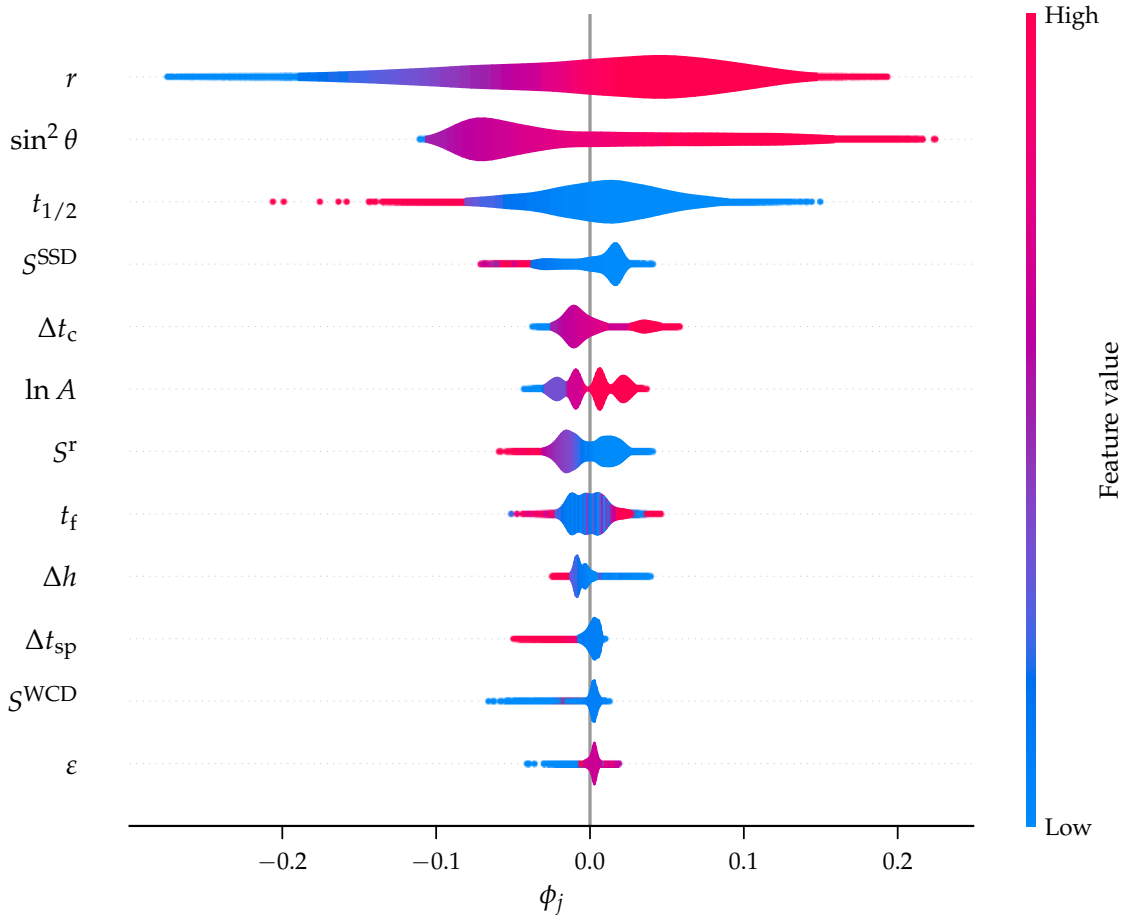


Figure 5.9.: *SHAP values* ϕ_j describe the additive attribution of each feature for the predictions of the LGBMRegressor, with respect to the expectation value of the predictions $\mathbb{E}[\hat{f}_\mu] = 0.51$. The features are ordered by their global importance $\langle|\phi_j|\rangle$. The color code indicates the feature value relative to its minimum and maximum value, while the width of the horizontal violins highlights the distribution of the feature with respect to its value. In general widely spread SHAP values indicate a strong attribution of a feature to the predictions and the feature is thus important. This is the case for the zenith angle θ and the shower-plane distance r. Additionally their feature values contribute smoothly to a gradual change in the corresponding SHAP values. Higher feature values give rise to positive SHAP values and thus add on top the expectation value and vice versa. This translates into stations far from the shower core or showers with high inclination, contribute to a higher muon fraction prediction, what is physically expected. Furthermore, the model learned a relation between the primary mass, which has been directly introduced as a feature, and the relative muon content. The four distinct blobs in $\ln A$ correspond to the four different primaries, which are just spread out by the Gaussian kernel of the violin plot.

improving the performance of his already very complex DNN only by tiny fractions with millions of parameters, while a classical linear regression could provide similar performance metrics. In this case the simpler model is always preferred, as stated by *Occam's razor* [91]. Another major benefit is that the increase in performance due to complexer models is explicitly quantifiable.
In the following, the baseline or reference models are linear regression and ensemble models. Since the target f_μ on the quality-selected data follows roughly a normal distribution with a mean of $\mu(f_\mu) = 0.511$ and a standard deviation of $\sigma(f_\mu) = 0.171$, a constant predictor suits as the absolute baseline any other model has to outperform. The input features are normalized according to Secs. 5.3.5 and 5.3.6. Starting from the matrix formalism $f_\mu = a + bS^{\mathrm{SSD}}/S^{\mathrm{WCD}}$, additional station and shower variables are added as input features. By iteratively adding new features and tracking the corresponding changes in the evaluation metrics, an additional feature importance measure is provided. In case of the linear model, polynomial features up to a degree of $P = 2$ are generated to allow it, to catch up interaction and non-linearities. This loosely corresponds to a data driven approach for the parametrization of the matrix parameters. The naming convention for the regression models is oriented by the degree of the polynomial input features. Thus the model with features up to degree $P = 1$ or $P = 2$ are referred to as the *linear* or *quadratic* models respectively. The ensemble models *XGBRegressor* and *LightGBM* are trained on the same feature sets, but without polynomial expansion, because they are inherently able to model non-linearities. They have been initialized with 500 estimators and default parameters, so no hyperparameter tuning is performed.

5.7. Neural Networks

5.7.1. Training Strategy

The loss function for the training of the NNs is chosen to be the MSE

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N}\sum_{i}^{N}(y_i - \hat{y}_i)^2, \tag{5.21}$$

where $y = f_\mu$. The NNs are trained with a batch size of 128 for a maximum number of 60 epochs. Adam [63] is used as an optimization algorithm with its default values in tensorflow [92]. To prevent the model to overfit on the training data, evaluation metrics are computed and tracked by callbacks on the validation data set. The independent validation data set is sampled randomly without replacement from 10% of the training data before the training start. The ReduceLROnPlateau[3] and EarlyStopping[4] callbacks monitor the validation loss at the end of each epoch. If the validation loss has not decreased since 3 epochs, the learning rate is decreased by a factor of 0.8. This helps to explore minima during optimization. The training is stopped if the validation loss has not decreased since 7 epochs. If the validation loss increases again, the weights of the best epoch are restored.

[3] https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/ReduceLROnPlateau
[4] https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/EarlyStopping

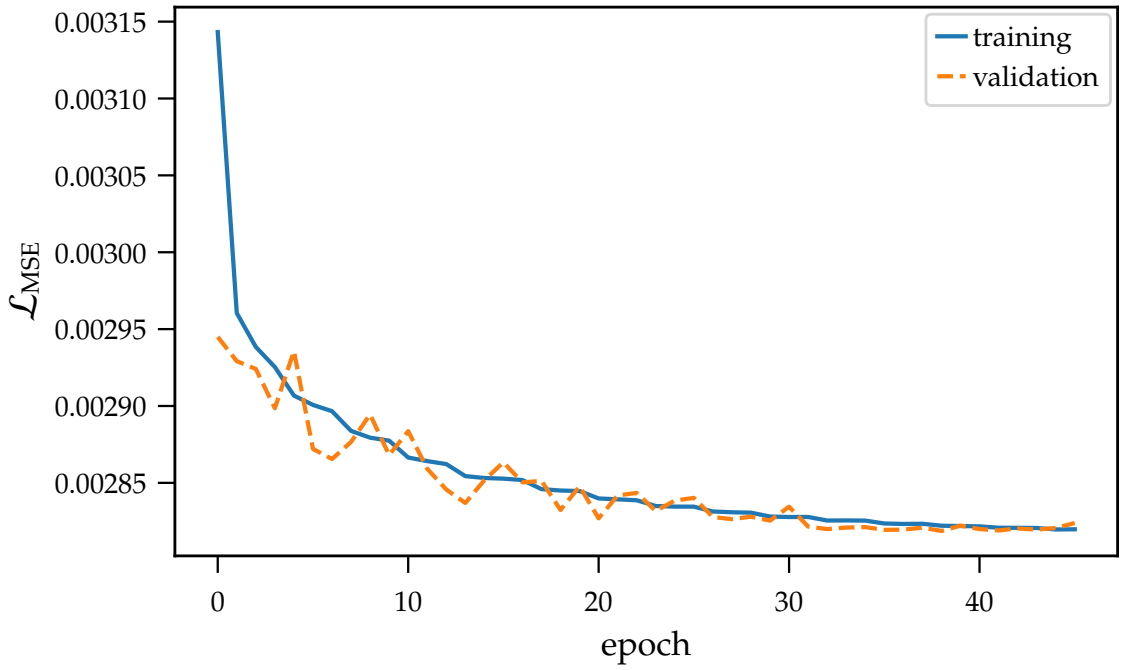


Figure 5.10.: Training and validation loss $\mathcal{L}_{\mathrm{MSE}}$ curves for a simple dense network. The training stops already at the 45^{th} epoch because the validation loss has not improved over the last 7 epochs. If trained further, the model would overfit, which would result in an increasing validation loss.

This training strategy results in a stable training process for which the model does not overfit and computing resources are not wasted.

5.7.2. Data Normalization

The data preprocessing is incorporated into the NN models. This greatly increases the transportability of the model, as the normalization is done on the fly during prediction and parameters are saved together with the model weights. It enables one to experiment with many different normalization techniques for the same data set more easily. Instead of creating a new duplicate data set each time a normalization is changed and keeping track of it, a new model with the changed normalization is trained. The drawback of this approach is a slightly increased training time. Nevertheless, this increase is acceptable because of the moderate size of the models.
The normalization is performed by custom layers implemented in Tensorflow [92]. The normalization in the *MetaNormalization* layer is done on a per feature basis, which means that each feature is normalized separately. The initialization of the *MetaNormalization* layer is done by adapting it to the training data before the training starts. Later, as the time traces are included as additional inputs, an additional custom *TraceNormalization* layer is added that implements the logarithmic transformation by default. The time dependent signals in the WCD and SSD traces are normalized with a constant offset $c = 1$, by the

transformation commonly applied in previous studies [31, 93, 94, 95],

$$s = \frac{\lg\left(s/(\text{VEM or MIP}) + c\right)}{\lg\left(100 + c\right)}. \tag{5.22}$$

5.7.3. Architectures

Dense Network

The simplest building blocks of NNs are fully connected dense layers. Dense layers can be represented by a number of units which are connected to every single unit in the preceding layer, creating a dense or full connection between the layers. In previous studies [31, 96] it was shown, that they are a powerful tool to predict the muon signal in the WCD. Hence, it perfectly serves as a starting point as a NN-based reference model. The architecture (Fig. C.1) is based on previous studies [31, 96]. The network consists of a *MetaNormalization* layer and four dense layers with 32 units each. For each dense layer a ReLU activation function was chosen. The output is a single unit that predicts the muon fraction. Since $f_\mu \in [0, 1]$, the output is followed by an additional modified ReLU function which is limited to the interval $[0, 1]$. The model complexity, given by roughly 3 500 trainable parameters, is low if compared to state-of-the-art NN models, such as BERT [97] which has around 340 million parameters.

Convolutional Neural Network

In order to exploit the information hidden in the signal traces, the dense architecture is unsuitable. Instead of manually extracting additional trace features, we let the network itself do the work. The general architecture of the CNN can be split into three sub-networks: the *Trace Analyzer*, the *Meta Analyzer*, and the *Predictor* (Fig. 5.11).
The *Trace Analyzer* (Fig. C.2) uses the first 400 bins of the detector traces after the signal start bin as input. The trace is normalized by a logarithmic transformation (Eq. (5.22)) by default and followed by *convolutional blocks* which successively extract higher level features while compressing the trace. The extracted features are flattened[5] to one dimension. Each convolutional block contains a convolutional layer with a kernel size of 3, a stride of 1, and a ReLU activation function. The downsampling is accomplised by a max pooling layer with a kernel size of 2 and a stride of 2. Alternating convolutional layers and pooling layers is a common used technique in the field of computer vision, e.g., AlexNet [98]. The idea behind this architecture is based on the increasing *receptive field*[6] with every convolutional layer, while still being able to capture the smaller features through the small kernel sizes.
The *Meta Analyzer* uses the scalar features as input. It resembles the previously introduced dense NN architecture without the final dense layer that contains only a single

[5] A flattening layer reshapes a multi-dimensional output from a preceding layer into a one-dimensional vector, allowing it to be processed by subsequent fully connected layers.
[6] The receptive field in a neural network refers to the region of the input data that is directly or indirectly connected to a particular node and influences its activation.

unit (Fig. C.1). Thus, it consists of the custom *MetaNormalization* layer, normalizing the scalar inputs, which is followed by three dense layers with a ReLU activation function each. The intention of this sub-network design is to integrate relevant characteristic properties of the shower event, which would not be available to the network if only traces are used. The *Predictor* (Fig. C.3) combines the extracted features from the *Trace Analyzer* and *Meta Analyzer* by concatenating them. The *Predictor* consists of three dense layers with 64 units and ReLU activation functions each. The output is a single unit that is followed by a modified ReLU activation function to predict the muon fraction. The model complexity is given by roughly 20 000 trainable parameters.

5.8. Convolutional Neural Network

5.8.1. Comparison to Reference Models

What is the amount of unused information in the complex and raw data structures and is it possible to utilize this information with special NN architectures? In order to answer this question, NN architectures with increasing complexity are evaluated and compared to the baseline models defined in Sec. 5.6 to validate the CNN (Sec. 5.7.3). In general, the model complexity is measured by the number of trainable parameters. The main goal is to explore, how increasing the complexity influences the predictions and if the increase in predictive power is significant.
The verification of the CNN architecture is done in two steps, by comparing performance metrics to other models. First, the CNN architecture is compared to the best performing baseline models. Then, the performance of the CNN architecture is crosschecked with another commonly used architecture that is used to process time dependent data, a LSTM-based network. Its architecture is obtained by substituting the iteratively stacked convolution and pooling layers in the *Trace Analyzer* of the CNN by a single LSTM layer. As a measure of performance, the binned biases and resolutions are used.

Baseline Comparison

Instead of optimizing the model architectures itself, as done in [96], we analyzed how the performance of each model class changes if different input feature sets are used. Starting from a feature set that consists of the total signal of each detector S^{WCD} and S^{SSD}, motived by the matrix formalism (Sec. 5.1.1), other features are successively added to find the best performing model. This can be viewed as a non-linear, higher order extension of the matrix formalism. For each feature set, the reference models are trained on the same training data set and the change in performance on the independent test data set is tracked. This procedure is a cross-check of the feature importance study done in Sec. 5.5.
The results for the linear and quadratic model are shown in Table B.1 and for the DNN in Table B.2. The decrease of the Mean Absolute Error (MAE) is consistent with the findings in the feature importance study done with ensembles (Sec. 5.5), θ and r are very important inputs to predict the muon fraction. Similar, h and $t_{1/2}$ significantly improve the predictive power. Both station-level variables provide information about upstream-downstream location of the station and the temporal structure of the signal. Again, the logarithmic

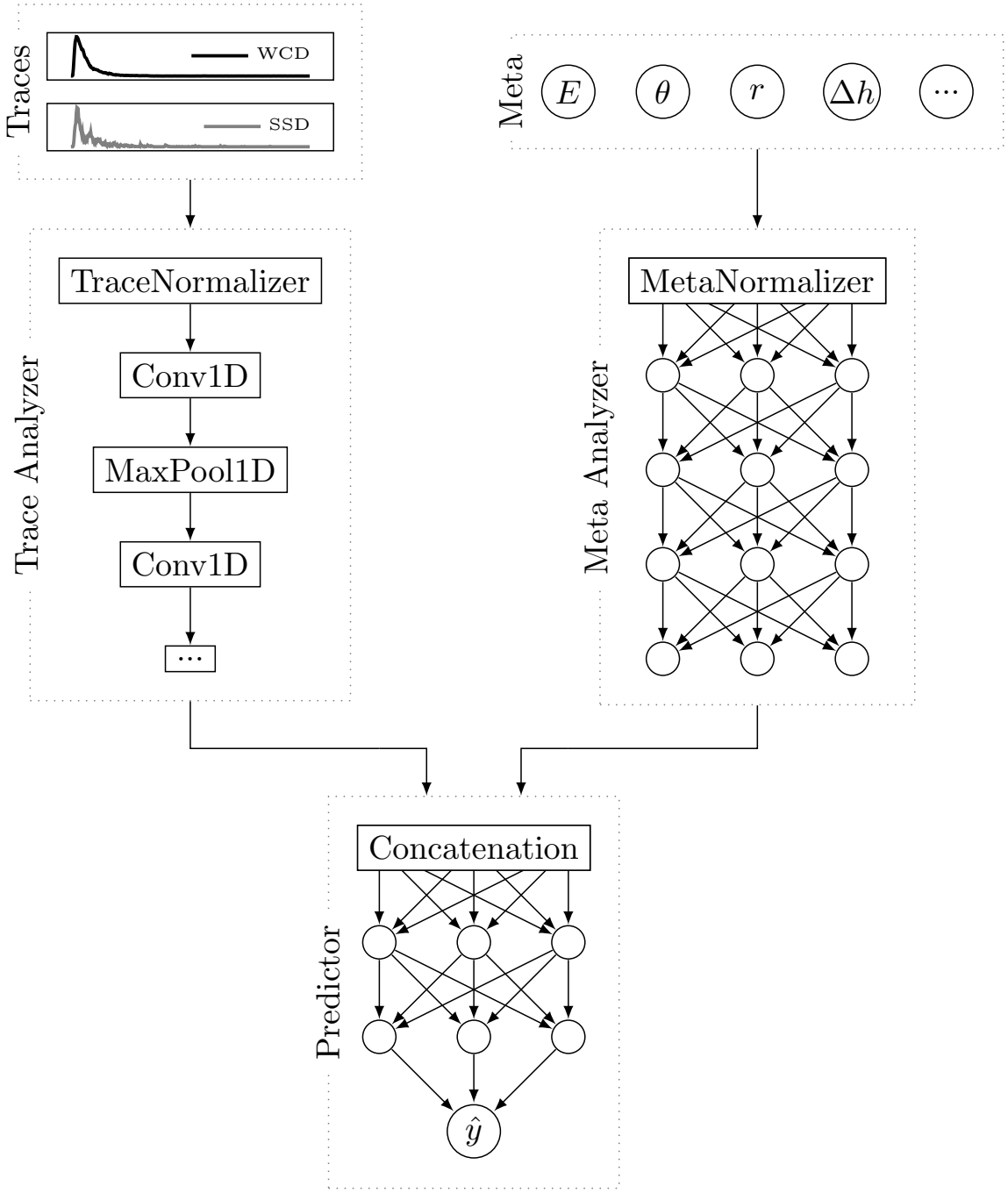


Figure 5.11.: Illustration of the base NN architecture, combining the time traces with station and shower level variables to predict the muon fraction. The main components are the *Trace Analyzer*, the *Meta Analyzer*, and the *Predictor*. The *Trace Analyzer* takes the first 400 bins of the detector traces as an input and extracts higher level features. It consecutively stacks convolutional and pooling layers while compressing the trace features and increasing the *receptive field*. The *Meta Analyzer* takes the scalar features as an input and resembles the former simple dense architecture (Fig. C.1) without the output layer. The *Predictor* concatenates the extracted features from the *Trace Analyzer* and *Meta Analyzer*, mixes their information through dense layers, and predicts the muon content $\hat{y} = \hat{f}_\mu$.

energy seems to not be important for the prediction of f_μ. The evaluation metrics are not altered by including it. In general, the performance increases with more features and higher model complexity per model class. Notable is also the significant enhancement from the first order features to the second order features in case of the regression models. It shows that there are non-linear relations between the input features and the muon fraction, which the linear model failed to represent. Studies, such as [31], indicate that substituting the S^{SSD} by the signal ratio $S^{\mathrm{r}} = S^{\mathrm{SSD}}/S^{\mathrm{WCD}}$ improves the performance, as this feature is more closely related to f_μ (Sec. 5.1.1). This is investigated in Table B.3. However, the result of this study does not support their claim since the global evaluation metrics do not change in comparison to Table B.2. Qualitatively, the attribution of a feature to the performance of the linear and quadratic models is similar to the results obtained by the simple dense NN. A major difference between the NN and the linear models is the better utilization of information contained in the feature sets. Less inputs are required for similar results in performances. This efficiency is due to the non-linearities of the activation functions in the NN.

After various features combinations on each baseline model have been tested, a summary of the best performing baseline models of each model class is given in Table 5.2. All models significantly outperform the linear predictor and thus verify the chosen methods and feature sets. As expected, the evaluation metrics improve with increasing model complexity, but the improvement is inherently diminishing between the quadratic, LightGBM and simple dense model. This saturation indicates that the information in the scalar features to predict f_μ is exhausted. The CNN is able to extract additional information directly from the temporal structure of the signal traces. Its global MAE of around $\langle|\Delta f_\mu|\rangle = 0.037$ can be translated into an lower limit on the average relative error on the prediction of the muon signal via $\langle|\Delta f_\mu|\rangle \approx \langle|\hat{S}_\mu - S_\mu|/S_{\mathrm{tot}}|\rangle \leq \langle|\hat{S}_\mu - S_\mu|/S_\mu\rangle$, and amounts to $\langle|\hat{S}_\mu - S_\mu|/S_\mu\rangle \geq 3.7\%$. Thus, the error is of the same order as the relative errors reported in previous studies [31, 96]. The additional information hidden in the traces, can be related to a decrease in the proton-iron bias of around $\Delta\langle\Delta f_\mu\rangle_{\mathrm{p,Fe}} = -0.0168$ and a increase in resolution of $\Delta\sigma(\Delta f_\mu)_{\mathrm{p}} = -0.0067$ and $\Delta\sigma(\Delta f_\mu)_{\mathrm{Fe}} - 0.0045$ for proton and iron, by comparing the global dense NN to the CNN.

Additionally to the global metrics, the bias and resolution in dependency of certain shower variables are of prime importance to further qualify the performance. In general, the LightGBM, dense NN and CNN consistently outperform the linear and quadratic models. They are not suspect to any significant bias over the selected shower and station variables. Figs. 5.12 to 5.13 show, that the complexity of the models, measured in their number of trainable parameters (Table 5.2), is correlated to the achievable resolution. The linear model fails to reproduce dependencies, which are caught up almost perfectly by the polynomial feature expansion in $\sin^2\theta$. As a consequence, the quadratic model reaches very close to the bias and resolution of the LightGBM and simple dense NN. This strongly indicates the exhaustion of information provided by the feature set and motivates to explore the information content of the signal time traces. The CNN model outperforms all other models due to the additional information contained in the signal traces. The increase in performance is reflected in a better resolution of around $\Delta\sigma(f_\mu) \approx -0.01$ in comparison to the LightGBM and NN models in all variables. The resolution of the CNN ranges from $\sigma(f_\mu)$=0.04 to 0.06, for r, θ and E. It indicates a promising discrimination capability

Table 5.2.: Performance metrics of the baseline models. The models are chosen as the models with the lowest MAE for each model class. All reference models exhibit a better performance than the constant predictor. The performance of the quadratic model is compatible with the more complex LightGBM and dense NN models. The number of trainable parameters is given in the second column.

Model	Parameters	$\langle \lvert \Delta f_\mu \rvert \rangle$	$\langle \Delta f_\mu \rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_\mathrm{p}$	$\sigma(\Delta f_\mu)_\mathrm{Fe}$
constant	1	0.1419	0.0819	0.1712	0.1638
matrix	2	0.1069	0.0349	0.1365	0.1276
linear	11	0.0550	0.0475	0.0730	0.0651
quadratic	66	0.0435	0.0350	0.0591	0.0529
LightGBM	500	0.0409	0.0299	0.0554	0.0504
dense NN	3 553	0.0408	0.0264	0.0550	0.0501
CNN	19 873	0.0359	0.0141	0.0483	0.0456

between proton and iron induced showers with respect to their muon fraction. Their average differences can reach $\langle f_\mu(r) \rangle_\mathrm{Fe} - \langle f_\mu(r) \rangle_\mathrm{p} \approx 0.1$ on a station level for certain energy and zenith bins (Fig. 5.1). Hence, the discrimination would be possible on a 1σ-level with an unbiased predictor. The monotonic increase in resolution with increasing signal strength, distances closer to the shower core, and higher energies, is similar for all models. For higher energies, the increased resolution can be attributed to an increase in statistics (Fig. 5.2).

CNN vs LSTM Trace Analyzer

Since the CNN outperforms the reference models in all tested metrics, a crosscheck to another architecture which is specifically designed to process sequential data has to be done. The sufficiency of the CNN architecture is validated by comparing its performance to a more complex NN using a LSTM-based *Trace Analyzer*. This is accomplished by substituting the stacks of convolutional and pooling layers with one LSTM (Sec. 4.3) layer in the *Trace Analyzer*. The LSTM layer is configured by its number of units to extract a total of 32 or 64 trace features to investigate possible improvements. Additionally a bidirectional LSTM layer, which extracts 64 trace features, is deployed instead of the normal LSTM layer in order to capture sequential information in both time directions. As a reference, the CNN analyzer extracts 96 trace features in its default configuration. All models use the same *Meta Analyzer* with the scalar inputs θ, E, r, Δh and Δt_c. The comparison in terms of their biases and resolutions is, for instance, depicted as a function of distance in Fig. 5.14. The performance comparison in the other binned quantity can be found in Fig. A.4 and Fig. A.5. All tested configurations show essentially the same resolution, while there are some minor differences in the bias. The same holds for the evaluation of global metrics in Table 5.3. Thus the used CNN analyzer is validated and all the following extensive studies are performed with the CNN analyzer, as training the LSTM is computationally much more expensive.

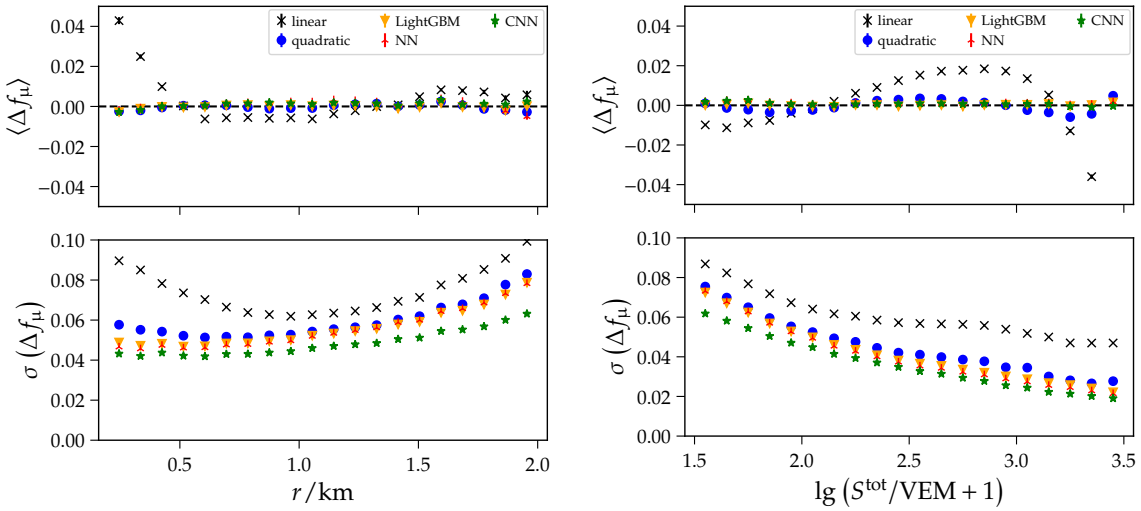


Figure 5.12.: Bias (*top*) and resolution (*bottom*) of the CNN and the reference models as a function of shower plane distance r (*left*) and logarithmic total signal $\lg(S/\mathrm{VEM}+1)$ (*right*). Only the linear model shows a noticeable bias in the whole phase space. The resolution monotonically increases with larger total signals and decreases with larger shower plane distances for all models. The decrease in resolution close to the shower core can be attributed to the limited statistics in this region. The CNN outperforms the resolution of the quadratic model by around 0.02.

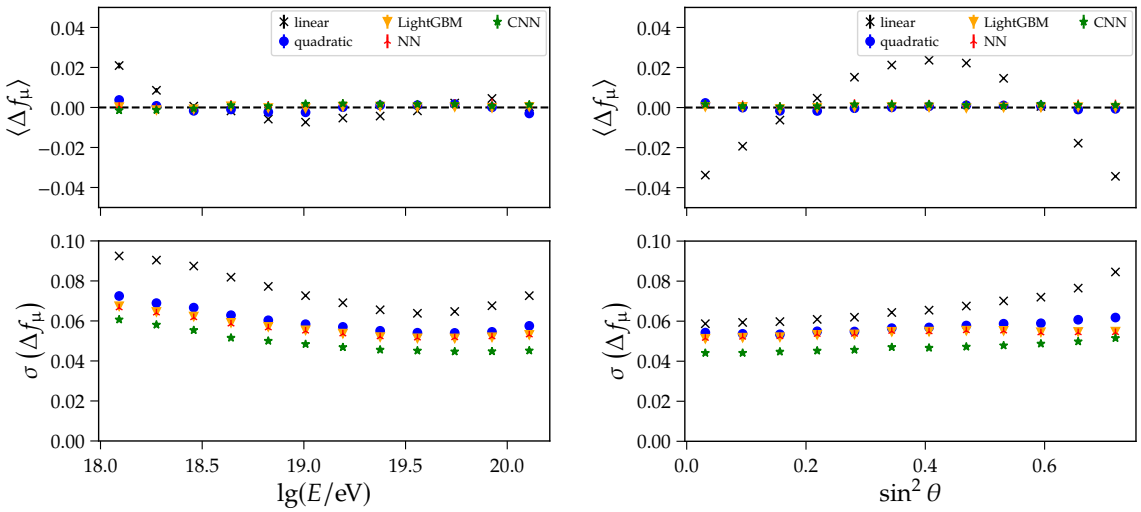


Figure 5.13.: Bias (*top*) and resolution (*bottom*) of the CNN and the reference models as a function of logarithmic energy $\lg E$ (*left*) and $\sin^2\theta$ (*right*). Only the linear model shows a noticeable bias in the whole phase space, and fails to catch the quadratic dependency in $\sin^2\theta$. The resolution of the CNN monotonically increases with larger energies until it is almost constant around $\sigma(\Delta f_\mu) \approx 0.042$, and gradually decreases with larger zenith angles. The CNN outperforms the resolution of the quadratic model by around 0.02 over the whole phase space.

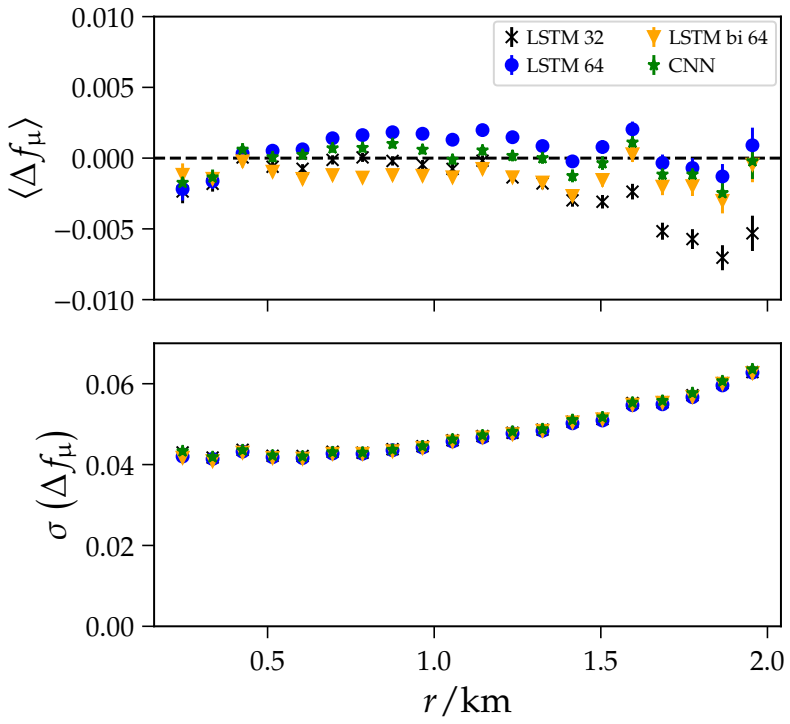


Figure 5.14.: Bias (*top*) and resolution (*bottom*) of a DNN using a convolutional, versus different LSTM-based *Trace Analyzers*, in dependence of the shower plane distance r. The LSTM analyzers consist of a single LSTM layer with 32 or 64 units, or a bidirectional layer with 64 units in total. While there are minor differences in the bias, the achieved resolution is essential the same over all trace analyzers. It validates the used CNN analyzer as a suitable choice.

Table 5.3.: Evaluation metrics for different LSTM-based *Trace Analyzer* architectures to validate the used CNN approach. As all metrics are essential the same to a sub-percent level, the CNN analyzer is validated as a suitable choice.

Model	Trace features	Parameters	$\langle \lvert \Delta f_\mu \rvert \rangle$	$\langle \Delta f_\mu \rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_{\mathrm{p}}$	$\sigma(\Delta f_\mu)_{\mathrm{Fe}}$
LSTM	32	16 193	0.0360	0.0166	0.0484	0.0454
LSTM	64	30 913	0.0356	0.0155	0.0480	0.0452
LSTM bi	64	22 721	0.0356	0.0149	0.0479	0.0453
CNN	96	19 841	0.0361	0.0160	0.0484	0.0457

5.8.2. Optimization of the Architecture

Since the CNN architecture reached the performance of LSTM-based architectures (Table 5.3), we investigated if it still has unreached potential to further increase the performance. Therefore, a hyperparameter scan is performed to find a possibly better performing configuration for the trace analyzer. Additionally, it is investigated if the log-normalization of the time traces has any impact on the convergence behavior of the model by comparing it to other possible normalization layers. Finally, the performances of different meta features are evaluated to examine the best performing input feature set. To reduce the randomness inherent to the training procedure, each network configuration is trained five times and the global metrics of the converged models are averaged.

Optimization of the Trace Analyzer

To optimize the used hyperparameters, a grid search over the number of convolutional blocks n_{b}, the number of convolutional layers per block $n_{\mathrm{b,l}}$ given as a list of their units, the pooling layer type (maximum and average) and the type of convolutional layer (Conv1D[7] or SeparableConv1D[8]) is performed. For this study, the input for the *MetaAnalyzer* is fixed to E, θ, r, Δh and Δt_{c}. The result is depicted in Table B.4. The normal Conv1D layer constantly outperforms the SeparableConv1D configuration. Increasing the number of convolutional layers per block $n_{\mathrm{b,l}}$, to increase the receptive field, has not significantly increased the performance. In fact, for some configurations it does the opposite as the MAE rises. It is attributed to th fact that the used architecture of the *TraceAnalyzer* is not suitable to be stacked to arbitrary depth. Substituting the maximum pooling layer with an average pooling layer lead to a decrease of ≈ 0.002 in $\langle|\Delta f_{\mu}|\rangle$ across all models. Further compressing down the number of extracted trace features n_{tf} by increasing the number of convolutional blocks did not improve the performance either. The major part of the configurations with the Conv1D layer consistently lead to $\langle|\Delta f_{\mu}|\rangle \in [0.0357, 0.0362]$. All of these results indicate that there is little space left for further improvements for this specific architecture.
In addition to the Trace Analyzers architecture search, the influence of the widely used log-normalization of the traces in Auger needs to be investigated and qualified. Therefore two further normalization layers are introduced. The maximum normalization takes each trace and divides it by its maximum, limiting the maximum value to 1 and removing any units. The summation normalization divides each trace by the sum of all values in the trace and thus, by the total deposited charge of the particles. These three possible trace normalization techniques are compared to not applying any normalization at all. The comparison is done in two steps. First, the convergence behavior during training is investigating by tracking the training and validation loss. Then, the trained models are evaluated with respect to their global metrics. All models are configured to train for a maximum of 100 epochs. The loss curves (Fig. 5.15) in case of no normalization and the logarithmic normalization converge more rapidly and smoothly than the other two. Furthermore, they are able exploit a more optimal model parameter space as their losses

[7] https://www.tensorflow.org/api_docs/python/tf/keras/layers/Conv1D
[8] https://www.tensorflow.org/api_docs/python/tf/keras/layers/SeparableConv1D

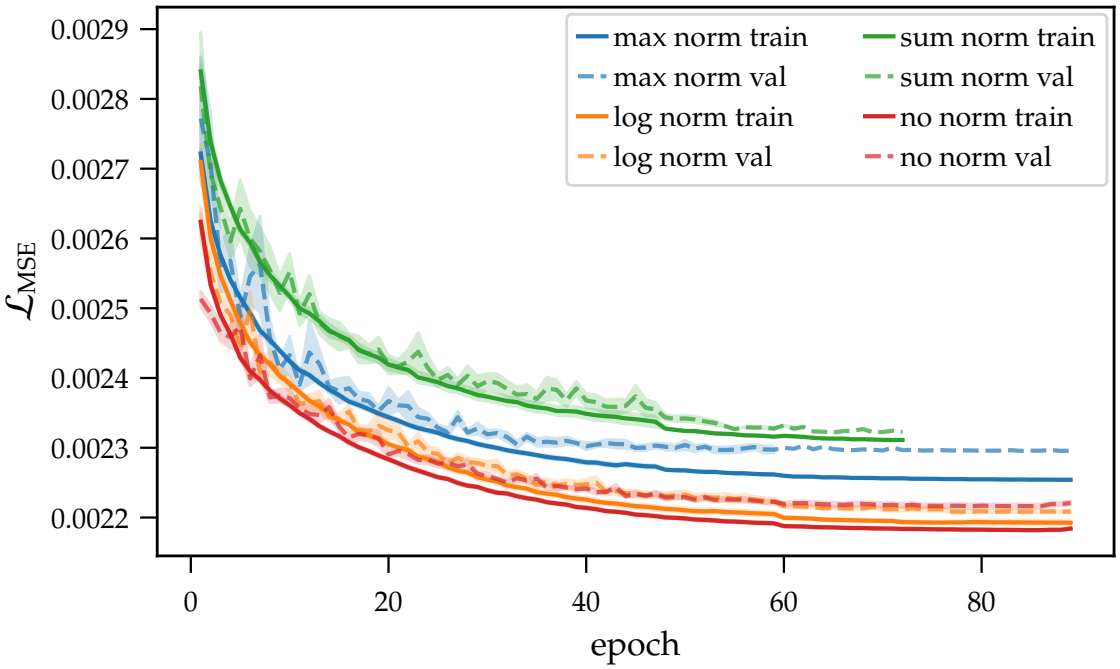


Figure 5.15.: Average loss curves and their standard deviation for an ensemble of five CNNs per unique trace normalization layer. No trace normalization and the logarithmic normalization provide a stable and quick convergence, while also clearly outperforming the maximum and summation layer.

are smaller as the training progresses. One has to keep in mind that the validation data set is randomly sampled from the training set before training start and thus, there are fluctuations in the validation losses which are indicated by the error bands. The networks with maximum and summation normalization layers stop training earlier, as the losses are not decreasing further. The effect of the different normalization techniques on an average global metric scale (Table 5.4) show, that applying no normalization at all lead to the best results. Even though, the actual differences to the log-normalization layer are of the sub-percentage-point level and could be attributed to statistical fluctuations inherent to the training procedure. The differences to the maximum and summation layer can be explained by the loss of information about the signal size, which is removed by dividing through the maximum or the sum.

Influence of Meta Inputs

Since the hyperparameter search of the CNN *TraceAnalyzer* did not yield any significant improvements, the performance of different meta input feature sets is analyzed. For the whole study, the architecture of the Trace Analyzer is fixed to its default configuration. The average global metrics for reconstructed input variables (Table 5.5) show, that the shower and station geometry defined by θ and r or θ and E provide almost the complete additional information content as the metrics only decrease on the fourth decimal place by adding further meta variables. In contrast to the feature importance study, the network

Table 5.4.: Evaluation metrics for different trace normalization layers, aggregated as the mean of the metrics of five CNNs. No trace normalization at all provides the best metrics on the test set.

Trace normalization	$\langle \lvert\Delta f_\mu\rvert \rangle$	$\langle \Delta f_\mu \rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_\mathrm{p}$	$\sigma(\Delta f_\mu)_\mathrm{Fe}$
none	0.0359	0.0141	0.0484	0.0456
log	0.0362	0.0144	0.0486	0.0459
max	0.0365	0.0151	0.0491	0.0463
sum	0.0370	0.0164	0.0498	0.0467

is able to utilize the energy as an important feature. Another important feature which induces a greater drop in the global metrics, is information about the curvature of the shower front and thus the shower age. It is provided interchangeably by Δt_sp, Δt_c and Δt_hs. By including the signal ratio S^r of both detectors, it shall be explored if the network somehow underweights the SSD signal trace. This is rejected on the basis of the global metrics. As the feature set θ, E, r, Δh and Δt_c provides the best global metrics, it is used for the following studies.
Since the reconstructed variables are subject to uncertainties and biases, an additional crosscheck on the global performance is done by using only MC variables as inputs. Therefore the total traces, which are simulated independently from the sub-component traces, are substituted by the sum of all sub-component traces and the reconstructed meta input features E, θ, r are replaced by the corresponding MC features. through the trace substitution, the random baseline fluctuations are gone and the baseline is constant which will lead to less noise propagating through the network. The resulting global metrics are depicted in Table 5.6. By comparing the global MAE of the reconstructed θ, E and r to the MC feature set, a decrease of about $\Delta\langle \lvert\Delta f_\mu\rvert \rangle = -0.001$ is found. The difference in the proton-iron primary bias, as well their resolutions are only marginally affected. This validates that the used reconstructed variables do introduce only a small fraction of uncertainty in the model predictions, and thus the Transferability of the trained model to measured data is enhanced. Furthermore, one could speculate that this indicates that the network has reached the absolute lower limit of achievable resolution, which is given by the inherent statistical fluctuations of the shower development and the detector responses.

5.8.3. Detector Study

Significance of the SSD

One of the main goals of the AugerPrime upgrade is the disentanglement of the electromagnetic and muonic sub-components of an air shower (Sec. 3.4). Until this section it has been assumed, that the complementary information provided by the SSD time trace is essential for a improved reconstruction of the muonic fraction, as it is the case in the classical matrix formalism (Sec. 5.1.1). To qualify the significance of the SSD, the CNN is trained on either only the SSD, or only the WCD trace, or on both detector traces in two

Table 5.5.: Evaluation metrics for different meta input feature combinations, aggregated as the mean of the metrics of five CNNs per feature set.

Meta Features	$\langle\lvert\Delta f_\mu\rvert\rangle$	$\langle\Delta f_\mu\rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_\mathrm{p}$	$\sigma(\Delta f_\mu)_\mathrm{Fe}$
θ	0.0430	0.0302	0.0566	0.0528
θ, r	0.0363	0.0165	0.0487	0.0460
θ, r, Δh	0.0362	0.0169	0.0487	0.0457
θ, r, Δh, Δt_c	0.0363	0.0166	0.0487	0.0459
θ, r, Δh, $\ln A$	0.0359	0.0032	0.0490	0.0457
θ, r, Δh, Δt_c, $\ln A$	0.0363	0.0011	0.0488	0.0455
θ, E	0.0364	0.0161	0.0489	0.0462
θ, E, r	0.0362	0.0158	0.0486	0.0460
θ, E, r, Δh	0.0361	0.0160	0.0485	0.0458
θ, E, r, Δh, Δt_c	0.0360	0.0140	0.0484	0.0457
θ, E, r, Δh, Δt_hs	0.0362	0.0159	0.0486	0.0458
θ, E, r, Δh, Δt_sp	0.0360	0.0143	0.0485	0.0456
θ, E, r, Δh, Δt_sp, S^r	0.0361	0.0148	0.0486	0.0458
θ, E, r, Δh, $\ln A$	0.0355	0.0006	0.0486	0.0453

Table 5.6.: Evaluation metrics for different meta input feature combinations, aggregated as the mean of the metrics of five CNNs per feature set. This time the input features are taken to be the MC instead of the reconstructed variables. Instead of the total signal traces, the sum of all sub-component traces is used which does not exhibit baseline fluctuations.

Meta Features	$\langle\lvert\Delta f_\mu\rvert\rangle$	$\langle\Delta f_\mu\rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_\mathrm{p}$	$\sigma(\Delta f_\mu)_\mathrm{Fe}$
θ_MC, E_MC	0.0378	0.0270	0.0504	0.0467
θ_MC, E_MC, r_MC	0.0350	0.0108	0.0472	0.0449
θ_MC, E_MC, r_MC, S^r	0.0350	0.0107	0.0472	0.0447
θ_MC, E_MC, r_MC, $\ln A$	0.0345	0.0006	0.0471	0.0442
θ_MC, E_MC, r_MC, S^r, $\ln A$	0.0344	0.0002	0.0469	0.0442

Table 5.7.: Evaluation metrics for different detector input combinations, aggregated as the mean of the metrics of five CNNs per feature set.

Detector	$\langle \lvert\Delta f_\mu\rvert \rangle$	$\langle \Delta f_\mu \rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_\mathrm{p}$	$\sigma(\Delta f_\mu)_\mathrm{Fe}$
SSD	0.0443	0.0340	0.0606	0.0538
WCD	0.0364	0.0163	0.0490	0.0460
WCD + SSD	0.0359	0.0139	0.0484	0.0457
WCD + SSD sep	0.0359	0.0137	0.0484	0.0457

different input configurations. In the first, the WCD and SSD traces are stacked depth wise with the dimensionality (400,2). One drawback of this method lies in the timing mismatch due to inherently different detector responses and signal decays. One attempt to solve this problem, is given by the second input configuration, where a separate second *Trace Analyzer* for the SSD trace is added to network. In this setup the traces are convolved separately and have an unique *Trace Analyzer* specialized on the feature extraction of each detector. The meta input feature set is fixed to θ, E, r, Δh and Δt_c.
The performance of the different models in terms of their binned bias and resolution, with respect to various shower and station variables, is shown in Figs. 5.16 and 5.17. The biases of all detector configurations are negligible small across the provided shower and station variables. In case the network is trained with the SSD trace, the resolution drops significantly for high energetic showers and almost across all inclinations. Especially far from the shower core $r \geq 1\,\mathrm{km}$, the resolution becomes worse which could be attributed to the limited cross sectional area of the SSDs, but this is however not reflected in the zenith angle behavior. Another problem probably also lies in the lack of causality, as particle traversing one detector are not necessarily traversing the other one and the particle density can exhibit fluctuations and absorptions locally and thus, deposit different signal sizes in both detectors. It is expected from (Sec. 5.1.1), that including the SSD trace will allow for the disentanglement of the electromagnetic and muonic sub-components and thus improve the performance. However, adding the SSD in any of the described ways, does not impact the resolution. Only in the case of low-energy showers a minor improvement becomes visible. But as the global primary-dependent bias (Table 5.7) decreases by including the SSD, a more detailed study on the primary-dependent bias and resolution with respect to proton and iron-induced air-showers is done (Figs. 5.18 and 5.19). For E, θ and S_tot the iron bias is reduced by a constant of $\Delta\langle\Delta f_\mu\rangle_\mathrm{Fe} \approx 2.5{\times}10^{-3}$ by including the SSD trace, while the resolutions are unaffected.
It is hard to draw a conclusion why the SSD does not improve the performance significantly, but it is likely that the information content that is hidden in the temporal signal of the WCD, together with the information through the meta variables is exhausted for the prediction of the muon fraction and the SSD only holds redundant information.

Gain Systematics

In this analysis we explored, if the CNN exhibits any systematics which are caused by the inherently different systematics of High Gain (HG) or Low Gain (LG) traces. Again, the

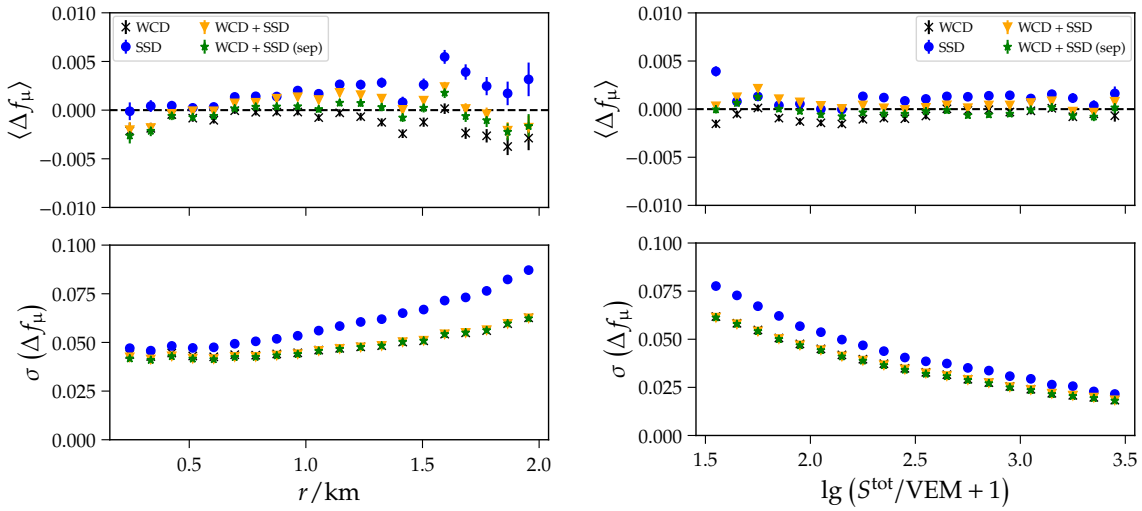


Figure 5.16.: Bias (*top*) and resolution (*bottom*) of CNN ensembles for different detector input configurations as a function of shower plane distance r (*left*) and logarithmic total signal $\lg(S/\mathrm{VEM}+1)$ (*right*). For each unique detector combination 5 networks are trained and the mean of their prediction is used for the evaluation. Because we predict the muon fraction in the WCD it is not surprising that the ensembles that only receive the SSD trace as an input perform worse. However adding the signal trace of the SSD, either as an separate input with an second *TraceAnalyzer* or as a color channel of the WCD trace, does not improve the resolution significantly.

meta inputs are fixed to θ, E, r, Δh and Δt_c.
In a first study, the networks are trained either separately on the HG/LG traces or the network explicitly receives the binary variable *SdRecStation:IsHighGainSaturated* as an additional meta input feature. These two configurations are then compared to the network being trained simultaneously on both HG and LG traces. As there is no significant difference in terms of bias and resolution, the results are moved into the appendix (Figs. A.6 to A.7). One can conclude that the network is capable of correctly handling both HG and LG traces in its default configuration, even though it does not explicitly receive information about it. Another crosscheck to reveal possible different systematics of HG/LG traces is done by investigating the bias and resolution suspect to the two gain types. The results are shown in Figs. 5.21 to 5.20. As the networks do not have any significant bias over the binned variables, the network is capable of handling both gain types and confirms the previous study. In general, the resolution of the LG traces is better than the one of the HG traces. This is expected, as the LG traces are less noisy and therefore the CNN is able to extract more information from them. Further, the resolution of HG traces continuously matches the resolution of the LG traces with increasing signal size. This validates, that there are no systematic errors introduced by training the network simultaneously on both gain types.

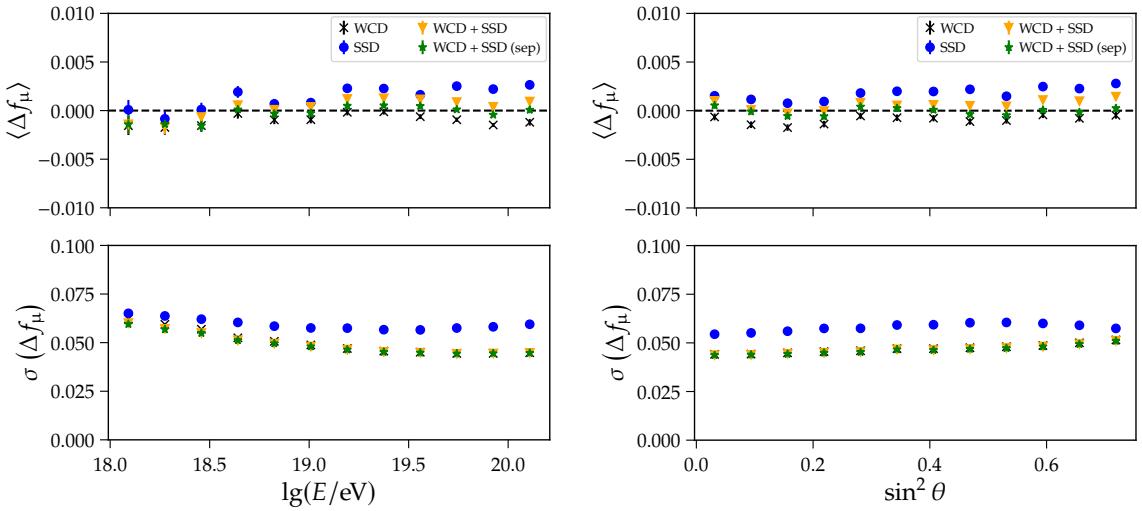


Figure 5.17.: Bias (*top*) and resolution (*bottom*) of CNN ensembles for different detector input configurations as a function of logarithmic energy $\lg E$ (*left*) and $\sin^2\theta$ (*right*). For each unique detector combination 5 networks are trained and the mean of their prediction is used for the evaluation. Because we predict the muon fraction in the WCD it is not surprising that the ensembles that only receive the SSD trace as an input perform worse. However adding the signal trace of the SSD, either as an separate input with an second *TraceAnalyzer* or as an color channel of the WCD trace, does not improve the resolution significantly.

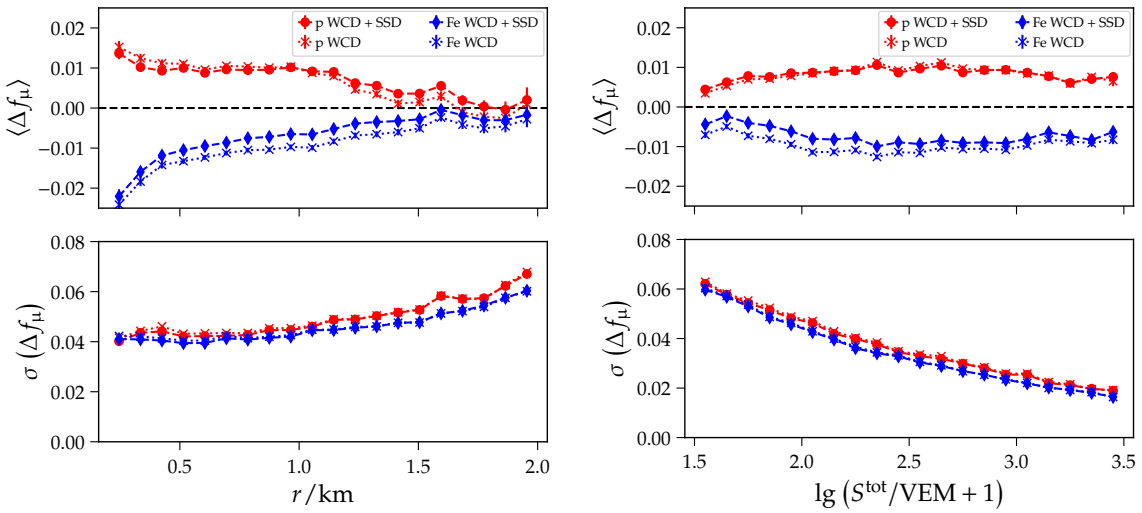


Figure 5.18.: Primary-dependent bias (*top*) and resolution (*bottom*) of CNN ensembles for different detector input configurations, as a function of shower plane distance r (*left*) and logarithmic total signal $\lg(S/\mathrm{VEM}+1)$ (*right*). For each unique detector combination, five networks are trained and the mean of their prediction is used for the evaluation. Including the SSD detector trace on top of the WCD does improve the primary-dependent bias a little, but for the resolution no significant improvement is observed.

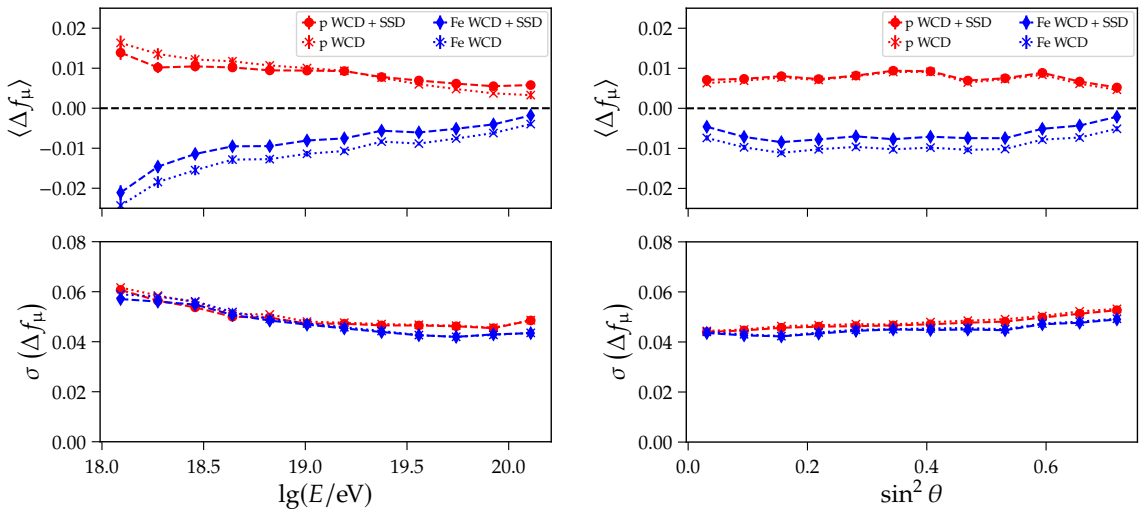


Figure 5.19.: Primary-dependent bias (*top*) and resolution (*bottom*) of CNN ensembles for different detector input configurations, as a function of logarithmic energy $\lg E$ (*left*) and $\sin^2\theta$ (*right*). For each unique detector combination, five networks are trained and the mean of their prediction is used for the evaluation. Including the SSD detector trace on top of the WCD does improve the primary-dependent bias a little, but for the resolution no significant improvement is observed.

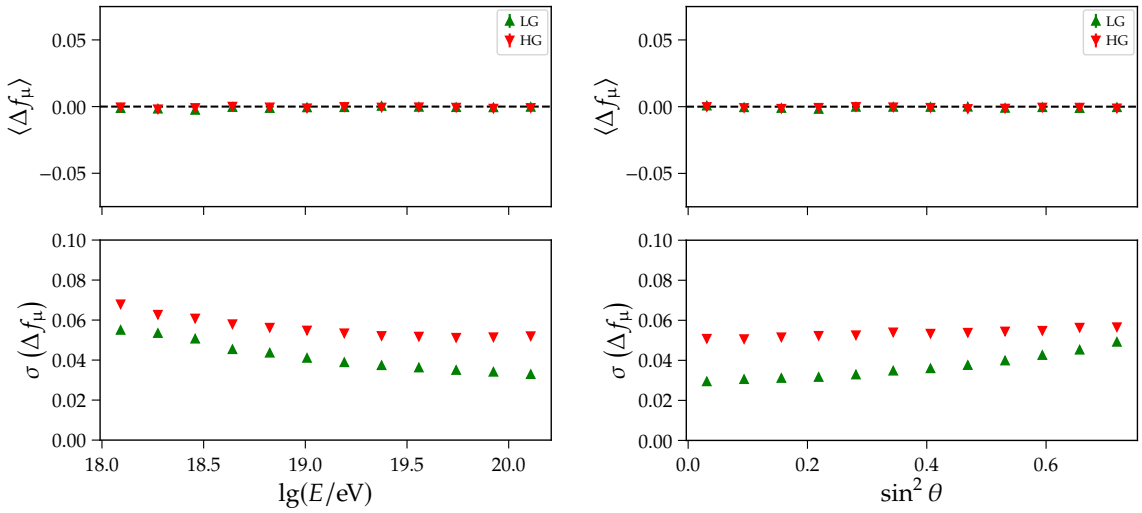


Figure 5.20.: Bias (*top*) and resolution (*bottom*) of the CNN network for LG (*green*) and HG (*red*) traces as a function of logarithmic energy $\lg E$ (*left*) and $\sin^2\theta$ (*right*). The predictions are given by the means of five CNN networks. The worse resolution of the HG traces is expected, as the LG traces are less noisy and therefore the CNN is able to extract more information from them.

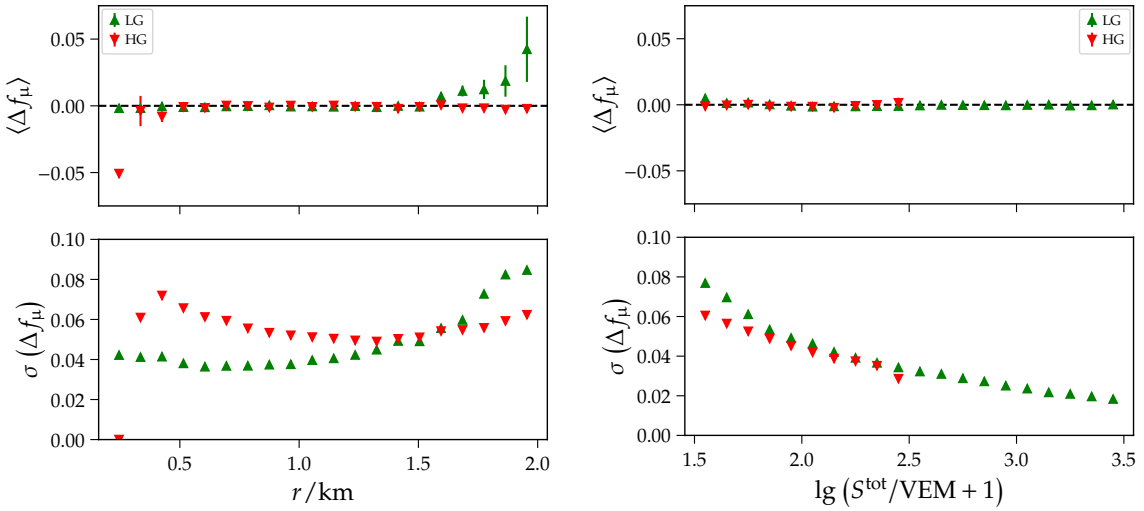


Figure 5.21.: Bias (*top*) and resolution (*bottom*) of the CNN network for LG (*green*) and HG (*red*) traces, as a function of shower plane distance r (*left*) and logarithmic total signal $\lg(S/\mathrm{VEM}+1)$ (*right*). The predictions are given by the means of five CNN networks. The matching resolution of HG and LG traces with increasing signal size validates, that there are no systematics introduced by training the network simultaniously on both gain types.

5.8.4. Composition Dependence

One of the main goals of this work is the construction of a composition-sensitive variable to discriminate on an event-to-event basis. Hence, the predictions of the network should not depend on the type of primary particle. In this section, it is investigated if the primary-dependent bias of the predictions of a NN can be reduced by directly penalizing it.

Composition dependent Loss

A second loss term $\mathcal{L}_\mathrm{c}$ is added to the loss $\mathcal{L}_\mathrm{MSE}$ with a weight λ_c, to penalize composition dependent biases as

$$\mathcal{L}_\mathrm{comp} = \mathcal{L}_\mathrm{MSE} + \lambda_\mathrm{c}\,\mathcal{L}_\mathrm{c} \tag{5.23}$$

$$\text{with} \qquad \mathcal{L}_\mathrm{c} = \frac{1}{4}\sum_{z}^{\mathrm{p,He,O,Fe}} \left[\frac{1}{N_z}\sum_{i=1}^{N_z}(y_i - \hat{y}_i)\right]^2 . \tag{5.24}$$

Where N_z is the number of elements in the current batch. The weight λ_c is scanned on a logarithmic scale and the biases of proton and iron induced showers are tracked during training of the CNN (Fig. 5.22). The meta-feature set used for this study is given through E, θ, r, Δh, and Δt_c. The primary dependent metrics (Table 5.8) are evaluated globally as well in dependence of shower and station variables (Figs. 5.23 and 5.24). In all cases a larger value of λ_c yields a smaller difference between proton and iron biases, but also decreases the resolution. In general, a well behaved flat and constant bias in every shower and station variable is preferred. Large weights $\lambda_\mathrm{c} \geq 10$ introduce random dependencies and up- and downward fluctuations in the biases. But for an equal weighting of the composition

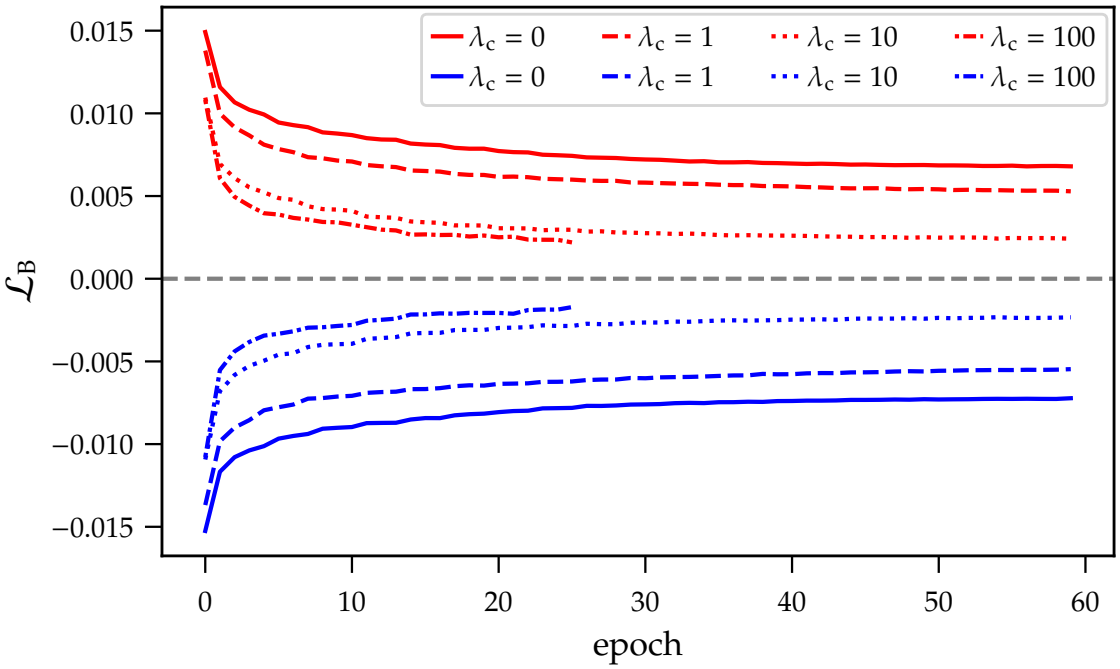


Figure 5.22.: Primary-dependent bias $\mathcal{L}_B$ of CNN predictions for stations triggered by showers induced by proton (*red*) and iron (*blue*) primaries for different weights λ_c, penalizing the composition loss. With increasing λ_c, the primary-dependent bias is reduced and thus the gap between proton and iron bias shrinks, but also leads to an earlier training stop.

loss $\lambda_c = 1$, the gap between the primary-dependent biases is reduced by 0.003, while the resolutions decrease only by around 0.0005 percentage-points when compared to $\lambda_c = 0$. Even though the biases remain flat and constant in the selected shower- and station variables, it is unclear if this holds for all variables. Other studies, such as [93], have shown that the composition dependent bias introduces unexpected dependencies in other variables. Therefore, the composition dependent loss is not used ($\lambda_c = 0$) in the following analysis and the possible reduction is dropped in favor of well behaved predictions.

Performance Evaluation of Leading Models

Since none of the various test yielded any significant improvement, it is from now on assumed that the chosen architecture is optimal to predict the muon fraction. Nevertheless, the network is configured to use the best performing configuration of each study. Namely, the CNN is trained with both traces of the SD being not normalized, and takes the meta feature set consisting of θ, E, r, Δh and Δt_c as inputs. To maximize the reachable performance, the maximum number of training epochs is increased to 100 epochs. The predictions of five networks out of ten are chosen by the best MAE on the test data set, averaged, and used for the future studies.
The performance of the CNN is evaluated in terms of correlation, bias and resolution plots

Table 5.8.: Evaluation metrics for different weights λ_c penalizing the composition loss, aggregated as the mean of the metrics of five CNNs per feature set. One can reduce the primary-dependent bias $\langle \Delta f_\mu \rangle_{p,Fe}$ arbitrarily by increasing λ_c at the cost of a worse resolution $\sigma(\Delta f_\mu)_p$ and $\sigma(\Delta f_\mu)_{Fe}$.

λ_c	$\langle \lvert \Delta f_\mu \rvert \rangle$	$\langle \Delta f_\mu \rangle_{p,Fe}$	$\sigma(\Delta f_\mu)_p$	$\sigma(\Delta f_\mu)_{Fe}$
0.01	0.0361	0.0138	0.0485	0.0458
0.1	0.0360	0.0138	0.0485	0.0458
0.0	0.0360	0.0141	0.0485	0.0458
1.0	0.0363	0.0110	0.0490	0.0462
10.0	0.0380	0.0056	0.0511	0.0482
100.0	0.0429	0.0046	0.0576	0.0542
1000.0	0.0428	0.0033	0.0573	0.0541

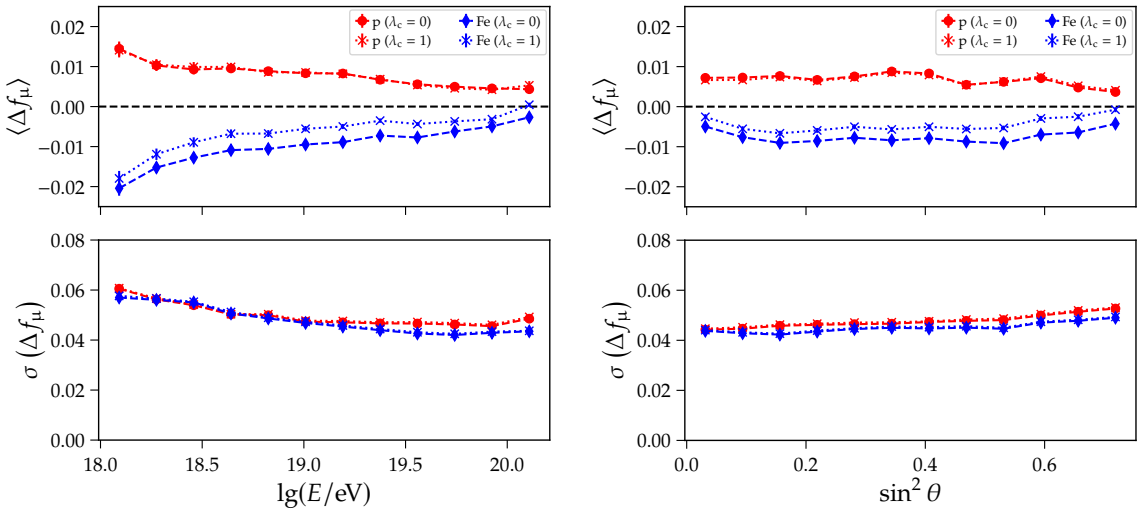


Figure 5.23.: Primary-dependent biases (*top*) and resolutions (*bottom*) of the CNN as a function of energy E (*left*) and zenith θ (*right*) for different weights λ_c penalizing the composition loss. The bias between proton (*red*) and iron (*blue*) is reduced by around 0.0003 for $\lambda_c = 1$, while no decrease in resolution is visible.

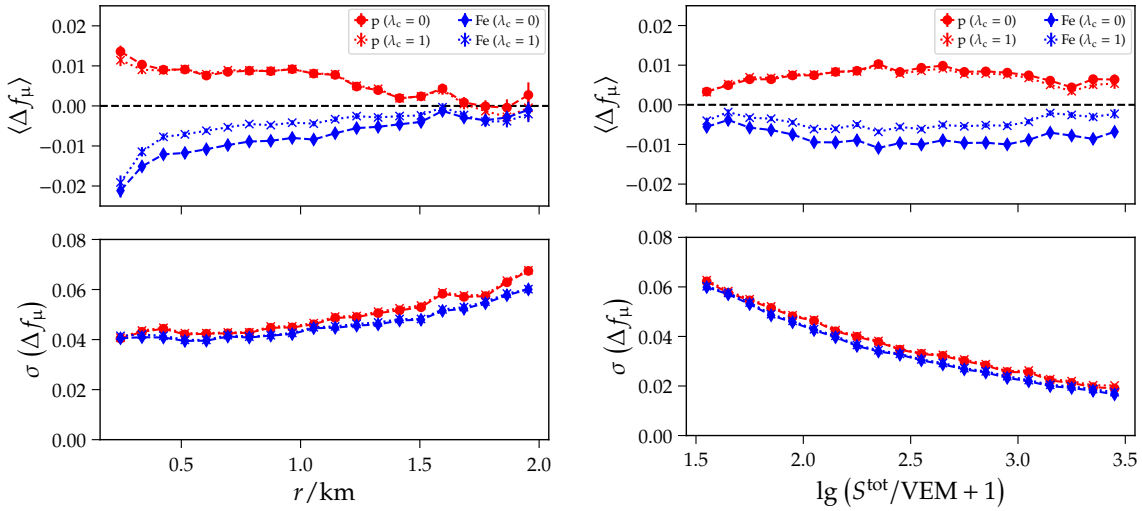


Figure 5.24.: Primary-dependent bias (*top*) and resolution (*bottom*) of the CNN as a function of shower plane distance r (*left*) and total signalS_{tot} (*right*) for different weights λ_c penalizing the composition loss. The bias between proton (*red*) and iron (*blue*) is reduced by around 0.0003 for $\lambda_c = 1$, while no decrease in resolution is visible.

in dependence of E, θ, r and S_{tot}. The global correlation of the CNN predictions is shown in Fig. 5.25 and the coefficient of determination amounts to $R^2 = 0.93$. The predictions lie on the diagonal and the spread of the residuals is equal to $\sigma(\Delta f_\mu) = 0.05$.
To highlight deviations from the diagonal in the correlation plot, each bin entry is normalized to the maximum number $N_{max}(f_\mu)$ of counts in each column, i.e. $N(f_\mu, \hat{f}_\mu)/N_{max}(f_\mu)$. The results for each primary are depicted in Fig. 5.26. In regions with low statistic the network predictions tend towards the mean value of f_μ, e.g., for high f_μ the network underpredicts the muon fraction. The primary-dependent bias gap between proton and iron decreases with increasing energy $\langle \Delta f_\mu \rangle_{p,Fe}(E \geq 10^{19}\,\text{eV}) \leq 0.02$ (Fig. A.9). At the same time the resolution of the CNN predictions increases $\sigma(\Delta f_\mu)_p(E \geq 10^{19}\,\text{eV}) \leq 0.05$. This could be attributed to increasing statistics in the data set or a decrease in the number of muons as the energy in the hadronic component decreases with energy. Bias and resolution tend to be almost constant over the whole zenith-angle range, except for small or large zenith angles. In this regions the bias between proton and iron-induced showers is reduced. It can be explained by the attenuation of the electromagnetic component, leading to an muon fraction close to one for large zenith angles, irrespective of the primary. A similar reasoning holds for the bias and resolution for high shower plane distances (Fig. A.8). The rapid increase in resolution with increasing signal size indicates that the network predictions are heavily influenced by the inherent statistical uncertainty. This can be quantified through the maximum error of the approximated error model (Eq. (5.9)) which shows a similar dependency. Iron has a slightly better resolution than the lighter particles, which also resembles the decreased statistical fluctuations of shower development due to the superposition principle. A more detailed analysis of the bias and resolution in dependence of two dimensional combinations in E, θ, and r is given in Figs. 5.27 to 5.30. While the bias per primary is almost constant in shower variables $\langle \Delta f_\mu \rangle \in [-0.01, 0.01]$, the resolution

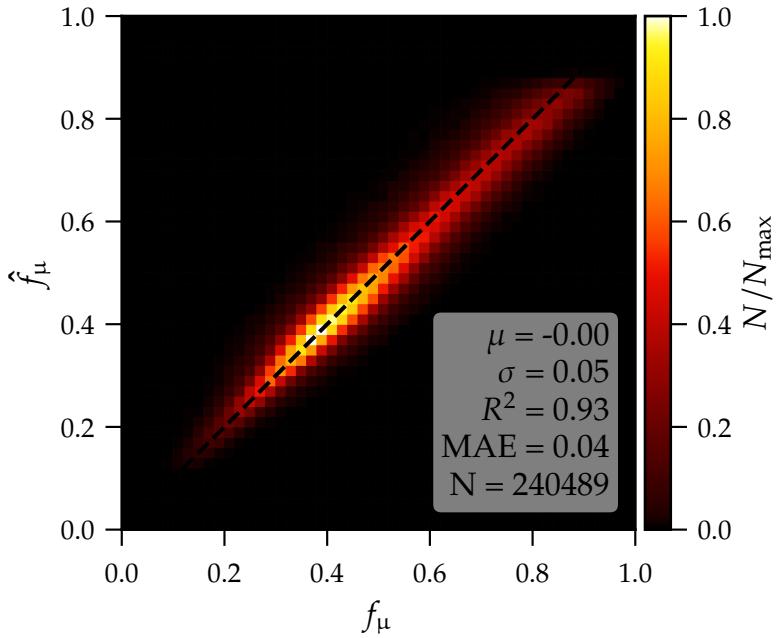


Figure 5.25.: Correlation of CNN predictions for all primaries, normalized to the maximum number counts of all bins $N_{\max}$.

drops significantly for low energetic and highly inclined showers $\sigma(\Delta f_\mu) \geq 0.06$. The network exhibits high primary-dependent biases $|\langle \Delta f_\mu \rangle| \geq 0.02$ very close to the shower core for highly inclined proton and iron showers. This can be attributed to limited statistics in this region of the phase space.

5.8.5. Mixture Density Network

As shown previously, the muon fraction exhibits large fluctuations, especially for small signal sizes. One architecture that qualifies the uncertainty of the NN prediction is given through Mixture Density Networks (MDNs) (Sec. 4.4). The CNN architecture can be modified by substituting the prediction unit with a Mixture Density Head (MDH), see Fig. C.4. It is assumed that the mixture is given by a single Gaussian ($\alpha = 1$). Hence, the network learns to predict the mean $\hat{\mu}(f_\mu) \equiv \hat{f}_\mu$ together with an uncertainty measure $\hat{\sigma}(f_\mu)$. To train this network, the negative log-likelihood of the Gaussian mixture is used as a loss function, see Eq. (4.4). An ensemble of ten networks is trained to gain insights into the spread of the predictions inside the ensemble $\sigma_{\text{ens}}(\hat{f}_\mu)$ and the error distribution of their predicted means $\langle \hat{f}_\mu - f_\mu \rangle$ (Fig. 5.31). As a cross-check, the predictions of an ensemble of ten CNNs, trained with the MSE loss, are provided. The mean and standard deviation of

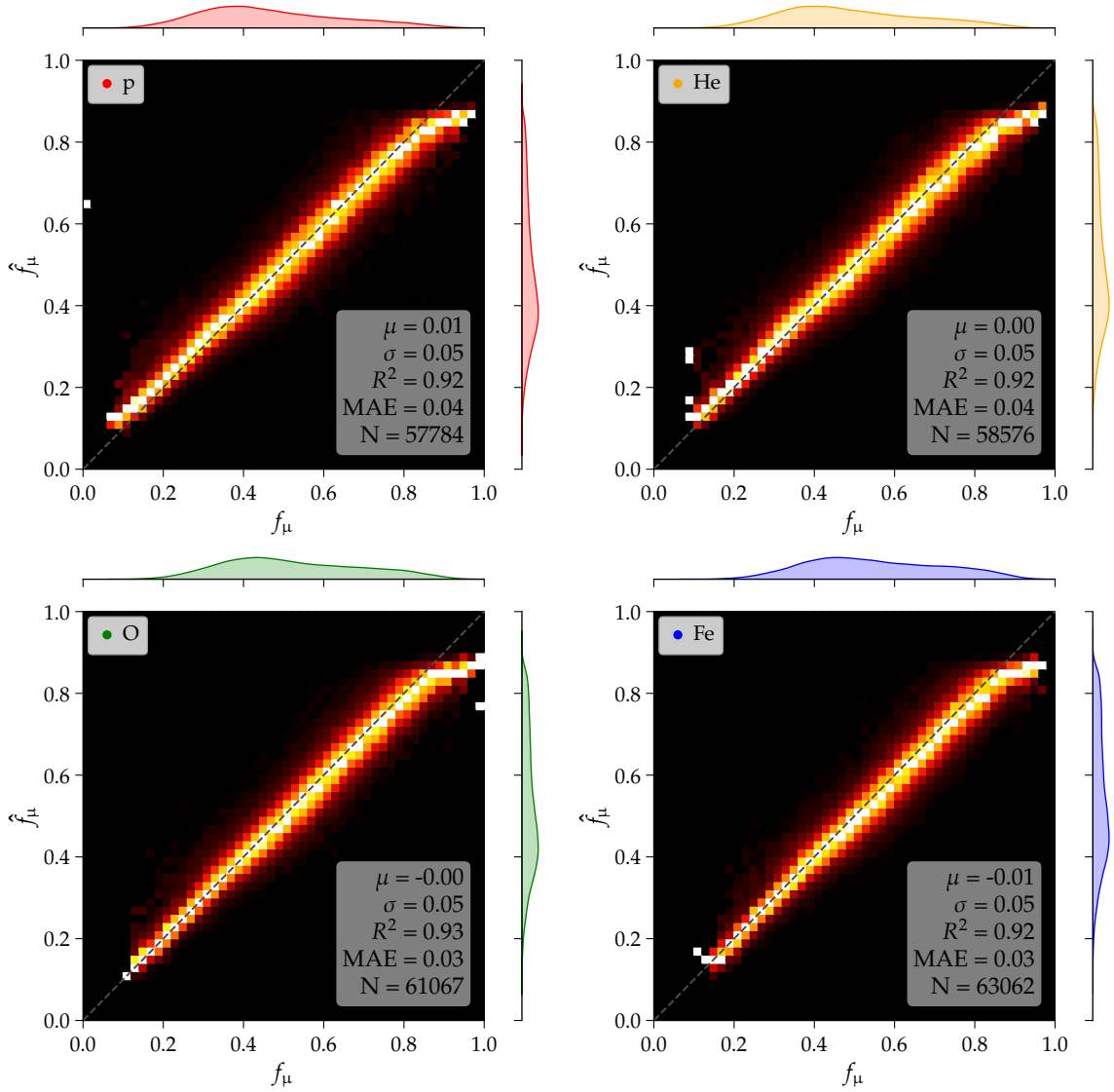


Figure 5.26.: Correlation Plots of CNN predictions for each primary, normalized to the maximum number $N_{\max}(f_\mu)$ of counts in each column. In regions of the phase space with only a few samples, for instance for large muon fractions, the DNN $\hat{f}_\mu$ underpredicts the fraction $f_\mu \geq 0.9$ in regions of the phase space that have limited statistics.

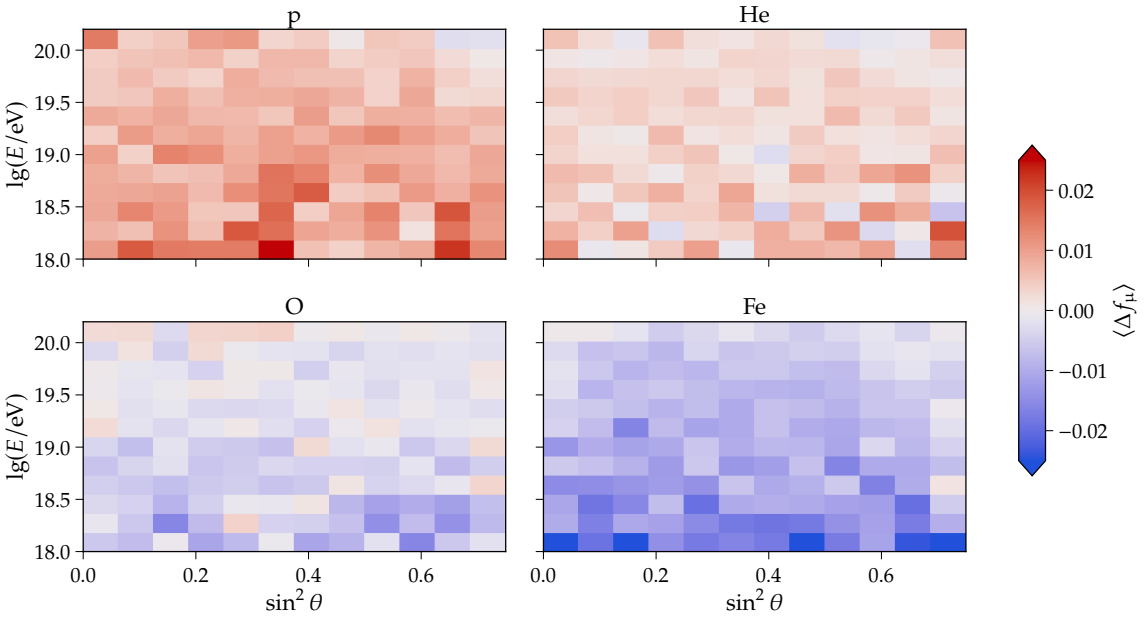


Figure 5.27.: Bias $\langle \Delta f_\mu \rangle$ of the DNN predictions for each primary, as a function of shower energy E and zenith angle θ. The primary-dependent bias is between 0.01 and −0.01 for proton and iron respectively for energies above 10^{19} eV.

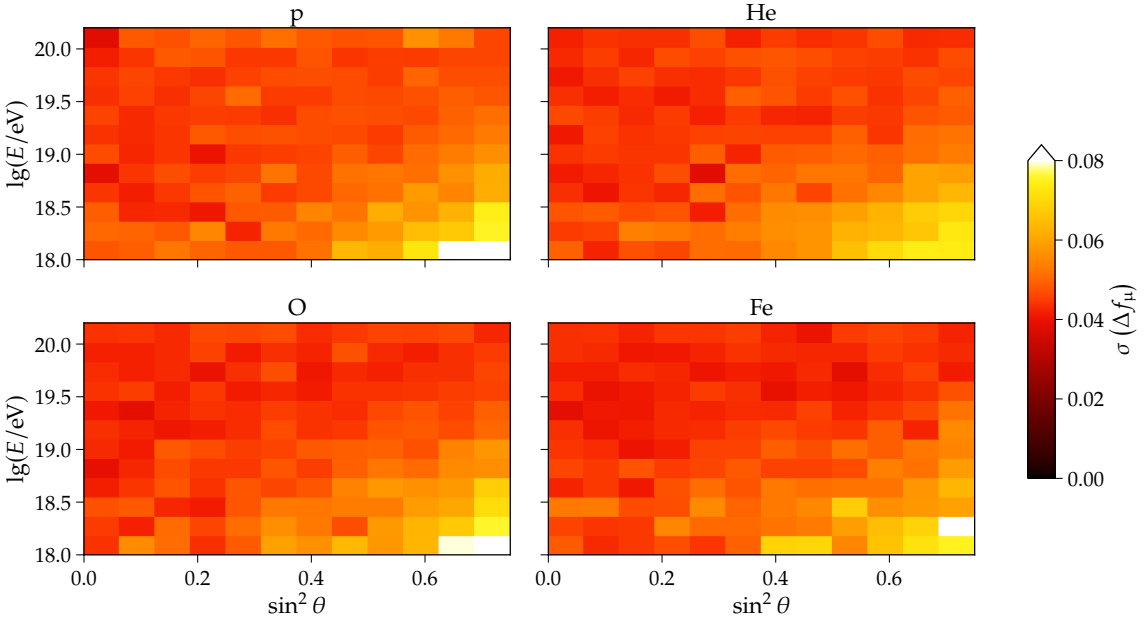


Figure 5.28.: Resolution $\sigma(\Delta f_\mu)$ of DNN predictions for each primary, as a function of shower energy E and zenith angle θ. The resolution is primary independent and constant around 0.05, except for low energetic and inclined showers.

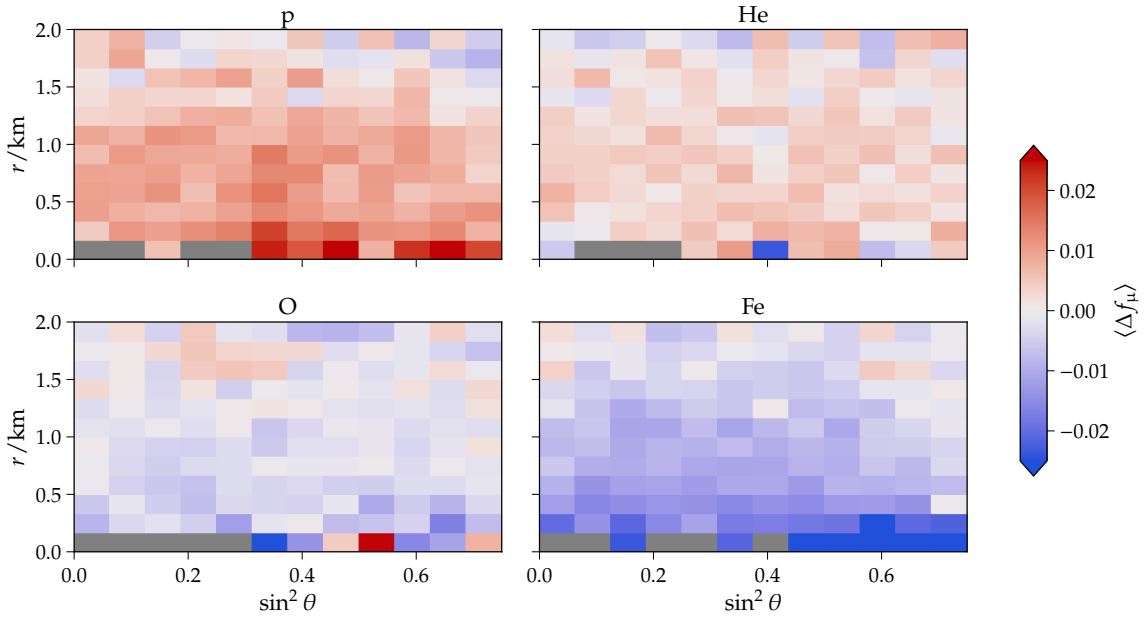


Figure 5.29.: Bias $\langle \Delta f_\mu \rangle$ of the DNN predictions for each primary, as a function of the shower zenith angle θ and the shower-plane distance r. The primary-dependent bias is flat in zenith, but its absolute value slightly increases for distances close to the shower core. This is attributed to limited number of stations or even missing data (*gray*) in vicinity of the shower core.

the MDN ensemble are given by [99]:

$$\hat{\mu}_{\text{ens}} = \frac{1}{10} \sum_{j}^{10} \hat{\mu}_j \tag{5.25}$$

$$\hat{\sigma}_{\text{ens}} = \sqrt{\frac{1}{10} \sum_{j}^{10} \left(\hat{\sigma}_j^2 + \hat{\mu}_j^2 \right) - \hat{\mu}_{\text{ens}}^2}. \tag{5.26}$$

The spread inside the MDN ensemble $\langle \sigma_{\text{ens}}(\hat{f}_\mu) \rangle = 0.008$ is slightly larger than for the reference ensemble $\langle \sigma_{\text{ens}}(\hat{f}_\mu) \rangle = 0.007$. An additional crosscheck to reveal potential systematics is done by comparing the distribution of the average residuals $\langle \hat{f}_\mu - f_\mu \rangle$ for both ensembles. In both cases they are well centered around zero, thus are bias-less and do exhibit the same width. Hence, the MDH ensemble is able to reproduce the mean of the f_μ distribution.
To investigate the quality of the predicted uncertainties, they are compared to the true absolute error of the predictions $|\hat{f}_\mu - f_\mu|$ (Fig. 5.32, *right*). If they corretly describe the uncertainty, the actual error divided by the predicted standard deviation should follow a standard normal distribution $z \overset{!}{\sim} \mathcal{N}(\mu = 0, \sigma = 1)$, which is also denoted as the z-

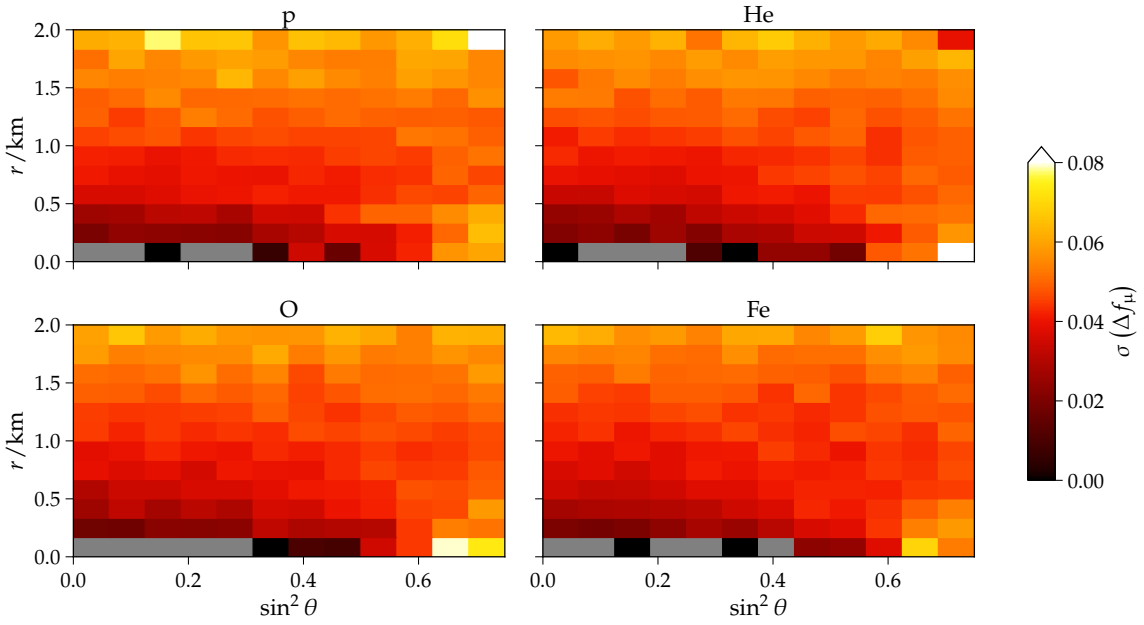


Figure 5.30.: Resolution $\sigma(\Delta f_\mu)$ of the DNN predictions for each primary, as a function of the shower zenith angle θ and the shower-plane distance r. The resolution is better than 0.05 except for distances closer than ≤ 200 m or stations far away from the shower core ≥ 1.8 km. This is attributed to limited number of stations or even missing data (*gray*) in vicinity of the shower core.

test (Fig. 5.32, *left*). The variable z is given by

$$z = \frac{f_\mu - \hat{\mu}_{\text{ens}}(f_\mu)}{\hat{\sigma}_{\text{ens}}(f_\mu)}. \tag{5.27}$$

For the MDN ensemble the mean of the z-distribution is given by $\mu(z) = -0.0741$ and its standard deviation by $\sigma(z) = 0.9645$. Especially for low (muon) signals the underlying statistics of the muon fraction are probably not Gaussian and thus introduce a systematic error. In a result, the MDN architecture will not be used for the final analysis as the predicted uncertainties are not reliable. However, the MDN architecture is a promising approach to estimate the uncertainty of the predictions and could be further investigated in the future with a correct underlying probability distribution of the muon signal fraction.

5.9. Hadronic Interaction Models

The current simulations of CR air showers consequently underestimate the muon content that is measured by Auger. The discrepancy of this *muon puzzle* (Sec. 3.5.1) depends on the hadronic interaction model that is used to simulate the air showers. Up to this point, the QGSJetII-04 model has been used explicitly to train and evaluate the NNs. In this section, the systematic impact of the hadronic interaction model on the predictions of the CNN is

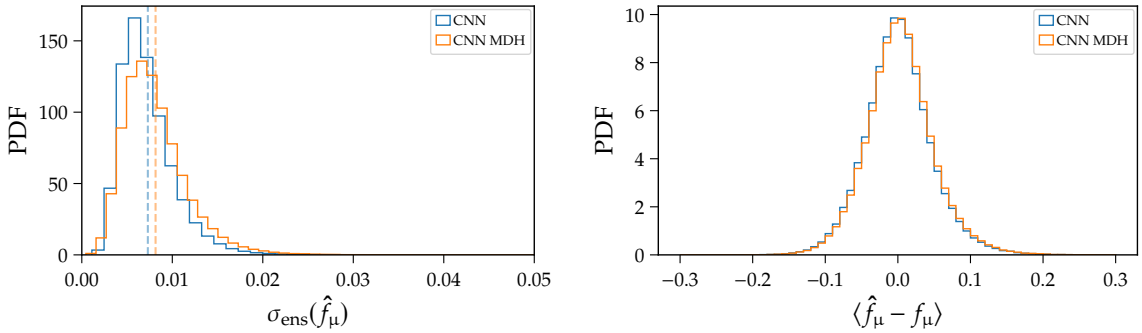


Figure 5.31.: Spread of the predictions inside each ensemble (*left*) and distribution of the mean ensemble residuals (*right*). As the distributions for both ensembles are very similar, the MDN ensemble is able to reproduce the mean of the f_μ distribution.

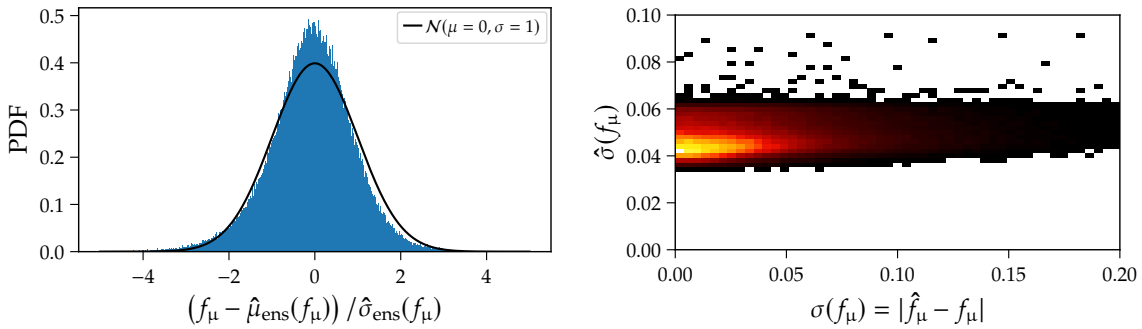


Figure 5.32.: Z-test (*left*) of the predicted mean and standard deviation of the MDH ensemble and correlation plot of the predicted uncertainties and the absolute residuals of the predicted means (*right*).

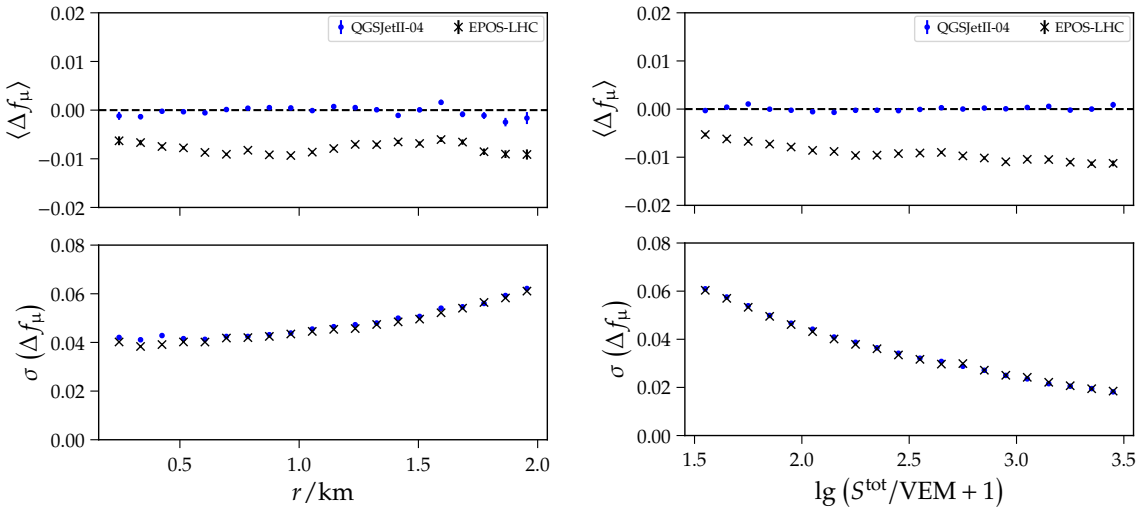


Figure 5.33.: Bias (*top*) and resolution (*bottom*) of the CNN network, trained on QGSJetII-04, for EPOS-LHC and QGSJetII-04 hadronic interaction models as a function of shower plane distance r (*left*) and logarithmic total signal $\lg(S/\mathrm{VEM} + 1)$ (*right*).

analyzed. Therefore, the data sets with EPOS-LHC and QGSJetII-04 simulations (Table 5.1) are used. The EPOS-LHC interaction model currently is favored by measurements [100]. Since the CNN models have been trained on QGSJetII-04 data, the systematic differences are explored by comparing their predictions on QGSJetII-04 with their predictions on EPOS-LHC data.
The results in terms of biases and resolutions are shown in Figs. 5.33 to 5.34. The CNN ensemble of Sec. 5.8.4 consistently underestimates the muon fraction of the EPOS-LHC simulations by -0.008 percentage points globally (Table 5.9). This is reflected in an almost constant bias shift with respect to r, E and θ. However, the bias exhibits a non-constant behavior in dependency of the total signal size. Its absolute value increases with higher total signal. The overall negative bias is caused by a higher muon content of showers simulated with the EPOS-LHC model in comparison to the QGSJetII-04 model. Even though, the obtained resolutions are almost identical for both interaction models.
Further insights into the different muon contents are investigated by the primary-dependent bias and resolution with respect to each hadronic interaction model (Figs. 5.35 to 5.36). The biases of proton and iron induced showers are shifted consistently by the same amount across all shower parameters. Only for large distances, the biases of the EPOS-LHC data behave differently, as the proton bias lies not on top of the iron bias anymore, as it is the case for the QGSJetII-04 data.
The average muon fraction between both data sets only changed slightly $\langle f_\mu \rangle_{\text{QGSJetII-04}} - \langle f_\mu \rangle_{\text{EPOS-LHC}} = -0.002$ and is below ≤ 0.02 per primary and per logarithmic energy interval. Thus, the muon fraction itself is little dependent on the hadronic interaction models, as a higher muon signal implies a higher total signal while their ratios stay the same.

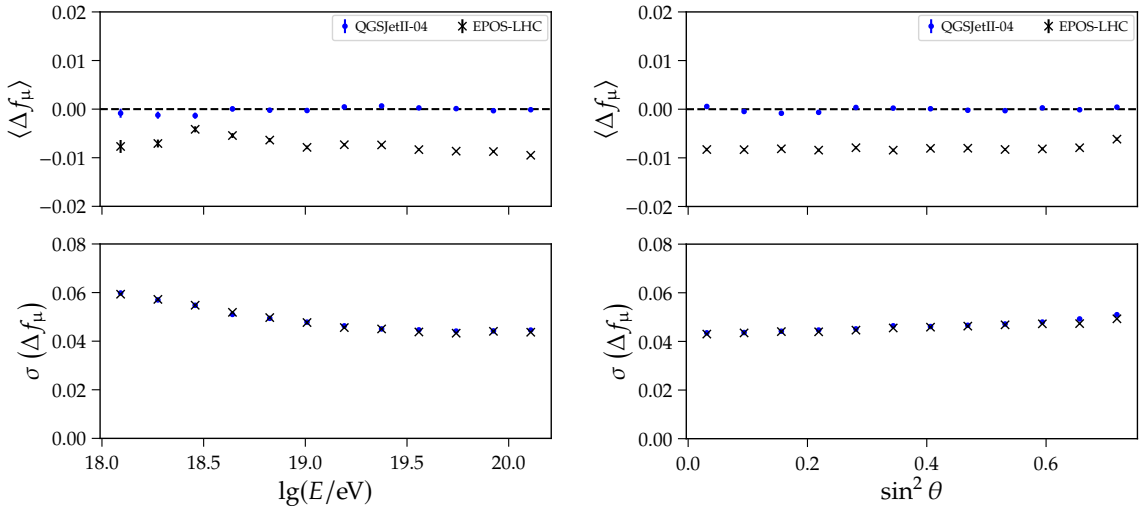


Figure 5.34.: Bias (*top*) and resolution (*bottom*) of the CNN network, trained on QGSJetII-04, for EPOS-LHC and QGSJetII-04 hadronic interaction models as a function of logarithmic energy lg E (*left*) and $\sin^2\theta$ (*right*).

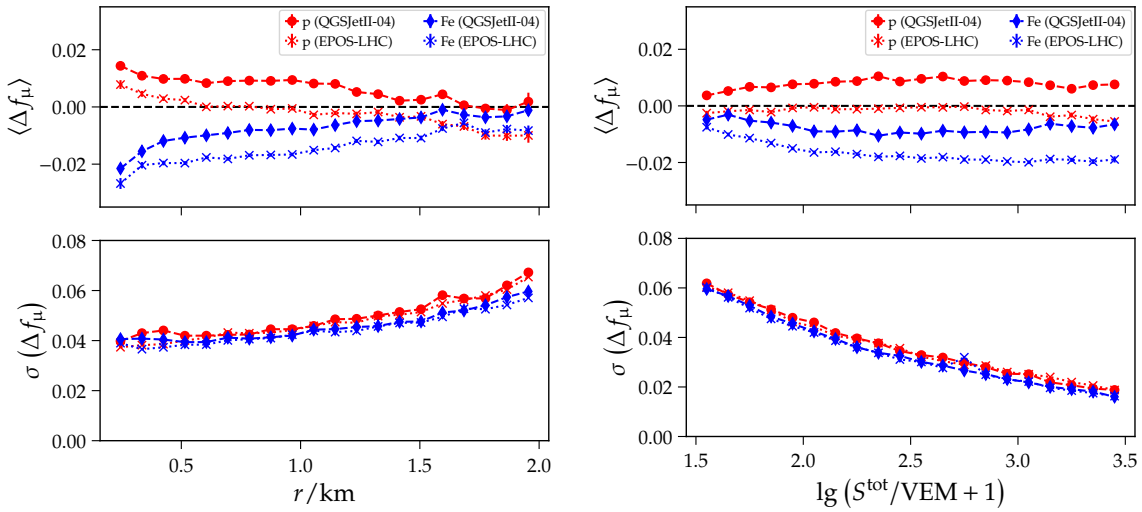


Figure 5.35.: Primary-dependent bias (*top*) and resolution (*bottom*) of the CNN network, trained on QGSJetII-04, for EPOS-LHC and QGSJetII-04 hadronic interaction models as a function of shower plane distance r (*left*) and logarithmic total signal lg(S/VEM+1) (*right*).

Table 5.9.: Evaluation metrics for a CNN ensemble on QGSJetII-04 and EPOS-LHC simulation data. The CNNs models are trained only on QGSJetII-04 simulation data.

Data set	$\langle \lvert \Delta f_\mu \rvert \rangle$	$\langle \Delta f_\mu \rangle$	$\langle \Delta f_\mu \rangle_{\mathrm{p}}$	$\langle \Delta f_\mu \rangle_{\mathrm{Fe}}$	$\sigma(\Delta f_\mu)_{\mathrm{p}}$	$\sigma(\Delta f_\mu)_{\mathrm{Fe}}$
QGSJetII-04	0.0354	0.0000	0.0070	-0.0070	0.0477	0.0450
EPOS-LHC	0.0357	-0.0080	-0.0017	-0.0144	0.0469	0.0445

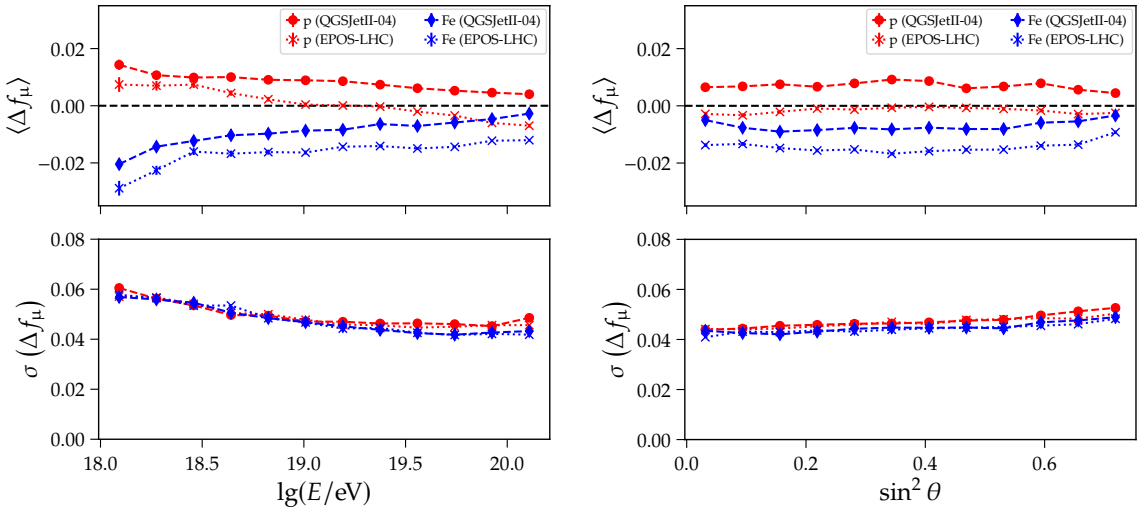


Figure 5.36.: Primary-dependent bias (*top*) and resolution (*bottom*) of the CNN network, trained on QGSJetII-04, for EPOS-LHC and QGSJetII-04 hadronic interaction models as a function of logarithmic energy $\lg E$ (*left*) and $\sin^2\theta$ (*right*).

5.10. Prediction of the Primary

To estimate the amount of available information related to the primary mass on station-level, the output of the CNN is switched to predict $\ln A$, or to predict the corresponding class of the primary. This study also serves as a reference for the mass dependent estimators that are derived on an event level. Ten CNNs models are trained and the five best performing ones on the test set are used for the predictions. In this section the whole data set, inclusive the training data, is used for the evaluation to increase the statistics in the energy bins.
The ReLU activation function in the output node is removed to allow the network for an unconstrained prediction of the mean masses. For instance, the distribution of the predicted proton mass would be shifted towards higher masses if $\ln A \geq 0$ is enforced. Despite this change, the training procedure stays the same as before. The distribution of the predicted $\ln A$ per station is shown in Fig. 5.37 on the *left* side. The separation between predicted proton and iron masses, given by the merit factor amounts to $\mathrm{MF}_{\mathrm{p,Fe}}(\ln A) = 1.50$. The mass centers of the predicted primary distributions are shifted towards the global mean. The distributions in different reconstructed-energy intervals is given in Fig. 5.38, where merit factors of up to $\mathrm{MF}_{\mathrm{p,Fe}}(\ln A) = 1.65$ are reached.
To adapt the network for the classification, the four primaries are one-hot encoded, so each primary is given by a one at the correspond class index: i.e. for iron [0,0,0,1] (Table 5.10). In order to transform the predictions of the network into a corresponding probability per class, a softmax[9] activation function is used. As a training loss, the categorical cross-entropy $\mathcal{L}_{\mathrm{CE}}$[10] is used. To evaluate the performance of a classification model, its most

[9] $\sigma(z_j) = \frac{\exp z_j}{\sum_{k=1}^{K=4} \exp z_k}$

[10] $\mathcal{L}_{\mathrm{CE}} = -\sum_{k=1}^{K=4} y_k \lg \hat{y}_k$

Table 5.10.: One-hot encoding of the primaries for classification.

primary	one-hot encoding
p	[1, 0, 0, 0]
He	[0, 1, 0, 0]
O	[0, 0, 1, 0]
Fe	[0, 0, 0, 1]

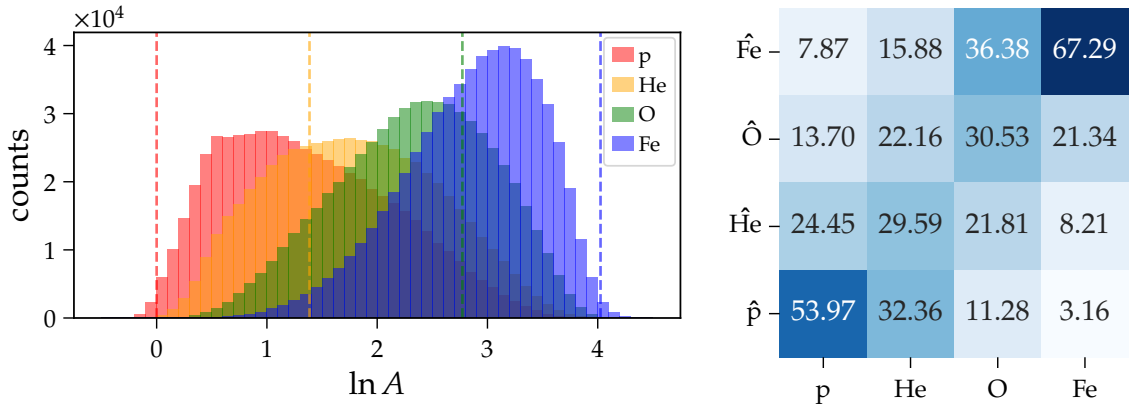


Figure 5.37.: Histogram of predicted logarithmic mass $\ln A$ (*left*) and confusion matrix for the different primaries, normalized to the true class (*right*) for each station. In case of the regression, a global merit factor of $\mathrm{MF}_{\mathrm{p,Fe}}(\ln A) = 1.50$ is achieved.

probable predicted class is compared to the truth class label in a confusion matrix. The confusion matrix is normalized column-wise to the total number of true labels per class representing the true/false positive/negative rates. In case of an ensemble classifier, a majority vote on their predictions is done. The baseline for the classification is given by a predictor that would randomly predict each class label with a probability of 25%. Hence, the classification network outperforms the baseline. The predicted iron classes are correctly classified in 67 out of 100 cases and only misclassified as protons in 3 out of 100 cases.

The logarithmic-mass predictions of the CNN on a single station level show, that the network is capable of deducing significant, composition-related information from the given detector traces and the meta data. Especially the obtained merit factor of the regression provides a cross-check to compare the mass-sensitive estimators obtained on an event level.

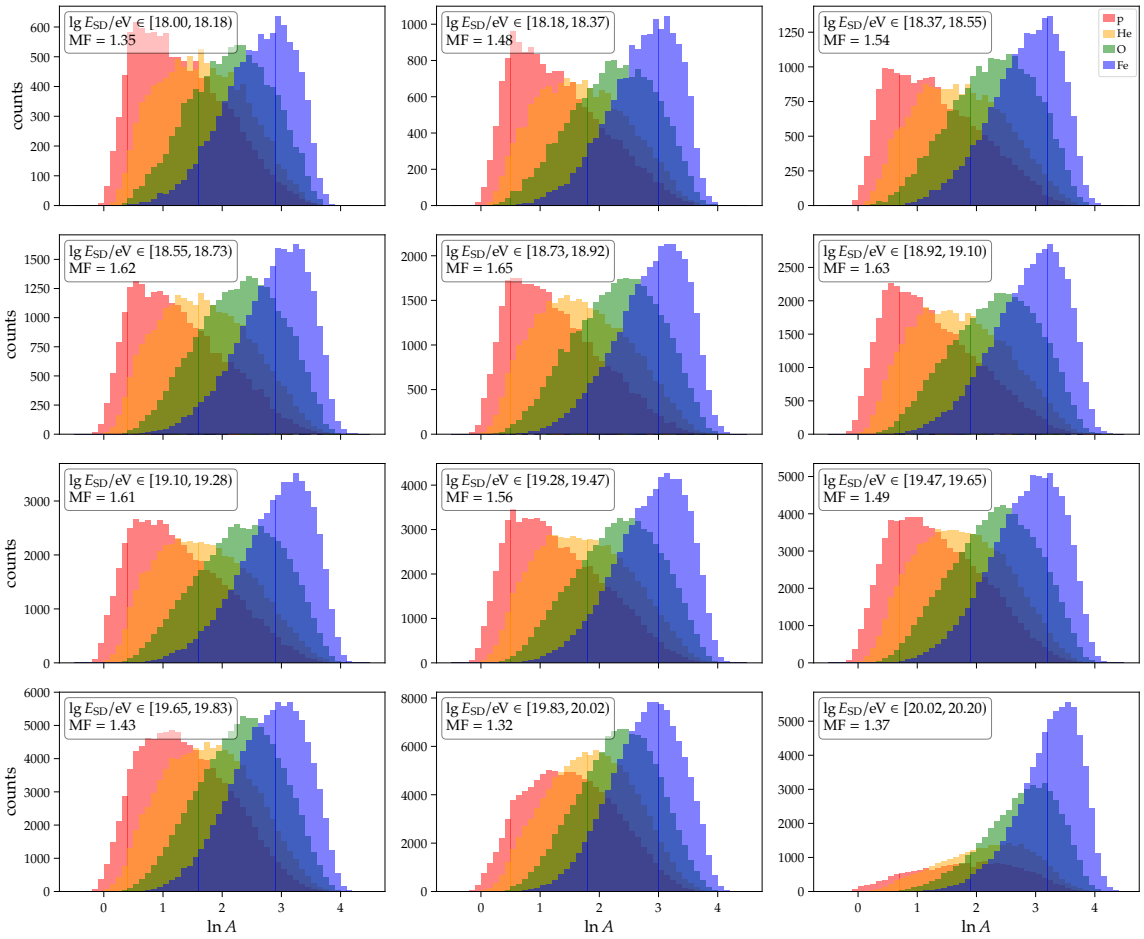


Figure 5.38.: Merit factor MF for the predicted $\ln A$ of the CNN on a station level, in dependency of different reconstructed energy intervals E_{SD}. A merit factor of up to $\mathrm{MF}(\ln A) = 1.65$ is obtained.

6. Reconstruction of a Mass-sensitive Estimator on Shower Level

Deriving a mass sensitive estimator for the muon signal fraction on event-level is non-trivial. Since the electromagnetic component is stronger attenuated than the muon component for more inclined showers, the azimuthal symmetry assumed in the SD-LDF fit is broken in case of the muon fraction. In the following, two methods to derive a mass sensitive estimator are explored and discussed in detail. In Sec. 6.1, an average muon fraction for each shower is estimated and then corrected for zenith and energy dependencies. In Sec. 6.2, a globally fit parametric model is scaled to the CNN predictions by fitting the logarithmic mass.

6.1. Lateral Fraction Estimator

As the muon fraction is introduced and motivated as a recent quantity in Auger and this thesis, still there exists no sophisticated theory describing the expected lateral muon signal fraction. A correct theoretical treatment is out of scope for this thesis, and thus the LFF is approximated by heuristically chosen logistic functions. The functional form of these muon LFFs is motivated by the following considerations:

1. $f_\mu(r) \in [0, 1]$. Therefore the function $f_\mu(r)$ needs to be bounded as well.
2. $f_\mu(r)$ should monotonically increase with r, as the attenuation for the electromagnetic component is stronger than for the muonic component.
3. $f_\mu(r = 0) \neq 0$, since for inclined showers the lateral fraction is expected to be constant.

These considerations are reflected in the following choice of functions under the transformation $f_\mu(r) \rightarrow c f_\mu(a(r - r_0))$:

$$f_\mu(r) = \begin{cases} \frac{\tanh(r)+1}{2} \\ \frac{2}{\pi}\arctan(r) \\ \frac{x}{\sqrt{1+x^2}} \\ \frac{x}{1+|x|} \end{cases} \tag{6.1}$$

An illustration of the functions is depicted in Fig. A.10, where each function is normalized to its corresponding derivative at $x = 0$ to reveal their inherently different curvature.
To find the parameters a, c, r_0, a two-step process was used. First, the functions have been

fitted in binned quantities of $\lg E_{\mathrm{MC}}$, $\sin^2\theta_{\mathrm{MC}}$ and $\cos\psi_{\mathrm{sp}}$ (Fig. A.11). Then, the resulting parameters itself have been re-parametrized by polynomial functions of the form:

$$a(E_{\mathrm{MC}}, \theta_{\mathrm{MC}}, \psi_{\mathrm{sp}}) \rightarrow \sum_k \xi_k (\sin^2\theta_{\mathrm{MC}})^k \sum_j \eta_j (\lg E_{\mathrm{MC}})^j \sum_l \rho_l (\cos\psi_{\mathrm{sp}})^l, \tag{6.2}$$

$$c(E_{\mathrm{MC}}, \theta_{\mathrm{MC}}, \psi_{\mathrm{sp}}) \rightarrow \sum_k \xi_k (\sin^2\theta_{\mathrm{MC}})^k \sum_j \eta_j (\lg E_{\mathrm{MC}})^j \sum_l \rho_l (\cos\psi_{\mathrm{sp}})^l, \tag{6.3}$$

$$r_0(E_{\mathrm{MC}}, \theta_{\mathrm{MC}}, \psi_{\mathrm{sp}}) \rightarrow \sum_k \xi_k (\sin^2\theta_{\mathrm{MC}})^k \sum_j \eta_j (\lg E_{\mathrm{MC}})^j \sum_l \rho_l (\cos\psi_{\mathrm{sp}})^l. \tag{6.4}$$

However, a globally performant re-parametrization could not be derived. As shown previously, all LFF perform equally because only a fraction of the functional shape is utilized, a polynomial model is equally suited as a potentially LFF. Thus the attempt of the global LFF re-parametrization is dropped in favor of a global polynomial model.

6.1.1. Up-/Downstream Correction

The major problem that arises when fitting the lateral fraction $f_\mu(r)$ is the up- and downstream dependence which breaks azimuthal symmetry. This dependency is caused by the different attenuation of the electromagnetic and muonic component. Therefore, its severity increases with increasing zenith angle and shower plane distance (Fig. 6.1). In first order the upstream-downstream dependency of f_μ, binned in E and θ, can be described by a linear function. To parametrize the up- and downstream dependence, three station level-variables have been identified: the shower plane azimuth $\cos(\psi_{\mathrm{sp}})$, the combined angle $\cos(\psi_{\mathrm{sp}})\tan(\theta)$, and the perpendicular distance to the shower plane at ground $\Delta h = r\cos(\psi_{\mathrm{sp}})\tan(\theta)$. As shown in (Fig. 6.2), all of them are equally suitable. However, due to the inherent relationship of the traveled path and the attenuation, Δh is chosen to be the natural complementary variable. By fitting the relation

$$f_\mu^{(i)}(r, \Delta h) = a_0 + a_1 \cdot \Delta h^{(i)}, \tag{6.5}$$

the $f_\mu^{(i)}(r, \Delta h)$ of each station i can be projected to a fraction that would occur at $f_\mu(r, \Delta h = 0)^{(i)}$. This is exemplary done by using the slope a_1 of the fit for the projection of the individual muon signal fractions to $\Delta h = 0$

$$f'_\mu(r, \Delta h = 0) = f_\mu^{(i)}(r, \Delta h) - a_1 \cdot \Delta h^{(i)}. \tag{6.6}$$

The muon fractions very close to or very far from the shower core are less dependent on up- and downstream than in the intermediate range, which is reflected in the change of the slope for different distance fits. This is especially prevalent for highly inclined showers (Fig. A.13), where Eq. (6.5) is not capable to capture the dependency across all distances with only one unique slope parameter. As a consequence, in a postceeding study, the correction for the up- and downstream dependency has to be refined. However, after this linear correction the upstream-downstream dependency is not as prevalent any more and the muon fraction for greater distances becomes far more narrow on a single shower

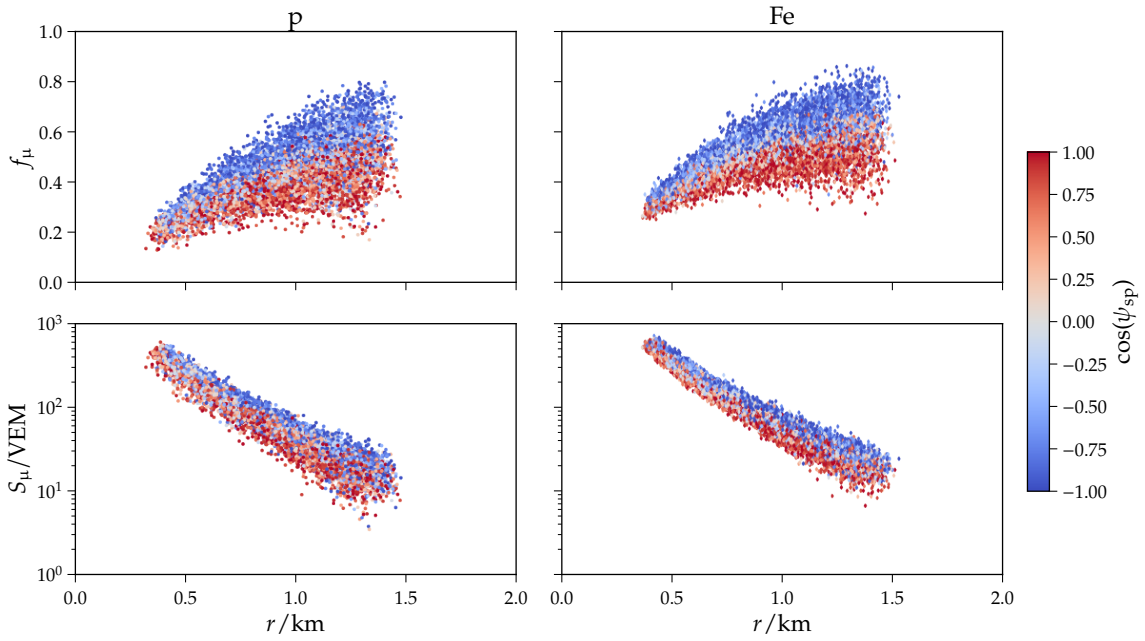


Figure 6.1.: Up-/downstream behavior of f_μ and S_μ, indicated by the color code for proton (*left*) and iron (*right*) as a function of shower plane distance r.

level (Fig. 6.3), as well on a global statistical level.
In practice, the radial and the up-/down-stream dependency is fit simultaneously with the function

$$f_\mu(r, \Delta h) = c\frac{ar}{\sqrt{1+(ar)^2}}(1 + b\Delta h). \tag{6.7}$$

By setting $\Delta h = 0$ after the fit, the LFF is projected to the reference of no up- and downstream dependency and is thus, equivalent to first projecting f_μ and then fitting the LFF.

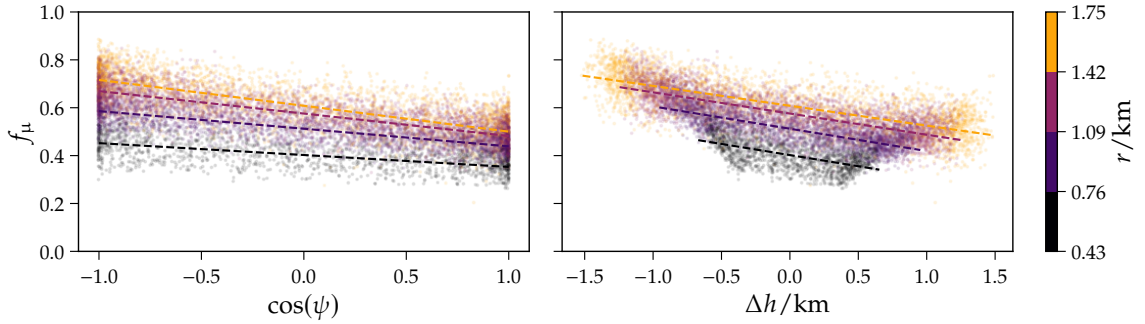


Figure 6.2.: Up-/downstream behavior of f_μ in dependency of the variables $\cos\psi_{sp}$ (*left*) and Δh (*right*) for iron induced showers in $\lg E \in [19.54, 19.76]$ and $\theta \in [37.76°, 42.13°]$. The lines represent linear fits in different distance intervals to track the change of slope.

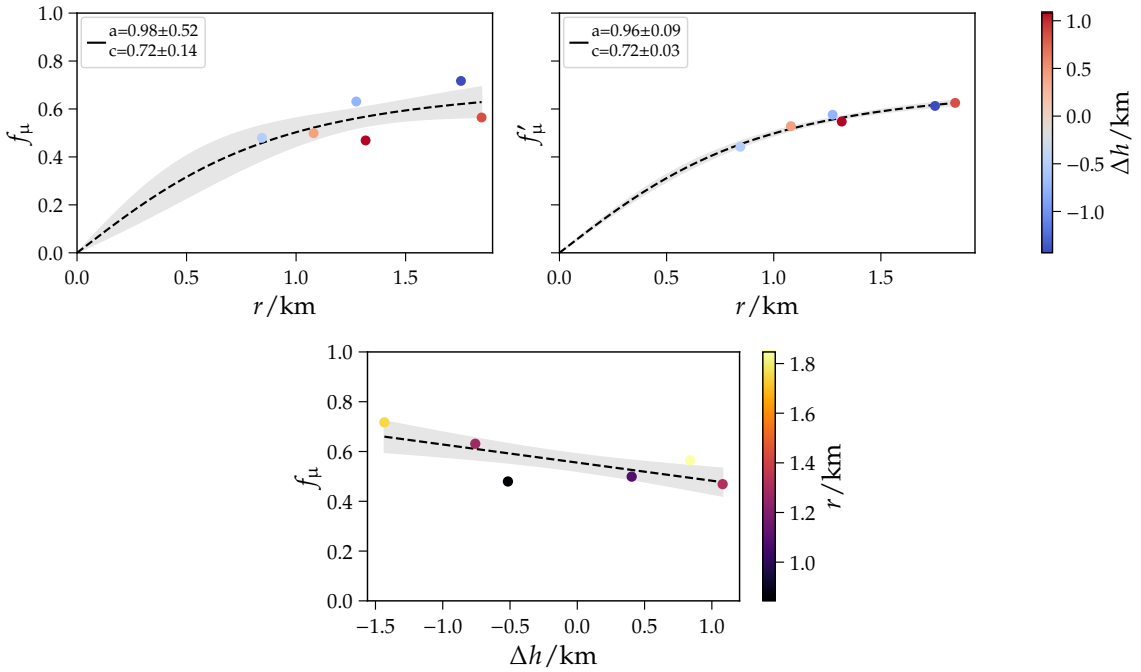


Figure 6.3.: Lateral muon fraction f_μ before (*left*) and after (*right*) the up-/downstream correction (*bottom*) for a oxygen induced shower of $\lg E/\mathrm{eV} = 20$ and $\theta = 39°$.

6.1.2. Energy and Zenith Correction

Because the muon fraction is rotationally invariant after the up- and downstream correction, it is possible to construct an event level muon fraction estimator $F_\mu^{(*)}$, by either integrating in between a fixed distance interval $r \in [r_0, r_1]$, or by inferring the value at a fixed reference distance r_{ref}:

$$F_\mu^{(*)} = \begin{cases} 2\pi \int_{r_0}^{r_1} f_\mu(r, \Delta h = 0) r dr, \\ f_\mu(r_{\mathrm{ref}}, \Delta h = 0). \end{cases} \tag{6.8}$$

The integral measure is motivated by averaging over the systematic uncertainty introduced by the heuristically chosen LFF, which probably will not correctly describe the true underlying LFF. One should ensure that the integration limits lie well inside the given phase-space of the data, as extrapolation will introduce systematic errors. On the other hand, one might loose discrimination power by averaging over a to broadly chosen distance range. Thus as a first approach, the value of the LFF at several distances is taken as an event level estimator, with an equidistant spacing of 100 m.
In both cases, the derived event level muon fraction estimator has to be corrected for the attenuation through the zenith angle and the energy dependency. This correction is

parametrized similar to the CIC method of the SD energy estimator:

$$F_\mu^{38} = \frac{F_\mu^{(*)}}{g_{\mathrm{CIC}}(x) h(E)}, \tag{6.9}$$

$$\text{where} \qquad h(E) = A \left(\frac{E}{10^{19}\,\mathrm{eV}} \right)^B, \tag{6.10}$$

$$g_{\mathrm{CIC}}(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3, \tag{6.11}$$

$$x = \sin^2\theta - \sin^2 38^\circ. \tag{6.12}$$

For convenience, the derived estimator is normalized to the average muon fraction of a proton-induced shower as

$$R_\mu = \frac{F_\mu^{(*)}}{\langle F_\mu^{38} \rangle_\mathrm{p}}. \tag{6.13}$$

Although the chosen parametrization is highly phenomenological, a reasonable separation between proton and iron-induced showers is accomplished at a distance of 1100 m (Fig. 6.4, *bottom*) with an merit factor around $\mathrm{MF}(R_\mu) \approx 1.34$ by utilizing the NN predictions. The integration of the LFF between r_0=800 m and r_1=1400 m leads to similar results Fig. A.14. The performance of the employed methods is affected by the primary-dependent bias of the NN predictions. Comparing the results obtained on the MC-muon fraction to the results obtained on the NN predictions, the performance decreases by $\Delta\mathrm{MF}(R_\mu) \approx 0.04$. Furthermore, the utilized parametrization through the protonic muon fraction is not universal. The muon fraction estimator of iron-induced air-showers has a different curvature with respect to the zenith angle. The developed method is most sensitive for zenith angles around 38°, as the separation between proton and iron-induced showers decreases for vertical or highly inclined showers.

6.2. Global Model

Another way to parametrize the muon content of an detected air shower is by using a global analytical fit. In addition, adding $\ln A$ as an input allows to explore the relative difference between different primaries on a station level basis. The sensitivity is given by the change in f_μ, when the mass is changed from iron to proton and serves as a measure for the discrimination power between different primaries. For a model that is linear in $\ln(A)$, this corresponds to an constant offset for different primaries. The fitting is done in a two-step process. First, the parametric model is fit by including the mass term as a dependent variable

$$\hat{f_\mu} = y(r, \theta, \Delta h, S, A) \tag{6.14}$$

$$= \beta_1 \sec(\theta) + \beta_2 r + \beta_3 \ln(A) + \beta_4 \Delta h \sec(\theta) + \beta_5 + \beta_6 \lg(S) + \beta_7 \Delta h + \beta_8 r^2. \tag{6.15}$$

The resulting parameters β_i are depicted in Table 6.1 and the performances, given by the bias and resolution, are shown in Fig. A.15 and Fig. A.16. The sensitivity on the mass,

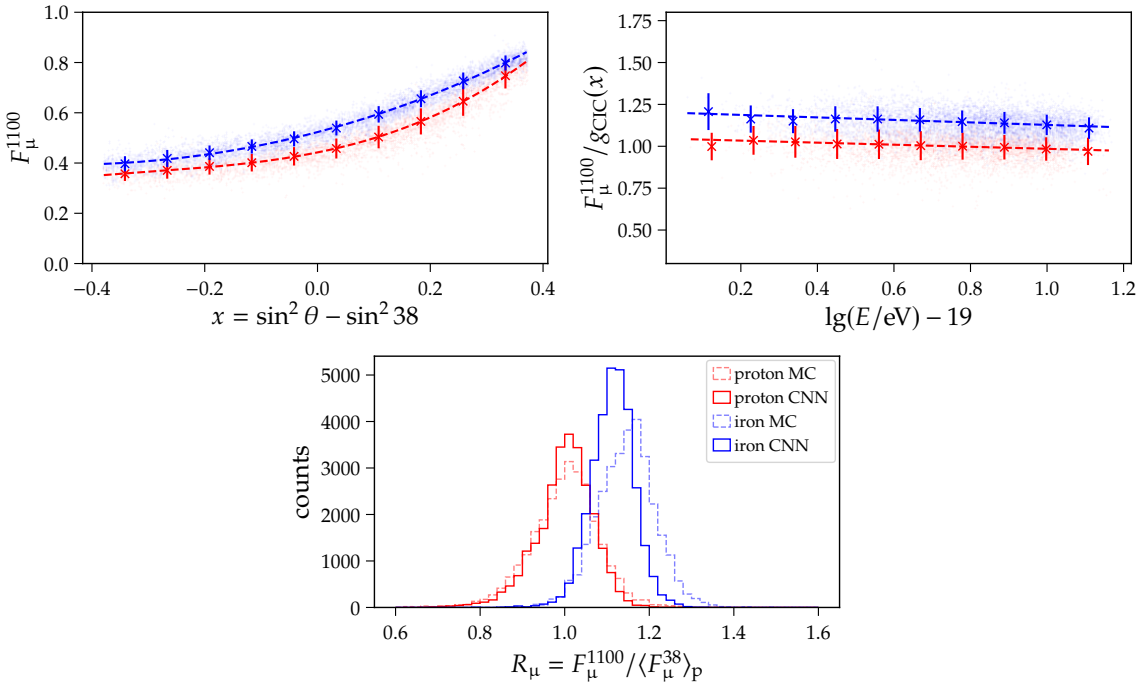


Figure 6.4.: Zenith angle (*top left*) and energy dependency (*top right*) of the muon fraction estimator at r_{ref}=1.1 km, derived by fitting the LFF to the predictions of the CNN (*solid*) or to the MC values (*dashed*). By normalizing the estimator to the average muon fraction of a proton-induced shower, a reasonable separation between proton and iron-induced showers is accomplished (*bottom*). The separation can be qualified by a merit factor of $\mathrm{MF}(R_\mu^{\mathrm{MC}})$=1.38 for the method deployed on the MC muon fraction, and $\mathrm{MF}(R_\mu^{\mathrm{MC}})$=1.34 in case of the CNN predictions.

Table 6.1.: Parameters of globally fitted polynomial model.

Parameter	Value
β_1	0.4783(2)
β_2	0.4010(5)
β_3	0.0191(0)
β_4	0.0638(2)
β_5	-0.3317(5)
β_6	-0.0509(1)
β_7	-0.1673(4)
β_8	-0.1245(2)

given by the parameter $\beta_3 \sim 0.2$, amounts to $\hat{f}_\mu^{\mathrm{Fe}} - \hat{f}_\mu^{\mathrm{p}} = 0.02(\ln 56 - \ln 1) \sim 0.08$.
After fitting, the global shape of f_μ is fixed by the fitted parameters. The mass variable $\ln A$ is turned into a parameter that is then explicitly fit to the predicted $\hat{f}_\mu$ of the DNN for each shower. In essence, the scaling of the fixed reference profile is directly coupled to the mass of the primary. The proton and iron reference lines are easily added to the predictions of the DNN by inserting their corresponding masses. The comparison of the predictions with respect to the showers induced by proton and iron are depicted in Fig. 6.5. The predictions of the CNN consistently lie above(below) the reference lines of iron(proton) for the given events.
In order to derive a statistical discrimination capability, the polynomial model is scaled to each shower with at least $N_{\mathrm{st}} \geq 2$ in the test data set. The resulting distributions of the derived $\ln A$ are depicted in Fig. 6.6. The discrimination power can be quantified by a merit factor of $\mathrm{MF}(\ln A) = 1.31(N_{\mathrm{st}} \geq 2)$ and $\mathrm{MF}(\ln A) = 1.43(N_{\mathrm{st}} \geq 5)$. More detailed insights about the discrimination capability with respect to a given reconstructed energy are given in Fig. 6.7. The width of the predicted distributions decreases with increasing energy until proton-induced showers exhibit limited statistics due to the energy reconstruction bias of the SD. Scaling a reference LFF-profile to the predictions of the CNN provides promising results for the discrimination of different primaries. Even though, the developed method has unreached potential, as the direct predictions of the primary mass with CNNs on station-level outperform the constructed event-level estimator.

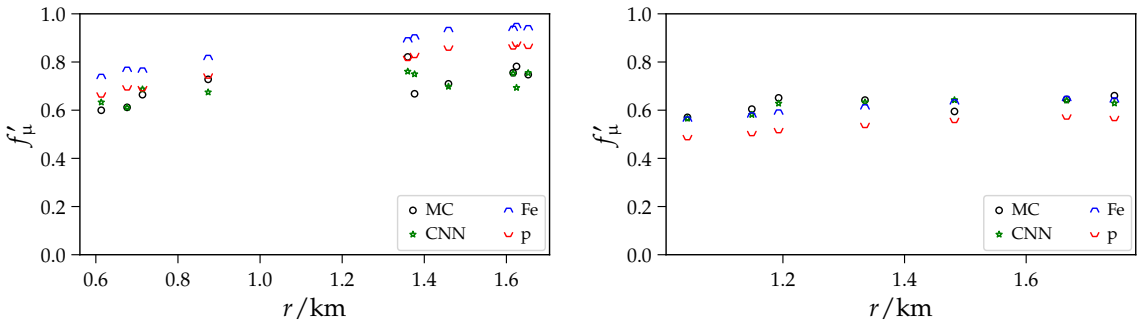


Figure 6.5.: Illustration of the polynomial model fit to the predictions of the CNN for a proton (*left*, $\ln A = -4.59(91)$) and iron (*right*, $\ln A = 4.37(33)$) induced shower. The muon fraction is projected to $f'_\mu = f_\mu(\Delta h = 0)$ to highlight the radial dependency. The polynomial references are given by proton (*red*) and iron (*blue*).

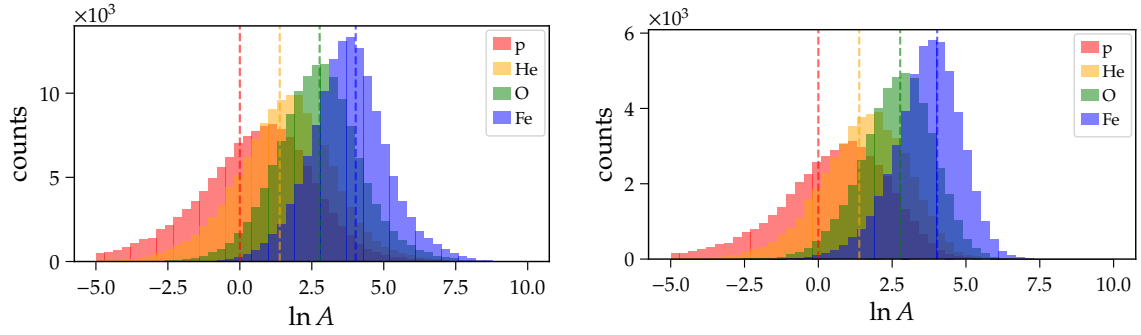


Figure 6.6.: Distribution of the derived primary masses by fitting the polynomial model to the CNN predictions. A merit factor of $MF = 1.31$ and 1.43 is achieved for showers triggering $N_{st} \geq 2$ (*left*) and $N_{st} \geq 5$ (*right*) stations after the quality cuts.

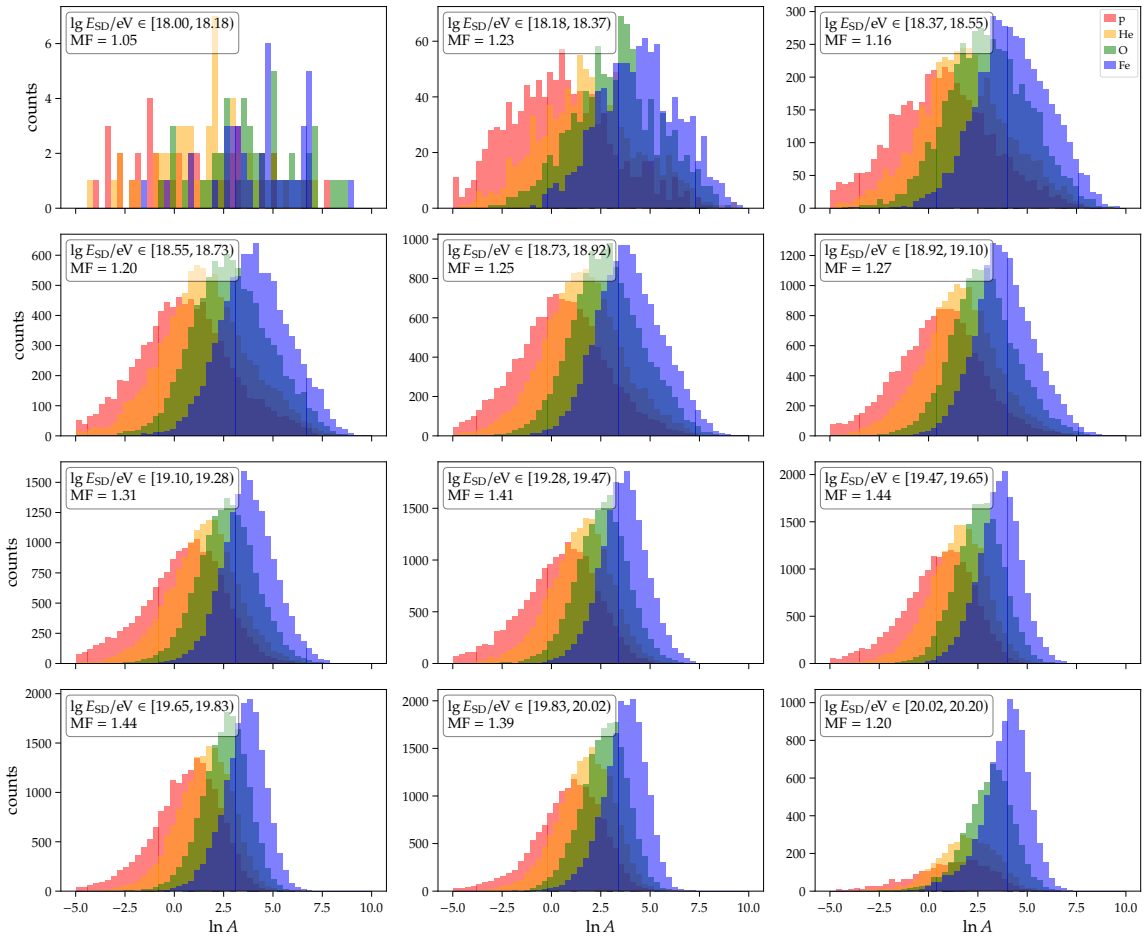


Figure 6.7.: Distribution of the derived ln A by scaling the polynomial model to the CNN predictions, in dependency of different reconstructed energy intervals E_{SD}. Showers that triggered at least two stations after the quality cuts are considered.

7. Summary and conclusion

In this work, I have introduced and motivated the muon signal fraction of the WCD as a well suited composition-dependent quantity. The muon fraction is almost independent of the *energy-migration effect* caused by inherent systematics of the SD energy estimator. Compared to the muon signal, the muon fraction retains its mass-sensitivity. One of the main objectives of the developed methods, is the identification of the primary particle on an event-by-event basis using the muon signal fraction. Therefore, I have developed a method that is compromised of a hybrid NN-analytical approach. On the station level, I have shown that NNs surpass the performance of muon fraction predictions of the selected classical analytical methods by a more than twice as good resolution. On event level, the NN-based predictions of the muon fraction are used to form an event-level estimator. Finally, this estimator is used to derive a mass-sensitive observable, which can be used to discriminate between proton and iron-induced showers.

By comparing the quality of the predictions of various models, I have determined that the CNN clearly surpasses all reference models in terms of evaluation metrics. Compared to the state-of-the-art matrix formalism, the resolution of the muon fraction improved from around 0.137 to 0.048, and the primary-dependent bias between proton and iron induced showers from 0.035 to 0.014. The saturation of the quality of the predictions of the complexer LightGBM and dense NN reference models indicate that the information content provided by tested scalar features is exhausted, while the CNN is capable of efficiently utilizing the information of the time dependent signal to outperform them. Moreover, the chosen CNN architecture is validated by cross-checks to LSTM-based NN architectures, which are particularly suited for analyzing one-dimensional, ordered time series data. The resolutions and biases of both architectures are essential the same. In various optimization attempts, such as a hyperparameter grid-search, I have pushed the CNN architecture to its limits, leading to some minor improvements in evaluation metrics. Including the SSD detector trace as an additional input to the CNN in any way, did not yield any significant improvement. However, the SSD information affects the primary-dependent bias between proton and iron, reducing it from 0.0163 to 0.0137, while the resolution is essentially unaffected. The information provided by the SSD is most likely redundant in the time signal of the WCD. By studying possible systematics inherent to the two gain types, I have shown that the CNN is capable in handling both gain types correctly and no biases are introduced by training the network simultaneously on both gain types. The muon fraction predictions of the CNN on QGSJetII-04 data, show that the prediction of f_{μ} only marginally depend on the employed hadronic interaction model used for the simulation of the air showers. The global bias on the predicted muon fraction between QGSJetII-04 and EPOS-LHC simulated air showers is around -0.008 and almost constant in E, θ, r, and S_{tot}, while the resolution is essentially the same.

The developed CNN methods, which directly predict the primary mass on single station-level, show promising discrimination capabilities with merit factors of up to MF=1.65 between the predicted proton and iron masses, depending on the reconstructed energy interval. In case the network is trained to classify each station by the particle that induced the shower, only around 3 out of 100 stations hit by an iron shower got falsely classified being triggered by proton showers.

The developed methods to form an event level muon fraction estimator which is linked to the mass of the primary have resulted in encouraging outcomes. Merit factors of up to MF=1.44 have been achieved by the polynomial method. Scaling a fixed lateral reference profile to the predicted muon fraction turns out to be suitable in progressing towards an event-to-event identification of the primary. However, a comparison with the results obtained on a single-station level show, that the employed analytical methods are not capable of fully utilizing the information content provided by the muon fraction. This is largely attributed to the heuristically developed methods. Especially the azimuthal asymmetry hampers a description of the muon fraction as a function of only the distance to the shower core. Consequently, a more sophisticated theory is required to fully gauge the potential of the muon fraction as a mass-sensitive observable in future studies.

In a first step, the LFF has to be investigated over an extended phase-space to deduce are more accurate functional shape. Simulations with dense-rings could be utilized to directly provide an event-level muon-fraction reference and reduce the uncertainty of the functional shape. The event level reference, at a given distance, could serve as a direct and less uncertain target for future NN-based studies. Additionally, the sensitivity of the developed methods can be further enhanced by taking into account the signals originating from muon decays into the calculation of the muon fraction. Of particular interest is the application of the developed methods to measured data. As the muon fraction is only minimally influenced by the energy migration effect and the deployed hadronic interaction models, f_μ provides a promising measure to test future hadronic interaction models aiming to resolve the discrepancy between measured data and simulations.

Acknowledgements

I am profoundly grateful to all those who have contributed to the successful completion of this master's thesis. Their support, encouragement, and guidance have been invaluable throughout this journey.

First and foremost, I would like to express my deepest gratitude to Prof. Dr. Ralph Engel and Prof. Dr. Guido Drexlin for refereeing this thesis and enabling me to explore new frontiers in Auger.

Many thanks to Dr. Markus Roth for all his encouragement and valuable feedback throughout the last year. The opportunities he has given me are for sure not to be taken for granted. It was a huge pleasure being able to visit Auger on-site and experience the ultimate la pampa and la vida de Argentinia. Consequently, despite broadening my scientific horizon, I was able to collect awesome life experiences which lead to a new perspective on the world.

An enormous appreciation to Dr. Darko Veberič and Dr. Steffen Traugott Hahn for their guidance and 24/7 valuable support throughout this whole thesis. I am deeply grateful for your patience, encouragement and all the mind-blowing discussions I could experience with you. The last year would have definitely not been the same without you.

Additionally, I want to thank Sabine Bucher for her excellent management of the institute. Her support has been invaluable, without it, I would remained puzzled while attempting to complete various business trip approvals. Most importantly, a huge thanks for organizing a large part of my trip to Argentinia.

Finally, I am very grateful for all my incredible friends who have given me the needed strength to push through difficult times. A huge thanks also to all the colleagues at the institute for the fun time, interesting discussions, last minute corrections and insights I could gain.

I want to conclude with my warmest gratitude to my family, which supported me throughout my whole life and especially during my studies.

Bibliography

[1] M. Walter and A. W. Wolfendale. "Early history of cosmic particle physics". In: *The European Physical Journal H* 37.3 (Aug. 1, 2012), pp. 323–358. ISSN: 2102-6467. DOI: 10.1140/epjh/e2012-30020-1.

[2] Charles TR Wilson. "On the leakage of electricity through dust-free air". In: *Proc. Cambridge Philos. Soc.* Vol. 11. 1900, p. 32.

[3] Charles Thomson Rees Wilson. "On the ionisation of atmospheric air". In: *Proceedings of the Royal Society of London* 68.442-450 (1901), pp. 151–161.

[4] Theodor Wulf. "Über den Ursprung der in der Atmosphäre vorhandenen γ-Strahlung". In: *Phys. Z* 10 (1909), pp. 997–1003.

[5] Theodor Wulf. "Beobachtungen über Strahlung hoher Durchdringungsfähigkeit auf dem Eiffelturm". In: *Physikalische Zeitschrift* 11 (1910), pp. 811–813.

[6] Albert Gockel. "Luftelektrische beobachtungen bei einer ballonfahrt". In: *Phys. Z* 11.280 (1910).

[7] Albert Gockel. "Messungen der druchdringenden Strahlung bei Ballonfahren". In: *Phys. Zeits.* 12 (1911), pp. 595–597.

[8] Domenico Pacini et al. *Penetrating Radiation on the Sea.* Mar. 3, 2011. DOI: 10.48550/arXiv.1101.3015.

[9] Victor F. Hess. "Über Beobachtungen Der Durchdringenden Strahlung Bei Sieben Freiballonfahrten". In: *Phys. Z.* 13 (1912), pp. 1084–1091.

[10] Bruno Rossi. "Directional Measurements on the Cosmic Rays Near the Geomagnetic Equator". In: *Physical Review* 45.3 (Feb. 1, 1934), pp. 212–214. DOI: 10.1103/PhysRev.45.212.

[11] Hans Geiger and Walther Müller. "Elektronenzählrohr zur messung schwächster aktivitäten". In: *Naturwissenschaften* 16.31 (1928), pp. 617–618.

[12] Bruno Rossi. "Über die Eigenschaften der durchdringenden Korpuskularstrahlung im Meeresniveau". In: *Zeitschrift für Physik* 82.3 (Mar. 1, 1933), pp. 151–178. ISSN: 0044-3328. DOI: 10.1007/BF01341486.

[13] Pierre Auger et al. "Extensive Cosmic-Ray Showers". In: *Reviews of Modern Physics* 11.3-4 (July 1, 1939), pp. 288–291. DOI: 10.1103/RevModPhys.11.288.

[14] A. M. Hillas. "The Origin of Ultra-High-Energy Cosmic Rays". In: *Annual Review of Astronomy and Astrophysics* 22.1 (Sept. 1984), pp. 425–444. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev.aa.22.090184.002233.

[15] Thomas K. Gaisser, Ralph Engel, and Elisa Resconi. *Cosmic Rays and Particle Physics*. 2nd ed. Cambridge University Press, June 2, 2016. ISBN: 978-0-521-01646-9. DOI: 10.1017/CBO9781139192194.

[16] Rafael Alves Batista et al. "Open Questions in Cosmic-Ray Research at Ultrahigh Energies". In: *Frontiers in Astronomy and Space Sciences* 6 (June 4, 2019), p. 23. ISSN: 2296-987X. DOI: 10.3389/fspas.2019.00023.

[17] ENRICO Fermi. "On the Origin of the Cosmic Radiation". In: *Physical Review* 75.8 (Apr. 15, 1949), pp. 1169–1174. DOI: 10.1103/PhysRev.75.1169.

[18] H. Alfvén. "Existence of Electromagnetic-Hydrodynamic Waves". In: *Nature* 150.3805 (3805 Oct. 1942), pp. 405–406. ISSN: 1476-4687. DOI: 10.1038/150405d0.

[19] Claus Grupen. *Astroparticle Physics*. Undergraduate Texts in Physics. Cham: Springer International Publishing, 2020. ISBN: 978-3-030-27339-2. DOI: 10.1007/978-3-030-27339-2.

[20] Jacco Vink. "Supernova Remnants and the Origin of Cosmic Rays". In: *Proceedings of the International Astronomical Union* 9.S296 (Jan. 2013), pp. 305–314. ISSN: 1743-9213, 1743-9221. DOI: 10.1017/S1743921313009630.

[21] Pasquale Blasi. "The Origin of Galactic Cosmic Rays". In: *The Astronomy and Astrophysics Review* 21.1 (Nov. 5, 2013), p. 70. ISSN: 1432-0754. DOI: 10.1007/s00159-013-0070-7.

[22] Kenneth Greisen. "End to the Cosmic-Ray Spectrum?" In: *Physical Review Letters* 16.17 (Apr. 25, 1966), pp. 748–750. DOI: 10.1103/PhysRevLett.16.748.

[23] G. T. Zatsepin and V. A. Kuz'min. "Upper Limit of the Spectrum of Cosmic Rays". In: *Soviet Journal of Experimental and Theoretical Physics Letters* 4 (Aug. 1966), p. 78.

[24] KASCADE-Grande Collaboration et al. "Kneelike Structure in the Spectrum of the Heavy Component of Cosmic Rays Observed with KASCADE-Grande". In: *Physical Review Letters* 107.17 (Oct. 20, 2011), p. 171104. DOI: 10.1103/PhysRevLett.107.171104.

[25] P. Abreu et al. "Search for Signatures of Magnetically-Induced Alignment in the Arrival Directions Measured by the Pierre Auger Observatory". In: *Astroparticle Physics* 35.6 (Jan. 1, 2012), pp. 354–361. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2011.10.004.

[26] THE PIERRE AUGER COLLABORATION et al. "Observation of a Large-Scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV". In: *Science* 357.6357 (Sept. 22, 2017), pp. 1266–1270. DOI: 10.1126/science.aan4338.

[27] A. Aab et al. "Large-Scale Cosmic-Ray Anisotropies above 4 EeV Measured by the Pierre Auger Observatory". In: *The Astrophysical Journal* 868.1 (Nov. 2018), p. 4. ISSN: 0004-637X. DOI: 10.3847/1538-4357/aae689.

[28] The Pierre Auger Collaboration et al. "Features of the Energy Spectrum of Cosmic Rays above 2.5×10^{18} eV Using the Pierre Auger Observatory". In: *Physical Review Letters* 125.12 (Sept. 16, 2020), p. 121106. DOI: 10.1103/PhysRevLett.125.121106.

[29] 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.

[30] H. Bethe and W. Heitler. "On the Stopping of Fast Particles and on the Creation of Positive Electrons". In: *Proceedings of the Royal Society of London Series A* 146.856 (Aug. 1934), pp. 83–112. DOI: 10.1098/rspa.1934.0140.

[31] M Steffen Traugott Hahn. "Methods for Estimating Mass-Sensitive Observables of Ultra-High Energy Cosmic Rays Using Artificial Neural Networks". Dec. 16, 2022.

[32] J. Matthews. "A Heitler Model of Extensive Air Showers". In: *Astroparticle Physics* 22.5 (Jan. 1, 2005), pp. 387–397. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2004.09.003.

[33] The Pierre Auger Collaboration. "The Pierre Auger Cosmic Ray Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (Oct. 2015), pp. 172–213. ISSN: 01689002. DOI: 10.1016/j.nima.2015.06.058.

[34] Paul Sommers. "Capabilities of a Giant Hybrid Air Shower Detector". In: *Astroparticle Physics* 3.4 (Aug. 1, 1995), pp. 349–360. ISSN: 0927-6505. DOI: 10.1016/0927-6505(95)00013-7.

[35] B. R. Dawson et al. "Simulations of a Giant Hybrid Air Shower Detector". In: *Astroparticle Physics* 5.3 (Oct. 1, 1996), pp. 239–247. ISSN: 0927-6505. DOI: 10.1016/0927-6505(96)00024-2.

[36] I. Allekotte et al. "The Surface Detector System of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 586.3 (Mar. 2008), pp. 409–420. ISSN: 01689002. DOI: 10.1016/j.nima.2007.12.016.

[37] J. Abraham et al. "Trigger and Aperture of the Surface Detector Array of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 613.1 (Jan. 21, 2010), pp. 29–39. ISSN: 0168-9002. DOI: 10.1016/j.nima.2009.11.018.

[38] Antonella Castellina. "AugerPrime: The Pierre Auger Observatory Upgrade". In: *EPJ Web of Conferences* 210 (2019), p. 06002. ISSN: 2100-014X. DOI: 10.1051/epjconf/201921006002.

[39] The Pierre Auger Collaboration et al. *The Pierre Auger Observatory Upgrade - Preliminary Design Report*. Apr. 12, 2016. DOI: 10.48550/arXiv.1604.03637.

[40] J. Abraham et al. "The Fluorescence Detector of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 620.2 (Aug. 11, 2010), pp. 227–251. ISSN: 0168-9002. DOI: 10.1016/j.nima.2010.04.023.

[41] T. K. Gaisser and A. M. Hillas. "Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers". In: 8 (Jan. 1, 1977), p. 353.

[42] R. M. Baltrusaitis et al. "The Utah Fly's Eye Detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 240.2 (Oct. 15, 1985), pp. 410–428. ISSN: 0168-9002. DOI: 10.1016/0168-9002(85)90658-8.

[43] T Abu-Zayyad et al. "The Prototype High-Resolution Fly's Eye Cosmic Ray Detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 450.2 (Aug. 11, 2000), pp. 253–269. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(00)00307-7.

[44] H. Kawai et al. "Telescope Array Experiment". In: *Nuclear Physics B - Proceedings Supplements*. Proceedings of the XIV International Symposium on Very High Energy Cosmic Ray Interactions 175–176 (Jan. 1, 2008), pp. 221–226. ISSN: 0920-5632. DOI: 10.1016/j.nuclphysbps.2007.11.002.

[45] The Pierre Auger Collaboration et al. "Data-Driven Estimation of the Invisible Energy of Cosmic Ray Showers with the Pierre Auger Observatory". In: *Physical Review D* 100.8 (Oct. 25, 2019), p. 082003. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.100.082003.

[46] A. Aab et al. "Reconstruction of Events Recorded with the Surface Detector of the Pierre Auger Observatory". In: *Journal of Instrumentation* 15.10 (Oct. 2020), P10021. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/10/P10021.

[47] Koichi Kamata and Jun Nishimura. "The Lateral and the Angular Structure Functions of Electron Showers". In: *Progress of Theoretical Physics Supplement* 6 (1958), pp. 93–155. ISSN: 0375-9687. DOI: 10.1143/PTPS.6.93.

[48] D. Newton, J. Knapp, and A. A. Watson. "The Optimum Distance at Which to Determine the Size of a Giant Air Shower". In: *Astroparticle Physics* 26.6 (Jan. 2007), pp. 414–419. ISSN: 09276505. DOI: 10.1016/j.astropartphys.2006.08.003.

[49] A. Aab et al. "Impact of Atmospheric Effects on the Energy Reconstruction of Air Showers Observed by the Surface Detectors of the Pierre Auger Observatory". In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02006. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/02/P02006.

[50] The Pierre Auger collaboration. "The Effect of the Geomagnetic Field on Cosmic Ray Energy Estimates and Large Scale Anisotropy Searches on Data from the Pierre Auger Observatory". In: *Journal of Cosmology and Astroparticle Physics* 2011.11 (Nov. 2011), p. 022. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2011/11/022.

[51] D. Veberic et al. *GAP 2015-065 Constant Intensity Cut: Unbinned Estimation of the Signal Attenuation Function.*

[52] The Pierre Auger Collaboration et al. "Measurement of the Cosmic-Ray Energy Spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory". In: *Physical Review D* 102.6 (Sept. 16, 2020), p. 062005. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.102.062005.

[53] Pierre Auger Collaboration. “Muons in Air Showers at the Pierre Auger Observatory: Measurement of Atmospheric Production Depth”. In: *Physical Review D* 90.1 (July 25, 2014), p. 012012. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.90.012012.

[54] Pierre Auger Collaboration et al. “Muons in Air Showers at the Pierre Auger Observatory: Mean Number in Highly Inclined Events”. In: *Physical Review D* 91.3 (Feb. 6, 2015), p. 032003. DOI: 10.1103/PhysRevD.91.032003.

[55] A. Aab et al. “Direct Measurement of the Muonic Content of Extensive Air Showers between 2×10^{17} and 2×10^{18}eV at the Pierre Auger Observatory”. In: *The European Physical Journal C* 80.8 (Aug. 18, 2020), p. 751. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-8055-y.

[56] The Pierre Auger Collaboration et al. “Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory”. In: *Physical Review Letters* 126.15 (Apr. 16, 2021), p. 152002. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.126.152002.

[57] Jose Bellido et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory: measurements above $10^{17.2}$ eV and composition implications”. In: *Proceedings of Science* 301 (2017), pp. 506–1.

[58] The Pierre Auger Collaboration et al. “Observation of the Suppression of the Flux of Cosmic Rays above 4×10^{19} eV”. In: *Physical Review Letters* 101.6 (Aug. 4, 2008), p. 061101. DOI: 10.1103/PhysRevLett.101.061101.

[59] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[60] OpenAI. *ChatGPT*. Version 3.5. Accessed on March 31, 2023. 2021. URL: https://openai.com/blog/chat-gpt-3-6b/.

[61] David Silver et al. “Mastering the Game of Go with Deep Neural Networks and Tree Search”. In: *Nature* 529.7587 (7587 Jan. 2016), pp. 484–489. ISSN: 1476-4687. DOI: 10.1038/nature16961.

[62] Léon Bottou. “Large-Scale Machine Learning with Stochastic Gradient Descent”. In: *Proceedings of COMPSTAT’2010*. Ed. by Yves Lechevallier and Gilbert Saporta. Heidelberg: Physica-Verlag HD, 2010, pp. 177–186. ISBN: 978-3-7908-2604-3. DOI: 10.1007/978-3-7908-2604-3_16.

[63] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. Jan. 29, 2017. DOI: 10.48550/arXiv.1412.6980.

[64] Aston Zhang et al. “Dive into Deep Learning”. In: *arXiv preprint arXiv:2106.11342* (2021).

[65] Zhuang Liu et al. *A ConvNet for the 2020s*. Mar. 2, 2022.

[66] Yann Lecun and Y. Bengio. “Convolutional Networks for Images, Speech, and Time-Series”. In: *The Handbook of Brain Theory and Neural Networks*. Jan. 1995.

[67] D. H. Hubel and T. N. Wiesel. "Receptive Fields of Single Neurones in the Cat's Striate Cortex". In: *The Journal of Physiology* 148.3 (Oct. 1959), pp. 574–591. ISSN: 0022-3751.

[68] D. H. Hubel and T. N. Wiesel. "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex". In: *The Journal of Physiology* 160.1 (Jan. 1962), pp. 106–154.2. ISSN: 0022-3751.

[69] D. H. Hubel and T. N. Wiesel. "Receptive Fields and Functional Architecture of Monkey Striate Cortex". In: *The Journal of Physiology* 195.1 (Mar. 1968), pp. 215–243. ISSN: 0022-3751. DOI: 10.1113/jphysiol.1968.sp008455.

[70] Kyunghyun Cho et al. *Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation*. Sept. 2, 2014. DOI: 10.48550/arXiv.1406.1078.

[71] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (Nov. 1, 1997), pp. 1735–1780. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/neco.1997.9.8.1735.

[72] Christopher M. Bishop. "Mixture density networks". Technical Report. Birmingham, 1994. URL: https://publications.aston.ac.uk/id/eprint/373/.

[73] Tim Miller. *Explanation in Artificial Intelligence: Insights from the Social Sciences*. Aug. 14, 2018. DOI: 10.48550/arXiv.1706.07269.

[74] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017.

[75] Christoph Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Second edition. Munich, Germany: Christoph Molnar, 2022. 318 pp. ISBN: 9798411463330.

[76] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York, 2009. ISBN: 978-0-387-84857-0 978-0-387-84858-7. DOI: 10.1007/978-0-387-84858-7.

[77] Leo Breiman. "Random Forests". In: *Machine Learning* 45.1 (Oct. 1, 2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324.

[78] Gilles Louppe et al. "Understanding Variable Importances in Forests of Randomized Trees". In: *Advances in Neural Information Processing Systems*. Vol. 26. Curran Associates, Inc., 2013.

[79] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. *All Models Are Wrong, but Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously*. Dec. 23, 2019. DOI: 10.48550/arXiv.1801.01489.

[80] Antoine Letessier-Selvon et al. "Layered Water Cherenkov Detector for the Study of Ultra High Energy Cosmic Rays". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 767 (Dec. 11, 2014), pp. 41–49. ISSN: 0168-9002. DOI: 10.1016/j.nima.2014.08.029.

[81] Pierre Billoir Antoine Letessier-Selvon. *GAP 2015-045 Separating the muonic and electromagnetic components of a shower through the matrix formalism for a two-fold detector*. 2015.

[82] David Schmidt. "Sensitivity of AugerPrime to the Masses of Ultra-High-Energy Cosmic Rays". Karlsruher Institute of Technology (KIT), Universidad Nacional de San Martìn (UNSAM), 2018. 109 pp.

[83] D Martello. *GAP 2019-033 Considerations about the method of the matrix*.

[84] Allan Payeras, David Schmidt, and Darko Veberič. *Matrix Formalism for AugerPrime*.

[85] S. Lafebre et al. "Universality of Electron–Positron Distributions in Extensive Air Showers". In: *Astroparticle Physics* 31.3 (Apr. 1, 2009), pp. 243–254. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2009.02.002.

[86] Roberta Colalillo, Fausto Guarino, and Alexey Yushkov. *GAP 2019-069 Napoli + Praha Library: A Brief Guide to the Summary Trees, Including the Correct Way to Calculate Components of Signals in SD Stations*.

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

[88] Tianqi Chen and Carlos Guestrin. "XGBoost". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016. DOI: 10.1145/2939672.2939785. URL: https://doi.org/10.1145%2F2939672.2939785.

[89] Guolin Ke et al. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[90] Scott M. Lundberg et al. "From Local Explanations to Global Understanding with Explainable AI for Trees". In: *Nature Machine Intelligence* 2.1 (Jan. 2020), pp. 56–67. ISSN: 2522-5839. DOI: 10.1038/s42256-019-0138-9.

[91] Anselm Blumer et al. "Occam's Razor". In: *Information Processing Letters* 24.6 (1987), pp. 377–380. ISSN: 0020-0190. DOI: https://doi.org/10.1016/0020-0190(87)90114-1. URL: https://www.sciencedirect.com/science/article/pii/0020019087901141.

[92] Martín Abadi et al. *TensorFlow, Large-scale machine learning on heterogeneous systems*. Nov. 2015. DOI: 10.5281/zenodo.4724125.

[93] Fiona Ellwanger. *GAP 2022-053 Deep-Learning Based Estimation of the Ultra-High Energy Cosmic Ray Spectrum Using the Surface Detector of the Pierre Auger Observatory*.

[94] Jonas Glombitza. "Deep-Learning Based Measurement of the Mass Composition of Ultra-High Energy Cosmic Rays Using the Surface Detector of the Pierre Auger Observatory". RWTH Aachen University, 2021, pages 1 Online–Ressource : Illustrationen. DOI: 10.18154/RWTH-2022-00759.

[95] The Pierre Auger Collaboration et al. "Deep-Learning Based Reconstruction of the Shower Maximum X_{max} using the Water-Cherenkov Detectors of the Pierre Auger Observatory". In: *Journal of Instrumentation* 16.07 (July 1, 2021), P07019. ISSN: 1748-0221. DOI: 10.1088/1748-0221/16/07/P07019.

[96] A. Guillén et al. "Deep Learning Techniques Applied to the Physics of Extensive Air Showers". In: *Astroparticle Physics* 111 (Sept. 1, 2019), pp. 12–22. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2019.03.001.

[97] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. May 24, 2019. DOI: 10.48550/arXiv.1810.04805.

[98] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Communications of the ACM* 60.6 (May 24, 2017), pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/3065386.

[99] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. *Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles*. Nov. 3, 2017. DOI: 10.48550/arXiv.1612.01474.

[100] Pierre Auger Collaboration et al. "Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory". In: *Physical Review Letters* 117.19 (Oct. 31, 2016), p. 192001. DOI: 10.1103/PhysRevLett.117.192001.

[101] Matei Zaharia et al. "Accelerating the machine learning lifecycle with MLflow." In: *IEEE Data Eng. Bull.* 41.4 (2018), pp. 39–45.

Acronyms

Symbols | A | B | C | D | E | F | G | H | I | K | L | M | N | P | Q | R | S | T | U | V | W | X

Symbols

Offline Offline Analysis Framework

A

Adam Adaptive Moment Estimation

AERA Auger Engineering Radio Array

AGN Active Galactic Nuclei

AI Artificial Intelligence

AMIGA Auger Muons and Infill for the Ground Array

ANN Artificial Neural Network

API Application Programming Interface

Auger Pierre Auger Observatory

B

BR Branching Ratio

C

CDAS Central Data Aquisition System

CIC Constant Intensity Cut

CL Confidence Level

CMB Cosmic Microwave Background

CNN Convolutional Neural Network

CORSIKA COsmic Ray SImulations for KAscade

CR Cosmic Ray

D

DAQ Data AcQuisition

DNN Deep Neural Network

E

EAS Extensive Air Shower

EPOS-LHC EPOS-LHC Hadronic Interaction Model

F

FADC Flash Analog-to-Digital Converter

FD Fluorescence Detector

FoV Field of View

G

GZK Greisen-Zatsepin-Kuzmin

H

HEAT High-Elevation Auger Telescopes

HG High Gain

I

ISM Interstellar Medium

K

KaLib Karlsruhe air shower Library

L

LDF Lateral Distribution Function

LFF Lateral Fraction Function

LG Low Gain

LHC Large Hadron Collider

LSTM Long Short-Term Memory

M

MAE Mean Absolute Error

MC Monte Carlo

MDH Mixture Density Head

MDN Mixture Density Network

MIP Minimum-Ionizing Particle

ML Machine Learning

MSE Mean Squared Error

N

NKG Nishimura-Kamata-Greisen

NLP Natural Language Processing

NN Neural Network

NP Nobel Prize

P

PDF Probability Density Function

PMT Photomultiplier Tube

Q

QGSJetII-04 QGSJetII-04 Hadronic Interaction Model

R

ReLU Rectified Linear Unit

RNN Recurrent Neural Network

S

SD Surface Detector

SGD Stochastic Gradient Descent

SNR Supernova Remnant

SPD Shower Plane Distance

SSD Surface Scintillator Detector

T

TA Telescope Array project

U

UHECR Ultra-High Energy Cosmic Ray

UMD Underground Muon Detector

UUB Upgraded Unified Board

V

VEM Vertical Equivalent Muon

W

WCD Water-Cherenkov detector

X

XAI Explainable Artificial Intelligence

Appendices

A. Additional Figures

In this chapter, additional figures which are referred in the text are shown.

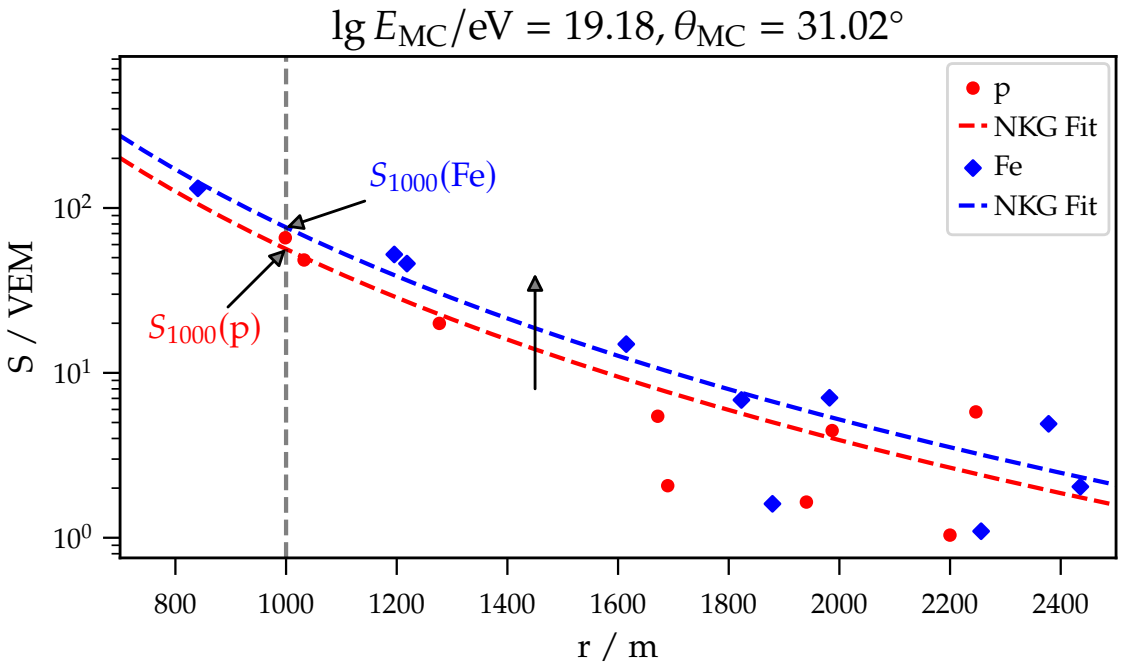


Figure A.1.: Illustration of the *energy migration effect* in the SD reconstruction. An iron (*blue*) induced shower contains more muons than a proton (*red*) induced one at the same MC true energy. As the electromagnetic component is universal, the signals in the WCD are driven by muons and thus the iron is reconstructed with a higher energy. The distance range is restricted for better visibility.

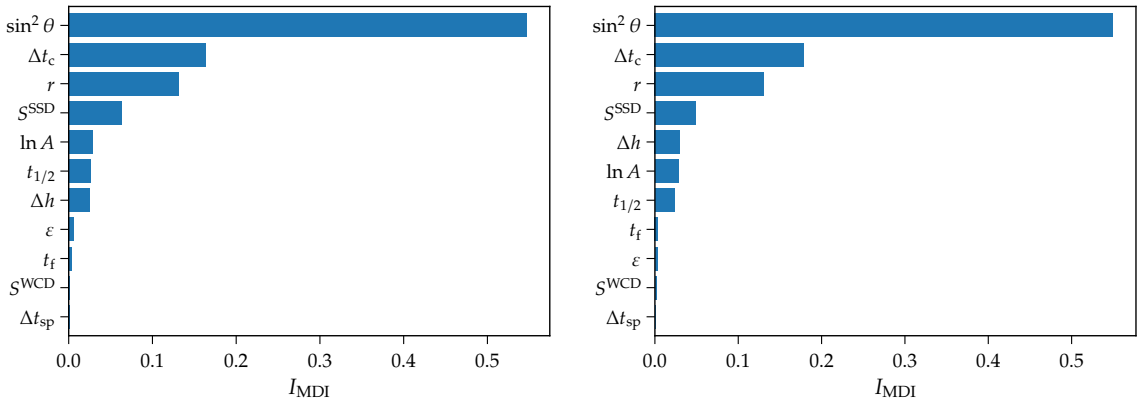


Figure A.2.: *Mean decrease in Impurity (MDI)* for LGBMRegressor (*left*) and XGBRegressor (*right*). The MDI feature importance I_{MDI} is normalized to the sum of all feature importances and sorted by their value. The ensembles are trained without the total signal ratio $S^r = S^{SSD}/S^{WCD}$ in comparison to Fig. 5.6. Even though the information about the signal in each detector is not redundant anymore, the ranking of the importances remains the same, with the total signal of the WCD being not important.

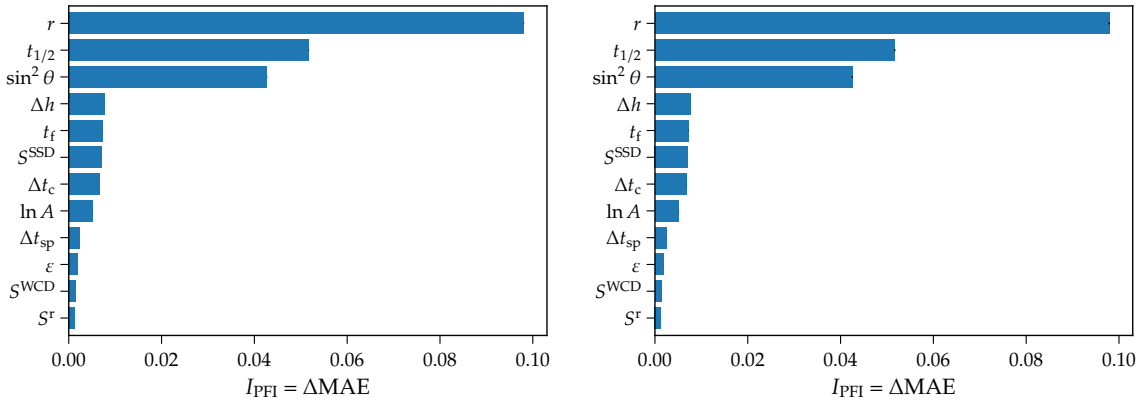


Figure A.3.: *Permutation Feature Importance (PFI)* $I_{PFI} = \Delta MAE$ for LGBMRegressor, calculated on the train (*left*) and test data (*right*). The MAE is chosen as the PFI metric because the muon fraction denotes a relative quantity. Because the PFIs of the train and test set are congruent, there is no indication that the model is overfitted.

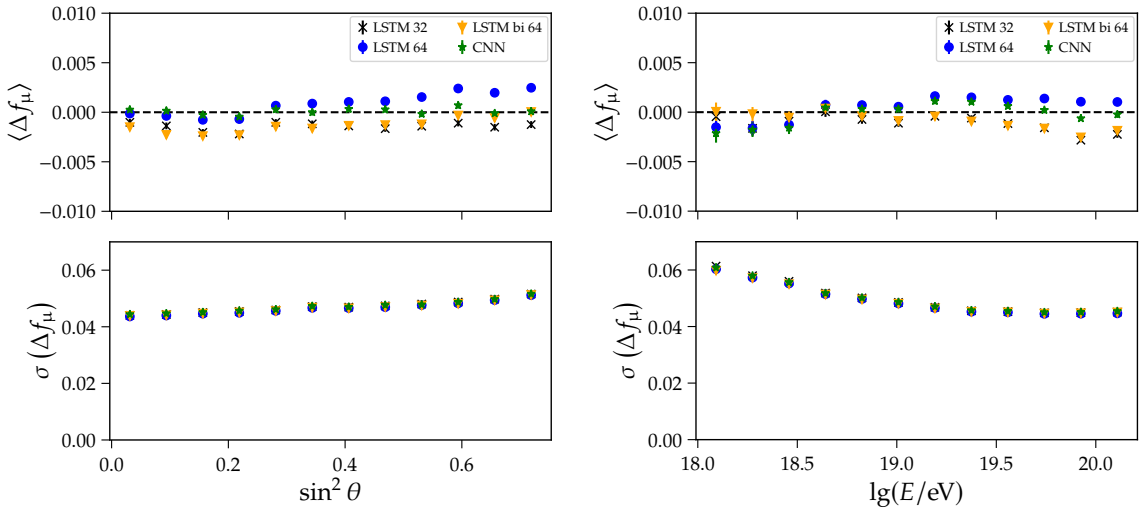


Figure A.4.: Bias (*top*) and resolution (*bottom*) of a DNN with a convolutional versus different LSTM *Trace Analyzers* in dependency of the uniformly distributed zenith angle $\sin^2 \theta$ and logarithmic energy $\lg E$. The LSTM analyzers consist of a single LSTM layer with 32 or 64 units, or a bidirectional layer with 64 units in total. While there are minor differences in the bias, the achieved resolution is congruent over all trace analyzers. It validates the used CNN analyzer as a suitable choice.

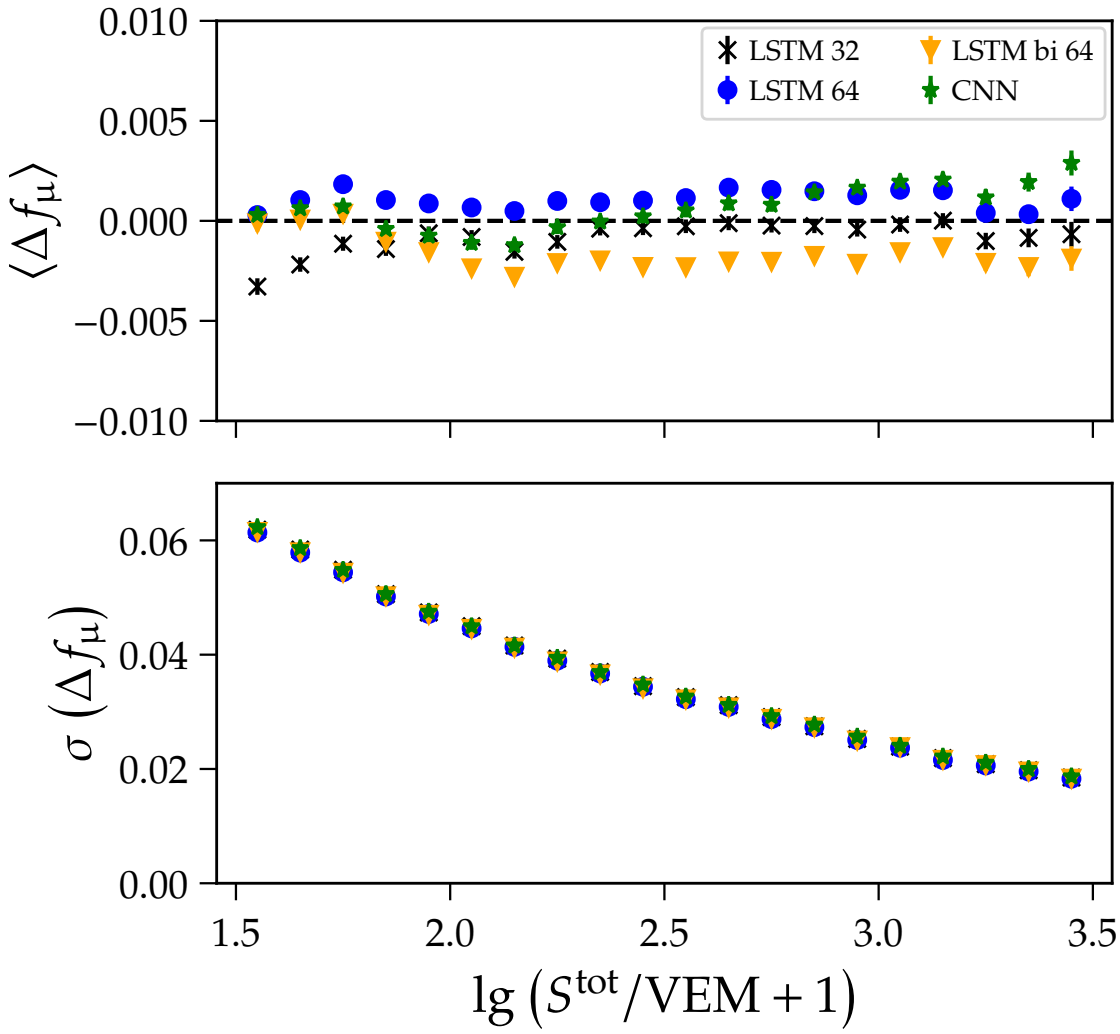


Figure A.5.: Bias (*top*) and resolution (*bottom*) of a DNN with a convolutional versus different LSTM *Trace Analyzers* in dependency of the logarithmic total signal $\lg(S/\mathrm{VEM}+1)$. The LSTM analyzers consist of a single LSTM layer with 32 or 64 units, or a bidirectional layer with 64 units in total. While there are minor differences in the bias, the achieved resolution is congruent over all trace analyzers. It validates the used CNN analyzer as a suitable choice.

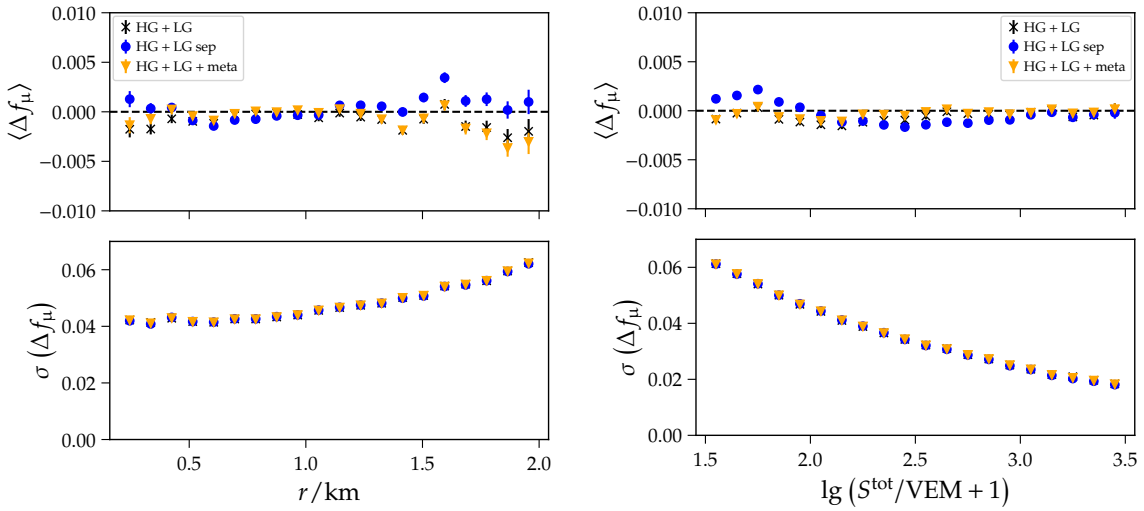


Figure A.6.: Bias (*top*) and resolution (*bottom*) of CNN ensembles trained on different high and low gain detectors as a function of shower plane distance r and logarithmic total signal $\lg(S/\mathrm{VEM}+1)$. For each combination, 5 networks are trained and the mean of their prediction is used for the evaluation. Either the networks are trained separately on HG or LG traces (*HG + LG sep*), or they are trained on both traces (*HG + LG*) with the variable *SdRecStation:IsHighGainSaturated* as an additional binary meta input (*HG + LG meta*). No significant difference in performance is observed, thus the *HG + LG* networks are capable of handle both gain types correctly.

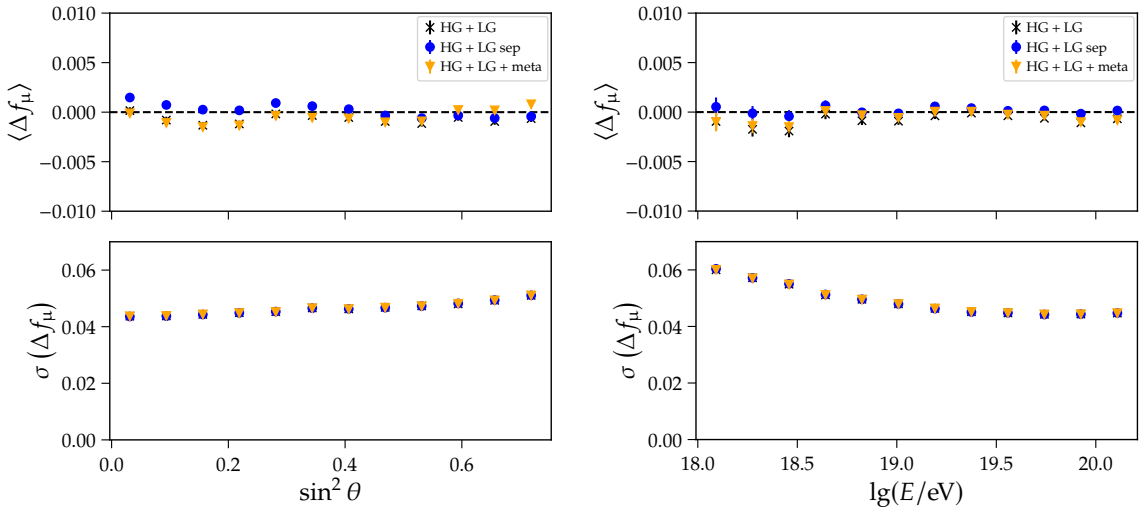


Figure A.7.: Bias (*top*) and resolution (*bottom*) of CNN ensembles trained on different high and low gain detectors as a function of $\sin^2\theta$ and logarithmic energy $\lg E$. For each combination, 5 networks are trained and the mean of their prediction is used for the evaluation. Either the networks are trained separately on HG or LG traces (*HG + LG sep*), or they are trained on both traces (*HG + LG*) with the variable *SdRecStation:IsHighGainSaturated* as an additional binary meta input (*HG + LG meta*). No significant difference in performance is observed, thus the *HG + LG* networks are capable of handle both gain types correctly.

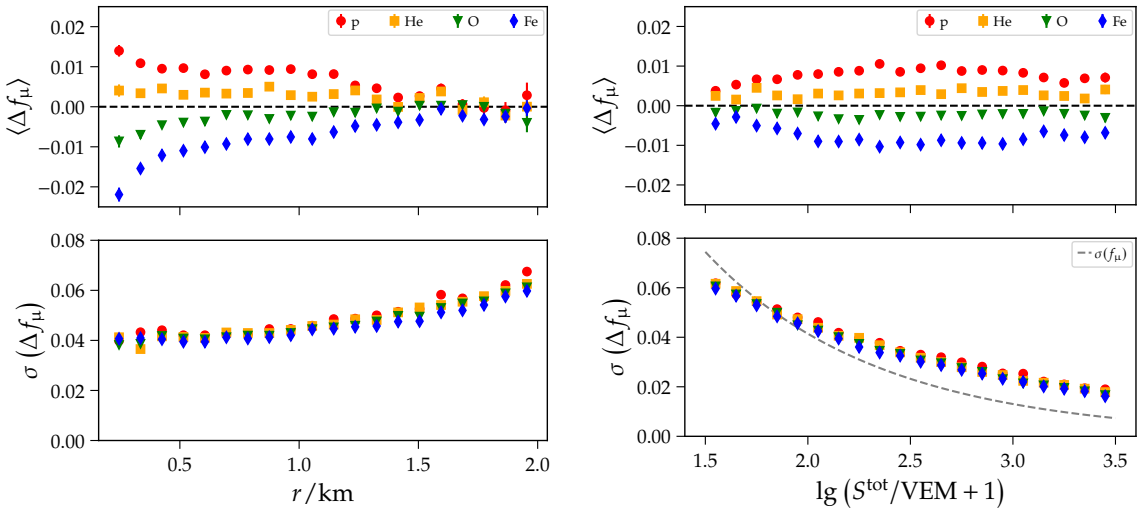


Figure A.8.: Primary-dependent bias (*top*) and resolution (*bottom*) of the best performing CNNs as a function of shower plane distance r (*left*) and total signal S_{tot} (*right*).

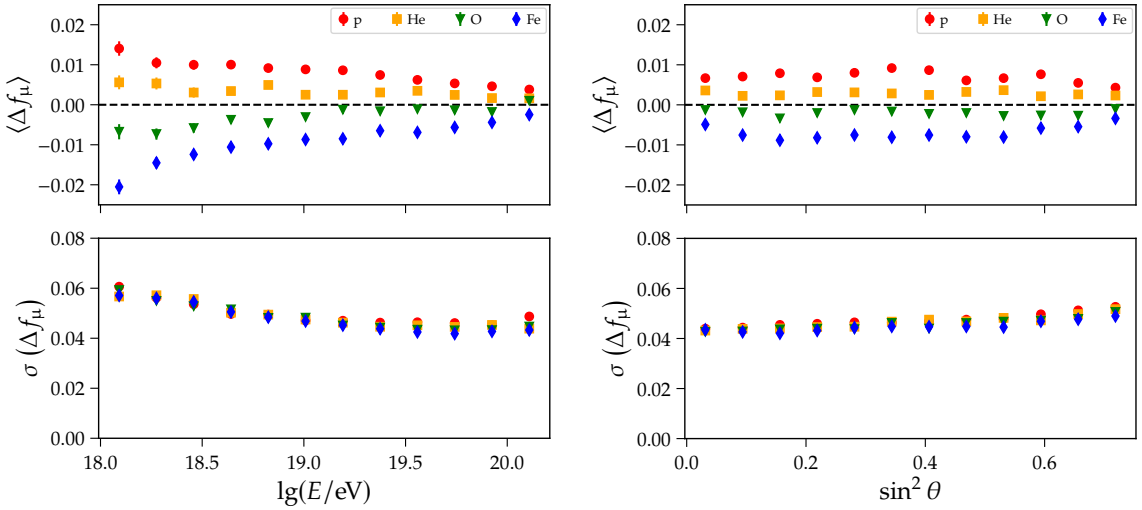


Figure A.9.: Primary-dependent bias (*top*) and resolution (*bottom*) of the best performing CNNs as a function of energy E (*left*) and zenith θ (*right*).

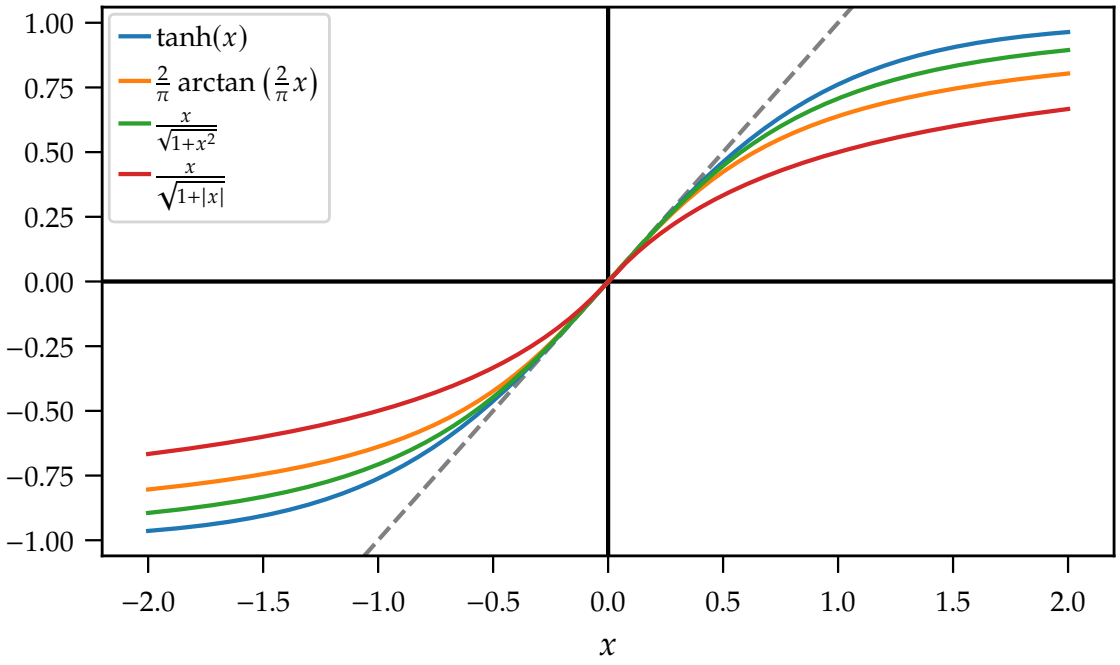


Figure A.10.: Graphs of the used logistic functions, transformed to the range $[-1, 1]$ and normalized to their derivative at $x = 0$.

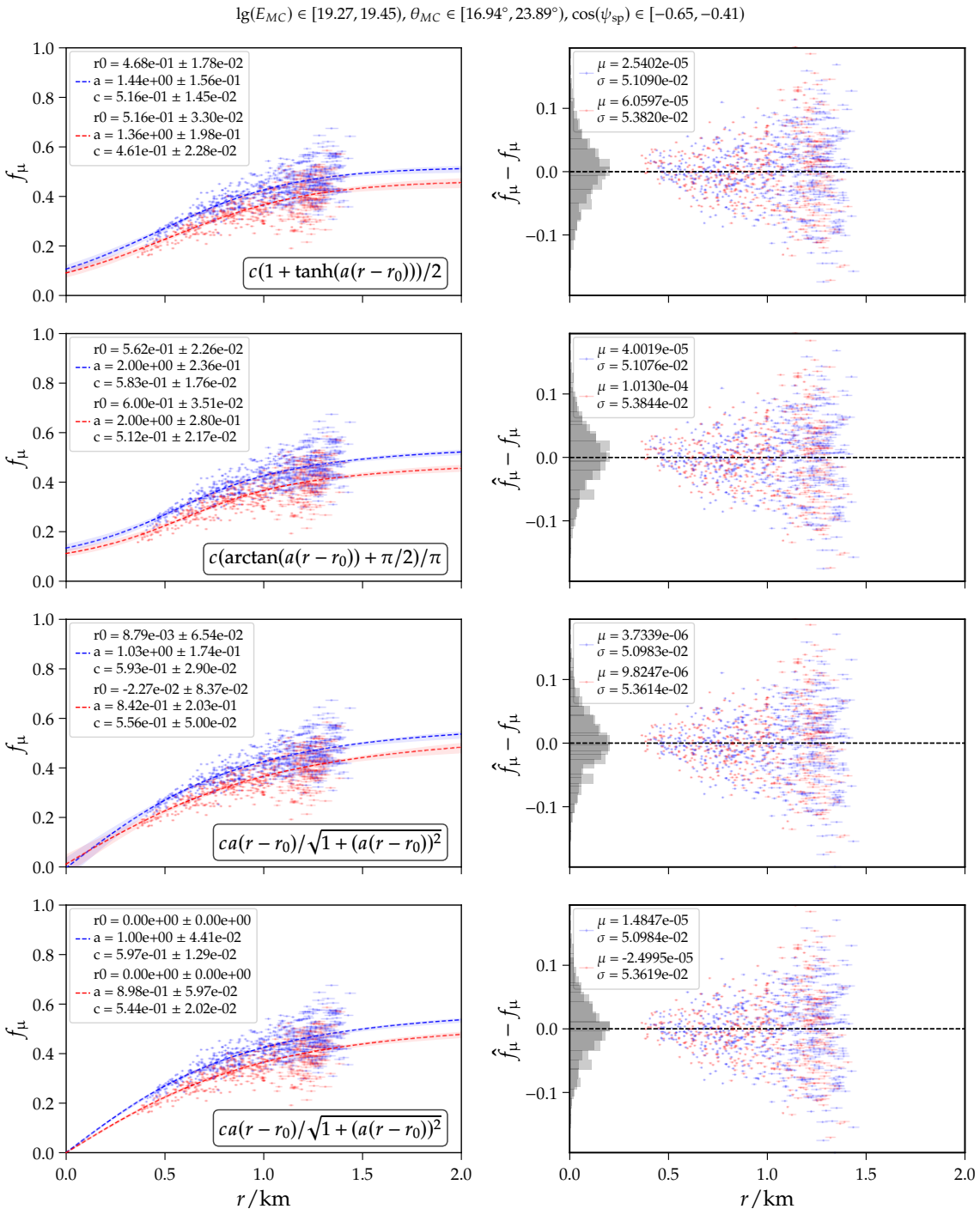


Figure A.11.: Different LFFs fits in dependence of the shower plane distance r for a given energy, zenith angle, and shower-plane azimuth angle. All utilized functions fit the lateral muon fraction equally well since their biases and uncertainties are essentially the same.

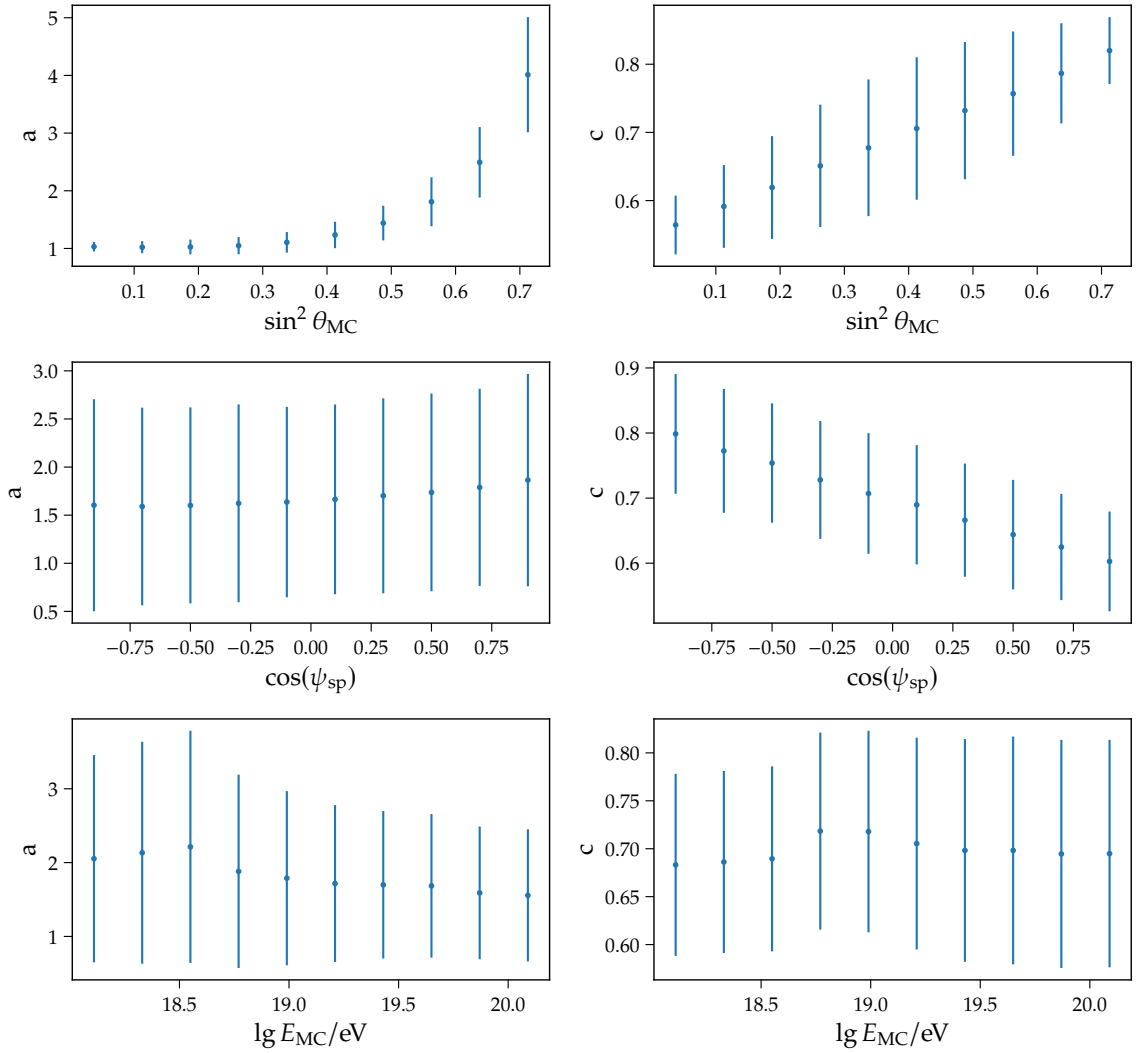


Figure A.12.: Fitted LFF parameters a and c in dependency of the energy, the zenith angle, and the azimuth angle.

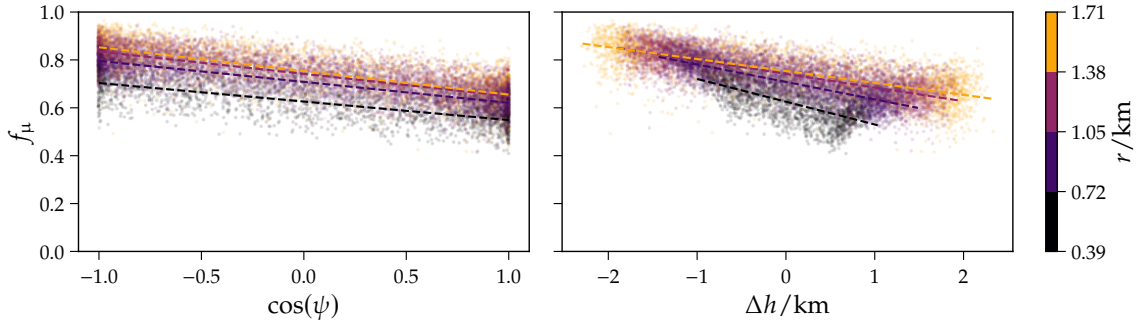


Figure A.13.: Up-/downstream behavior of f_μ in dependency of the variables $\cos \psi_{sp}$ (*left*), and Δh (*right*) for iron induced showers in $\lg E \in [19.54, 19.76]$ and $\theta \in [50.77°, 55.24°]$. The lines represent linear fits in binned distance slices, which exhibit now a change in slope for very inclined showers.

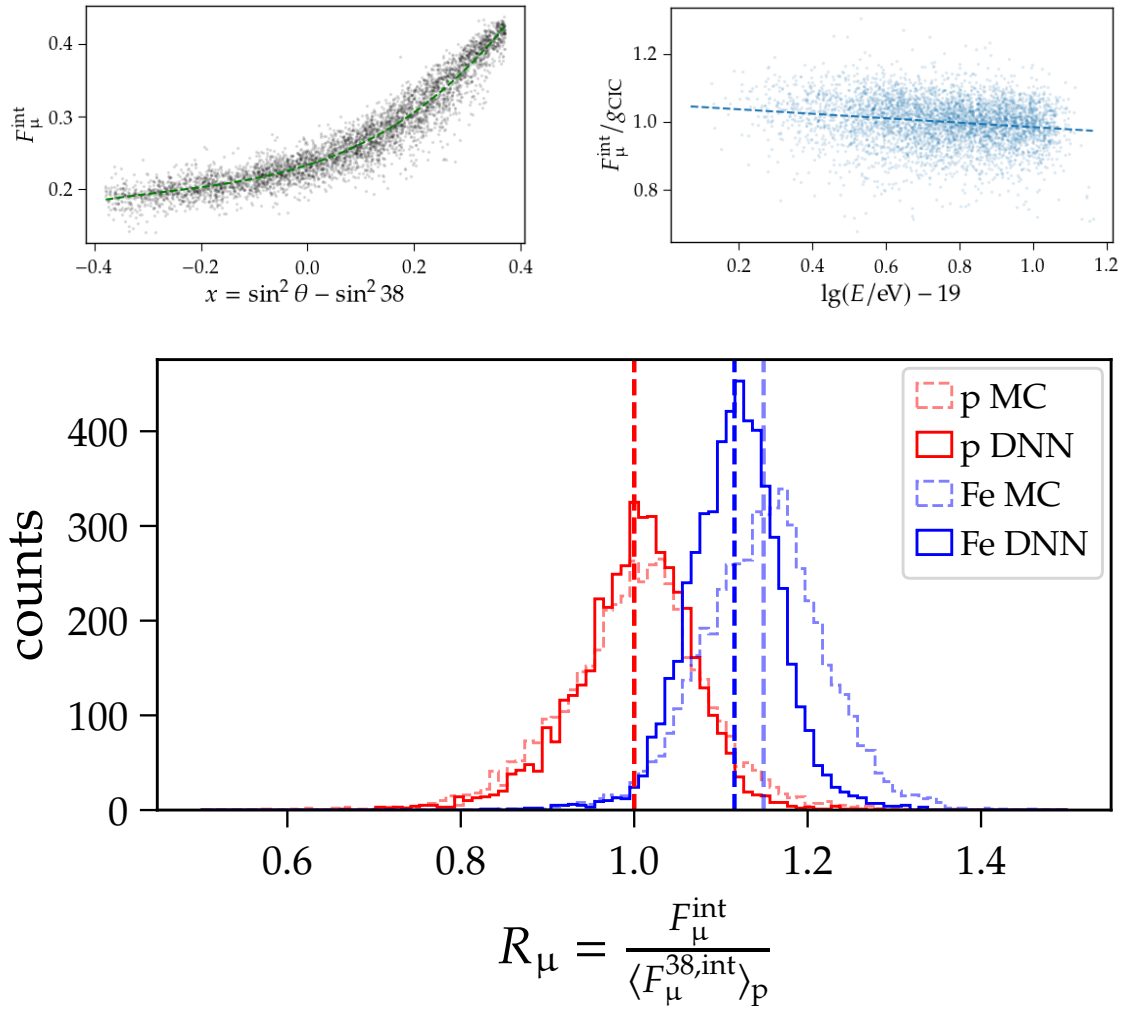


Figure A.14.: Mass-sensitive muon fraction estimator, derived by parametrizing the resulting integral from the LFF in terms of energy (*top right*) and zenith dependency (*top left*). A reasonable separation between proton and iron-induced showers is obtained. To gauge the performance of the estimator, showers that triggered at least five stations after the quality cuts are used.

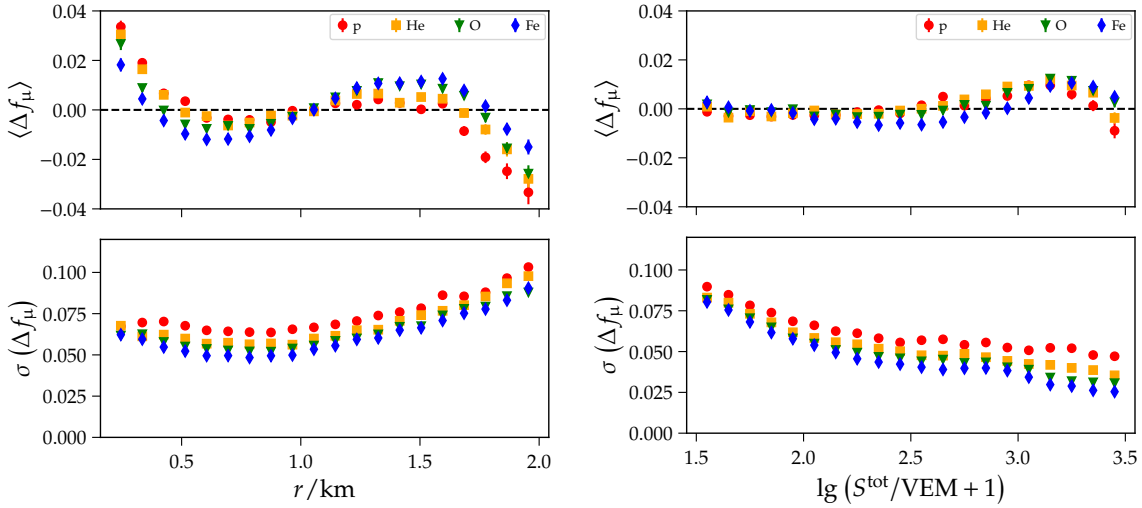


Figure A.15.: Bias (*top*) and resolution (*bottom*) of the global polynomial model as a function of shower plane distance r and logarithmic total signal $\lg(S/\mathrm{VEM}+1)$.

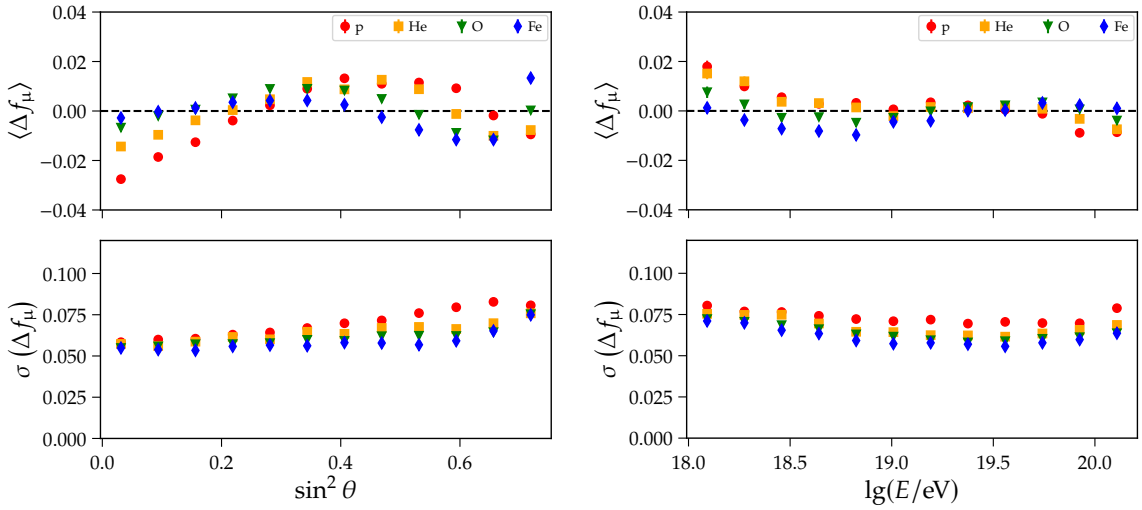


Figure A.16.: Bias (*top*) and resolution (*bottom*) of the global polynomial model as a function of $\sin^2\theta$ and logarithmic energy $\lg E$.

B. Tabular Data

In this chapter, the results of some extensive optimization studies are tabulated.

Table B.1.: Evaluation metrics for the linear regression model for various feature sets. Starting from an matrix formalism motivated feature set containing the integral signals of both detectors, further shower and station level features are added. To enable the linear model to capture interactions between features and non-linearities, polynomial features are created up to an polynomial degree of $P = 2$.

Features	P	$\langle \lvert \Delta f_\mu \rvert \rangle$	$\langle \Delta f_\mu \rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_\mathrm{p}$	$\sigma(\Delta f_\mu)_\mathrm{Fe}$
S^{WCD}, S^{SSD}	1	0.124	0.062	0.154	0.145
	2	0.113	0.051	0.143	0.133
+ θ	1	0.088	0.078	0.116	0.096
	2	0.077	0.070	0.103	0.087
+ r	1	0.077	0.070	0.101	0.086
	2	0.065	0.066	0.087	0.075
+ E	1	0.077	0.070	0.101	0.085
	2	0.064	0.062	0.086	0.074
+ h	1	0.065	0.072	0.086	0.072
	2	0.051	0.066	0.069	0.058
and + Δt_c	1	0.064	0.064	0.085	0.072
	2	0.051	0.059	0.069	0.058
or + Δt_sp	1	0.064	0.064	0.085	0.072
	2	0.051	0.059	0.069	0.058
or + t_f	1	0.063	0.066	0.084	0.071
	2	0.050	0.059	0.067	0.057
or + $t_{1/2}$	1	0.060	0.058	0.080	0.070
	2	0.046	0.043	0.063	0.055
or + Δt_c, Δt_sp, $t_{1/2}$, t_f	1	0.057	0.051	0.076	0.067
	2	0.043	0.037	0.060	0.054

Table B.2.: Performance metrics of the the simple dense network for different feature combinations. For each feature set five networks are trained. The evaluation metrics are computed on the independent test data set and are averaged over the converged models in each ensemble.

Features	$\langle\lvert\Delta f_\mu\rvert\rangle$	$\langle\Delta f_\mu\rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_\mathrm{p}$	$\sigma(\Delta f_\mu)_\mathrm{Fe}$
S^WCD, S^SSD	0.093	0.041	0.121	0.113
+ θ	0.072	0.060	0.094	0.082
+ r	0.061	0.058	0.082	0.071
+ E	0.061	0.057	0.082	0.071
+ h	0.048	0.059	0.064	0.055
and + Δt_hs	0.048	0.059	0.064	0.055
or + Δt_c	0.047	0.053	0.063	0.054
or + Δt_sp	0.047	0.053	0.063	0.054
or + t_f	0.046	0.050	0.062	0.054
or + $t_{1/2}$	0.042	0.032	0.056	0.051
or + Δt_c, Δt_sp, $t_{1/2}$, t_f	0.041	0.027	0.055	0.050

Table B.3.: Performance metrics of the the simple dense network for different feature combinations, this time with S^r instead of S^SSD. For each feature set five networks are trained. The evaluation metrics are computed on the independent test data set and are averaged over the converged models in each ensemble. The results resemble the ones in Table B.2.

Features	$\langle\lvert\Delta f_\mu\rvert\rangle$	$\langle\Delta f_\mu\rangle_{\mathrm{p,Fe}}$	$\sigma(\Delta f_\mu)_\mathrm{p}$	$\sigma(\Delta f_\mu)_\mathrm{Fe}$
S^WCD, S^r	0.093	0.039	0.121	0.113
+ θ	0.072	0.060	0.094	0.083
+ r	0.060	0.058	0.082	0.072
+ E	0.061	0.057	0.082	0.072
+ h	0.048	0.059	0.064	0.055
and + Δt_hs	0.048	0.059	0.064	0.055
or + Δt_c	0.047	0.052	0.063	0.055
or + Δt_sp	0.047	0.052	0.063	0.054
or + t_f	0.046	0.050	0.062	0.054
or + $t_{1/2}$	0.041	0.032	0.056	0.051
or + Δt_c, Δt_sp, $t_{1/2}$, t_f	0.041	0.027	0.055	0.050

Table B.4.: Hyperparameter Scan of the CNN Trace Analyzer.

Conv1D Layer	Pooling	n_b	$n_{b,l}$	n_{tf}	$\langle\lvert\Delta f_\mu\rvert\rangle$	$\langle\Delta f_\mu\rangle_{p,Fe}$	$\sigma(\Delta f_\mu)_p$	$\sigma(\Delta f_\mu)_{Fe}$
Conv	Avg	6	16, 16, 16	96	0.0359	0.0136	0.0482	0.0454
Conv	Avg	6	16, 16	96	0.0367	0.0148	0.0486	0.0459
Conv	Avg	6	16	96	0.0360	0.0138	0.0483	0.0458
Conv	Avg	7	16, 16, 16	48	0.0404	0.0303	0.0542	0.0489
Conv	Avg	7	16, 16	48	0.0357	0.0136	0.0481	0.0454
Conv	Avg	7	16	48	0.0362	0.0141	0.0487	0.0459
Conv	Avg	8	16, 16, 16	16	0.0358	0.0135	0.0481	0.0455
Conv	Avg	8	16, 16	16	0.0361	0.0137	0.0484	0.0456
Conv	Avg	8	16	16	0.0362	0.0139	0.0487	0.0460
Conv	Max	6	16, 16, 16	96	0.0358	0.0142	0.0482	0.0454
Conv	Max	6	16, 16	96	0.0357	0.0138	0.0481	0.0455
Conv	Max	6	16	96	0.0359	0.0142	0.0484	0.0456
Conv	Max	7	16, 16, 16	48	0.0357	0.0142	0.0482	0.0454
Conv	Max	7	16, 16	48	0.0358	0.0142	0.0482	0.0455
Conv	Max	7	16	48	0.0360	0.0146	0.0485	0.0456
Conv	Max	8	16, 16, 16	16	0.0361	0.0145	0.0485	0.0459
Conv	Max	8	16, 16	16	0.0358	0.0143	0.0483	0.0454
Conv	Max	8	16	16	0.0360	0.0145	0.0485	0.0458
Separable	Avg	6	16, 16, 16	96	0.0500	0.0660	0.0669	0.0560
Separable	Avg	6	16, 16	96	0.0497	0.0652	0.0667	0.0558
Separable	Avg	6	16	96	0.0367	0.0150	0.0492	0.0464
Separable	Avg	7	16, 16, 16	48	0.0500	0.0660	0.0667	0.0558
Separable	Avg	7	16, 16	48	0.0500	0.0658	0.0667	0.0559
Separable	Avg	7	16	48	0.0365	0.0146	0.0492	0.0464
Separable	Avg	8	16, 16, 16	16	0.0498	0.0653	0.0665	0.0558
Separable	Avg	8	16, 16	16	0.0502	0.0650	0.0666	0.0562
Separable	Avg	8	16	16	0.0431	0.0398	0.0579	0.0511
Separable	Max	6	16, 16, 16	96	0.0498	0.0656	0.0668	0.0558
Separable	Max	6	16, 16	96	0.0503	0.0665	0.0668	0.0559
Separable	Max	6	16	96	0.0365	0.0151	0.0491	0.0462
Separable	Max	7	16, 16, 16	48	0.0498	0.0654	0.0666	0.0559
Separable	Max	7	16, 16	48	0.0502	0.0669	0.0666	0.0560
Separable	Max	7	16	48	0.0365	0.0152	0.0491	0.0462
Separable	Max	8	16, 16, 16	16	0.0500	0.0657	0.0668	0.0558
Separable	Max	8	16, 16	16	0.0500	0.0655	0.0665	0.0560
Separable	Max	8	16	16	0.0365	0.0150	0.0491	0.0463

C. Neural Network Architectures

In this chapter, some of the used basic building blocks, sub-networks, custom layers and NN architectures are depicted in detail.

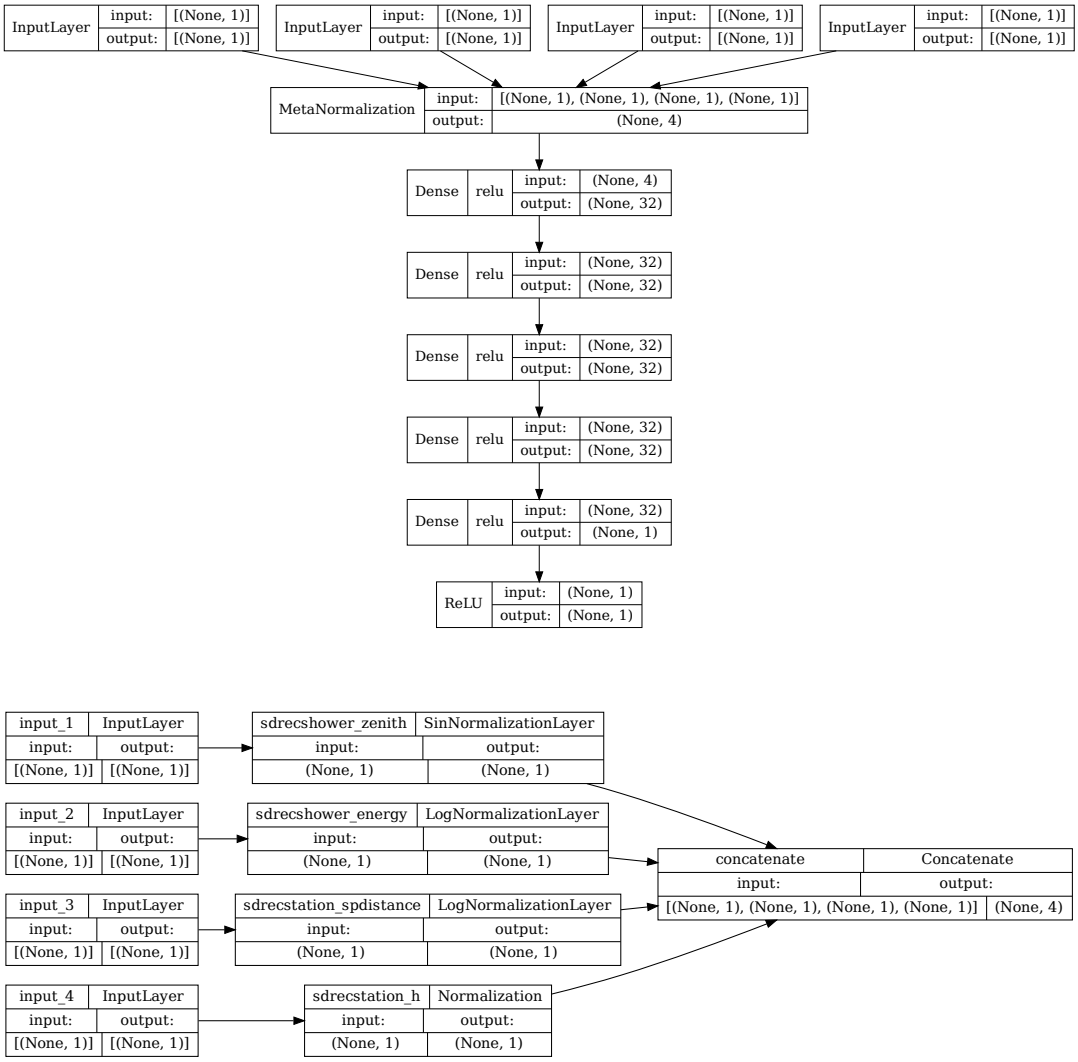


Figure C.1.: Simple dense NN architecture (*top*) to predict f_μ. In this model four features are used as inputs, where each feature is normalized separately and concatenated together afterwards in the *MetaNormalization* layer (*bottom*). The MetaNormalization layer is followed by four fully connected dense layers, each with a ReLU activation and 32 units. The output layer is used for the prediction of f_μ and thus has one unit followed by a ReLU with a maximum value of 1.

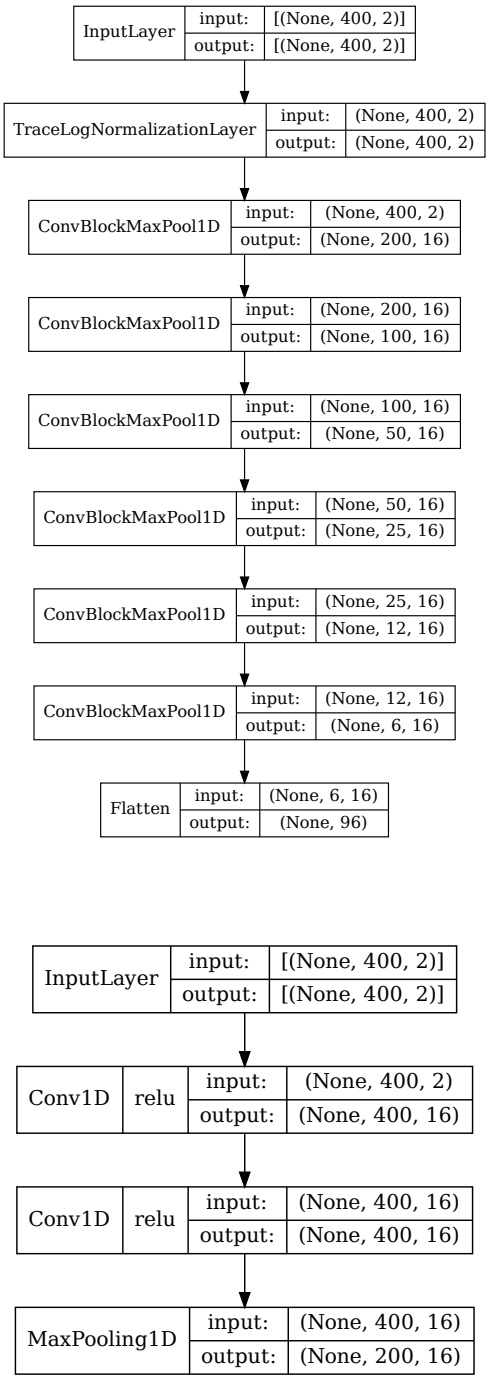


Figure C.2.: Trace analyzer (*top*) of the CNN model architecture. It consists of consecutively stacked convolution blocks (*bottom*).

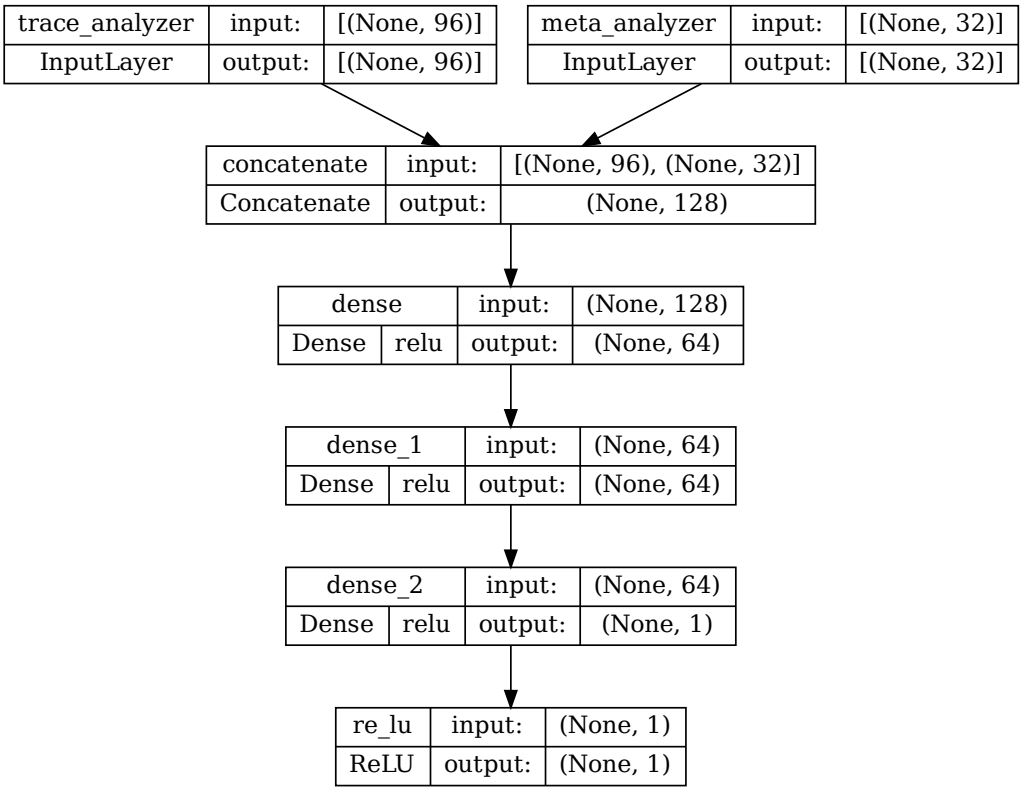


Figure C.3.: Simple predictor of the CNN. It unifies the outputs of the Trace Analyzer and the Meta Analyzer by concatenation. Three dense layers extract further information to ultimately predict the f_μ target. Because $f_\mu \in [0, 1]$ the output is given by a ReLU that has a maximum value of 1.

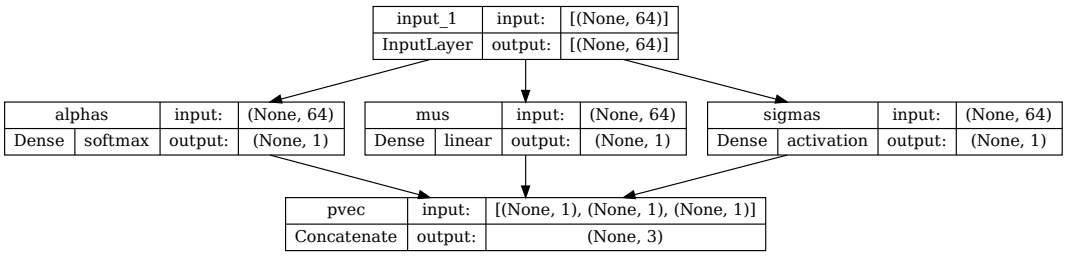


Figure C.4.: Illustration of the Mixture Density Head that is substituted as the new output of the base Predictor. It separately predicts the parameters of a Gaussian mixture model in order to learn the underlying distribution of f_μ. The parameters are given by the mean $\mu(f_\mu)$, the standard deviation $\sigma(f_\mu)$ and the mixture weights α_k. The number of components is chosen to be $k = 1$.

D. Experiment Tracking

In the machine learning lifecycle it is indispensable to keep track of experiments and their results, as an immense number of combinations among datasets, model classes or hyperparameters are possible. This is especially important when many different experiments are conducted with several runs each, and their results should be compared with each other. In this work the MLflow [101] framework is used extensively to accomplish an organized and central tracking system for all carried out experiments. MLflow is an open source platform for the machine learning lifecycle, which enables one to track experiments, package and share models, manage and reproduce runs, register and deploy models. The tracking of an experiment is accomplished by logging the associated hyperparameters, metrics and their history, and artifacts of each run. It is capable of handling almost any machine learning library and programming language, and is thus easy to integrate into existing code snippets with minimal changes. The MLflow tracking server can be hosted locally and can either be accessed via a user interface through the web browser or programmatically via an API. The framework heavily simplifies the experimentation process and increases reproducibility, transparency and progress in machine learning research.