Classical and Machine Learning Approaches for Determination of the Energies of Ultra-High-Energy Photons using the Surface Detector of the Pierre Auger Observatory

Master's Thesis of

Daniel Rech

At the KIT Department of Physics
Institut für Astroteilchenphysik (IAP)

First examiner:	Prof. Dr. Ralph Engel
Second examiner:	Priv. Doz. Dr. Roger Wolf
Advisors:	M.Sc. Fiona Ellwanger
	Dr. Steffen Hahn
	Dr. Markus Roth

07.06.2023 – 07.06.2024

Karlsruher Institut für Technologie
Fakultät für Physik
76128 Karlsruhe

I declare that I have developed and written the enclosed thesis completely by myself. I have not used any other than the aids that I have mentioned. I have marked all parts of the thesis that I have included from referenced literature, either in their original wording or paraphrasing their contents. I have followed the by-laws to implement scientific integrity at KIT.

Karlsruhe,

.....	.....
(Daniel Rech)	(Prof. Dr. Ralph Engel)

Abstract

All measured cosmic rays above 10^{18} eV are believed to be composed of protons or atomic nuclei. However, the potential existence of ultra-high-energy photons would be significant, as it would put strong constraints on our current assumptions of the mass composition of cosmic rays at ultra-high energies. Additionally, the production of these ultra-high-energy photons is suggested in some beyond-the-Standard-Model theories. Given that direct detection methods are not viable for cosmic ray energies above 10^{12} eV, the primary particle must be reconstructed from extensive air showers, with the atmosphere serving as a calorimeter for the cosmic rays. This thesis presents two methods for predicting the exact energy of non-preshowering photons by exclusively analyzing extensive air showers in the range from 10^{18} eV to $10^{20.2}$ eV using Monte Carlo simulations.

Both methods utilize the water-Cherenkov-based surface detector of the Pierre Auger Observatory. The first method is based on the same classical reconstruction technique used for hadron-induced air showers. It is adapted to photon-induced air showers by optimizing the parameterization of the lateral distribution function and the direct energy calibration. An additional parameterization, not used in the energy reconstruction, shows that the signal of photon-induced air showers exhibits a significant asymmetry depending on the azimuth angle to the shower axis.

The second method leverages deep learning techniques. Two distinct neural network architectures are tested: one based mainly on two-dimensional separable convolutions, the other based on a Vision Transformer architecture. The hyperparameters of both implementations will be fine-tuned via Bayesian optimization. It is found that the Vision Transformer exhibits a slight advantage over the other architecture.

A comparison of the developed models reveals that the Vision Transformer-based energy reconstruction provides accurate results for Monte Carlo energies between $10^{18.2}$ eV and $10^{19.75}$ eV. The machine-learning algorithms show edge effects, manifesting as under- and overpredictions at low and high energies, which will be investigated further. The classic reconstruction exhibits a bias linked to both energy and the zenith angle of an event, suggesting that the direct energy calibration used in this thesis does not fully model the characteristics of photon-induced showers. However, the classic reconstruction performs reasonably well in the calibration interval, and potential improvements are suggested for future studies. When compared to an Lookup Table based approach as developed in Reference [1], the Vision Transformer-based reconstruction yields better results.

Finally, the applicability to preshowers is tested, with the finding that the energy of preshowering events can be reconstructed remarkably well. Furthermore, a network trained on a data set that includes preshowers achieves the best overall performance.

Zusammenfassung

Alle gemessenen kosmischen Strahlen oberhalb von 10^{18} eV bestehen vermutlich aus Protonen oder Atomkernen. Die mögliche Existenz von ultrahochenergetischen Photonen wäre jedoch bedeutsam, da sie Annahmen über die Massenkomposition von ultrahochenergetischen kosmischen Strahlen stark beschränken würden. Darüber hinaus wird die Produktion dieser Photonen in einigen über das Standardmodell hinausgehenden Theorien vorgeschlagen. Da direkte Nachweismethoden für Energien oberhalb von 10^{12} eV nicht praktikabel sind, müssen Primärteilchen aus Luftschauerexperimenten rekonstruiert werden, welche die Atmosphäre als Kalorimeter benutzen. Diese Arbeit präsentiert zwei Methoden zur Vorhersage der exakten Energie von Photonen ohne *Preshower* durch die ausschließliche Analyse von Monte Carlo Luftschauern zwischen 10^{18} eV und $10^{20.2}$ eV.

Beide Methoden nutzen den Wasser-Cherenkov basierten Oberflächendetektor des Pierre-Auger-Observatoriums. Die erste Methode basiert auf der klassischen Rekonstruktionstechnik für hadroninduzierte Luftschauer, welche durch Optimierung der Parametrisierung der lateralen Verteilungsfunktion und der direkten Energiekalibrierung an photoninduzierte Luftschauer angepasst wird. Eine zusätzliche Parametrisierung, die nicht in der Energierekonstruktion verwendet wird, zeigt, dass das Signal von photoninduzierten Luftschauern eine signifikante Asymmetrie in Abhängigkeit vom Azimutwinkel zur Schauerachse aufweist.

Die zweite Methode basiert auf Deep-Learning. Zwei verschiedene neuronale Netzwerkarchitekturen werden getestet: eine basiert auf zweidimensionalen separierbaren Faltungen, die andere auf einer Vision Transformer Architektur. Die besten Hyperparameter beider Implementierungen werden mittels Bayes'scher Optimierung gefunden. Der Vision Transformer weist eine leicht besser Leistung gegenüber der anderen Architektur auf.

Ein Vergleich der entwickelten Modelle zeigt, dass die Energierekonstruktion des Vision Transformers genauere Ergebnisse für Monte-Carlo-Energien zwischen $10^{18.2}$ eV und $10^{19.75}$ eV liefert. Die maschinellen Lernalgorithmen zeigen Randeffekte, die sich als Unter- und Überschätzungen bei niedrigen und hohen Energien manifestieren, welche näher untersucht werden. Die klassische Rekonstruktion zeigt eine Verzerrung, die sowohl mit der Energie als auch mit dem Zenitwinkel eines Ereignisses verbunden ist, was darauf hindeutet, dass die in dieser Arbeit verwendete direkte Energierekalibrierung die Eigenschaften photoninduzierter Schauer nicht vollständig modelliert. Die klassische Rekonstruktion funktioniert jedoch im Kalibrierintervall recht effizient und potenzielle Verbesserungen für zukünftige Studien werden vorgeschlagen. Im Vergleich zu einem Ansatz basierend auf einer Lookup-Tabelle, entwickelt in Referenz [1], liefern die Schätzungen des Vision Transformers bessere Ergebnisse.

Schließlich wird die Anwendbarkeit auf *Preshower* getestet, wobei festgestellt wird, dass die Energie dieser bemerkenswert gut rekonstruiert werden kann. Darüber hinaus erreicht ein Netzwerk, trainiert mit einem Datensatz der *Preshower* einschließt, die beste Gesamtleistung.

Contents

Abbreviations

ADC Analog to Digital Converter. 15, 16
ALP Axion Like Particle. 9
AMIGA Auger Muon and Infill for the Ground Array. 12, 13
AoP Area over Peak. 34, 51
BSM Beyond the Standard Model. 1, 8, 9
CIC Constant Intesity Cut. 21, 75
CNN Convolutional Neural Network. 24, 25
CR Cosmic Ray. 1, 3, 4, 8, 11, 32
DEC Direct Energy Calibration. 21, 33, 45, 48, 49, 64, 71–79, 81, 82, 100, 102, 103
EAS Extensive Air Shower. 1, 4–6, 11, 15, 19
FD Fluoroscence Detector. 11–13, 15, 21
HE High-Energy. 4, 11, 13
HEAT High Elevation Auger Telescopes. 11, 12
LDF Lateral Distribution Function. 1, 16, 17, 31, 33, 34, 37–47, 65, 71, 72, 77, 78, 81, 83, 84, 102
LPM Landau, Pomeranchuk, and Migdal. 5
LUT Lookup Table. i, 2, 71, 76–79, 84
MC Monte Carlo. 31–33, 38, 41, 42, 51, 62, 65, 67, 72, 75–77, 98
ML Machine Learning. 2, 31, 51, 64, 68, 75, 76, 81, 83–85
MLP Multi-Layer Perceptron. 29, 56, 57
MSA Multiheaded Self Attention. 27, 56, 57
NN Neural Network. 23, 24, 26, 31, 32, 51, 53, 63, 65, 66, 71, 72, 76, 77, 81
OOT out-of-time. 68–70
PMT Photomultiplier Tube. 11–13, 16, 34, 51, 52, 65
SD Surface Detector. 11–13, 17, 18, 41
sepConv2D two-dimenesional separable convolution. 54–57, 59–61, 63, 64, 76, 83, 84, 97, 98
SHDM Super Heavy Dark Matter. 9
UHE Ultra-High-Energy. 3, 5, 6, 8, 9, 11, 15, 83
UHECR Ultra-High-Energy Cosmic Ray. 1, 4, 6, 8, 11, 12
UMD Underground Muon Detector. 12, 13
UUB Upgraded Unified Board. 13
VEM Vertical Equivalent Muon. 15, 16, 34, 51, 52
ViT Vision Transformer. 28, 29, 54, 56, 57, 59–66, 68–70, 72, 74–79, 81, 82, 84, 85
WCD Water Cherenkov Detector. 12, 13, 15, 21, 31

1. Introduction

Since their discovery 102 years ago by Victor Hess [2], Cosmic Rays (CRs) have become invaluable tools for modern science. The majority of CRs are positively charged nuclei ranging from single protons to heavier elements such as iron. They arrive at Earth from both inside and outside our galaxy. Charged Ultra-High-Energy Cosmic Rays (UHECRs) ($E \geq 10^{18}$ eV) are of great interest, as the particles that arrive at Earth did not experience major deflections due to magnetic fields and still point to their source within some degrees. However, this thesis focuses on UHECRs with photons as primary particles.

To date, no γ-rays with energies above 1.4×10^{15} eV have been identified [3]. Nonetheless, certain Beyond the Standard Model (BSM) models, like Lorentz invariance breaking interactions [4, 5], super heavy dark matter models [6, 7], or Axion Like Particles [8], predict a significant flux of photons above 10^{18} eV. Even in the absence of BSM physics, a small flux of photon-UHECRs is expected. The detection of photons at these high energies would impose significant constraints on the mass composition of charged UHECRs in the case of cosmogenic photons. Moreover, the detection of ultra-high-energy γ-rays is important for theories that predict the photon production in point sources, as the photons or their BSM propagators are not affected by the galactic magnetic fields.

Given the low rate of γ-rays expected by the models, direct detection methods are not viable. The Pierre Auger Observatory can overcome this limitation by using the atmosphere as a calorimeter for the photons. When the γ-rays interact with the atoms in the air or the geomagnetic field (a process known as preshowering), they initiate an Extensive Air Shower (EAS) in the atmosphere, detectable by the observatory. Reconstructing the properties of the primary particle from an EAS is a non-trivial task, requiring detailed knowledge of the detector response and high-energy shower development processes. Specialized estimators for the energy of photon-induced showers are necessary, as the unaltered standard reconstruction has been found to underestimate their energy by approximately 70 % [9, Chapter 4.5] since it was optimized for hadron-induces air showers.

Assuming γ-rays with energies above 10^{18} eV can be successfully identified, methods to estimate their exact energies are essential. This thesis presents two methods for estimating the energy of photon-induced showers from the shower footprint as measured by the Pierre Auger surface detector by employing Monte Carlo simulations. The first method adapts the classical reconstruction used for hadron primaries, optimizing it for photon-induced showers. This includes a new parameterization of the Lateral Distribution Function (LDF) for ultra-high-energy γ-ray showers and the direct energy calibration based on the shower size and its zenith angle. Additionally, the amplitude of the shower signal asymmetry is parameterized as a function of the azimuth angle relative to the shower axis, although this is not included in the energy reconstruction developed in this thesis due to time

constraints. However, it may be useful for future works on the topic as the asymmetry was found to be non-negligible.

The second method involves Machine Learning (ML) techniques for the energy reconstruction. To achieve the best possible performance, two different architectures are tested and optimized regarding their hyperparameters. The first relies partly on two-dimensional separable convolutions while the second employs a more involved Vision Transformer architecture [10]. The limitations of the ML estimators and possible improvements are discussed in depth.

Both methods are compared to each other and to a previous approach as detailed in Reference [1] which employs a look-up table as the energy estimator. Moreover, as both methods in this work are calibrated or trained without preshowers, the applicability to preshower events will be investigated. In the case of the ML reconstruction, additional networks will be trained with expanded data sets that include preshowering events.

The following section outlines the structure of this thesis. Chapter 2 starts by providing a general introduction to cosmic rays and follows with a more in-depth look at ultra-high-energy γ-rays. Chapter 3 introduces the Pierre Auger Observatory and its hybrid detector layout. Following this, Chapter 4 focuses on the general classical reconstruction as it is performed for hadron-induced showers. Chapter 5 provides an overview of ML methods, with a more in-depth explanation of convolutional layers and the Vision Transformer architecture. An adaptation for the reconstruction of photon-induced and the parameterization of the azimuthal signal asymmetry can be found in Chapter 6, followed by the application of the ML model in Chapter 7. A general comparison of the here-developed energy estimators, as well as the Lookup Table approach is conducted in Chapter 8. Lastly, the applicability to preshowers is discussed in Chapter 9.

2. Cosmic Rays

2.1. Energy Spectrum

Cosmic Rays (CRs) are high-energy particles from outer space. They were first discovered by Victor Franz Hess in the year 1912 by measuring the radiation intensity at great heights above sea level with balloon experiments. The common belief at the time was that the background radiation was purely terrestrial due to radioactive decays in Earth's crust and should only decrease with height. However, Hess observed that after a certain threshold, the radiation intensity started to increase again after decreasing at first [2].

Since then, the study of CRs has become an integral part of particle physics and astroparticle physics due to the high energies achieved by them. One example would be the discovery of the muon in 1936 by Carl D. Anderson and Seth Neddermeyer who found particles in cosmic radiation that have a e/m value that is much smaller than for electrons but are not as ionizing as protons [11].

Although the majority of CRs are composed of protons or heavier nuclei, stripped from all their electrons, the term is inclusive to all space-born high-energy particles. The energy-dependent flux of CRs follows a broken power law: $F(E) \propto E^{\gamma(E)}$, see Figure 2.1. The negative spectral index γ is not constant over the whole energy range and changes at multiple points of the spectrum [13]:

- $E \approx 4{\times}10^{15}$ eV: *The (first) knee* - a decrease in the slope from $\gamma \approx -2.7$ to $\gamma = -3.1$ [14]. The reason for this is probably the start of a mass-depended cutoff for most galactic accelerators, starting with the light elements [15][12, Chapter 5: Cosmic Rays].
- $E \approx 400 \times 10^{15}$ eV: *The second knee* - a further steepening of the slope to $\gamma \approx -3.32$ [16, 14] with the cutoff for heavy nuclei, especially the iron component [12, Chapter 5: Cosmic Rays].
- $E \approx 3 \times 10^{18}$ eV: *The ankle* - a slight hardening of the spectrum to $\gamma \approx -2.6$ [14]. It is probably caused by a new population of Ultra-High-Energy (UHE) particles that overtake the lower energy population. This is often interpreted as the change from galactic CRs sources to extragalactic CR sources [17][12, Chapter 5: Cosmic Rays].
- $E \approx 1.3 \times 10^{19}$ eV: *The instep* - a change to $\gamma \approx -3.1$. An explanation for this feature would be the cutoff due to the photodisintegration effect of the heavier chemical components and the spectrum emitted by the sources. However, an undisputed explanation for the instep feature has not yet been found [18].

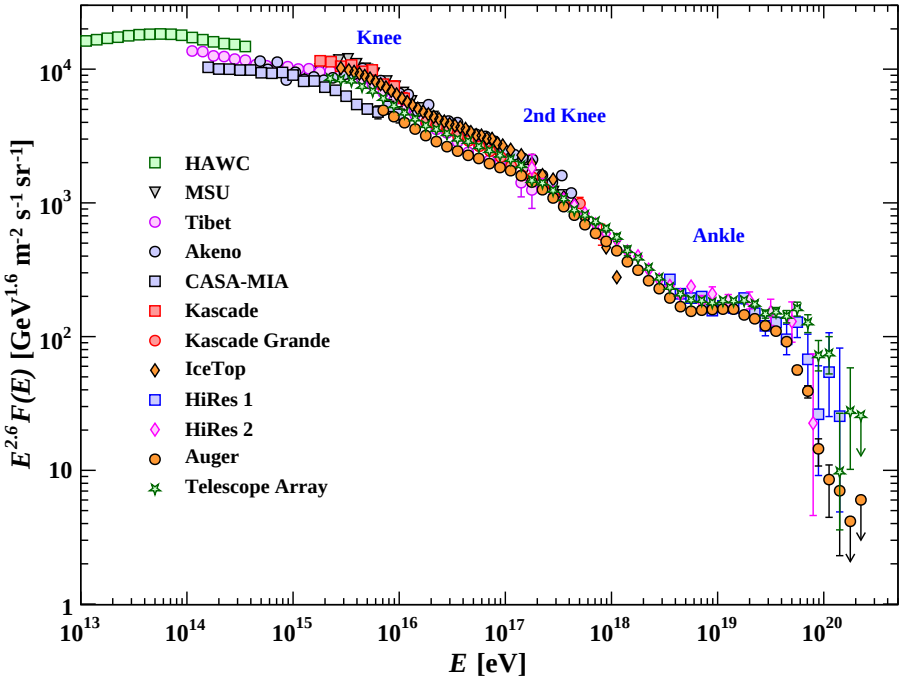


Figure 2.1.: **All particle CR spectrum as a function of the energy.** Accumulated data from different experiments, distinguished by the different markers. The differential spectrum has been multiplied by $E^{2.6}$ to help differentiate the marked features. The instep and the suppression are not labeled in this representation. Figure from Reference [12].

- $E \gtrapprox 5.0 \times 10^{19}$ eV: The suppression - Softening the spectrum to $\gamma \approx -5.1$ [18]. When cosmic rays produced by extragalactic accelerators reach the energies where the GZK effect for light nuclei and the photodisintegration for heavier nuclei are relevant and the effective mean free travel length is greatly reduced—hindering the UHECRs from reaching Earth (see *scenario 1* in Section 2.3).

The highest energy of a single CR ever observed is estimated to exceed 3×10^{20} eV, which is far beyond the reach of any human-made particle accelerator. It was measured by the Fly's Eye experiment on October 15, 1991 [19] and given the title *Oh-My-God particle*.

2.2. Extensive Air Showers

With increasing energy the flux of CRs gets progressively smaller, ruling out the use of direct measurement methods like the AMS-02 experiment. This led to the development of ground-based detectors like the Pierre Auger Observatory (see Chapter 3) which use Earth's atmosphere as a calorimeter. This is possible, as HE-CRs and Ultra-High-Energy Cosmic Rays (UHECRs) produce a so-called Extensive Air Shower (EAS). EASs are a cascade of secondary particles, induced by the interaction of the primary particle with the molecules of the atmosphere. Note that in this thesis, *primary particle* denotes the particle,

that arrives from outer space and causes the shower, whereas *secondary particle* denotes a particle produced in the air shower. If the primary particle is a hadron, the showers will have a collimated hadronic core [12]. This core is the starting point for electromagnetic sub-showers, mostly induced by the decay $\pi^0 \rightarrow \gamma\gamma$ [12]. The electrons and positrons from these sub-showers make up the majority of the charged particles produced in an air shower [12]. Even more numerous are the photons from the electromagnetic cascades, with 10 times as many particles. On the other hand, the number of muons, which are produced by the decays of charged mesons, is one order of magnitude lower [12].

The shower can then be measured at the ground level, either by using detectors that are sensitive to individual particles from the shower like scintillation panels and water Cherenkov tanks, or via secondary effects like radio signals produced by the showers and the fluorescence light from air molecules that get excited by the shower particles. By analyzing the data, the trajectory, energy, and type of the primary particle can be estimated. The algorithm of how this is accomplished with water Cherenkov detectors for the Pierre Auger Observatory is explained in Chapter 4.

2.3. Ultra-High-Energy Photons

Photon-induces showers at UHEs have a signature significantly different from hadron-induces showers. The showers from photon primaries contain fewer muons and produce mainly electromagnetic cascades, given the electromagnetic nature of the photon. Furthermore, they have a deeper shower maximum on average. At 10^{19} eV the difference is approximately $200\,\mathrm{g\,cm^{-2}}$ to proton-induced showers [1].

A deeper shower maximum results in a more curved showerfront when reaching the surface. The differences in the muonic and electromagnetic content, as well as the difference in the shower maximum and the curvature, are illustrated in Figure 2.4. The depth of the shower maximum increases further for primary energies above 10^{19} eV when the Landau, Pomeranchuk, and Migdal (LPM) effect [24, 25] starts to become relevant. It describes the destructive interference from multiple scattering centers as the photon energy increases. This effectively decreases the Bethe-Heitler cross-section for the pair production, leading to a deeper shower maximum on average. The LPM effect is opposed by the preshower effect [26]. In a preshower the initial interaction of the photon does not occur inside the atmosphere, but rather outside the atmosphere in the form of a pair production with the Earth's magnetic field. The products of the initial preshower then reach the atmosphere and start an EAS. The likelihood for a photon to start a preshower increases with the energy of the latter. A review of the relevant effects and the in-depth phenomenology of photon-induced showers can be found in [27].

Figure 2.3 provides further quantitative information about the average shower maximum. It shows the simulated X_{max} for photon-induced showers in comparison to showers initiated by protons (red) or iron (blue), simulated for different hadronic interaction models. All showers are simulated with a zenith angle of 60°. The approximate X_{max} difference of

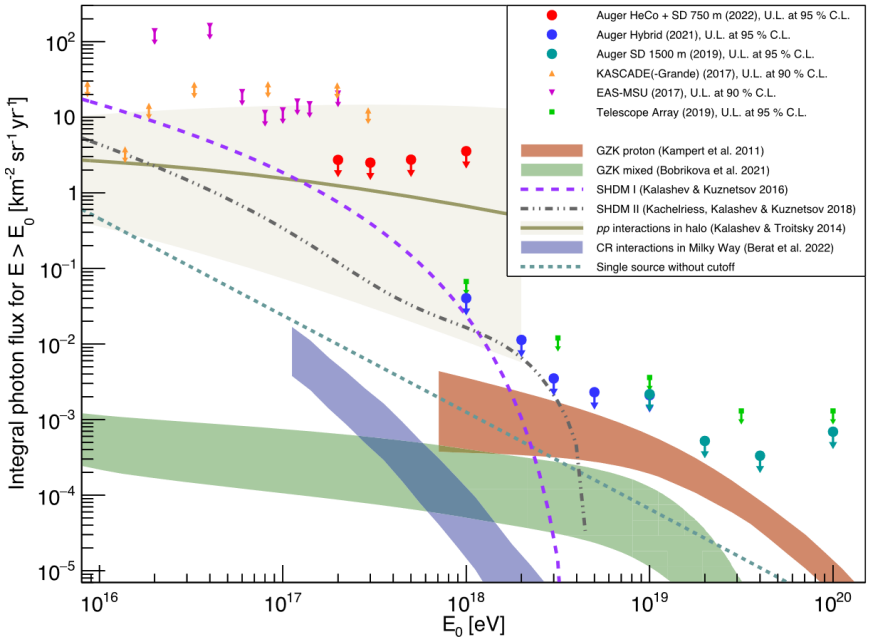


Figure 2.2.: **Integrated flux limits beyond energies of 10^{16} eV for photon primaries.** The arrows and markers show the limits of the flux determined by different experiments. The colored bands show the expected flux for different UHE photon production processes, the gray and purple dashed lines show the expected flux for the decay of super-heavy dark matter and the doted turquoise line shows the extension of a putative source where no cutoff in the spectrum exits, modeled after HAWC J1825-134. The graphic originates from [20].

$100\,\mathrm{g\,cm^{-2}}$ between photons and protons at energies of 10^{18} eV is about the same as the difference between proton and iron showers. For larger energies, the average X_{max} of the photon-induced showers grows more rapidly than that of showers initiated by protons or iron. This is attributable to the increasing influence of the LPM effect. Showers from the North are affected more strongly by preshowering, which explains the difference in X_{max} compared to showers from the South or showers in the absence of the geomagnetic field. At energies above $\approx 5 \times 10^{19}$ eV, the preshowering effect overcomes the LPM effect and the average X_{max} value decreases again for higher primary energies.

Until today, no EASs from photon primaries were observed at UHEs. Multiple photons with energies at or below 1.4×10^{15} eV were recorded by the LHAASO experiment [3], but in the region beyond 1 EeV only upper limits are set for the flux of Ultra-High-Energy (UHE) γ's. Many searches are conducted for UHE photons [7, 20, 1, 23, 9, 28, 29], resulting in strict limits on the photon flux. The integral flux limit as a function of threshold energy can be seen in Figure 2.2.

In the UHE region, a small diffuse flux of γ-rays is expected by the standard model of astrophysics. However, this diffuse flux is not composed of photons directly accelerated at the sources, but rather originates from the interactions of UHECRs with the background photon fields, interspersing the intergalactic medium [20, 30]. Additional photons are

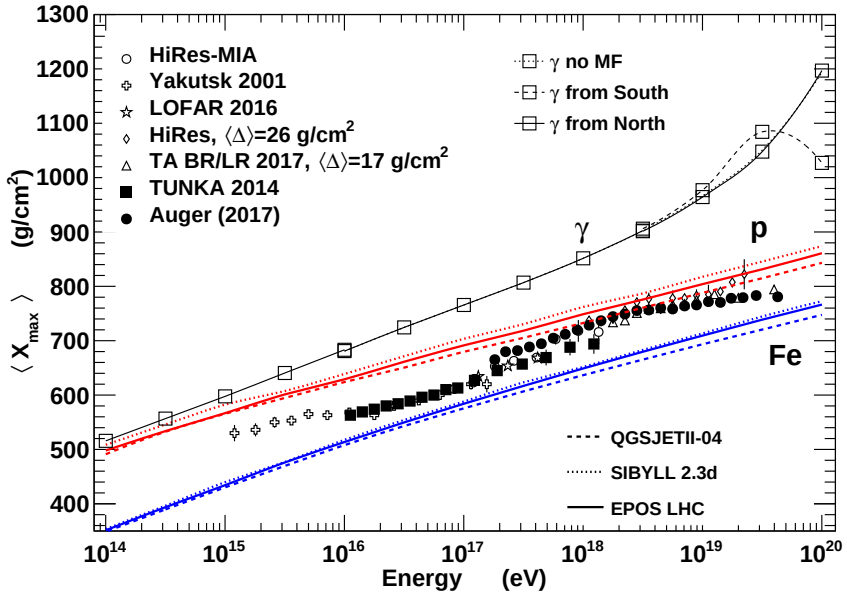


Figure 2.3.: **Measurements of the average X_{max} as a function of the energy obtained from different experiments (disconnected markers) and the average X_{max} predicted by simulations for different primaries and hadronic interaction models (lines).** The simulations for the photon-initiated showers are created with the CONEX tool [21] for 60° inclined showers at fixed energies. They match the conditions of the Auger site and include three different configurations for photons: no geomagnetic field, photons from the North and photons from the South. The figure was kindly provided by Tanguy Pierog after private communication. An older version of the figure with further information can be found in [22].

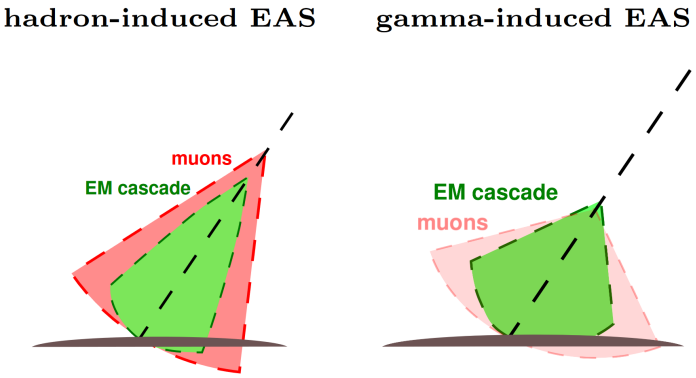


Figure 2.4.: **Illustrative interpretation of hadron-induced showers (left) and photon-induced showers (right).** Gamma-showers start later in the atmosphere and display a more curved shower front at ground level. The muon content of gamma showers is reduced compared to hadronic showers. Figure from Reference [23].

expected to be produced by the interaction of UHECRs with the interstellar matter in the Galactic disk [20, 31]. As the uncertainties of the models that predict the flux of photons are rather high and highly dependent on the mass composition at UHEs, only an upper bound is given as $10^{-3.5}\,\mathrm{km}^{-2}\,\mathrm{sr}^{-1}\,\mathrm{yr}^{-1}$ for energies above 10^{18} eV [20]. This is more than two orders of magnitude lower than the sensitivity of current experiments, marking the expected diffuse flux a negligible background effect for the search of photons emitted by point sources [20]. For this reason, any identified photon above 10^{18} eV would be of great interest to the astroparticle community, as it could imply great constraints on the mass composition of UHECRs or strongly hint at Beyond the Standard Model (BSM) physics if the energy exceeds 10^{19} eV. A list of possible production and propagation scenarios for UHE photons follows.

Scenario 1: Protons with energies below $\approx 5 \times 10^{19}$ eV interact with the cosmic background radiation by producing e^+e^- pairs, while protons above this energy interact predominantly in the form of the GZK process [32] and produce a Δ^+ resonance that subsequently decays into pions. Heavier primaries at energies of $\approx 10^{20}$ eV are affected by photodisintegration processes that significantly migrate the propagation effects, making them comparable to those experienced by protons at lower energies. All products that emerge out of these reactions (e^+, e^-, and pions) then proceed to be the source of *cosmogenic photons* via inverse Compton scattering for the $e^\pm$ or via pion decays. The results are UHE γ-rays where, in extreme cases, energies of 10^{21} eV can be exceeded [33, 30]. As the flux of cosmogenic photons is strongly dependent on the mass of primaries (higher flux for lighter nuclei), the observation of a photon in this energy range would impose strong limits on the mass composition of UHECRs [33, 30].

Scenario 2: UHE photons could be emitted by point sources: If such a source produces ultra-energetic particles they interact with the ambient radiation and matter and produce photons with energies of approximately one order of magnitude less energy. The relevant mechanism for this would be pion photoproduction or inelastic nuclear collisions. Assuming a Fermi acceleration at the source and negligible energy losses during the propagation, a E^{-2} spectrum is expected [28]. This would match the spectrum, measured by TeV γ-ray detectors. Suppose that a source of this acceleration capacity exits and is located within the propagation length of the UHE photons, a flux of photons pointing back to the source direction would be observed. So far no significant source could be identified [33, 29, 28].

Scenario 3 (BSM): In pursuit of combining particle physics with gravity, many quantum gravity models suggest the breaking of the Lorentz invariance. If the Lorentz invariance is broken, the dispersion relation would be affected for photons, leading to the suppression of $\gamma\gamma \rightarrow e^+e^-$ at UHEs. Under the assumption that the UHECR spectrum at the sources is proton rich, GZK-photons are produced plentifully and can reach Earth from cosmological distances. In this case, the fraction of diffuse photons in cosmic rays above 10^{19} eV would be significant. The allowed parameter space of Lorenz invariance violating models is therefore tightly bound to the UHE photon fraction [4, 5].

Scenario 4 (BSM): Multiple studies established a strong correlation between the arrival direction of CRs with energies $> 10^{19}$ eV and the distribution of multiple BL Lac objects—a type of active galactic nucleus. As the particles point directly to the source, a deflection due

to magnetic fields is not happening, suggesting particles of neutral nature [8]. Since the distance to the BL Lac sources is far too great for neutral particles to traverse the distance to Earth without destructive interactions, a direct propagation of photons is impossible. A piece solving this puzzle without introducing the violation of Lorentz invariance would be the propagation via Axion Like Particles (ALPs). ALPs are particles with a similar Lagrangian structure as the Peccei-Quinn axion but they do not solve the strong-CP problem. Said Lagrangian includes a term that allows the mixing of photon and ALPs within the presence of strong magnetic fields. A set of parameters for the galactic magnetic fields and the ALP-model can be found that allows for the conversion of photons to ALPs close to the source and then the back conversion to real photons close to the observer [33, 8].

Scenario 5 (BSM): The photon flux in the UHE region is of great importance for the parameter limits of Super Heavy Dark Matter (SHDM) models. SHDM in the mass region 10^{13} GeV to 10^{14} GeV could be the key to the dark matter problem and simultaneously be a *subdominant* contributor to the UHE photon flux. The SHDM in question would be generated during the inflation area and is quasi-stable with a lifetime of $\geq 10^{21}$ yr. As pions are the dominant decay product of the SHDM particles, a flux of UHE neutrinos and UHE photons is expected from the subsequent decay of the former. The currently measured Auger photon flux limits do not exclude the SHDM hypothesis [6, 7].

3. The Pierre Auger Observatory

The Pierre Auger Observatory is a large-scale hybrid detector located in Argentina. It is named after the scientist Pierre Victor Auger, one of the forefathers of cosmic ray physics [34]. The main physics task of the installation is to measure and study the characteristics of High-Energy (HE) and Ultra-High-Energy Cosmic Rays (UHECRs). To accomplish this the observatory has two main instruments: the Surface Detector (SD) and the Fluoroscence Detector (FD), which will be explained in the sections below. After over 19 years of continuous runtime, the Pierre Auger Observatory reached an accumulated exposure of over 80 000 km sr yr [35] for energies above 6×10^{15} eV. With this data, the Pierre Auger Observatory inter alia discovered a dipole structure in the arrival direction of CRs at energies above 8×10^{18} eV [36] and sets strong limits for the accumulated photon flux above $2{\times}10^{17}$ eV [20]. Furthermore, the Observatory has an integral role in the definition of spectrum features at UHEs [37, 18] and the discovery of the migration to heavier primaries above $\approx 2 \times 10^{18}$ eV [38].

3.1. The Fluorescence Detector

The FD is a collection of 24 fluorescence telescopes at four locations around the Pierre Auger perimeter. Each location (Los Leones, Los Morados, Loma Amarilla, and Coihueco) has six telescopes with a field of view window of $30° \times 30°$. The six telescopes are each aligned so that their combined horizontal field of view adds up to 180° and they observe the volume above the SD. When the particles of an EAS propagate through the air, they excite molecular nitrogen—leading to the isotropic emission of ultraviolet light. The fluorescent glow of EASs is then captured by the Photomultiplier Tubes (PMTs) after traversing a bandpass filter and the mirror array that focuses the light. As the FD can see the emission of a shower during its development, a lateral distribution profile can be reconstructed. By involving the known slant depth X of the atmosphere, the profile is often written as the longitudinal development profile $\mathrm{d}E/\mathrm{d}X(X)$. A major disadvantage of the FD is its limited uptime. Measurements can only happen during nights without moonlight or clouds [39]. The FD is designed to measure showers above 10^{18} eV. To lower the sensitive energy region to 10^{17} eV, where the transition of galactic to extragalactic cosmic rays is expected, the detector was complemented by the three High Elevation Auger Telescopes (HEAT) in 2009. The HEAT detectors are tiltable fluorescence telescopes that work identically to the other fluorescence detectors, but the inclination of the field of view can be altered dynamically. This is necessary because low-energy showers cannot be observed with the standard fluorescence telescope alone [40, 41].

3.2. The Surface Detector

The surface detector of the Pierre Auger Observatory is an array of 1660 surface stations in a triangular configuration setup. All stations use the same cylindrical design (10 m surface area and 1.2 m height) and contain 12 m^3 of ultrapure water as the detector volume [42]. The detection mechanism is based on the detection of charged high-energy particles that exceed the local speed of light as they pass through the tank and emit Cherenkov light. Similar to a sonic boom, the light from the de-excitation of atoms adds up to a shock front. To detect the light from this process, every cylindrical tank is equipped with three 9 ” PMTs and has a reflective inner surface [42]. This part of the station is referred to as a Water Cherenkov Detector (WCD). Currently, the stations are being upgraded to AugerPrime (see the following section) with additional radio and muon detection capabilities, broadening the detector response beyond the capabilities of the WCD alone. Without an operational dependency on the weather or the daytime, the SD achieves an uptime close to 100 % [43]. The SD stations are positioned according to a triangular grid with a spacing of 1500 m between tanks. This way the grid covers an area of approximately 3000 km^2 [43]. As of September 2011, the surface detector is complemented by the infill and HEATLET extension of the Auger Muon and Infill for the Ground Array (AMIGA) project [41], which is known as the SD750 array. The SD750 spans over an area of 27 km^2 and increases the density of stations by reducing the station spacing to 750 m [41]. In total, the SD750 added 50 additional WCD detectors which brings to total station count up to 71 in the SD750 area [44]. The location was chosen so that the SD750 array is close to HEAT and within its field of view, which provides a calibration source for the array [41]. Additionally to the SD750 stations, the AMIGA upgrade extended the Pierre Auger Observatory by seven scintillation counters buried under 2.3 m of soil—the Underground Muon Detectors (UMDs). Another extension of the SD detector is the SD433 array. The SD433 array is located within the SD750 region and subdivides the triangular grid with 12 additional WCDs which have a spacing of 433 m. This sub-array lowers the threshold of detectable shower energies to approximately $10^{16.5}$eV [45].

3.3. AugerPrime

The AugerPrime upgrade is the next step in the history of the Pierre Auger Observatory. It is an ongoing expansion of the existing detector capabilities to enhance the mass sensitivity for UHECRs. Since the uptime of the FD is only at around 15 % [39], an estimation of the shower maximum X_{max} from the SD alone is desirable. This can be achieved by disentangling the muonic and electromagnetic components of the shower at ground level. The new detection channels of the AugerPrime listed below help with this distinction between electrons and muons. Additionally, the new detector channels greatly improve the amount of gathered data for every event—boosting the precision of the reconstruction. The following list provides an overview of the improvements addressed by the upgrade:

- **Adding surface scintillation detectors.** Two symmetrical modules are installed on top of nearly every tank. While the WCDs are more sensitive to the muons in a shower, the scintillation panels are about equally sensitive to the electromagnetic and muonic components. This helps to differentiate the muon content, leading to the described sensitivity boost for the mass estimation [46].
- **Erecting additional Radio Antennas.** Due to geomagnetic deflecting and the Askaryan effect [47, 48], air showers emit weak radio signals. Using an array of radio antennas, air showers can be reconstructed from first principles, increasing the amount of information recorded for each shower. For this reason, the Auger Engineering Radio Array (AREA) is expanded to the whole array. The short aperiodic loaded loop antennas will be deployed on top of the tanks, alongside the scintillation modules [46].
- **Upgrade of the Underground Muon Detector.** Currently, the Auger Observatory possesses seven AMIGA Underground Muon Detectors (UMDs) modules. The difference to the above-ground scintillation modules is that the layer of soil blocks the electromagnetic component, leaving only muons to produce a signal. The number of UMDs will be extended to 64, covering the whole infill array [46].
- **Increasing FD uptime.** By reducing the requirements on the night sky background, the duty cycle of the FD can be increased up to $\approx$ 25 % without a change to the hardware. The downside is a degradation of data quality for high background events. But especially for HE events, there will be close to no tradeoff in neither selection efficiency nor X_{max} and energy resolution [46].
- **Increasing Number of WCD - PMTs to four.** Additionally to the existing three PMTs a smaller fourth one with a 1 ” diameter will be installed in the WCDs. This addition boosts the dynamic range of the array in the direction of large signals - where the large PMTs are saturated, even in low gain mode, the new PMT is still operational, enabling the non-saturated measurement of large showers close to the shower core [49].
- **New Readout Electronics.** The AugerPrime upgrade includes a new electronics frontend, the Upgraded Unified Board (UUB). It handles the data acquisition of the SD and coordinates local and global triggering commands. The new board does the readout for the four WCD-PMTs and the scintillation module at 120 MHz, increasing the current rate by a factor of three [49].

4. Standard Reconstruction with the Surface Array

The following section describes the standard reconstruction for hadronic primaries. As this work focuses on the WCDs, the reconstruction for showers which are recorded by the FD is not explained. Further information can be found in Reference [39].

This thesis uses the software framework $\overline{\text{Off}}\underline{\text{line}}$ [50, 51] for the standard reconstruction of events. Furthermore, it is used to simulate the WCD detector response.

4.1. Geometry Reconstruction for the Surface Detector

EASs at ground level can be described as a thin curved disk of particles propagating through the atmosphere. The highest particle density is found in the center and falls off radially. The travel direction of the disk is the same as the trajectory of the initial particle and the diameter of the disk increases with the age of the shower till the maximum amount of particles is reached. The curve of the disc comes from the fact that a shower can be modeled in first order as a radially expanding sphere shell, propagating from an origin point $\vec{x}_o$ located somewhere along the shower axis. The signal measured at the ground plane is then the intersection of the plane with the expanding sphere [51]. As the amount of hadronic showers at UHEs is quite low with around 1 shower per km^2 for energies $> 10^{19}$ eV, the area of the detector needs to be quite large. Luckily, UHE showers have a large footprint of i.e. 10 km^2 at 10^{19} eV. For this reason, the Pierre Auger array samples a shower with a set of distributed stations at ground level [51]. As the shower does not hit all stations simultaneously, every station perceives the shower at a different development stage additionally to the difference in distance to the shower core.

An involved reconstruction chain is needed to recover the properties of the primary particle from the measured signals. This starts with the estimation of the station-wise signal and the determination of the geometrical properties of the showers. Signals measured by the WCDs are initially given as Analog to Digital Converter (ADC) channels and must be calibrated to a physically meaningful unit: a Vertical Equivalent Muon (VEM). A VEM charge is the charge deposited by a single vertical muon traversing a WCD at the center. The deposited energy equivalent is ≈ 240 MeV [51]. A calibration from ADC channels to VEM charge is achieved by permanently measuring atmospheric muons and using them as a constant calibration source [52]. If a histogram with the charge deposited by the individual muons is created in units of integrated ADC channels, a local maximum is

visible. The position of this peak is proportional to one VEM charge. The same is done in the form of a pulse height histogram, which is a histogram of the muon trace peaks in ADC channels. The location of the local maximum in the pulse height histogram produced by the atmospheric muons is related to the peak current that is induced by a single vertical through-going muon (VEM peak). This calibration current is used to set the correct gain of the PMTs, which is important to achieve the correct trigger threshold and calibrate the gains of the individual PMTs so that the VEM calibration peak is approximately at 50 channels for all PMTs [52].

In the next step, a selection process is applied that removes all the stations that are likely not part of the shower. This can happen either when stations are triggered by a lightning storm or when stations are triggered by accidental muons that did not emanate from the same shower. The latter cases are filtered by doing a preliminary reconstruction with the seed triangle, a triangle of three non-aligned stations which must include the *hottest* station (station with the most signal) and otherwise chooses the configuration with the most accumulated signal [53]. Firstly the shower barycenter $\vec{x}_\mathrm{b}$ is obtained as the signal-weighted mean of the station positions $\vec{x}_i$ of the involved stations. This is later used as the starting parameter for determining the precise impact point of the shower, $\vec{x}_\mathrm{c}$. Secondly, a first estimation of the shower axis $\hat{a}$ is given by solving the equation

$$ct_\mathrm{sh}(\vec{x}) = ct_\mathrm{b} - (\vec{x} - \vec{x}_\mathrm{b})\hat{a}, \tag{4.1}$$

with the timing information of the three seed stations, where $t_\mathrm{sh}(\vec{x})$ is the time when the shower front passes through the location $\vec{x}$. With this, the preliminary reconstruction defines the shower front as a plane perpendicular to $\hat{a}$, traveling with the speed of light in the direction of $-\hat{a}$ and intersecting the ground at the barytime t_b at the barycenter $\vec{x}_\mathrm{b}$ [51]. This initial seed reconstruction is used to identify out-of-time stations that do not fit the approximate geometry of the shower. The fit of the shower axis is then updated with all candidate stations and is also used as the starting point of the global fit with all remaining stations later on. For a more complex fit of the shower that includes the curvature, a more accurate estimate of the impact point is required. Therefore, an approximation of the shower size is calculated. Using this estimation as a starting value, a simultaneous fit for the shower size and the impact point is performed (see Section 4.2) using the current value of $\hat{a}$.

This is followed by a more precise fit of the shower geometry. The fit approximates the shower as described above, a spherical shell emitted from an origin point $\vec{x}_\mathrm{o}$ with the particle density concentrated around an axis $\hat{a}$, inflating approximately at the speed of light. A description of the expanding sphere is given by

$$ct_\mathrm{sh}(\vec{x}) = ct_\mathrm{o} + |\vec{x} - \vec{x}_\mathrm{o}|, \tag{4.2}$$

where c is the speed of light and t_o the starting time of the shower. This formulation gives four free parameters, which are given by the parameters u and v as components of the vector $\hat{a} = (u, v, \sqrt{1-u^2-v^2})$, time t_o and the shower curvature radius R_o. The latter is defined as $R_\mathrm{o} = (\vec{x}_\mathrm{o} - \vec{x}_\mathrm{c})/|\vec{x}_\mathrm{o} - \vec{x}_\mathrm{c}|$ where $\vec{x}_\mathrm{c}$ is determined during the LDF fit. Accordingly, the impact time is given by $t_\mathrm{c} = t_\mathrm{o} + R_\mathrm{o}/c$.

Since only three or four stations are triggered and accepted for events with lower energy, it is not feasible to resolve the curvature and the impact point for these events. To resolve this, if not at least five stations are used in the geometry fit, R_o is fixed to a parameterized value, optimized from events with a higher station-multiplicity [51]. The fit is performed as a minimization of χ^2 where the uncertainty of the start time for every station is considered [51].

4.2. Shower Size

The shower size for the main Auger SD array (1500 m spacing), used in this thesis, is defined as the signal in 1000 m distance to the shower core. It is a proxy for the shower energy after correcting for the zenith dependency. Assuming a perfect shower and reconstruction, the signal measured by the stations solely depends on the distance r to the shower core in shower plane coordinates (a plane perpendicular to $\hat{a}$ with a supporting point at $\vec{x}_c$) and the azimuth angle ζ of the stations with respect to $\hat{a}$. The latter is as of now not parameterized for the Offline framework [51]. For photon-induced showers, the azimuthal-asymmetry will be examined in Section 4.2.2. The estimation of the shower size is done by regressing the signal $S(r)$ of the triggered stations to the Lateral Distribution Function (LDF) and further constraining it with stations close to $\vec{x}_c$ that did not trigger, the silent stations [51].

As the amount of triggered stations and their spatial distribution is often sparse, the shape of the LDF cannot be determined on an event-by-event basis and must be approximated by a data-driven parameterization: $f_{\mathrm{LDF}}(r)$. The signal at the stations in shower plane coordinates can then be described by

$$S(r) = S(r_{\mathrm{opt}}) f_{\mathrm{LDF}}(r), \tag{4.3}$$

where $S(r_{\mathrm{opt}})$ is the shower size estimator. The distance r_{opt} is chosen so that the variation from shower to shower fluctuations is minimized. The optimal distance for the main array was found to be $r_{\mathrm{opt}} \approx 1000\,\mathrm{m}$ [54]. Naturally the function $f_{\mathrm{LDF}}(r)$ is defined to hold $f_{\mathrm{LDF}}(r_{\mathrm{opt}}) = 1$. As $S(r_{\mathrm{opt}}) = 1000\,\mathrm{m}$ is just the signal at 1000 m distance to the shower core in shower plane coordinates, it is shorthandedly written as S_{1000}. The functional form of f_{LDF} is a modified version of the NKG function [55, 56]:

$$f_{\mathrm{LDF,\,NKG}}(r; \beta, \gamma) = \left(\frac{r}{r_{\mathrm{opt}}}\right)^{\beta} \left(\frac{r + r_s}{r_{\mathrm{opt}} + r_s}\right)^{\beta+\gamma}, \tag{4.4}$$

where $r_s = 700\,\mathrm{m}$. The parameters β and γ determine the slope and shape of the LDF. Since the event-by-event fit of β and γ is generally not possible, these geometry parameters are described in a data-driven parameterization, which depends on S_{1000} and the zenith angle θ of the shower [51].

With the parameterized geometry parameters, a likelihood fit is carried out to obtain the value of S_{1000} and $\vec{x}_c$. The likelihood function $\mathcal{L}$, which should be maximized, is a product

of station-wise probabilities P for values of S_{1000} and the shower core coordinates $\vec{x}_c$ given the observed station signals S_i and the station coordinates $\vec{x}_i$. Written in the form of the log-likelihood, it looks like this:

$$\ln \mathcal{L} = \sum_i \ln P(S_{1000}, \vec{x}_c | S_i, \vec{x}_i). \tag{4.5}$$

In the used $\overline{\text{Off}}\underline{\text{line}}$ framework, the form of P depends on the type of signal. Small signals are modeled according to Poissonian probability for the number of measured particles while large signals are modeled as Gaussian distributions with a zenith-dependent uncertainty model. The signal of unrecoverable saturated stations is treated as a lower signal limit. The probability of a station being not triggered can be described by the inverse probability of a station being triggered, given the expected signal $S(\vec{x}_i)$. This is how non-triggered, so-called silent stations, contribute to the likelihood function. Further information can be found in Reference [51], which describes the full SD reconstruction procedure in detail.

4.2.1. Hadronic LDF Model

As previously discussed, the parameters β and γ of equation (4.3) must be determined in a data-driven way. For hadronic showers, events are selected for the β parameterization if they have at least two, three, or more stations within a distance r of $400\,\text{m} < r < 1600\,\text{m}$ from the shower core. Additionally, at least two of these stations must be separated by distances of 900 m, 800 m, or 700 m. This distance criterion is applied respectively if the initial selection is met by two, three, or more stations [51]. Estimating γ is more difficult, as both parameters are strongly correlated. The parameter describes the difference between a raw power law and equation (4.3) for large r. The second criterion for events that are used for fitting γ is the same as for β. The search radius of the first criterion however is changed to $1000\,\text{m} < r < 2000\,\text{m}$ [51]. The functions used to describe beta and gamma are defined as

$$\beta(s, \theta) = a_0 + a_1 s + \sec\theta(b_0 + b_1 s + \sec\theta(c_0 + c_1 s)) \tag{4.6}$$

$$\begin{aligned}\gamma(s, \theta) = {} & \alpha_0 + \alpha_1 s \\ & + \frac{\beta_0 + \beta_1 s}{\exp((\gamma_0 + \gamma_1 s)(\cos^2\theta - (\delta_0 + \delta_1 s))) + 1} \\ & + \frac{\epsilon_0 \left(\cos^2\theta\right)^{\zeta_0}}{\exp(\eta_1(s - \eta_0)) + 1} - \beta(s, \theta),\end{aligned} \tag{4.7}$$

where $s := \lg(S_{1000}/\text{VEM})$ and the letters a_i through c_i as well as the Greek letters α_i through η_i with $i \in \{0, 1\}$ are the coefficients of the fit. Their numerical values can be found in Table A.1 of the Appendix.

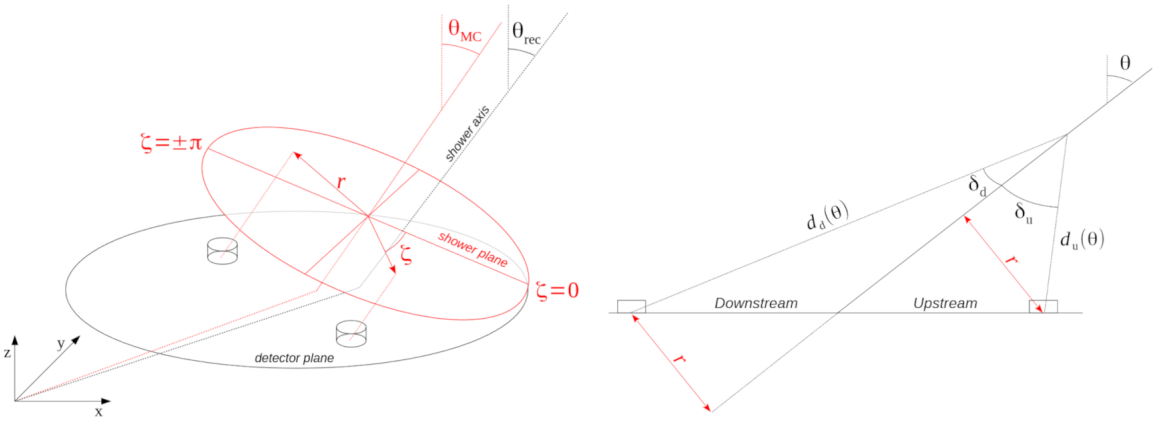


Figure 4.1.: **Schematic display of the shower geometry.** *Left:* 3D view of the ground plane with two stations and the true shower plane (red). The subscript *MC* refers to the Monte-Carlo shower axis whereas *rec* is the reconstructed axis. *Right:* side view of a shower with up and downstream stations. Figures from Reference [57].

4.2.2. Azimuthal Asymmetry in Showers

The sampling of air shower signals at the ground plane for showers with a non-zero zenith angle is not symmetrical around the shower axis. This asymmetry can be captured by expanding equation (4.3) according to

$$S(r,\zeta) = \overbrace{S_{1000}\, f_{\mathrm{LDF}}(r)}^{S'(r)} \overbrace{(1+\alpha(r,\theta,S_{1000})\cos\zeta)}^{\equiv S_a(\zeta,\alpha)}\,, \tag{4.8}$$

where $S(r,\zeta)$ would be the shower signal measured by the stations, ζ is the azimuth angle between the stations and the shower axis, $S'(r)$ is the shower signal at a distance r, independent of ζ, S_a is the function that describes the asymmetry, and α is the amplitude of the asymmetry. The convention prescribes that the stations hit first by the shower are located at $\zeta = 0°$. A schematic view of the ground plane, the shower plane, and the shower axis can be seen on the left side of Figure 4.1. Photon-showers can start very late and especially at high energies, their particle maximum X_{max} is below ground. Due to having mainly electromagnetic interactions and changes in the showering behavior, the asymmetry of photon showers is different from that of hadron showers and will be parameterized in Section 6.2. The main reasons for the emergence of asymmetry in the azimuth angle are explained below.

4.2.2.1. Geometric Asymmetrie

Let's imagine one would look at the detector from the point of origin of an EAS with a low zenith inclination. The stations between the point where the shower hits the ground and the observer would be the upstream stations and the stations behind the shower core

are the downstream stations. This is illustrated on the right side of Figure 4.1. Since the upstream stations are closer (distance $d_{\mathrm{u}}(\theta)$) to the observer they would appear larger than the downstream stations (distance $d_{\mathrm{d}}(\theta)$) which means that they feature a larger effective area. Additionally, if one draws an imaginary line $\vec{x}_{\mathrm{so}}$ between the stations $\vec{x}$ and the shower origin $\vec{x}_{\mathrm{o}}$, $\vec{x}_{\mathrm{so}} := \vec{x}_{\mathrm{o}} - \vec{x}$, the angle of this line with respect to the ground plane is different for up and downstream stations. This can be understood, as looking at the stations from different angles. As the stations are cylindrical, this affects the effective area. This effect is also visible on the right side of Figure 4.1: the upstream station is close to being directly under the point of first interaction, and would be seen from atop whereas the downstream station would be seen from a much steeper angle.

Another geometrical effect is the difference in the angular distance of stations to the shower-core axis. As seen on the right side of Figure 4.1, the upstream stations have a larger angular distance δ_u to the core axis than the downstream stations with δ_d. This means, that the stations sample the shower cone with different angular distances to the shower axis. This results in different signals since the particle density decreases with increasing values of δ: ADF $\propto (\delta/\delta_0)^{-\gamma}$ [58, 57]. Here, ADF is the angular distribution function, and γ and δ_0 are parameters of the ADF.

4.2.2.2. Shower Age

The age of the sampled shower is different for the upstream and downstream stations. As the shower propagates through the detector array, the upstream stations are hit first, followed by the downstream stations, assuming a reasonable distance of X_{max} to the ground. During this time difference, the particle energy and density are not constant, and therefore the expected signal changes with distance and time. The average energy per particle decreases with the shower age as the energy gets distributed between more and more particles and high-energy particles produced at the start of the shower are attenuated over time. The particle density increases till reaching X_{max} and then decreases again [57]. This mixture of attenuation effects, particle density, and particle energy generates a complicated shower footprint.

4.2.2.3. Magnetic Field

Earth's geomagnetic field influences the trajectory of the charged shower particles and introduces an additional factor to the asymmetry. This effect does not depend on ζ but rather on the angle of the shower to the magnetic field vector [59]. The effect is only mentioned for completeness reasons. It is neither reflected in equation (4.8) nor will it be considered in the analysis of the azimuthal asymmetry in Section 6.2.

4.3. Energy Reconstruction

The energy of the primary particle is closely connected to the measured value of S_{1000}. Before doing the energy calibration, some small bias corrections are applied to S_{1000} which correct influences from atmospheric fluctuations and the effects of the local geomagnetic field, depending on the orientation of the shower axis [51]. The next steps revolve around the removal of the zenith angle dependence of S_{1000}. This is done via the Constant Intesity Cut (CIC). Assuming a constant, isotropic flux for cosmic rays arriving at Earth, the measured intensity is expected to be constant after correcting for the attenuation caused by the propagation through different layers of the atmosphere. The shape of this attenuation curve $f_{\text{att}}(\theta)$ can be determined using the CIC method by first identifying a constant intensity of cosmic rays in equally sized bins of $\sin^2\theta$ [43]. This zenith-dependent constant intensity is then described in relation to the constant intensity at 38° by $f_{\text{att}}(\theta)$. The value of 38° is chosen, as it is the median value for θ. Mathematically, this is described by

$$\mathrm{S}_{38} = \frac{S_{1000}}{f_{\text{att}}(\theta)}, \tag{4.9}$$

where

$$f_{\text{att}}(\theta) = 1 + ax + bx^2 + cx^3 \tag{4.10}$$

and

$$x = \cos^2\theta - \cos^2 38^\circ.$$

The value of S_{38} can be understood as the shower size that would have been detected if a shower had entered the atmosphere at a zenith angle of 38° [43]. The attenuation function is modeled as a third-degree polynomial with a fixed constant term and the coefficients a, b, and c.

As the last calibration step, the value of S_{38} is used to calculate the shower energy by applying

$$E = A\left(\frac{S_{38}}{\mathrm{VEM}}\right)^B. \tag{4.11}$$

The energy E used for the fitting of the parameters A and B is provided by caloric measurements of the FD in hybrid events, where the shower is simultaneously observed by the FD and the WCDs [43]. The model assumes a linear dependency of shower size and shower energy for fixed zenith angles. For simulated showers, the CIC can be replaced by the Direct Energy Calibration (DEC) [60], where equation (4.9) and equation (4.11) are simultaneously fitted with the known Monte-Carlo Energy.

5. Machine Learning

5.1. Neural Networks

Neural Networks (NNs) are heuristics that can successfully perform complex tasks that are difficult or impossible to model with traditional computing methods while learning all relevant information by themself. The brain with its synaptic connections poses as an inspiration to the node base structure of NNs. The nodes of a NN are called neurons and are connected to each other with different weights. One of the simplest examples for NNs is a deep feedforward network. Their neurons are grouped into layers and every neuron in every layer is fully connected to every neuron in the previous layer. The first and last layers are the input and output layers of the network and the layers in the middle are referred to as hidden layers. In a feed-forward network, the hidden layers are also referred to as *dense* layers. The term *deep* expresses that there is more than one hidden layer. In a feed-forward network, each neuron receives the scalar output values from the neurons of the previous layer as an input. The neuron then processes the inputs and produces a scalar output, according to a predefined algorithm. This algorithm takes the weighted average, adds a *learnable bias*, which constant offset, and applies an activation function [61, 62]. The *learnable weights* used in average, as well as the learnable bias, are the parameters optimized during the network training process.

An activation function can be any function that maps $\mathbb{R} \mapsto \mathbb{R}$ but should be continuously differentiable and nonlinear. Two of the most well-known activations are the *ReLU* function, often used for the hidden layers and the (scaled) *sigmoid* or *thanh* function for the output layer in classification tasks. Learning happens when the trainable parameters in the network are tuned so that a predefined loss function J is minimized. This is the training process of a network. The loss functions J can vary depending on the problem, but i.e. for regression problems, the mean squared error loss

$$J_{\text{MSE}} = \sum_i \left(y_{\text{pred}}^{(i)} - y_{\text{true}}^{(i)} \right)^2 \tag{5.1}$$

is often used. Here the i are the individual samples of the data set, *true* denotes the ground truth and *pred* is the predicted value of the network.

Minimizing the loss function by tuning the weights and biases is a complex and non-trivial process and is often done with backpropagation for supervised learning tasks. Supervised learning means that the network maps a complex input to a ground truth, the *label*. As the goal of this work is to reconstruct the energy out of simulated events with photon primaries, the problem is categorized as a supervised learning task. A key part of the

training process is the backpropagation, where the gradient g_i of J with respect to the weights w_i and biases b_i is calculated with the chain rule. Then the weights and biases are shifted in the direction of $-g_i$ to minimize the loss in the next epoch: $w_i \rightarrow w_i - \alpha_{\mathrm{LR}}\, \partial J / \partial w_i$ [62]. This is only a simple example of the weight adjustments, in reality, more sophisticated algorithms like *ADAM* [63] are used. The parameter α_{LR} is the learning rate which is the step size for each adjustment of the weights. The training of the NNs is divided into epochs. In each epoch, the network is trained with the entirety of the data set. The number of data points evaluated in the loss function is not necessarily the same as the length of the data set. In the event that only a subset of training samples is evaluated at a given time, this is referred to as minibatching. The number of samples in a batch is referred to as the batch size and the weights, and biases are updated after each batch.

One problem with this approach is the possibility of overfitting the network. In this case, the NNs perform great on the training data, but fail to generalize to the underlying data behavior. Now, if the network predicts unseen data after the training is finished, the performance is worse than observed during training. Intuitively, this can be viewed as the network just *memorizing* the inputs labels and their respective outputs. To recognize this during training time, the data is split into a training and validation set, where training only happens on the training data. After each epoch, the network performance is tested on the unseen validation set. If the training loss and the validation loss diverge significantly, overtraining is happening. This can be mitigated by applying a weight regularization [61] during the training or by applying dropout layers [64]. Briefly described, the regularization of weights during training appends a term to the loss function that penalizes large values of the absolute weights. However, this thesis only implements dropout layers. A dropout layer sets a fraction r_{drop} of randomly selected inputs to 0 while rescaling the other inputs by a factor of $1/(1 - r_{\mathrm{drop}})$, which keeps the sum of the inputs unchanged. The dropout mask is updated after each batch is finished. This can be conceptualized as turning off different neurons at each training step, which forces the networks to generalize.

5.2. Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a special kind of NN, tailored to recognize spatially correlated data. The default example of such a task would be image recognition, where the spatial correlation is the correlation of neighboring pixels. CNNs look similar to feed-forward networks, with the exception that they can have convolutional layers as hidden layers. A convolutional layer has one or more learnable convolutional kernels K where $\dim(K)$ matches the number of dimensions of the previous layer. I.e. if the input of a CNN would be a picture of size $(N \times M)$, the input dimension would be $(N \times M \times 3)$, where the last dimension are the RGB components. The last dimension is referred to as the *channel* dimension and needs to be accounted for in the convolution even though it is not a spatial dimension. Therefore it would hold $\dim(K) = 3$ in this example. Note that the number of channels equals the length of K in the channel dimension. The convolution layer works, by calculating a convolution of the input with K and taking the sum as the

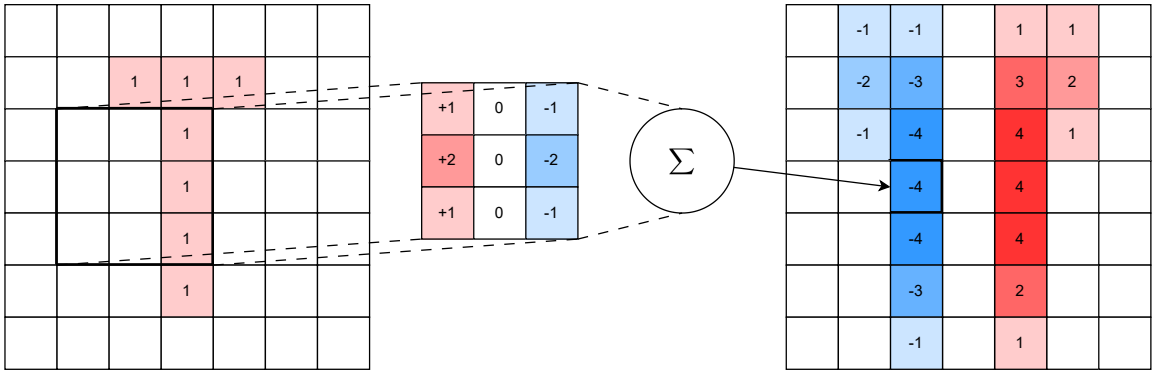


Figure 5.1.: **Example of a two-dimensional convolution operation.** The shown exemplary kernel is the Sobel operator in the horizontal direction, which is often used in signal processing tasks. The differential operator calculates the approximation of the horizontal gradient of the image intensity.

output of every step for the convolution. The output S at coordinates (i, j) with input I is therefore defined as

$$S(i,j) = a((I * K)(i,j) + b) = a\left(\sum_m \sum_n \sum_l I(m,n,l)K(i-m, j-n, l) + b\right), \tag{5.2}$$

where a is the activation function and b the bias [61, 65]. Here, the index l iterates through the channels. An example of a simple 2-dimensional convolution with one filter and one channel is shown in Figure 5.1. When a convolutional layer has o kernels [1], the process is repeated for all of them and the output tensor has length o in the channel direction. Typical parameters of a convolutional layer are:

- `filters`: number of kernels
- `kernel size`: the size of the kernel in the spatial input dimensions
- `strides`: modifying the stride length of the convolution
- `padding`: whether to zero-pad the input (if zero-padding is used and `strides = 1`, then the output size is the same as the input size in the spatial dimensions. This is often referred to as `"same"`-padding)

The advantages of CNNs for spatial data are twofold. For once, the data extraction with the convolution is invariant under the translation of features, as the same kernel is applied at all positions of the input. The second advantage is that the number of weights is minimal compared to the case where one applies a normal dense layer to a (flattened) image, which reduces the probability of overfitting the network [66].

Convolutional layers are often followed by other convolution layers or pooling layers. The latter defines a window similar to the convolution window, where a function is applied to all values within the window. A typical function would be $\max x_i$ or $\langle x_i \rangle$. Typically,

[1] Often also referred to as the number of filters.

the output is then piped into another convolutional layer or flattened and sent to a dense feed-forward network [66].

5.3. Vision Transformers

Transformers are arguably one of the most influential architectures in recent years, powering influential artificial intelligence agents like ChatGPT [67]. The Google-developed architecture was first published in [68] and greatly accelerated in sequence-to-sequence translation tasks. By relying solely on the *attention mechanism*, the architecture designed for natural language processing is equal to the quality of previous networks, but dispenses the need for recurrences and convolutions, leading to a big improvement in training time and parallelizability [68].

Many NNs rely on the efficient encoding of the inputs into the *latent space* or *embedding space*. The latent space is an abstract self-learned representation of the data, often with a high information density. In sequence-to-sequence tasks, the encoder part of a network maps the inputs to the latent space and the decoder the latent space to the output.

In sequence processing tasks, the *attention mechanism* is a method, where all hidden states $\{h_j\}$ of the input encoding as well as the hidden state s_{i-1} of the output, generated at the $(i-1)$-th time step, is considered by the decoder for the prediction of the i-th hidden output state [69]. This is achieved by having an attention matrix, which assigns a weight or score α_{ij} to all the hidden states from the input encoding. These weights are calculated, for example, by a feed-forward network. The so-called context vector used in the calculation of the i-th hidden output is then given by $c_i = \sum_{j=0}^{T_x} \alpha_{ij} h_j$, where T_x is the length of the input sequence. The output y_i at timestamp i is a function of the hidden state h_i and the context vector c_i. The attention score α_{ij} can be understood as a measure of how important the input x_j is for the creation of the i-th output [69].

For Transformers the concept of attention is extended and implemented as a so-called scaled dot-product attention. The input consists of three tensors, the query Q and key K tensor of dimension d_k and the value tensor V of dimension d_k. The output of the scaled dot-product attention is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{5.3}$$

$$\text{where} \quad \text{softmax}(\vec{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}.$$

The left panel of Figure 5.2 shows the scaled dot-product attention as a flowchart.

The next building block in the hierarchy of transformers is the *multi-head attention*. As explained in Reference [68] it is beneficial to linearly project the queries, keys and values h-times from the model dimension d_m to the dimensions used in the multihead Attention Block, d_k and d_v. This way, the attention mechanism is performed for h *heads* in parallel,

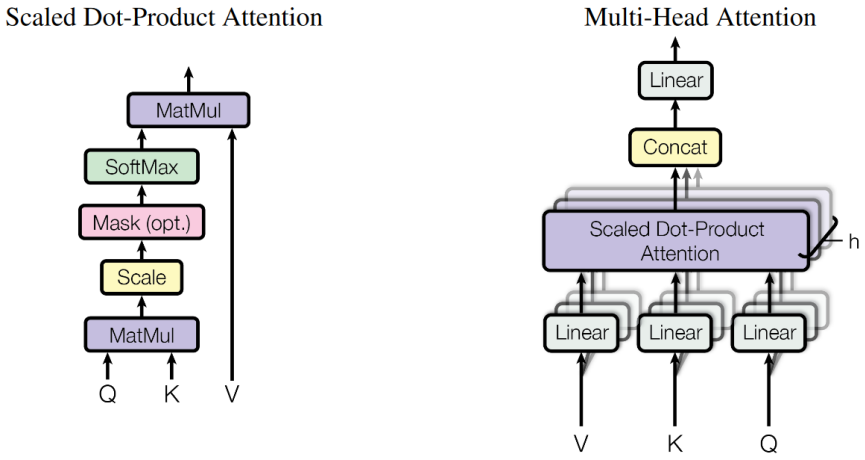


Figure 5.2.: Scaled dot-product attention and multi-head attention mechanism of the transformer architecture as flow diagrams
Figure from Reference [68].

as shown on the right panel of Figure 5.2. The projections are learned during training and are respectively defined as

$$W_i^Q \in \mathbb{R}^{d_m \times d_k}, W_i^K \in \mathbb{R}^{d_m \times d_k}, \text{ and } W_i^V \in \mathbb{R}^{d_m \times d_v}. \tag{5.4}$$

In the next step, the output of the heads is concatenated and projected back to size d_m by the learned weight matrix $W^O \in \mathbb{R}^{hd_v \times d_m}$. The scaled dot-product attention and the multi-head attention are shown as a flow diagram in Figure 5.2. The multi-head attention can be described as

$$\begin{aligned} \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \ldots, \text{head}_\text{h}) W^O \\ \text{where } \text{head}_i &= \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right). \end{aligned} \tag{5.5}$$

If all inputs V, K and Q of a multi-head attention layer have the same values it is called self-attention or intra-attention. The idea is that the network develops a method to dynamically weigh the importance of input elements depending on the value and position of all inputs. The following multiplication by W^O can then be understood as an update of the representation V_i in the latent space with the context from the other values V_j, where $j \neq i$. The whole operation is a Multiheaded Self Attention (MSA) block,

$$\text{MSA}(X) = \text{MultiHead}(Q = K = V). \tag{5.6}$$

To conceptualize this better, a simple analogy within the framework of natural language processing is given. Suppose the input of the network would be: *The Queen song Bohemian Rhapsody is an all-time classic.* If one only looks at the words *(The Queen)*, they would probably be connected to the meaning of a female monarch, but as it is surrounded by the context of the words *(song)* and *(Bohemian Rhapsody)*, the attention head would probably have high attention scores for the matrix elements connecting *(The Queen)*, *(song)* and

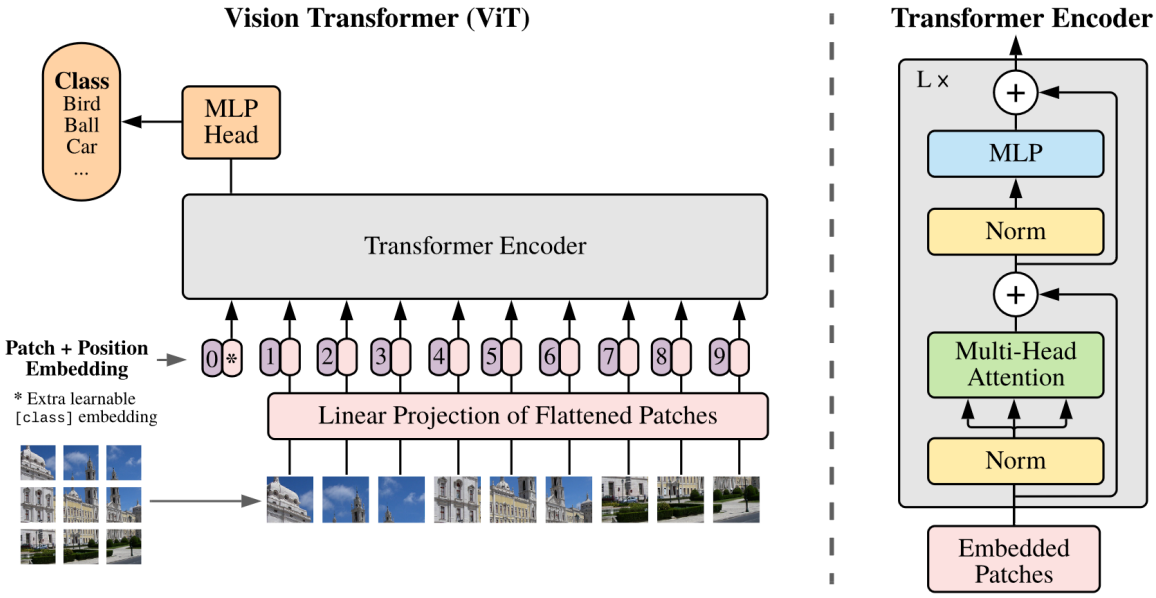


Figure 5.3.: **Schematic overview of the *Vision Transformer* architecture.** The vision transformer is depiced as defined in Reference [10] (also source of the graphic). The input image is split and encoded into several *patches*. Apart from the raw image information, patches contain information on their position in the original image. A learnable *class token* which is later used for the classification is prepended to the vector of embedded patches.

(Bohemian Rhapsody) and would update the representation V of *(The Queen)* so that it points to the band instead.

Building upon the success in natural language processing, a new architecture called Vision Transformer (ViT) was developed by the *Brain Team* of *Google research* and published in the paper *An Image Is Worth 16x16 Words: Transformers For Image Recognition at Scale* [10]. The goal of the architecture is to achieve image processing with a transformer network. As the ViT is a great alternative to convolutional networks, it will also be applied to the photon energy prediction task in this thesis.

The basic flow diagram of the ViT as defined in Reference [10] is presented in Figure 5.3. The input of the ViT is the full RGB picture $\mathbf{x}$ that is separated into N rectangular segments, which are then flattened to a one-dimensional input vector $\vec{x}_p^i$ each. The next step is the so-called *patch-embedding*, where the $\vec{x}_p^i$ are reduced from size d_p to size d_{emb} by multiplying with the learned embedding matrix $\mathbf{E}$. Additionally, a learnable class token $\vec{x}_{\text{cls}}$ is prepended to the vector of visual patches. As the network has no information on the position of the patches in the original image, a learnable positional encoding $\mathbf{E}_{\text{pos}}$ is added to the whole input as the last step of the patch embedding process. The result obtained from the patch embedding process is referred to as $\mathbf{z_0}$. The embedded patches are now used as the input for the transformer encoder. A flow diagram of the transformer encoder, as used in Reference [10], is shown on the right panel of Figure 5.3. The `Norm` operation is a standard layer normalization. Although similar, the encoder of the ViT is not the

same as the encoder used in the original transformer-paper [68]. In the ViT the norm is applied before the other operations, in the original transformer, the norm is applied after the other operations. The encoder is stacked L times and applied to $\mathbf{z}_0$. As the last step of a transformer encoder block and for the final classification, a Multi-Layer Perceptron (MLP) [2] is used. When used for the final prediction the MLP receives the first element of the transformer encoder-output $\mathbf{z}_L^0$ as its input. The full Vision Transformer (ViT) network can be described as

$$\mathbf{z}_0 = \left[\vec{x}_{\text{class}}\,; \vec{x}_p^1\mathbf{E}; \vec{x}_p^2\mathbf{E}; \cdots ; \vec{x}_p^N\mathbf{E}\right] + \mathbf{E}_{\text{pos}}, \qquad \mathbf{E} \in \mathbb{R}^{d_p \times d_{\text{emb}}}, \mathbf{E}_{\text{pos}} \in \mathbb{R}^{(N+1)\times d_{\text{emb}}} \tag{5.7}$$

$$\mathbf{z}'_\ell = \text{MSA}(\text{LN}(\mathbf{z}_{\ell-1})) + \mathbf{z}_{\ell-1}, \qquad \ell = 1 \ldots L \tag{5.8}$$

$$\mathbf{z}_\ell = \text{MLP}\big(\text{LN}\big(\mathbf{z}'_\ell\big)\big) + \mathbf{z}'_\ell, \qquad \ell = 1 \ldots L \tag{5.9}$$

$$\mathbf{y} = \text{LN}\big(\mathbf{z}_L^0\big), \tag{5.10}$$

where LN is the layer normalization and MLP a generic Multi-Layer Perceptron. The idea of a vision transformer is that during the training the class token $\vec{x}_{\text{cls}}$ is updated with the given input information encoded by the other tokens. Since $\vec{x}_{\text{cls}}$ is fully learnable, it can store some prior information. The MLP at the end then uses this updated information of the class token to generate the final prediction.

[2] Fully connected feed-forward network with at least three layers and a nonlinear activation function.

6. Photon Energy Reconstruction with Classical Methods

Prior to evaluating the performance of a neural network-based reconstruction, we investigate the applicability of the classical method described in Chapter 4 for photon-induced showers. As the footprint of photon-induces showers deviates significantly from hadron-induces showers, the classic reconstruction method, if not optimized for photon-induces showers, underestimates the photon energy by approximately 70 %, as demonstrated in Chapter 4.5 of Reference [9].

In this work, we posit that a hypothetical photon-induced shower has already been successfully identified. Consequently, our focus is only on the reconstruction of the energy. The principal distinction between the reconstruction of photon-induced showers and that of hadron-induced showers is the utilization of a parameterization of the shape of the LDF tailored to photon induced-showers and a change of the parameters in the energy calibration. The change of the LDF parameterization is important because we expect the shape of the LDF to be steeper for photon-induced showers [7]. As in the case of a LDF optimized for hadron-induced showers, the input parameters for the LDF parameterization in this work are given by, S_{1000}, the estimated signal of the shower at 1000 m core distance, and the zenith angle, θ. In addition to the LDF, we provide a parameterization of the azimuthal asymmetry amplitude. Due to time and software constraints, the azimuthal asymmetry is not used in the event reconstruction for this thesis. However, this analysis could be significant for future studies, as we find that the relative asymmetry amplitude of simulated photon-induced showers is quite large, with peak values of 0.7.

6.1. Shower Simulations

The data used for the fitting of the energy calibration parameters in the case of a classical reconstruction, as illustrated in this chapter, and the training of the ML-reconstruction model described in the next chapter, consist of the same set of Monte Carlo (MC) showers. All showers have been simulated with the `CORSIKA` [70] simulation software (version 7.7420), using the hadronic interaction model `EPOS LHC` [71] for high energy interactions and `FLUKA-INFN` [72, 73] for low energy interactions. The WCDs simulation for the detector response is done in Offline using the `GEANT4` [74] toolkit for particle physics simulation. In total, 5500 `CORSIKA` showers are simulated. The energy of the showers ranges from 10^{18} eV to $10^{20.2}$eV with a spectral index of -1; ensuring an oversampling of the highest energies compared to the expected real distribution. For the nonlinear NNs, it is crucial to cover

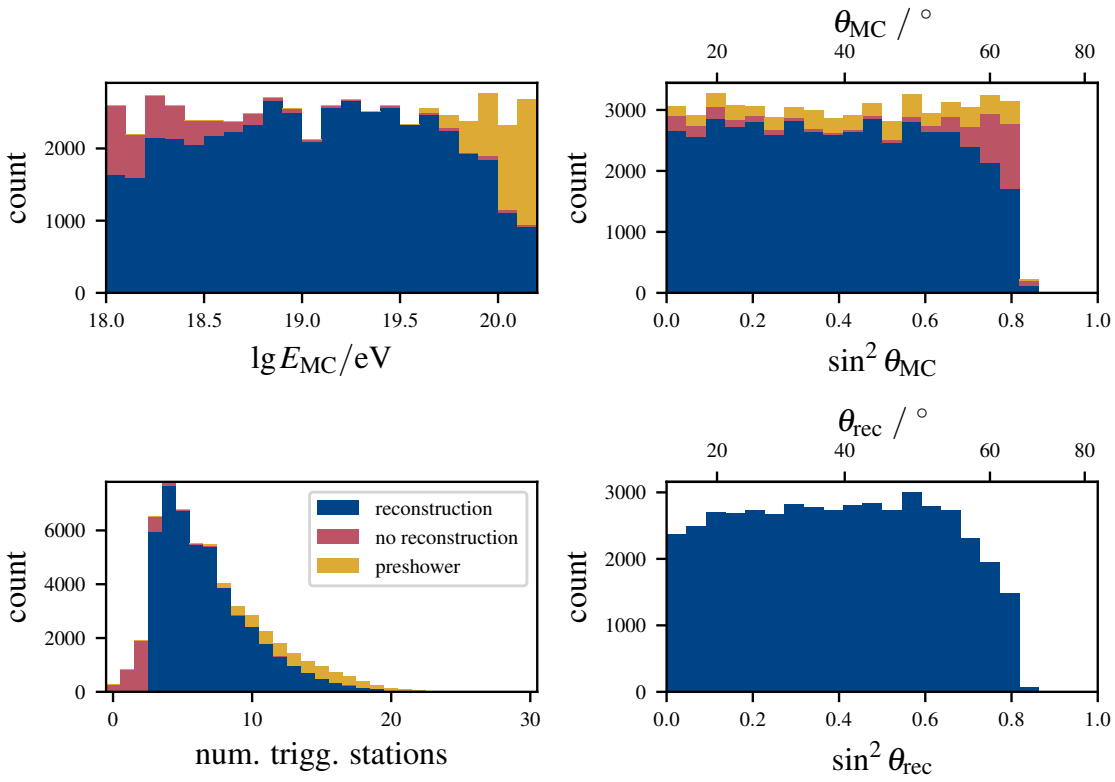


Figure 6.1.: **Properties of the data set.** Shown are distributions for the Monte Carlo energy (upper right), the number of stations that triggered in the event (lower right), the Monte Carlo shower zenith angle (upper left), and the reconstructed zenith angle (lower left, reconstruction with the photon-LDF derived in this thesis). It is differentiated between events that were reconstructed successfully, events that did not fulfill the reconstruction criterion, and preshower events.

the whole range of desired inputs and outputs, as failure to do so may yield unreliable results in the affected parameter regions. This is due to the fact that NNs generally cannot extrapolate beyond the provided training data. Moreover, the models can only interpolate along a given parameter axis with a high degree of reliability if the distance between the data points in the training data set is sufficiently small. A parameter axis in this context could be the zenith angle or the MC energy of the showers.

The zenith angle distribution of the showers is uniform in $\sin^2 \theta$, as this is the distribution seen by the observatory if the arrival direction of CRs is isotropic and the detection efficiency is perfect. The simulated range is $0° \leq \theta \leq 65°$. All showers are simulated with the same atmospheric conditions and without accidental muons. Accidental muons are atmospheric muons, that coincide with a shower signal by random chance. To increase the size of the statistic, every shower is simulated 10 times at different positions in the array, leading to a total of 55 000 simulated showers. The `CORSIKA` showers include events with preshower effects. These events are removed from the data set for the analysis, as their signature differs significantly from that of non-preshowering events. It is nonetheless

important to know the distribution of preshowering events, to find a parameter region at which the classical reconstruction can be applied without any major contamination of preshowering events. If not explicitly stated otherwise, preshowering events are not included in any analysis throughout this thesis. The distribution of all events can be seen in Figure 6.1. The showers are separated into preshowering and non-preshowering events as well as events where a reconstruction is possible and not possible, using the LDF form this work. After the exclusion of preshowers, there are 50 550 events left, out of which 46 424 can be fully reconstructed, although many do not fall into the parameter region at which the parameterization of the LDF and DEC is carried out. On top of the regular stations from the triangular array, a set of 12 so-called dense stations is additionally simulated for each event. The dense stations are virtual stations, arranged in a ring in shower plane coordinates with a distance of 1000 m to the MC shower axis. The angular spacing between the dense stations is equidistant and the offset is chosen, such that one station is positioned at $\zeta = 0$. The rationale behind the dense rings is that an estimation of S_{1000} can be derived from the signals measured by the dense stations. The reconstructed value of S_{1000} with the classical method will be compared to the values estimated with the dense rings.

A second dataset, comprised of 550 events with identical properties is created to characterize the asymmetry in the following section and to parameterize the LDF function in Section 6.3. The difference solely exists in the configuration of dense stations. The data set includes 15 dense rings with 24 stations each, equidistantly positioned at distances ranging from 100 m to 1500 m. This is also the reason for the small size of the data set, as a large amount of stations close to the shower core is computationally expensive.

6.2. Azimuth Asymmetry of Photons

As previously discussed in Section 4.2.2, the measured signals may exhibit an asymmetry depending on the azimuth angle of the station relative to the shower axis for a given distance r in shower plane coordinates. In this section, an examination of the amplitude of this asymmetry in photon-induced showers is performed. The key element will be a parameterization of the asymmetry magnitude α, as defined in equation (4.8), with respect to the distance r in shower plane coordinates, the zenith angle θ of the event and the shower size S_{1000}. As the extraction of α is independent of the efficiency of the observatory, all events that are non-preshowering are included. Prior to parameterizing the amplitude of the asymmetry, it is necessary to determine it on a per-dense ring basis. To this end, equation (4.8) is fitted to the unsaturated station signals in each dense ring, where the parameters α and S' are free. In other words, the amplitude of the azimuthal asymmetry and the azimuthal-asymmetry-corrected signal is estimated for each dense ring. An illustrative example of such a fit can be found in Figure 6.2.

This fit is also possible for small distances to the shower core, as the usually simulated saturation effects are bypassed, since the goal is to parameterize the photon showers as well as possible from MC data, for a precise fitting procedure. This is achieved by accessing

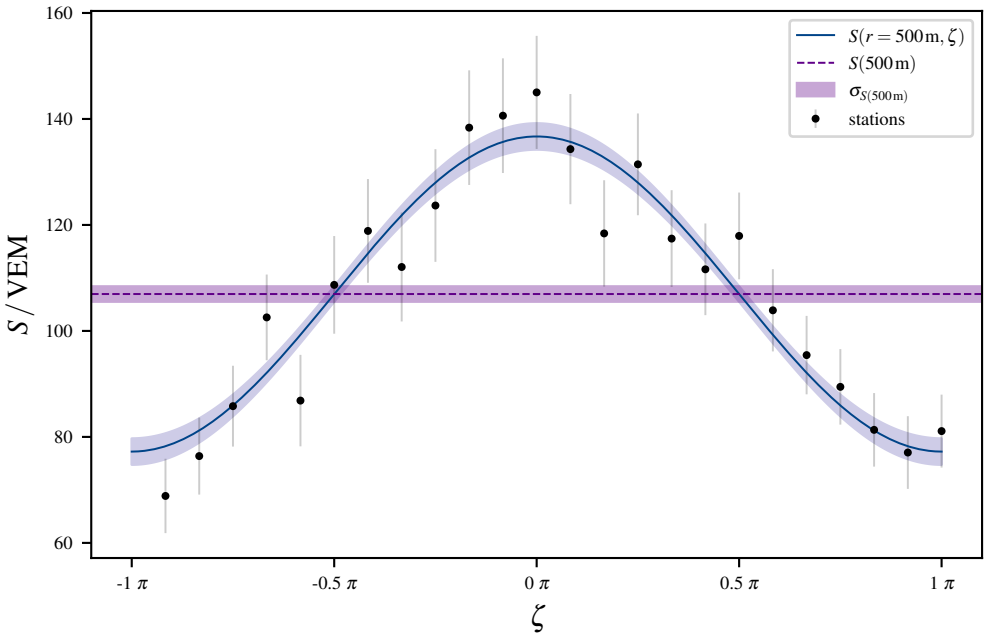


Figure 6.2.: **Exemplary fit of the azimuthal asymmetry in showers.** The fit of equation (4.8) (blue) for one dense ring at 500 m distance to the impact point in shower plane coordinates is shown, where the parameters α and S' are free. The estimated value of S' with its uncertainty is plotted too (purple). The gray uncertainties of the station signals are not used in the fit. The event has an energy of 4.2×10^{18} eV and a zenith angle of 44.3°.

the unsaturated traces recorded by the PMTs and calculating the signal calibrated to VEM charge via integrating the traces and dividing by the PMT-specific Area over Peak (AoP) value. The AoP is the fraction of the VEM charge to the VEM peak, as defined in Section 4.1. This conversion with the AoP is necessary because the accessed traces are calibrated to the VEM peak. For the total charge, we need to change the calibration to VEM charge, which is achieved by dividing by the AoP value. The AoP value is fixed to be approximately 3.2 in the case of Monte Carlo simulations. For real stations, the AoP values are known to decrease with the age of the tanks, as the reflectivity of the tank liner lessens and the absorption length of the photons in the water changes. The LDF fit in Section 6.3.1 also uses the unsaturated signals. However, one problem is that the stations close to the core can receive simulated particles with very high thinning weights. The concept of thinning in simulations reduces the number of particles by combining multiple particles into one, which is injected into the detector multiple times. Further explanations can be found in [70, 75]. A result of this, are artificially high fluctuations of the signals for small values of r, which are also not reflected in the signal uncertainty model. The variance of the number of particles in a station is given by

$$\mathrm{Var}[N] = \mathrm{E}[N_{\mathrm{thin}}]\left(\mathrm{Var}[W] + \mathrm{E}[W]^2\right), \tag{6.1}$$

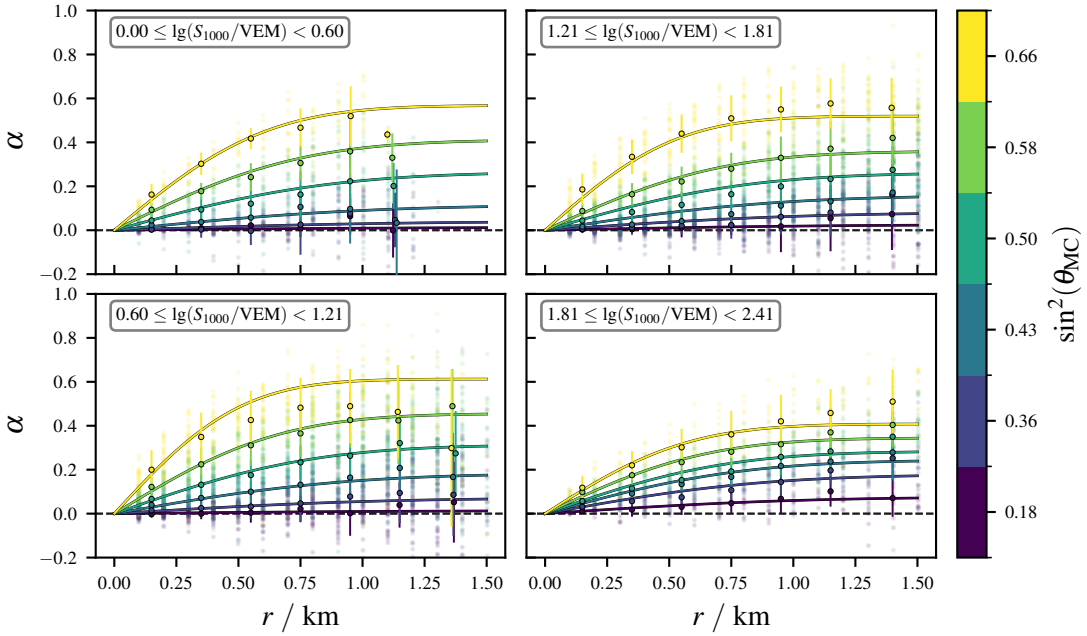


Figure 6.3.: **Photon asymmetry fit as a function of r.** Additional binning is done in $\sin^2\theta_{\mathrm{MC}}$ on the color axis and in $\lg(S_{1000}/\mathrm{VEM})$ by dividing the data into four plots.

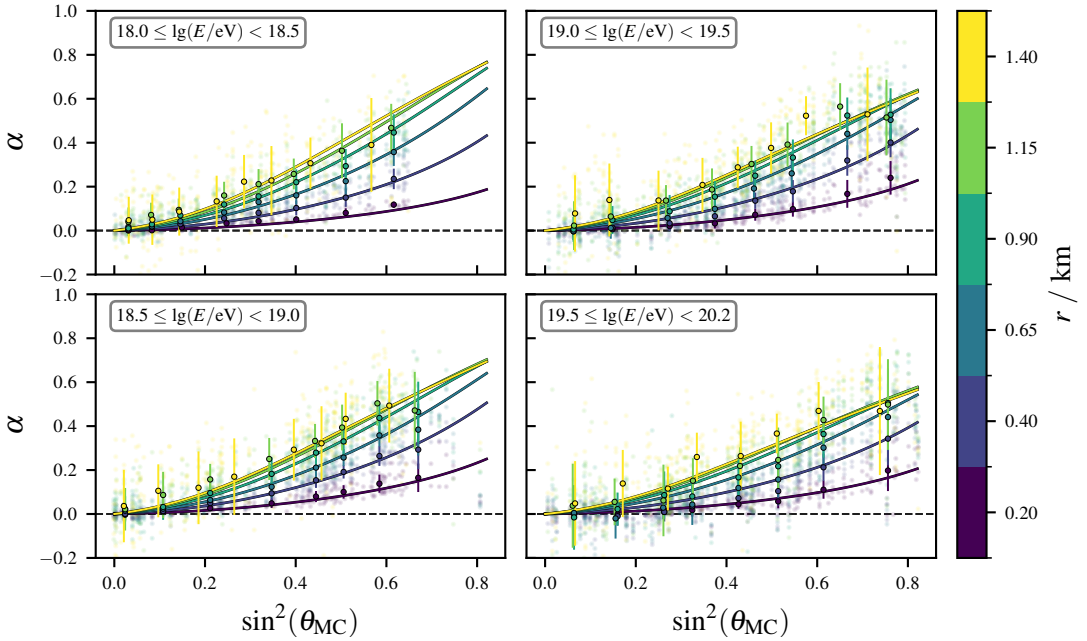


Figure 6.4.: **Photon asymmetry fit as a function of $\sin^2\theta_{\mathrm{MC}}$.** Additional binning is done in r on the color axis and in $\lg(E/\mathrm{eV})$ by dividing the data into four plots.

where N_{thin} is the number of thinned particles and W is the distribution of the weights in the thinning. For the derivation see Reference [76]. To derive equation (6.1) from what is written in Reference [76], we additionally applied the assumption that N_{thin} follows a Poisson distribution. At distances close to the core, the expectation value of the thinning weights $\mathrm{E}[W]$ is very high on average, leading to an increase in the variance of particles and therefore in the variance of the signal. The normal signal uncertainty model is not designed for these high signals and thinning levels, which introduces artificial fluctuations into the estimated signal errors. For this reason, the individual signal uncertainty of the stations is not considered when fitting equation (4.8) to the dense rings. A plot that showcases the fluctuations of the signal uncertainty at small core distances is presented in Figure A.2 in the appendix. However, with enough stations in one dense ring, the uncertainty of the parameter α is reflected by the statistical fluctuations of the station signals. In addition to the recovery of the uncertainty of α, there is another reason to demand that at least 80 % of stations in a ring are triggered and contribute to the fit. If too few stations with a similar value of ζ contribute to the fit, the asymmetry cannot be determined with sufficient precision. This is demonstrated by the following extreme example: consider an event that exhibits a very pronounced asymmetry, but only the three most upstream dense stations measure a signal. As the cosine in equation (4.8) is approximately equal to one in the first order approximation, $\cos(\zeta)|_{\zeta \ll 1} \approx 1$, the estimation of α is not sufficiently precise in this case. The 80 % requirement ensures the quality of the fitted α.

For the parameterization of α, the following function is employed:

$$\alpha(r, \theta, S_{1000}) = a(\theta, S_{1000}) \operatorname{erf}\left(\frac{r}{r_0(\theta)}\right), \tag{6.2}$$

where

$$a(\theta, S_{1000}) = \frac{(a_0^{(\mathrm{a})} + a_1^{(\mathrm{a})} \lg(S_{1000}/\mathrm{VEM})) \sin^2\theta}{1 + \exp\left(-\frac{\sin^2\theta - a_2^{(\mathrm{a})}}{a_3^{(\mathrm{a})}}\right)},$$

and

$$r_0(\theta) = r_0^{(\mathrm{a})} + r_1^{(\mathrm{a})} \sin^2\theta.$$

Parts of the functional form are found by using the symbolic regression technique implemented in the PySR [77] python library while the main structure was inspired by the parameterization used in Reference [57]. Symbolic regressors try to solve a minimization problem by finding multiple analytical solutions of varying complexity, given a set of allowed operations. In this thesis, the use of numeric parameters, the Gaussian error function, as well as addition and multiplication operations are allowed. As the choice of function (6.2) is empirical, one of the solutions from the symbolic regression was chosen that follows the data nicely while not overfitting the data. It was furthermore ensured that α converges to zero for $r \to 0\,\mathrm{km}$ or $\theta \to 0^\circ$ as this is mandated by the geometry of the problem. Then a last least squares fit with the MINUIT algorithm [78], using the iminuit framework [79], of equation (6.2) is performed to ensure the optimal choice of numerical values for the parameters $a_0^{(\mathrm{a})}$, $a_1^{(\mathrm{a})}$, $a_2^{(\mathrm{a})}$, $a_3^{(\mathrm{a})}$, $r_0^{(\mathrm{a})}$ and $r_1^{(\mathrm{a})}$. In the fit, the Monte Carlo zenith angles θ_{MC} are used. This is done, since if the asymmetry would be used in

the event reconstruction, the reconstructed shower axis does depend on the asymmetry model and the reconstructed angles would be only obtainable after the implementation of the asymmetry parameterization. A convergence of the fit is reached with a reduced χ^2 value of 3.4. The numerical fit result for the parameters can be found in the appendix in Table A.2.

In Figure 6.3 the parameterization is plotted as a function of r with binning in $\sin^2\theta_{\mathrm{MC}}$ for the color axis and a further division into four different panels, depending on the value of $\lg(S_{1000}/\mathrm{VEM})$. In this plot, the individual dense rings are displayed as points with a high transparency factor in the background. The markers with a black edge and error bars help to interpret the plot by displaying the mean and the standard deviation for the individual points for bins in $\sin^2\theta_{\mathrm{MC}}$. A different representation of the same data is depicted in Figure 6.4. The distinction between this figure and the previous one is the swap of the x and color axes, and the splitting into the four panels depends on E_{MC} instead of the estimated value of S_{1000}.

It is obvious that although the reduced χ^2 is relatively high, the fit follows the data well. The reason for this could be strong fluctuations in the asymmetry amplitude from shower to shower. By looking at the plot, it does become apparent, that some showers show extreme asymmetry amplitudes with $\alpha > 0.7$, even at distances $r < 1\,\mathrm{km}$, which is quite significant, as this can considerably influence the estimated energy. Utilizing the asymmetry parameterization for future works is therefore advised. Even the fitted parameterization reaches values of $\alpha \approx 0.6$. This significant asymmetry amplitude likely arises from the strong absorption of the electromagnetic shower component, which dominates photon-induced showers. Simulations of proton-initiated events at a zenith angle of $\theta = 55°$, yield an asymmetry amplitude of approximately 5.8 [57], which is only marginally smaller than the here determined values. However, it is known that simulations of proton-induced showers underestimate the muonic component compared to observed hadron-induced showers [80, 81]. An increase in the muonic component in the simulations would decrease the amplitude of the azimuthal signal asymmetry [57]. Consequently, the observed asymmetry for proton-induced showers is likely overestimated. Another noteworthy effect is the existence of data points with values of $\alpha < 0$. Although the theory does not dictate that the asymmetry amplitude must be positive, the observed effect is probably mainly due to statistical fluctuations in showers with low zenith angles.

6.3. Lateral Distribution Function of Photons

6.3.1. Parameterization of the LDF

For the description of the LDF for photon-induced showers we use the same functional form as used in the hadronic case—the modified NKG function, as defined in equation (4.3). This is a reasonable choice, as the original NKG function was derived for electromagnetic showers. What is changed is how the parameters β and γ are parameterized: finding new forms of equation (4.6) and (4.7). For the parameterization, it is necessary to select a subset

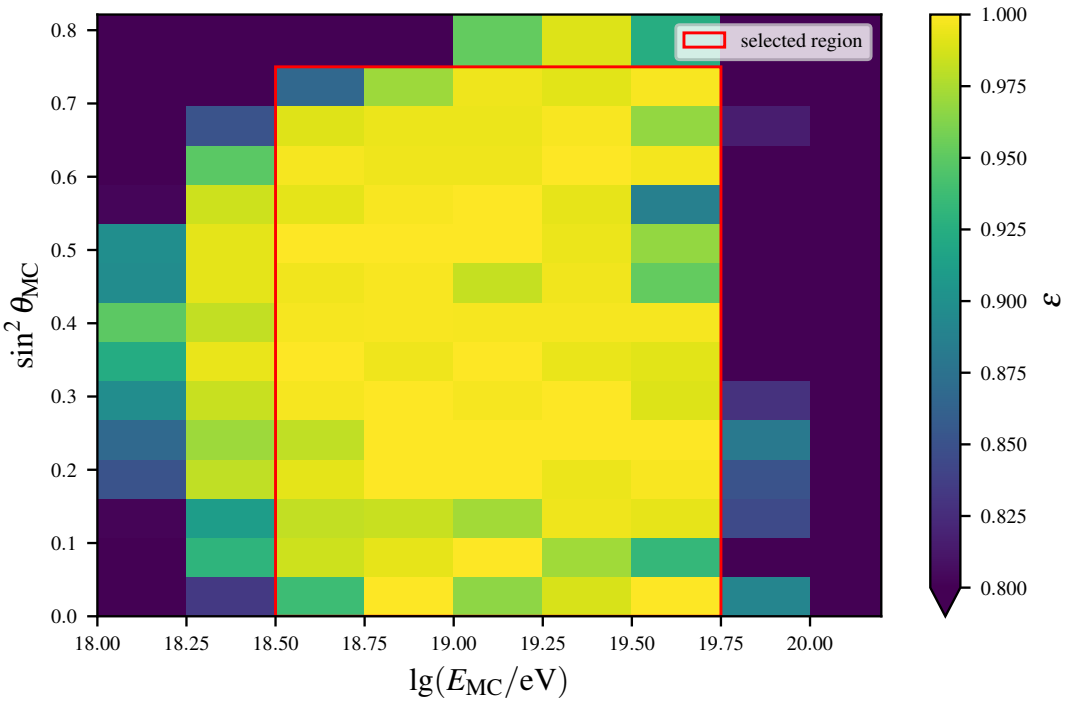


Figure 6.5.: **Two-dimensional histogram with the fraction of usable events to total events with bins in $\lg(E_{MC}/\mathrm{VEM})$ and $\sin^2\theta_{MC}$.** The parameter ε is defined as the fraction of non-preshowering events passing a T5 and 6T5 trigger compared to the total number of events in a bin. The red rectangle shows the parameter space selected by the cuts.

of the data set where the fraction of non-preshowering events and the efficiency of the main array of the Pierre Auger Observatory are sufficiently high for photon showers. The efficiency cut ensures that no bias is introduced into the fits due to upward fluctuations of showers. If the full efficiency is not reached, showers with upward fluctuations would dominate the fit, as they are more likely to be captured by the array. For this reason, cuts for low energies and for high zenith angles are put into place with the values of $10^{18.5}\mathrm{eV} < E_{MC}$ and $\theta_{MC} < 60°$. We are cutting on the MC angles, as the angular reconstruction is not yet optimized for photons and we use the MC angle during the LDF fit. At the upper end of the energy spectrum, the relative number of preshowers increases. As this parameterization is aimed towards non-preshowering photons, an additional cut is put in place, selecting only events that satisfy $E_{MC} < 10^{19.75}\mathrm{eV}$. By using only showers that satisfy the above conditions, a fraction of at least $\varepsilon = 80\,\%$ suitable showers for the LDF fits is ensured. A value of 80 % is a good compromise for covering a large parameter space, while not being too influenced by upward fluctuations and preshowers. The fraction is measured as the quotient of non-preshowering events that fulfill the T5 and 6T5 trigger criterion [1] divided by the total amount of events, for a two-dimensional binning in $\lg(E_{MC}/\mathrm{eV})$ and $\sin^2\theta_{MC}$,

[1] Energy and angles of an event can be properly reconstructed and all 6 stations around the station with the highest signal exist and are operational.

as seen in Figure 6.5. The latter shows ε for a two-dimensional histogram together with the selected cuts.

For the LDF, we want the signal of the shower at distance r to the shower axis without the influence of the azimuthal asymmetry or saturation effects. Therefore, the same fit as in Section 6.2 to equation (4.8) is performed for each dense ring. Instead of α, we now focus on the fitted $S'(r)$. With the 15 dense rings, the LDF is effectively sampled by the $S'(r \in \{100\,\mathrm{m}, 200\,\mathrm{m}, ..., 1500\,\mathrm{m}\})$ at up to 15 distances to the shower axis (rings that have less than 80 % of their stations triggered are discarded again). The uncertainties on $S'(r)$ are gained from the statistical fluctuation of the unsaturated signals, analog to the asymmetry amplitudes in Section 6.2. In the next step, the azimuth-independent signals $S'(r)$ from the dense rings are used to fit equation (4.3) to the events with the least-squares method, where the parameters $S(r_{\mathrm{opt}} = 1000\,\mathrm{m})$, β and γ are free. The LDF for this is given by equation (4.4). This minimization is performed on the remaining 279 events with scipy's [82] `curve_fit` method.

Theoretically, it would be possible to find parameterizations for β and γ with the just-fitted values, but β and γ are highly correlated. This is problematic, as a small shift in the shower can greatly influence γ where β is compensating for the shift and vice versa. To eliminate this behavior, γ is not directly fitted as a function of S_{1000} and θ, but rather as a function of the fitted β. On a statistical average, the fluctuations in the β-γ correlation cancel out, resulting in the mean behavior of the function $\gamma(\beta)$ being recovered. Initially, it was planned to parameterize γ as a function of β, S_{1000}, and θ. However, it became evident that the inclusion of S_{1000} and θ did not provide any additional information. Consequently, a less complex parameterization is achieved by solely relying on β. However, future works on this topic may expand on this. Especially when a broader parameter space for the energy is used, a more complex functional form for γ could be needed. For the fit, the function

$$\gamma(\beta) = a_\gamma + b_\gamma \beta + c_\gamma \exp(d_\gamma \beta) \tag{6.3}$$

is used, were the free fit parameters are a_γ, b_γ, c_γ, and d_γ. The numerical parameter values can be found in Table A.3 and the fit is displayed in Figure 6.6. The uncertainties of the γ values, taken from the initial fit, are used for this fit, and the β values are modeled without errors.

Now that γ is described as a function of β, equation (4.4) can be in itself described as only a function of β

$$f_{\mathrm{LDF,\,photon}}(r;\beta) \coloneqq f_{\mathrm{LDF,\,NKG}}(r;\beta,\gamma(\beta)). \tag{6.4}$$

The event-wise LDF fit is then repeated with equation (6.4) for the LDF shape, where the shower size and β are free in the fit. Figure 6.7 shows some examples of the fits, namely the two worst and the two best fits, as measured by their χ^2 values. The worst fits demonstrate that function (6.4) is generally capable of following the shape of the LDF to a margin of error with only one free parameter.

The finishing step towards a parameterization of a photon LDF is the parameterization of β. The parameterization of beta and gamma is in each case a function of S_{1000} and θ. However, the dependency on S_{1000} was found to be negligible for photons.

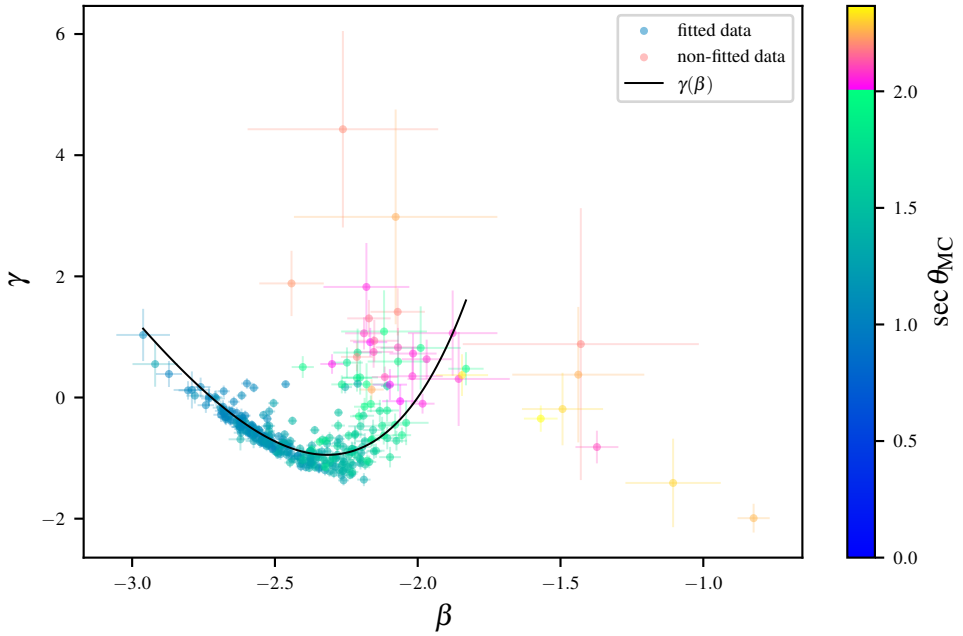


Figure 6.6.: **Fitting γ as a function of β for the LDF parameterization.** Only data points with $10^{18.5}\text{eV} < E_{MC} < 10^{19.75}\text{eV}$ are shown and used in the fit. The selection for the fit is further limited to events with $\theta < 60°$ (blue-green colors). Events with $\theta \geq 60°$ are not used in the fit and displayed in magenta-yellow colors.

The simple linear function

$$\beta(\theta) = a_\beta + b_\beta \sec\theta \tag{6.5}$$

is chosen to describe the behavior of β. Here, a_β and b_β are the fitted parameters, whose values can be found in Table A.4. The small uncertainties stem from the small propagated uncertainties on the values of the β-s. A plot with the fit to the data can be seen on the left side of Figure 6.8. The linear behavior is apparent for values of $\theta_{MC} > 45°$, but for lower values, the bulk of data is below the fitted function. This could eventually be a physical effect, but the data points do not influence the shape of the fit greatly, as the uncertainties in β are much larger for these data points. To investigate this further, two alternative fits with a second-order polynomial are shown in red in the upper left panel of Figure 6.8. One fit uses the uncertainties of the data points and is shown as a solid red line. The other is drawn as a dashed red line and ignores the data uncertainties. It is evident, that the latter follows the bulk of the data, while the polynomial fit that uses the uncertainties of the data points behaves similarly to the linear fit. Nevertheless, it was decided to use the linear fit, which propagates the errors correctly and has a minimum amount of parameters. The importance of this decision is lessened by the fact that the influence of β for the LDF shape in the affected region is not as significant as for higher values of β, as it can be seen in the right side of Figure 6.8, where function (6.4) is plotted for different values of β.

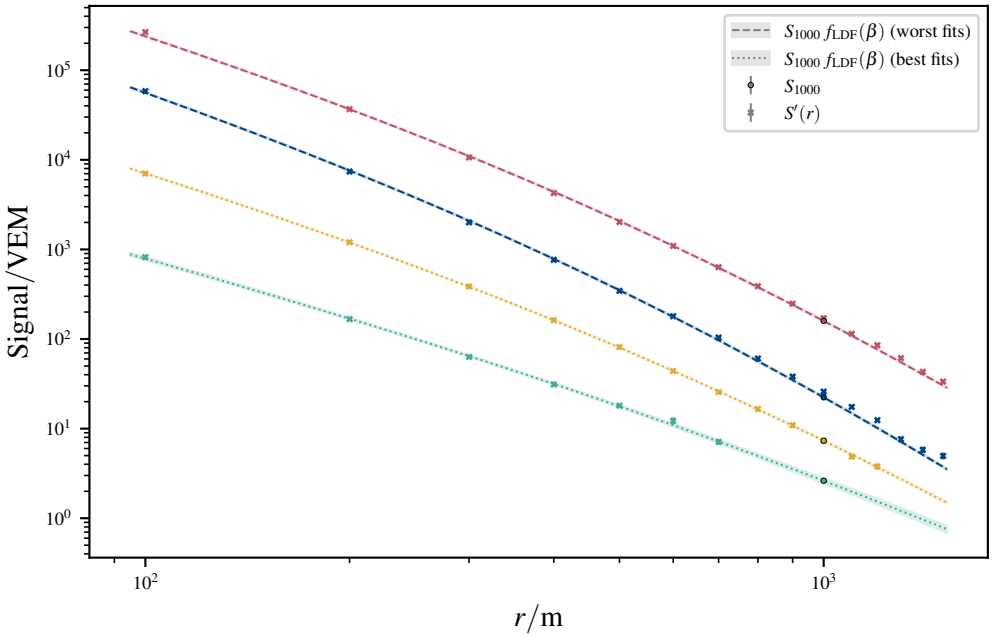


Figure 6.7.: **Examples of the event-wise fits of the photon LDF with β as the free parameter.** The two best fits (dashed) and the two worst fits (dotted) are displayed together with the uncertainty band of the fit. Additionally, the estimated values for S_{1000} are displayed as dots.

The individual uncertainties of the β's are relatively small, although the overall spread is large. A partial explanation for this is a systematic shift of β with the value of X_{max}. This makes sense since the LDF of a shower with the shower maximum closer to the ground looks different. The correlation between β and X_{max} is discernible from the lower panel of Figure 6.8. Showers with a higher X_{max} have lower values for β, which can be interpreted, as a shallower LDF, as seen in the right panel on the figure. The here displayed values of X_{max} are determined by a fit of the Gaisser–Hillas function to the MC longitudinal shower profile and are very imprecise for deep showers since the profile is only given down to the detector height. Therefore, it does not reflect the actual physical value and must be solely understood as the only possible loose estimation. For more information refer to Section 7.6. Currently, no method for the estimation of X_{max} from the SD-recorded shower footprint for photon-induces showers exists, therefore the value of X_{max} cannot be used in a LDF parameterization. However, it could be possible for future studies to expand on this by using machine learning for the X_{max} approximation (see e.g. Reference [83] for hadrons) or develop a new method to estimate X_{max} from the footprint of photon-induces showers to use it in the LDF parameterization. Another part of the explanation for the spread is shower-to-shower fluctuation, where the values of β can be accurately estimated for each shower, but the LDF shape changes between events, even for identical shower energies and zenith angles.

For the Offline reconstruction, an estimation of the β-uncertainty is expected for the propagation of systematic uncertainties, which can be a function of S_{1000} by implementation.

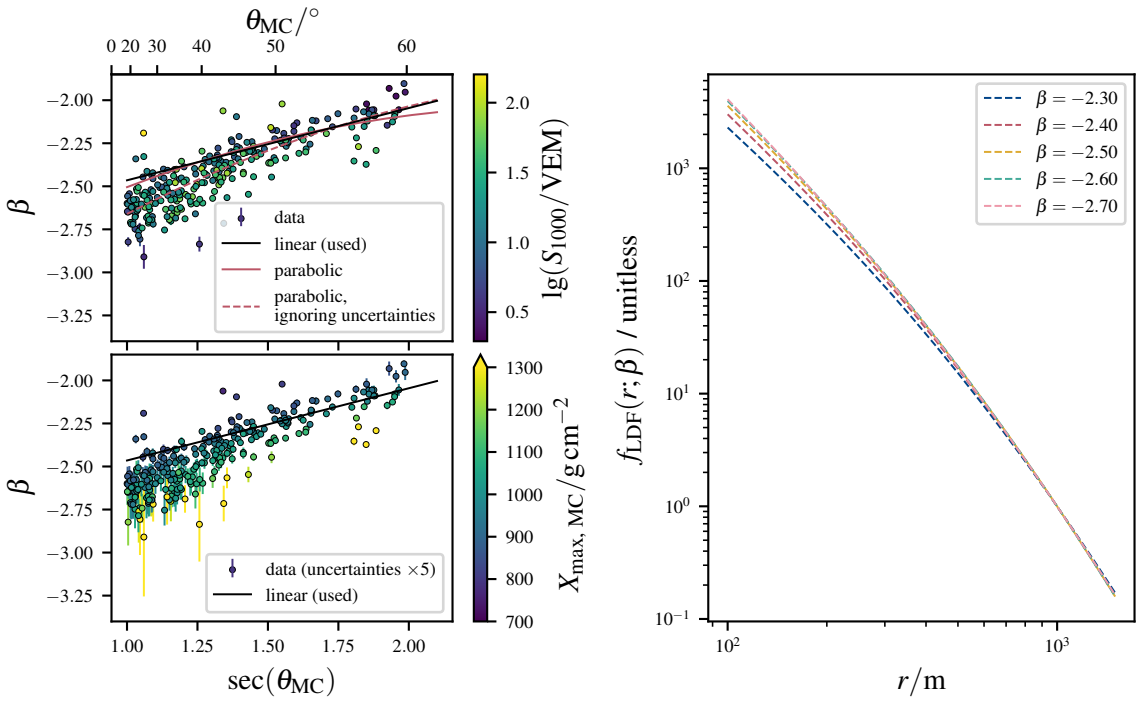


Figure 6.8.: **Fitting β as a function of θ for the LDF parameterization.** The upper left plot shows the linear fit of equation (6.5) to the data and $\lg(S_{1000}/\mathrm{VEM})$ is represented on the color-axis. Alternative fits are shown in red. The lower plot shows the same, except the color presents the X_{max} of the events and the uncertainties are enlarged by a factor of 5 for better visibility. Note that the X_{max} from MC data is imprecise for large values and does not necessarily match the actual physical value. Please refer to the text for more information. The right plot shows equation (6.4) for different values of β.

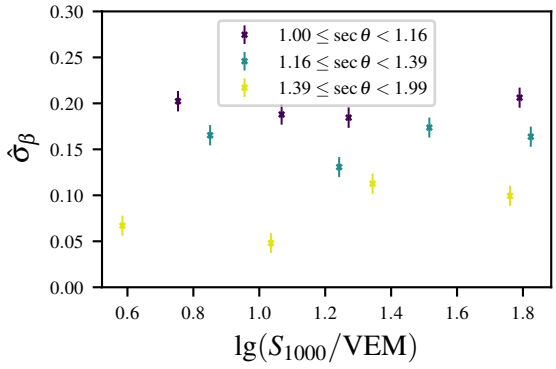


Figure 6.9.: **Uncertainty estimation for the LDF parameterization.** The evaluation of equation (6.6) for different binnings in θ_{MC} and $\lg(S_{1000}/\mathrm{VEM})$ is depicted. The uncertainties of the $\hat{\sigma}_\beta$ are determined via bootstrapping.

The uncertainty does not influence the nominal values of the reconstruction process and will not be utilized further in this work. However, it could be of interest for future studies. The deviation from the event-by-event fit to the general parameterization is calculated to gauge the uncertainty of the photon LDF parameter. With these residuals, a modified standard deviation $\hat{\sigma}$ of N events can be calculated according to

$$\hat{\sigma}_\beta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\beta_i - \beta(\theta_{\mathrm{MC}}))^2}, \tag{6.6}$$

where the β_i are the parameters from the event-by-event fit and $\beta(\theta_{\mathrm{MC}})$ is the parameterization. By doing this for bins in $\lg(S_{1000}/\mathrm{VEM})$ and $\sec\theta$, possible dependencies can be seen. This respective plot is shown in Figure 6.9. Both, the bin edges for $\sec\theta_{\mathrm{MC}}$ and for $\lg(S_{1000}/\mathrm{VEM})$, are determined through the quantiles of the data distribution so that every depicted data point features the same amount of events. Despite the limited quantity of data due to the high cost of simulations with numerous stations in close proximity to the core, a clear dependency of $\hat{\sigma}_\beta$ with respect to $\sec\theta_{\mathrm{MC}}$ is noticeable. The possibility to reflect this in Offline does not exist at the moment of writing, and implementing it would be beyond the scope of this thesis. Opposite of that, no significant dependency is observed for $\lg(S_{1000}/\mathrm{VEM})$, and the value of $\hat{\sigma}_\theta$ is therefore modeled as constant. A constant value can be calculated over the whole dataset, yielding $\hat{\sigma}_\theta = 0.1538 \pm 0.0110$.

6.3.2. Performance Evaluation of the Parameterization

Prior to calibrating the energy, the effects of the photon LDF created in this work are evaluated. For this, the developed LDF is compared to a reconstruction run, where photon-induced showers are reconstructed with the LDF that is optimized for hadrons. Each reconstruction is executed for all 55 000 showers, with the same event-wise seeds for the simulation and reconstruction randomizers. With this, the showers are equal in position and the detector simulation is the same as well, so that the only remaining difference is the used LDF. To compare them, a subset of events is selected, with the aforementioned 80 % criterion for preshowers and detector efficiency. Furthermore, only events that pass a T5 and 6T5 trigger and that can be properly reconstructed with both LDFs are selected, leaving 27 892 events. It was found that 32 fewer events could be fully reconstructed with the photon LDF than with the hadron LDF (27 892 vs. 27 924).

The first improvement is a reduced error of the reconstructed core position $\vec{x}_c$. The respective histogram is shown in the upper panel of Figure 6.10, where the blue bins show the reconstruction with the here developed photon LDF parameterization and the red bins show the reconstruction with the default LDF parameterization for hadrons. The 68.27 % percentile is reduced from 98.52 m to 85.62 m. Thus, the reconstruction of the core position for photon-induced showers is significantly improved by using the here developed LDF. It is believed, that an additional inclusion of the azimuthal signal asymmetry would further improve the accuracy of the reconstruction.

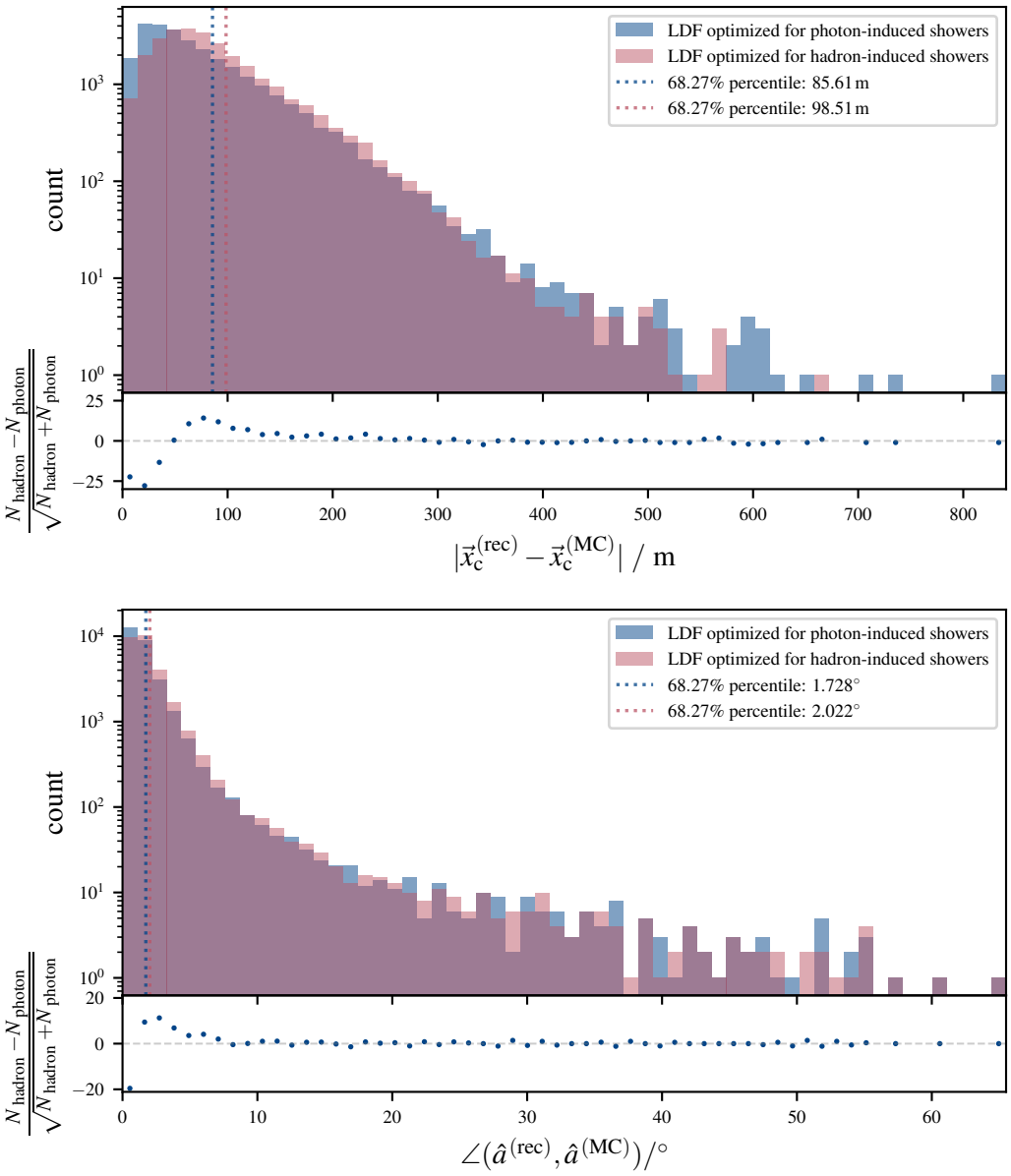


Figure 6.10.: **Improvements of the impact point and shower axis with the LDF optimized for photon-induced showers over the LDF optimized for hadron-induced showers for events initiated by photons.** The upper plot shows the distribution of the error on the reconstructed (rec) impact point to the Monte Carlo (MC) impact point. The lower plot shows the distribution of the error on the reconstructed shower axis to the Monte Carlo shower axis. The dotted lines in every histogram are the 68.27 % percentile of the data sets. Below each histogram, the pull values for every bin are displayed. N_{photon} and N_{hadron} are amount of events in the bins given the respective LDFs.

Another improvement is observed in the angular resolution of the shower axis $\hat{a}$. This is presented in the lower panel of Figure 6.10, where it can be seen that the 68.27 % percentile is enhanced from 2.023° to 1.728°. Although not as significant as the improvement seen for the core reconstruction, a better angular resolution is essential for potential photon-induced events, as the original direction of the photon is of great interest. We believe that the inclusion of the signal asymmetry model into the reconstruction would further increase the angular accuracy.

As each event is simulated with a dense ring at a distance of 1000 m, an LDF independent estimation of the shower size can be made. By removing the effects of the asymmetry as described above, a direct estimation of the shower size can be obtained from the dense rings. The reconstructed values can now be compared to this value. The result is shown in Figure 6.11, where the upper subfigure shows the relative bias and the relative resolution in dependence of the Monte Carlo energy, and the lower subfigure shows the dependency as a function of the Monte Carlo zenith angle. It should be noted that only events are included, where at least 9 of the 12 stations in the dense ring did trigger, to ensure the proper removal of asymmetry biases. It can be seen, that in all cases where the photon LDF is used, a slight underestimation is present. A possible cause for this could be the non-inclusion of asymmetry effects in the reconstruction chain. Nevertheless, the here developed photon LDF yields better results for bias and resolution across the entire parameter space compared to the LDF for hadron-induced showers.

6.4. Direct Energy Calibration with Monte Carlo Data

This section describes how the Direct Energy Calibration (DEC) is applied to the reconstructed events. The DEC is the simultaneous fit of equation (4.10) and (4.11) to the events, in order to determine the parameters a, b, c, A, and B. Fitting the DEC is not feasible without applying additional cuts on the zenith angle. Vertical photon showers at high energies can penetrate the atmosphere deeply and have a very deep shower maximum, as discussed in Section 2.3. Since the footprint of these showers can look significantly different, the assumption of a linear behavior between the zenith-corrected shower size and the energy does not hold anymore. By discarding events that do not suffice $30° < \theta_{\mathrm{rec}} < 60°$ on top of the 80 % cut for efficiency and non-preshowering photons, it is ensured that the majority of the showers reach their X_{max} before hitting the detector. Here, θ_{rec} is the reconstructed zenith angle, where the here developed photon LDF is used. By restricting the zenith range, the median of the distribution changes to approximately 45°. Consequently, the reference point of the attenuation function is altered accordingly to $\theta_{\mathrm{ref}} = 45°$. With this, the energy of the classic reconstruction for this thesis is given by

$$\mathrm{E}_{\mathrm{DEC}}(S_{1000}, \theta) = A \left(\frac{S_{1000}}{f_{\mathrm{att}}(\theta) \cdot \mathrm{VEM}} \right)^{B} \cdot \mathrm{eV}, \tag{6.7}$$

where $$f_{\mathrm{att}}(\theta) = 1 + ax(\theta) + bx(\theta)^2 + cx(\theta)^3$$

and $$x(\theta) = \cos^2\theta - \cos^2\theta_{\mathrm{ref}}.$$

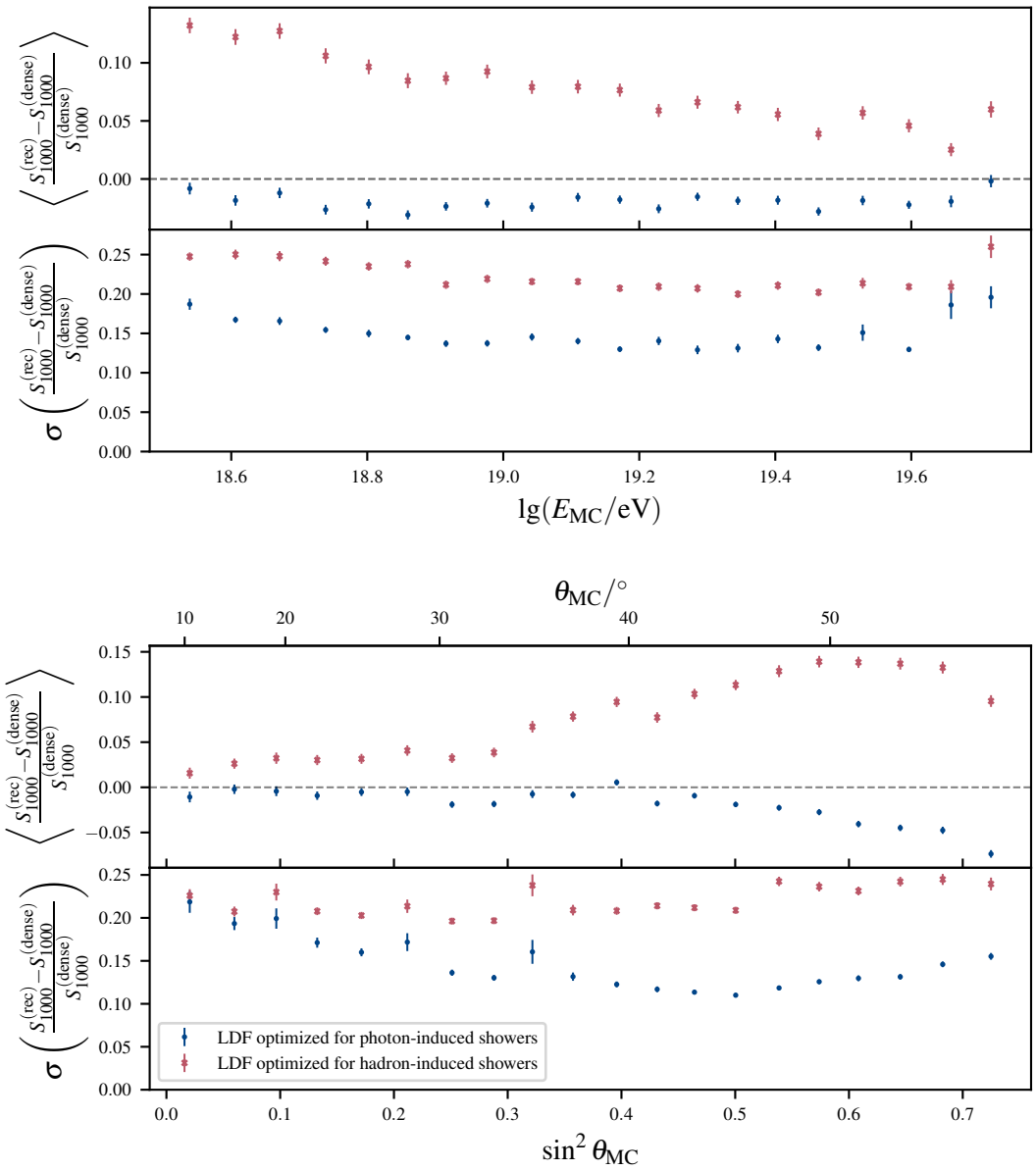


Figure 6.11.: **Estimation of S_{1000} with reconstructed with different LDFs compared to the estimation with dense rings.** The plot shows the relative bias and the relative resolution of the shower sizes for the LDF optimized for photon-induced showers and the LDF optimized for hadron-induced showers compared to the estimation with the dense ring at 1000 m. The upper panel shows the values as a function of $\lg(E_{\mathrm{MC}}/\mathrm{eV})$ and the lower panel as a function of $\sin^2 \theta_{\mathrm{MC}}$. The variable $S_{1000}^{(\mathrm{dense})}$ is the shower size estimation from the dense rings.

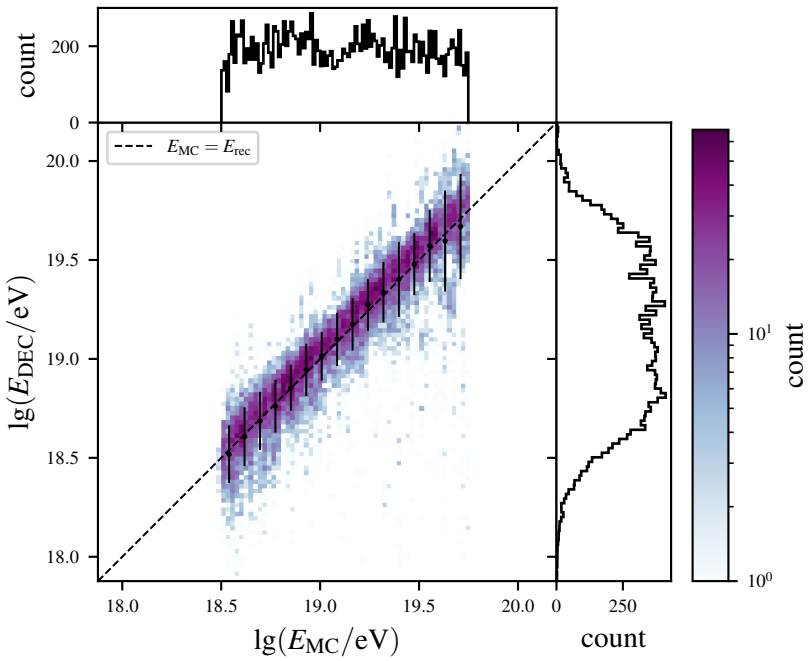


Figure 6.12.: **Direct energy calibration of photon-induced showers.** The plot shows the distribution of the predicted energy with the direct energy calibration and the photon LDF as a function of the Monte Carlo energy. The markers with the error bars show the mean and the standard deviation of the histogram for discrete bins in the Monte Carlo Energy and are just a visual guide. The dashed line represents the prediction of a perfect predictor.

For the fitting process, the equation is logarithmized, rearranged and the parameter A is replaced by $A' = \lg A$:

$$\lg(S_{1000}(E_{MC}, \theta)/\text{VEM}) = \frac{\lg(E_{MC}/\text{eV}) - A'}{B} + \lg(f_{att}(\theta)). \tag{6.8}$$

The transformation of A to A' is beneficial for the convergence of the minimization problem, as the anticipated value is expected to be in the range of $A \approx 10^{18}$. The inversion of the function was done, to enable an easy method to incorporate the uncertainties of the S_{1000}, although it was ultimately decided to not use them. Using scipy's minimization capabilities, equation (6.8) is fitted to the 19 313 events selected by the cuts, according to a least squares approach. The individual uncertainties of the S_{1000}, given by the Offline reconstruction, are not used in the minimization, as they would bias the fit toward events with larger shower sizes. The propagated uncertainty of $\lg(S_{1000}/\text{VEM})$ is proportional to $\sigma(S_{1000})/S_{1000}$, where $\sigma(S_{1000})$ is the uncertainty of S_{1000}. Since $\sigma(S_{1000})/S_{1000}$ decreases for increasing values of S_{1000}, showers with upward fluctuations of S_{1000} would receive a larger weight in the fit.

An additional constraint was provided to the minimization algorithm, that limits the parameter space to those parameters, that yield only positive values for $f_{att}(\theta)$, on the

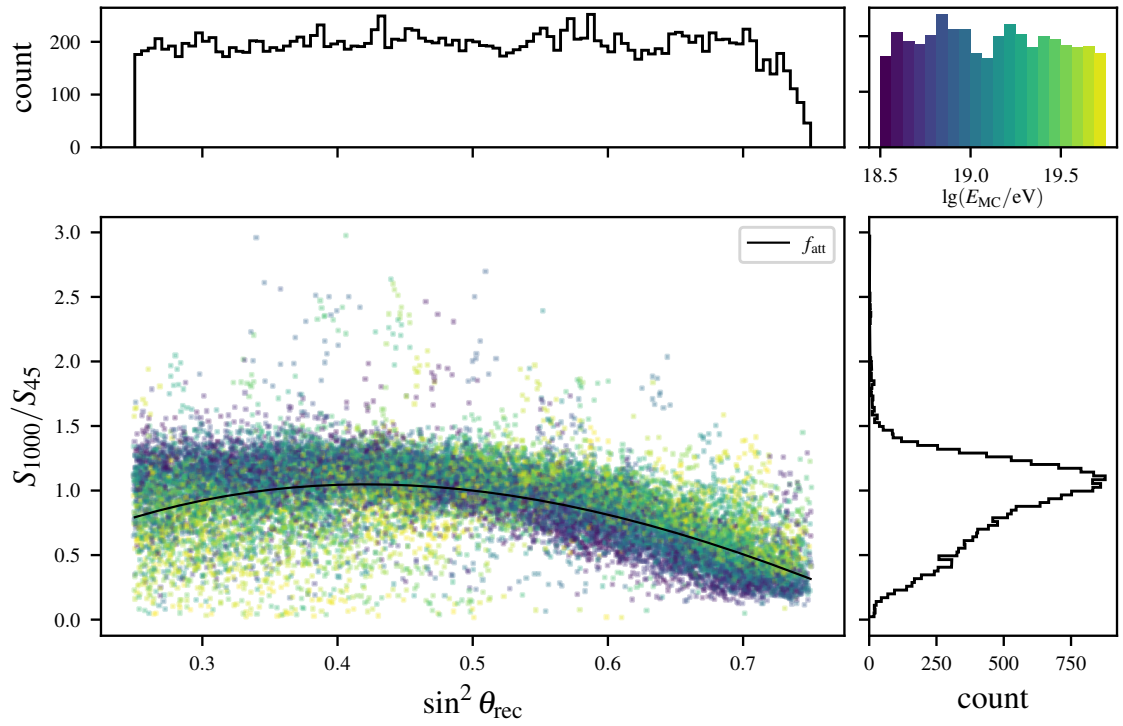


Figure 6.13.: **The attenuation function for photon-induced showers.** The black line is the fitted function $f_{\text{att}}(\theta; a, b, c)$ while the histogram shows the value S_{1000}/S_{45} evaluated for each event, where S_{45} is inversely calculated from the Monte Carlo energy and the parameters A' and B. The color-coding of the data points is determined by the Monte Carlo energy.

range $30° < \theta < 60°$. This improves the fit's robustness, as negative values of f_{att} are non-physical. To ensure a successful convergence, the minimum is checked with iminuit and the uncertainties of the parameters are calculated with the hesse algorithm and rescaled [2]. The found parameters are provided in the appendix in Table A.5. A two-dimensional histogram showing the final result as the relation between the Monte Carlo energy and E_{DEC} is depicted in Figure 6.12.

The energy reconstructed with classic methods follows a linear behavior up to energies of around $E_{\text{MC}} = 10^{19.6}$eV, above this threshold the energy is slightly underestimated. The reason for this is most likely an increasing population of events with a deep X_{max}, which are visible as outlier events. A caveat of the DEC method for photon-induced showers, is that equation (6.8) is not fully able to entangle the energy prediction from correlation effects of the shower size and the zenith angle. This can be seen when plotting the fit of f_{att} together with the value S_{1000}/S_{45}, where $S_{45} = S_{1000}/f_{\text{att}}$ is calculated for each event from the Monte Carlo energy using the optimized parameters. The result can be seen in Figure 6.13. The x-axis is given as a function of $\sin^2 \theta_{\text{rec}}$ and the data points are color-coded

[2] As the absolute uncertainties of the data points are not used in the fit, each data point is modeled with an uncertainty of $\sigma = 1$ for the minimization. To recover a correct scaling of the uncertainties in the end, the covariance matrix is multiplied by a constant factor, which is determined by the constraint that the reduced χ^2 value of the fit should be equal to one.

according to their Monte Carlo energy. If the assumptions that S_{45} effectively removes the zenith dependency of the shower size and that the mapping from S_{45} to the energy is completely linear hold, then no ordering with $\lg(E_{\mathrm{MC}}/\mathrm{VEM})$ should be visible. This, however, is not the case when looking at the distribution shown in Figure 6.13, since a clear ordering is discernible, which crosses over for increasing values of $\sin^2\theta_{\mathrm{rec}}$. This may be a hint to the fact, that the normal DEC equation does not fully capture the underlying data distribution. A more in-depth discussion follows in Section 8.1.

7. Photon Energy Reconstruction with Machine Learning

Using Machine Learning (ML) for the reconstruction of air showers can yield impressive results as already shown in the References [83, 84, 85, 86, 87]. This chapter presents the development, optimization, and application of two different network architectures to the energy reconstruction problem. The implementation of all presented neural networks was conducted using TensorFlow [88] and the Keras API [89].

7.1. Data Preprocessing

For the training and the evaluation of the neural networks, the same data set is used as for the classical reconstruction, described in Section 6.1. We again apply the 6T5 trigger and non-preshowering-criterion, without any restrictions regarding the MC energy or the zenith angles. The simulation data set is split multiple times, beginning with the train-test split. Around 20 % of events are randomly selected and reserved for a test data set. This is the data, on which the evaluation of the NNs will be happening. The remaining 80 % are split again into the actual training data set and the validation data set.

The prediction target y for the training is defined as

$$y_{\text{true}} = \lg(E_{\text{MC}}/\text{eV}) - 19.35. \tag{7.1}$$

With this transformation, the values of y are approximately zero-centered and reach values between $-1 \lessapprox y_{\text{true}} \lessapprox 1$. The exact value of 19.35 was chosen, to be consistent with previous works about the energy reconstruction with ML, which focused on hadrons [85]. Choosing the same convention allows for the utilization of the same prepossessing functions.

Every network presented in this thesis uses the traces recorded by the PMTs as part of their input. For every trace, the first 120 time bins after the start of the signal are used. The start of the trace is derived by Offline and the time duration of one bin is 25 ns. All PMT traces are calibrated to the VEM peak, as explained in Section 4.1. Normally, traces calibrated to the VEM charge are used for energy estimation, but since the AoP is fixed in simulations, the difference is a constant factor. Consequently, we can use the traces as is, since a small constant scaling factor is not relevant for the network training. If the networks are supplied with real data, this has to be modified, as described in Reference [85]. The mean of all triggered PMT traces is then used as the input to reduce the dimensions

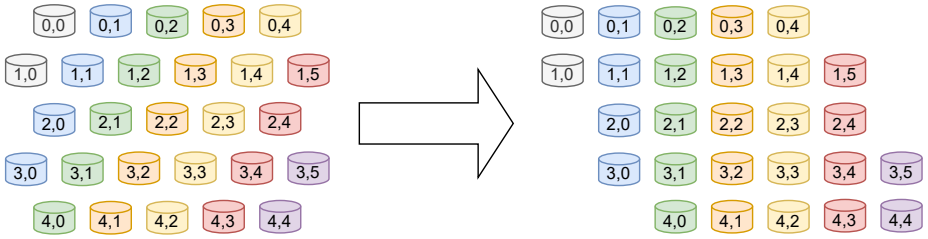


Figure 7.1.: **Transformation from the triangular grid to the rectangular grid.** The rectangular grid is used as the memory layout for the networks, as described in [86].

of the latter. By doing this, the proper behavior for missing PMTs is also ensured. Future works may expand on this approach since the differences of the recorded PMT traces may contain additional shower information. All traces are then transformed according to

$$\tilde{S}_i(t) = \frac{\lg(S_i(t)/\mathrm{VEM} + 1)}{\lg(100/\mathrm{VEM} + 1)}, \tag{7.2}$$

following the convention of Reference [83]. The $\hat{S}_i(t)$ is the unit-less network input for station i at time t and the $S_i(t)$ is the signal trace of station i calibrated to the VEM peak. Dividing by $\lg(100/\mathrm{VEM}+1)$ normalizes a signal of 100 VEM to unity. The addition of one in both logarithms satisfies the condition, that when $S_i(t) = 0$, it holds $\hat{S}_i(t) = 0$. Stations that did not trigger or are rejected are represented as traces filled with zeros.

To boost the network performance, it is beneficial to narrow down the possible values of the azimuth angle of showers, as the networks do not have to learn all representations of the full azimuth range. For this, all distance-conserving transformations that preserve the hexagonal lattice structure of stations around a single grid point are leveraged to narrow the azimuth angle of all showers down to a 30 degree range. This includes the combinations of six rotational symmetries with rotations in 60° increments around a grid point and a mirror symmetry for an axis through the grid points. This thesis uses the implementation developed by Steffen Hahn and an exact explanation can be found in Reference [86].

Convolutional network architectures expect an input that conforms to a rectangular mapping. For this reason, the mapping of the tanks is transformed from a triangular grid to a rectangular representation by shearing the station vectors to the left. For this, we follow again the example of Reference [86]. A simplified schematic of this transformation can be found in Figure 7.1. We call the azimuth standardization and the shearing together the footprint standardization.

All networks in this thesis use a slice of 5×5 stations around the hottest station with the aforementioned 120-time bins as input, accounting for 3 µs of data per station. The influence of the number of time bins and the optimal number of stations have already been investigated for hadronic showers in Reference [85]. These found hyperparameters are also applied in this thesis, as the tasks of hadron or photon energy reconstruction are similar.

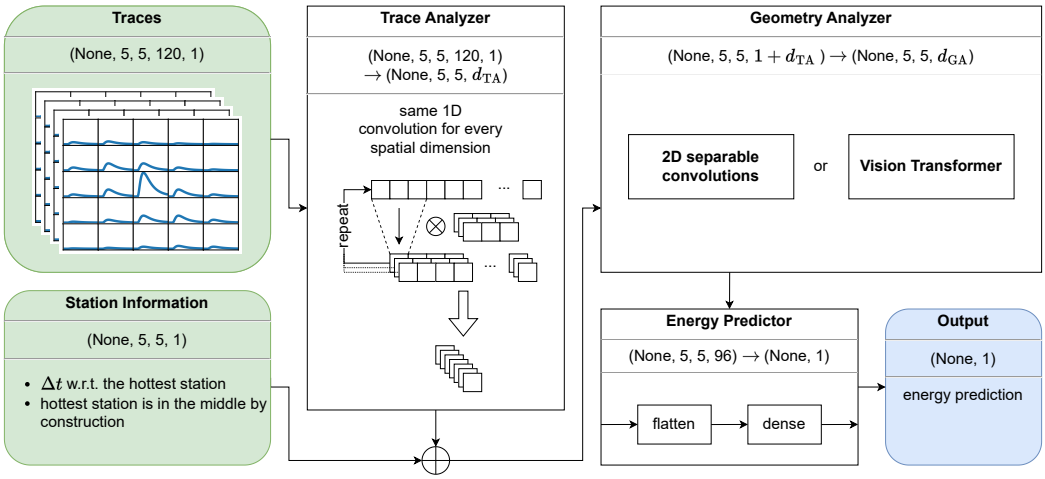


Figure 7.2.: **Schematic overview of the network architecture used in this thesis.** The inputs, consisting of the 5×5 grid of traces and the timing information, are shown in green on the left side, while the energy output is positioned at the bottom right in blue. The *None*-dimension symbolizes the batch size of the data, which is also indicated by the stacking of the traces.

Hyperparameters are network-specific parameters that are constant for one network but can be changed from network to network for a fixed architecture. Additionally, the footprint of photon-initiated showers is expected to be smaller than for hadron-initiated showers. Therefore, the 5×5 slice of stations should be sufficient.

Training the networks exclusively on the raw output of the detector is advantageous. This approach prevents the propagation of potential biases and uncertainties associated with classically reconstructed values into the network. Therefore, the only input of the networks, apart from the traces, is the trigger time relative to the hottest station in the middle. This allows the NNs to infer information about the shower axis and the curvature of the plane front, provided that more than three stations are involved. The network timing input $\Delta\tilde{t}_i$ of a station i is defined as

$$\Delta\tilde{t}_i = \Delta t_i / 4987.1775\,\text{ns}, \tag{7.3}$$

where the Δt_i is the time delta in ns. The division by the constant ensures that the standard deviation of all Δt_i in the data set is equal to one. In theory, the network does not require supplementary information, as classical reconstruction methods also do not utilize additional inputs.

7.2. Network Architecture

The network architecture used in this thesis, visualized in Figure 7.2, is split into three diffent parts: the trace analyzer, the geometry analyzer, and the predictor. The exact implementation of each subelement will be explained in the following sections. While the

Table 7.1.: **Characteristic values for the layers of the trace analyzer.**

layer	filters	kernel	stride	output shape
convolution 1	64	(1, 1, 7)	(1, 1, 3)	(5, 5, 38, 64)
convolution 2	32	(1, 1, 7)	(1, 1, 4)	(5, 5, 8, 32)
convolution 3	f_{TA}	(1, 1, 8)	(1, 1, 1)	(5, 5, 1, f_{TA})
reshaping				(5, 5, f_{TA})

structure of the trace analyzer and the predictor is always the same, two entirely different architectures are implemented and tested for the geometry analyzer.

7.2.1. Trace Analyzer

The trace analyzer is comprised of three stacked one-dimensional convolution layers. The objective of the trace analyzer is to extract f_{TA} features from the traces. These features are task-specific and learned in the training process. In each layer, the same convolution with the same kernel is applied to each trace. Consequently, the operation applied by the trace analyzer is entirely independent of the station position with only the shape and amplitude of the trace being relevant.

As for the exact implementation, the trace analyzer works by leveraging a three-dimensional convolution layer, but the first two kernel dimensions are fixed to one. The same is true for the first two dimensions of the stride length. The remaining dimension of the kernel size and stride length are carefully chosen to reduce the length of the convolved trace so that is one at the end of the three convolutions. For the first two convolution layers, the number number of filters is fixed to 64 and 32 respectively. In the last convolution, the number of filters f_{TA} is a hyperparameter, as it is equal to the number of resulting features for a single trace. The best set of these hyperparameters for every architecture will be optimized in Section 7.4.

The trace analyzer features only one hyperparameter, f_{TA}. In the last step of the trace analyzer, the data is reshaped in a manner that discards the dimension of the trace length, as it does not contain any information anymore. An overview of the characteristics of each layer is provided in Table 7.1. Every convolution layer uses the `"valid"` padding and the `"relu"` activation function.

7.2.2. Geometry Analyzer

Two distinct variants have been implemented for the geometry analyzer. One employs two-dimenesional separable convolutions (sepConv2Ds) layers, while the other is a modified version of the ViT architecture. The overarching idea of the geometry analyzer is to take the features of the trace analyzer and extract information about the whole event. In this

step, the networks should be able to extract information on the shower geometry. Prior to the application of any geometry analyzers, the trace analyzer output is concatenated with the station timing information, resulting in an input dimension of $(5, 5, f_{\mathrm{TA}} + 1)$ for the geometry analyzer.

7.2.2.1. 2-dimensional Separable Convolution

A sepConv2D-operation consists of two consecutive convolutions. First, a *depthwise* convolution is applied to the two spatial dimensions, producing an intermediate output. Then a second *pointwise* convolution is applied to the remaining channel dimension, mixing the features. The last step of a sepConv2D layer is the application of the nonlinear activation function [90]. The sepConv2D is an often-used alternative to a full three-dimensional convolution, as the latter would increase the number of trainable weights significantly. The whole geometry analyzer is defined by

$$\mathbf{s}_i = \mathrm{concat}(\mathrm{sepConv2D}_i(\mathbf{s}_{i-1}), \mathbf{s}_{i-1}), \qquad i = 1 \dots L_{\mathrm{GA}}, \tag{7.4}$$

$$\mathbf{s}_{L_{\mathrm{GA}}+1} = \mathrm{sepConv2D}_{L_{\mathrm{GA}}}(s_{L_{\mathrm{GA}}}), \tag{7.5}$$

where $\mathbf{s}_i$ is the state of the geometry analyzer after i iterations and the hyperparameter L_{GA} is the depth of iterations in the geometry analyzer. Each iteration of the sepConv2D layer, designated by the index in the subscript, is characterized by a unique set of weights and the number of filters, which are specific to that iteration. The concatenation operation in equation (7.4) stacks the output of the sepConv2D layer on top of its original input tensor along the first axis. These skip connections help with the vanishing of gradients during the backpropagation process of the training. In the absence of these connections, the length of the gradient chains for earlier layers becomes exceedingly large, which results in small gradients, due to the stacking of multiple multiplications with values smaller than one. Consequently, this results in a slow learning process in the early layers, and in extreme cases, the gradient is reduced to zero, thus preventing the initial layers from being updated at all.

The final step of the geometry analyzer is a last sepConv2D layer without the prior concatenation of the inputs. The number of filters used in the sepConv2D layers doubles for every occurrence and the starting value is given by the hyperparameter f_{GA}, with the resulting layer configurations provided in Table 7.2. Moreover, every layer uses the `"same"`-padding with (3, 3)-strides and is initialized with `"he_normal"`-distributed kernel weights.

In addition to the previously mentioned hyperparameters of the geometry analyzer, two further parameters exist. The first is the choice of the activation function $a_{\mathrm{GA}} \in \{$`relu`, `gelu`, `leaky_relu`$\}$ and the second is a Boolean variable $b_{\text{hex mask}}$, which determines whether or not a masking is applied to the depthwise kernels of all sepConv2D layers. The mask in question is defined as

$$m_{\mathrm{hex}} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \tag{7.6}$$

Table 7.2.: **Characteristic values for the layers of the geometry analyzer (sepConv2D).**

layer	filters	output shape
$\mathbf{s}_1$	f_{GA}	$(5, 5, f_{\mathrm{TA}} + 1 + f_{\mathrm{GA}})$
$\mathbf{s}_2$	$2f_{\mathrm{GA}}$	$(5, 5, f_{\mathrm{TA}} + 1 + 3f_{\mathrm{GA}})$
$\mathbf{s}_3$	$4f_{\mathrm{GA}}$	$(5, 5, f_{\mathrm{TA}} + 1 + 7f_{\mathrm{GA}}))$
$\vdots$	$\vdots$	$\vdots$
$\mathbf{s}_{L_{\mathrm{GA}}}$	$2^{L_{\mathrm{GA}}-1}f_{\mathrm{GA}}$	$(5, 5, f_{\mathrm{TA}} + 1 + (2^{L_{\mathrm{GA}}} - 1)f_{\mathrm{GA}})$
$\mathbf{s}_{L_{\mathrm{GA}}+1}$	$2^{L_{\mathrm{GA}}}f_{\mathrm{GA}}$	$(5, 5, (2^{L_{\mathrm{GA}}+1} - 1)f_{\mathrm{GA}})$

with the effect, that the depthwise convolution operates only on one crown of rank one. A crown of rank m is defined as all stations that are reachable with m jumps, starting at the central station. The hypothesis is that a kernel that operates only on a single crown could improve performance by aligning with the underlying symmetry of the array.

7.2.2.2. Vision Transformer

The here proposed alternative to the stacking of sepConv2D layers is the implementation of the ViT architecture, as presented in Section 5.3. In the original work [10], the flattened image patches are directly used as the input $\vec{x}_p^i$ in equation (5.7). However, in this thesis, the self-learned features of the trace analyzer are used instead. It should be noted that some preliminary testing was conducted in which the input of the ViT was composed of the raw traces plus the timing information, effectively eliminating the trace analyzer. This configuration performed significantly worse, and thus it was not further pursued. No further modifications are necessary for the ViT architecture to be applicable, as it has been designed to work on a three-dimensional input, comprising two spatial and one feature dimension.

We again define the depth of the subnetwork as the hyperparameter L_{GA}, which is equivalent to L in equation (5.7). Furthermore, the embedding dimension $d_{\mathrm{emb,\,GA}}$, from equation (5.7), and the size of the query dimension $d_{\mathrm{a,\,GA}}$ used in the MSA layer, as defined in (5.4), are also hyperparameters. As the ViT employs self-attention, it follows that $d_{\mathrm{a,\,GA}} = d_q = d_k = d_v$. The multilayer perceptron from equation (5.9) has one hidden layer of size h_{GA} and one output layer of size $d_{\mathrm{emb,\,GA}}$ to regain the original shape. Instead of using y from equation (5.10), as the final output, a last MLP is implemented before the predictor processes the output. This final MLP has again one hidden layer of size f_{GA} and one output layer of size $f_{\mathrm{GA}}/2$ [1]. Apart from h_{GA} and f_{GA}, there are three remaining hyperparameters. The first one is the number of heads $h_{\mathrm{heads,\,GA}}$ in the MSA layers. The

[1] When optimizing f_{GA} later on, the value is constrained to be a multiple of 4, therefore it cannot happen that f_{GA} equates to a non-integer value.

second one being a Boolean variable $b_{z_0,\,\mathrm{GA}}$ that decides if equation (5.10) is used unaltered ($b_{z_0,\,\mathrm{GA}} = \texttt{True}$), or if it is changed to

$$\mathbf{y} = \mathrm{LN}(\mathbf{z}_L), \tag{7.7}$$

with the difference being that the norm layer receives the whole vector $\mathbf{z}_L$ instead of only the first component. The last hyperparameter is the rate of dropout r_{drop} used during the training and is shared with the predictor. Dropout with the shared rate r_{drop} is applied to the MSA layer and after every fully connected layer in the MLPs.

The vector $\vec{x}_{\mathrm{class}}$ is realized as a set of fully trainable weights, which are initialized according to a Glorot-uniform distribution. The positional encoding is also fully trainable and one-dimensional, following the example of Reference [10].

7.2.3. Predictor

The predictor is the final step in the computational graph and has the same structure for both instances of the geometry analyzer. It can be conceptualized as a final interpreter that translates the geometry analyzer output to the predicted energy. It is defined by

$$y_{\mathrm{pred}} = \mathrm{dense}(\mathrm{dropout}(\mathrm{flatten}(s_{\mathrm{GA}}))), \tag{7.8}$$

where s_{GA} is the output state of the geometry analyzer. The flatten-operation stacks the whole data into one dimension. This is in theory not necessary for the ViT, but required for the sepConv2D output. The dense layer is a fully connected layer of size one, which directly predicts the energy of the event, $E_{\mathrm{pred}} = 1\,\mathrm{eV} \cdot 10^{\,y_{\mathrm{pred}} + 19.35}$. The dropout layer uses r_{drop} for the dropout rate.

7.3. Model Training

For the training of all models, the Huber loss [91], defined as

$$J_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \leq \delta, \\ \delta \cdot \left(|a| - \frac{1}{2}\delta\right) & \text{otherwise,} \end{cases} \tag{7.9}$$
$$\text{where } a = y_{\mathrm{pred}} - y_{\mathrm{true}},$$

is utilized. Some preliminary testing was performed to find an optimal value of δ, with the result that the TensorFlow default of $\delta = 1$ yields the best results. The Huber loss is used in cases of robust regression and is less sensitive to outliers compared to the mean squared error. Since the photon footprint is expected to fluctuate, i.e. for showers with extremely high X_{max}, this is a reasonable choice. Loss minimization is done with the *ADAM* [63] algorithm, using the default arguments of TensorFlow. Only the starting learning rate α_{LR} is optimized as a hyperparameter.

Although the training cycle aims for a maximum of 100 epochs, this number is rarely reached. This is due to the early stopping callback, which terminates the training if the validation loss has not decreased for seven epochs in a row and restores the weights of the epoch with the best validation loss. Callbacks are routines that are executed at predefined moments during training. The early stopping callback in question is called after each training epoch and compares the validation loss of the current epoch to previous epochs. The other important callback routine, utilized in the training, is the gradual reduction of α_{LR}. Every time the validation loss does not decrease in five consecutive epochs, the learning rate is reduced by a factor of 0.2. This reduction can prevent the minimization process from "jumping" around the minimum.

For each new configuration of a network, a whole ensemble of 10 networks is trained. Since the training depends on the randomly initialized starting weights, this is useful to see if the training is robust and converges independently of the initialization state. The train-validation split of the dataset is not statically assigned but changes for each network in the ensemble. This gives the ensemble indirect access to all events that are not in the test set.

7.4. Hyperparameter Tuning

This section only analyzes the effects of the hyperparameter tuning, the qualitative properties of the network predictions will be discussed in Section 7.5. All hyperparameters of both networks are optimized with a Bayesian optimization algorithm. To do this efficiently, the *Tuner* framework [92] of Keras with the *BayesianOptimization* tuner is used. Bayesian optimization is an iterative minimization technique that applies to black-box models with up to approximately 20 dimensions. It does not require any prior knowledge of the target function.

It was decided, to assess the hyperparameter tuning on events that fulfill the same selection criteria as the training data set without any additional restrictions on energy or shower zenith angle. The reasoning behind this is, that some highly optimized networks may excel at the energy reconstruction of events for very high energies or for events that develop their maximum deep underground. By cutting away these parameter regions, possible improvements there would not be visible. The *Tuner* framework iteratively trains a network architecture for different hyperparameters to find the best set of parameters according to a loss function. The loss function is given by

$$J_{\mathrm{HP}} = J_{\delta}^{(\mathrm{val})} + 0.001 \ \lg(N_{\mathrm{param}}), \tag{7.10}$$

where $J_{\delta}^{(\mathrm{val})}$ is the best-recorded validation loss during training and $\lg(N_{\mathrm{param}})$ is the base-10 logarithm of the number of trainable parameters in a network. The last term slightly punishes networks with a higher parameter count. During the tuning process, every network uses the same events for the validation, to avoid statistical fluctuations. A

Table 7.3.: **Results of the hyperparameter tuning process.** The left side of the table shows the hyperparameters in question. The minimum and maximum values, step sizes, and initial values are presented in the center of the table. If a step size value is marked with a star, it indicates that the step size is not defined by addition/subtraction, but rather by multiplication/division. If the values are divided by a slash, different values are used for sepConv2D and ViT. For non-numeric values, the set of possible choices is presented. The right side shows the optimal hyperparameter found by the tuning process. The final validation loss $J_\delta^{(\text{val})}$ and the final hyperparameter tuning loss J_{HP} are quoted at the bottom.

HP	min	max	step size/ choices	inital	sepConv2D	ViT
f_{TA}	6	48	4	10	14	38
f_{GA}	2/16	10/128	2/2*	6/16	6	64
L_{GA}	2	5	1	4	4	3
a_{GA}	-	-	{`relu`, `gelu`, `leaky_relu`}	`relu`	`leaky_relu`	-
$b_{\text{hex mask}}$	-	-	{`True`, `False`}	`True`	`False`	-
h_{GA}	8	128	4	16	-	76
$d_{\text{emb, GA}}$	8	32	4	16	-	24
$d_{\text{a, GA}}$	8	32	4	12	-	12
$h_{\text{heads, GA}}$	2	8	1	4	-	8
$b_{z_0,\text{ GA}}$	-	-	{`True`, `False`}	`True`	-	`True`
r_{drop}	0.0	0.5	0.125	0.0	0.25	0.0
α_{LR}	0.00001	0.1	10*	0.001	0.001	0.001
N_{batch}	16	256	2*	64	32	64
N_{param}				35 142 / 33 867	37 094	70 107
$J_\delta^{(\text{val})}$					0.0110	0.0103
J_{HP}					0.0154	0.0144

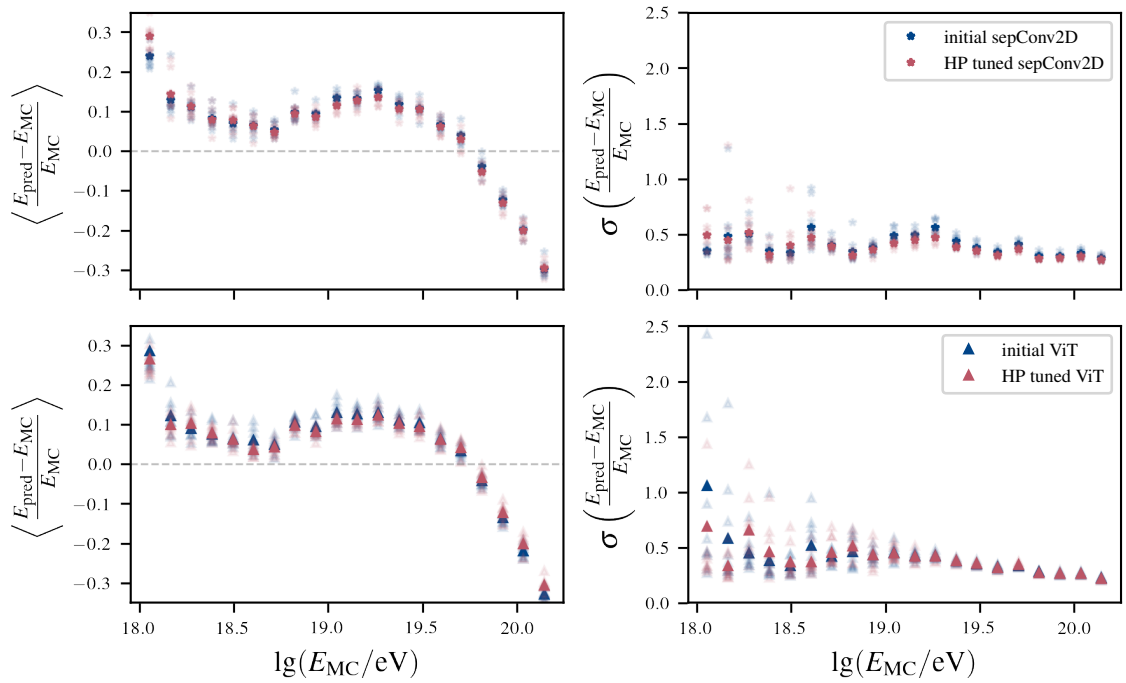


Figure 7.3.: **Relative bias and resolution of neural networks with and without tuned hyperparameters.** The plots show networks trained with the initial hyperparameters and the tuned values as defined in Table 7.3. The upper two plots show networks with sepConv2D layers and the lower plots make use of a ViT architecture. In each case, an ensemble of 10 networks is trained and the individual networks are plotted translucently while the bin-wise mean of the ensemble is presented as a solid marker.

summary of the final results of the hyperparameter optimization, including the parameter limits, defaults, and step sizes, is presented in Table 7.3.

The final hyperparameter N_{batch} at the end of Table 7.3, represents the batch size during the training process. This parameter was also included in the tuning process, as it is linked to the learning rate. The geometry analyzer with sepConv2D layers is optimized for 80 iterations, while the ViT architecture is tuned for 100 iterations, as there are more hyperparameters.

What is surprising, is the large number of trace analyzer features for the ViT geometry analyzer, that was found in the tuning process. It was originally hypothesized, that all the information in the traces could be condensed into a small subset of features. The results of the tuning now call this into question, although it could be possible, that this is an artifact of the ViT architecture. Another interesting observation is the fact, that the $b_{\text{hex mask}}$ converged to `False`, which means, that the two-dimenesional separable convolution performs better with a full convolution kernel, even if it breaks the symmetry that the kernel spans over exactly one crown.

To closely examine performance, Figure 7.3, illustrates the difference in relative bias and resolution for both network types using the initial and the optimized hyperparameters plotted against the base-10 logarithm of the Monte Carlo energy. The relative bias and the relative resolution of N events are defined as

$$\left\langle \frac{E_{\text{pred}} - E_{\text{MC}}}{E_{\text{MC}}} \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \frac{E_{\text{pred}}^{(i)} - E_{\text{MC}}^{(i)}}{E_{\text{MC}}^{(i)}}, \tag{7.11}$$

$$\sigma\left(\frac{E_{\text{pred}} - E_{\text{MC}}}{E_{\text{MC}}} \right) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{E_{\text{pred}}^{(i)} - E_{\text{MC}}^{(i)}}{E_{\text{MC}}^{(i)}} - \left\langle \frac{E_{\text{pred}} - E_{\text{MC}}}{E_{\text{MC}}} \right\rangle \right)^2}, \tag{7.12}$$

where the superscript (i) iterates through all events. The figure shows an ensemble of 10 networks for each architecture. The training set includes all 6T5 events without preshowers and no restrictions on the zenith angles. For the evaluation, additional requirements are imposed on the events. Namely, events must pass a T5 quality trigger [2] and only events with a Monte Carlo zenith angle below 60° are shown.

All networks are evaluated on the same independent test data set. However, the initial random weights of the networks and the subset of events selected for the training and validation set differ from network to network, explaining the deviation on a per-network basis. Moreover, as the training is conducted on GPUs, the training process is statistical, even networks with the same initial weights would not be identical. Additionally to the individual networks, which are plotted as translucent points, the means of the ensemble in each bin are plotted as solid markers. By visualizing the individual networks, the spread between them can be studied.

In the case of networks with sepConv2D layers as the geometry analyzer, the relative bias for energies close to 10^{18} eV is slightly worse, but for energies in the region $\approx 2 \times 10^{19}$ eV, the bias improves with the optimized hyperparameters. As for the relative resolution of sepConv2D networks, an improvement can be seen in nearly all energy bins. For the ViT architecture, the opposite is true, the relative resolution is nearly unchanged in the upper half for the energies. For the lower energies, the spread between the networks is large, and the deviation between architectures is therefore not significant. The improvement for the ViT architecture does come in the form of a small improvement of the relative bias over nearly the whole energy range.

In terms of raw validation loss, the ViT model exhibits superior performance to the sepConv2D model. However, this comes at the cost of having significantly more parameters with approximately 70×10^3 compared to the 37×10^3 in the case of the sepConv2D model. Moreover, the discrepancy is sufficiently minimal to justify the continued use of both architectures. In particular, as each iteration of the tuning process trains only one network, the difference could be further diminished by statistical fluctuations from network to network. The networks with the tuned hyperparameters exhibit better performance for both architectures. Therefore, the found hyperparameters will be used exclusively for the configuration and the training outside this section.

[2] Geometry and energy of an event can be reconstructed successfully by Offline.

7.5. Qualitative Properties of the Network Predictions

The raw energy estimate of the network, as seen in Figure 7.3 possesses some unwanted artifacts, which will be discussed and partly removed in this section. We start by addressing the fact, that the relative bias is slightly positive for a large range of Monte Carlo energies. This can be explained easily by the particular choice of representing the network performance as a relative bias and a relative resolution with the non-logarithmic energies. On one hand, this is a good representation, as one can see, that i.e. for a value of $E_{MC} \approx 19\,\text{eV}$ the energy is overestimated by $\approx 10\,\%$. However, this overestimation exists, since the network is trained with residuals that scale with $\lg(E_{\text{MC}}/\text{eV}) - \lg(E_{\text{pred}}/\text{eV})$, rather than $E_{\text{MC}} - E_{\text{pred}}$. When $\left\langle (E_{\text{pred}} - E_{\text{MC}})/E_{\text{MC}} \right\rangle$ is calculated, it is not equal to zero, even if $\left\langle \lg(E_{\text{MC}}/\text{eV}) - \lg(E_{\text{pred}}/\text{eV}) \right\rangle = 0$, since the first moment is not invariant under the transformation from a logarithmic scale to a linear scale. Therefore, it is beneficial to plot $\left\langle \ln\left(E_{\text{pred}}/E_{\text{MC}}\right) \right\rangle$, rather than $\left\langle (E_{\text{pred}} - E_{\text{MC}})/E_{\text{MC}} \right\rangle$, as it holds, that when the first moment of the distribution of residuals in log space is zero, $\left\langle y_{\text{pred}} - y_{\text{true}} \right\rangle = 0$, we also find that $\left\langle \ln\left(E_{\text{pred}}/E_{\text{MC}}\right) \right\rangle = 0$. Here, y_{true} is the training target for the events, as defined in equation (7.1). Furthermore, does it hold in first-order approximation that

$$\ln\left(\frac{E_{\text{pred}}}{E_{\text{MC}}}\right)\Bigg|_{(E_{\text{pred}} \approx E_{\text{MC}})} \approx \frac{E_{\text{pred}} - E_{\text{MC}}}{E_{\text{MC}}}. \tag{7.13}$$

Additional information can be found in the appendix in Section A.1, where Figure A.1 showcases the difference between the representations.

The energy bias and resolution for the hyperparameter-tuned networks for the logarithmic energy are shown in the upper two panels of Figure 7.4. The networks and events are identical to what is pictured in Figure 7.3, with the only change being the presentation. Again, 10 individually trained networks for each architecture are shown as translucent data points, while the means of the 10 individual biases and resolutions are shown as solid markers.

The bias (left plot) still shows an overestimation of the logarithmic energy of approximately 0.2 for $E_{\text{MC}} \approx 10^{18}\,\text{eV}$ for both networks. In the first-order approximation, this equals an 20 % overestimation. By dividing $\ln\left(E_{\text{pred}}/E_{\text{MC}}\right)$ by $\ln(10)$, it is possible to express the bias as the number of orders of magnitude by which the predictions are off. For $\ln\left(E_{\text{pred}}/E_{\text{MC}}\right) \approx 0.2$, the network predictions are approximately 0.086 orders of magnitude too high. At the high-energy end, $E_{\text{MC}} \approx 10^{20}\,\text{eV}$, both networks show a large underprediction of around 0.5, which is equal to underestimating the shower energy by 0.20 orders of magnitude. Overall, the bias can be described as a sideways S-shape, with an overestimation of energies of about 0 % to 5 % in the middle region.

The resolution of the logarithmic energy (right plot), is significantly lower at small MC energies and does not show the same extreme outlier networks for the ViT architecture as seen in the lower right plot of Figure 7.3. The reason for the outliers is, that overestimations of $\lg(E/\text{eV})$ lead to very high values of $(E_{\text{pred}} - E_{\text{MC}})/E_{\text{MC}}$. Overall, the resolution of both network architectures prediction gets worse with increasing energy, till reaching values

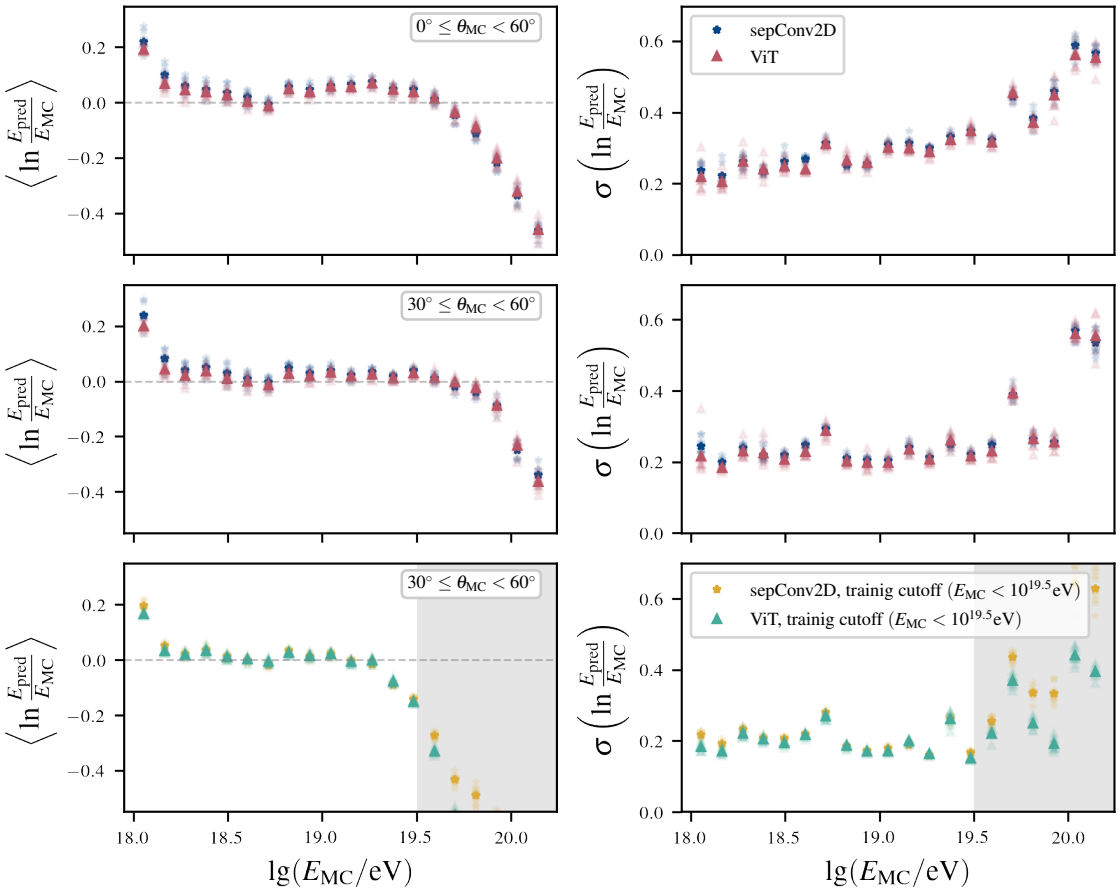


Figure 7.4.: **Bias (left) and resolution (right) of predicted logarithmic energy for the NN models.** The upper plots show the 10 sepConv2D and 10 ViT models which are trained on all events and evaluated on events with $\theta_{MC} < 60°$. In the second row, the same is shown for an additional lower limit of $\theta_{MC} > 30°$. The last plots show models with the same configuration, but the training was conducted only with events that have $E_{MC} < 10^{19.5}$eV (angle cuts identical to middle row). The 10 networks are evaluated on the whole energy range, which includes events beyond the energy training cutoff (gray band). Translucent points are individual networks, and the solid markers show the bin-wise mean of the ensemble biases and the ensemble resolutions.

of 0.6 at for $E_{\mathrm{MC}} \approx 10^{20}$ eV. The large variation in the energy prediction emanates most probably from shower-to-shower fluctuations with X_{max}

The observed spread between the individual networks is small for bias and resolution, for both the ViT architecture and for the use of sepConv2D in the geometry analyzer, with a maximum spread of $\sigma\left(\ln\left(E_{\mathrm{pred}}/E_{\mathrm{MC}}\right)\right) \approx 0.15$ in the case of the ViT at high energies. This is an indication that networks with different weight initializations converge to comparable minima during the training process. This is a desired behavior since then the performance of individual models is independent of the initial weights. Such convergence is indicative of robust training. Moreover, the fact that both architectures exhibit similar performance across the whole energy range suggests that they are sensitive to the same features and that the networks are approaching an upper performance limit for the given input features.

The performance of the networks is suboptimal, particularly in regard to the S-shaped bias curve, which is undesirable. There are three main reasons for the non-ideal performance, which can be partly avoided. The first improvement is achieved by the exclusion of very vertical showers during the evaluation, analogous to the DEC-method. The hypothesis is again, that events with an extremely deep shower maximum exhibit a footprint that is harder to interpret. During the training, the showers are still included, as ML models are of a nonlinear nature. Therefore the inclusion of the events should not negatively impact the overall performance and it helps to avoid edge effects, which will be discussed at the end of this section.

For this reason, the bias and resolution of the predicted logarithmic energy are assessed for the same networks, but only evaluated for events that fall into a limited range of zenith angels, $30^\circ \leq \theta_{\mathrm{MC}} < 60^\circ$. This is displayed in the second row of plots in Figure 7.4. For the low energies, no significant change in bias or resolution is observed. In the region of $E_{\mathrm{MC}} \approx 10^{19}$ eV the relative bias is reduced to only around 3 % or 4 %, while the resolution of the logarithmic energy is reduced from approximately 0.3 to 0.2. At the high end of Monte Carlo energies, the predicted energy is still underestimated, but the downturn of the bias starts only at $E_{\mathrm{MC}} \approx 10^{19.75}$eV instead of $E_{\mathrm{MC}} \approx 10^{19.55}$eV as before. The resolution at these energies is reduced from around 0.4 to 0.2, except for Monte Carlo energies larger than 10^{20} eV, where we find $\sigma\left(\ln\left(E_{\mathrm{pred}}/E_{\mathrm{MC}}\right)\right) \approx 0.6$ in both cases. From these slight improvements, it is clear, that the networks perform better on steeper events, where showers with very deep X_{max} are not present. For a further discussion, please refer to Section 7.6.

Another key factor in explaining the non-optimal performance in the low and high-energy regions is the scarcity of showers for the training. This is visible in the upper left histogram in Figure 6.1. At low energies, the detector efficiency is only around 40 % for photon-induced showers, rendering the majority of showers unusable for the training. At high energies, the relative number of preshowers increases, with fractions up to 70 % at the very end of the energy spectrum. With fewer showers in a parameter region, the networks are also expected to perform worse. A possible solution to this predicament in future studies could be to oversample the showers in the affected energy regions. This can be done either by completely simulating more `CORSIKA` showers or by using existing `CORSIKA` shower more often with random locations for the detector response simulation in `Offline`. Another

possibility is to simply insert showers in the affected energy regions more frequently during the training process, thereby increasing their perceived importance. Similarly, the weight of high and low-energy showers could be increased when computing the loss function. Although the last proposition is the simplest, it is discouraged, as it is expected that networks would learn better by having more unique examples of showers.

The final reason for the unfavorable performance at the low and high energies is intrinsic to NNs. It is possible for NNs to exhibit edge effects at the boundaries of the parameter phase space that was used during training. This can be understood as a local regression towards the mean effect in the affected parameter regions. The effect is additionally amplified by the scarcity of data at high and low energies and the large shower-to-shower fluctuations of photon-induced showers. To ascertain that the S-shape is at least partly a result of the edge effect, networks with identical configurations were trained only on events with MC energies below $10^{19.5}$eV. It is established that a network trained on the full energy range performs well and exhibits only a minimal bias at the cutoff energy of $10^{19.5}$eV. Therefore, if the networks with the energy cutoff during training exhibit a decline in performance at $E_{\mathrm{MC}} \approx 10^{19.5}$eV, it can only be due to edge effects. The remaining two plots of Figure 7.4 demonstrate that this is the case for both network architectures if only trained with events up to the cutoff energy. A clear negative bias trend is visible, starting at $\approx 10^{19.5}$eV, which is not visible in the networks trained on the energy range, confirming that the S-curve behavior originates partly from edge effects. If evaluated on events with energies larger than the training cutoff (gray band), the networks trained without the highest energies perform poorly, which is of course expected due to the non-extrapolative nature of NNs.

This non-extrapolative behavior can be seen nicely in Figure 7.5, where a two-dimensional histogram is presented, relating Monte Carlo energy to the predicted energy. The predictions of one ViT that was trained with the full energy range (left) and one that was trained on events with energies below $10^{19.5}$eV (right) are presented. The networks with the lowest loss on the test data set are selected from each ensemble. Both networks are shown for events with zenith angles between 30° and 60°, making Figure 7.5 comparable to Figure 6.12. The network without the cutoff shows the edge effects at the beginning and end of the training range in the form of over and under-predictions. The effect can also be observed shortly before the cutoff (gray band) in the right plot. Events with energies above the cutoff are just migrated to the highest energies in the training data set since the network is not able to extrapolate.

We anticipate further improvements in photon energy estimation by using the enhanced functionality of the AugerPrime upgrade in future studies. The small Photomultiplier Tube should enable us to observe the LDF closer to the shower core, which provides the networks with more useful information in otherwise saturated traces. Additionally, the scintillation detectors, with their sensitivity to the electromagnetic shower component, should be able to greatly improve the amount of information observed for every photon shower. Apart from utilizing new hardware, it is likely that a reduction of the generalization error could be achieved with the here-trained network ensembles. This is possible by employing model averaging or more sophisticated ensemble methods like those described in Chapter 7.11 of "Deep Learning" [61].

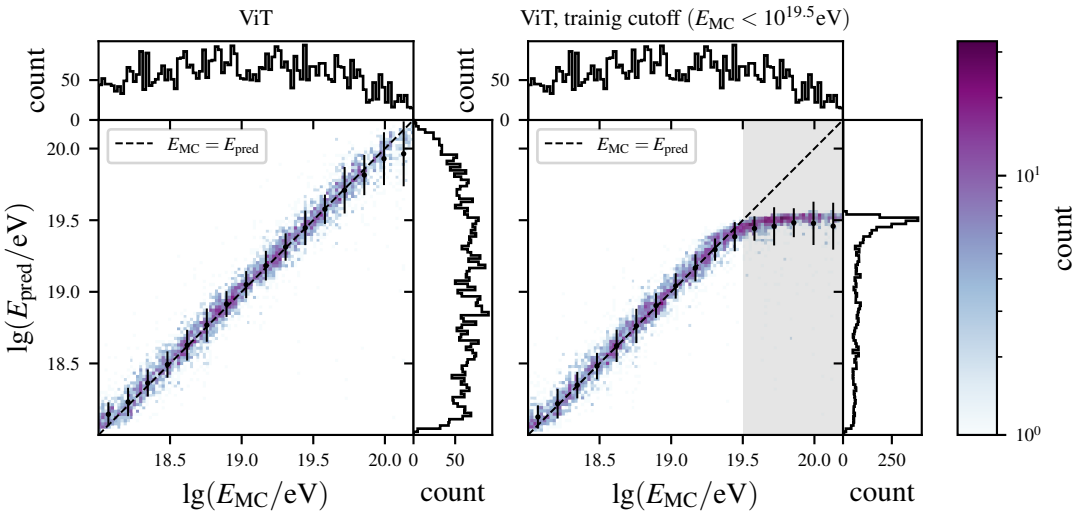


Figure 7.5.: **Demonstration of the edge effects for NNs.** In the two-dimensional histograms the relation of the predicted energy and the Monte Carlo Energy is displayed. Both show a single ViT network. While the network in the left plot is trained with all events, the one on the right has only seen events with energies lower than $10^{19.5}$eV during training. While the here shown evaluation is restricted to events with zenith angels between 30° and 60°, the training was done on the full zenith range.

7.6. Out of Time Stations

In comparison to hadron-initiated showers, those induced by photons penetrate deeper in the atmosphere before reaching the shower maximum, X_{max}. A plot that showcases this behavior in our simulated showers is depicted in Figure 7.6. The left side of the figure shows $\Delta X^{(MC)}$ which is defined as the height of $X_{max}^{(MC)}$ above the ground in units of g cm^{-2} while the right side shows the plain $X_{max}^{(MC)}$ distribution. Figure 7.6 also compares the behavior of photon-induced events to proton-induced events. The histograms use different y-axes for the different primaries, as the number of simulated events does not align. The value of $\Delta X^{(MC)}$ can be calculated as

$$\Delta X^{(MC)} = X_{max}^{(MC)} - 875\,\mathrm{g\,cm^{-2}}/\cos\theta_{MC}, \tag{7.14}$$

where 875 g cm^{-2} is the atmospheric overburden of the Pierre Auger Observatory, assuming a completely vertical shower. With this convention showers with $\Delta X^{(MC)} < 0$ develop their $X_{max}^{(MC)}$ above ground. The value of $X_{max}^{(MC)}$ for the individual showers is derived from the Monte Carlo data, where a Gaisser-Hillas distribution [93] is fitted to the longitudinal shower profile. As the shower is only simulated up to the detector plane, the precision of the estimated X_{max} diminishes as showers go deeper underground, as only the tail at the beginning of the Gaisser-Hillas distribution can be fitted. With this in mind, the distributions have to be treated not as the underlying physical distributions, but rather as the best feasible approximation of X_{max} given the `CORSIKA` showers. This is the reason, why

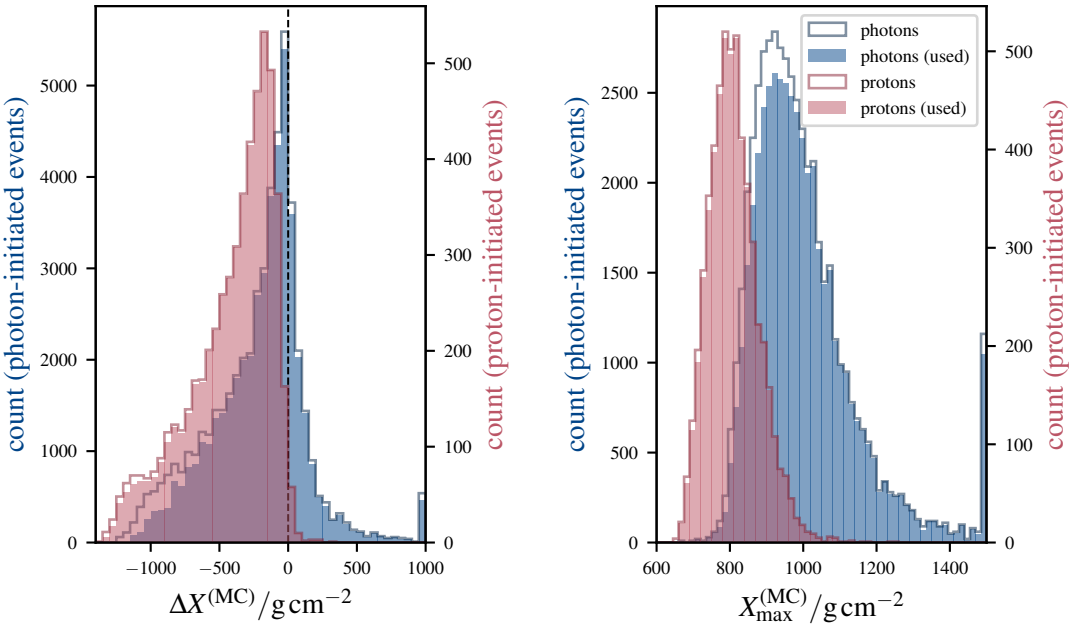


Figure 7.6.: **$\Delta X^{(\mathrm{MC})}$ and $X_{\mathrm{max}}^{(\mathrm{MC})}$ estimations of non-preshowering photon-induced events and proton-induced events, computed from the longitudinal MC shower profile.** The left histogram shows the distribution of $\Delta X^{(\mathrm{MC})}$, which is the height of $X_{\mathrm{max}}^{(\mathrm{MC})}$ in units of radiation lengths. The right histogram shows the same data distributed as a function of X_{max}. The rightmost bin in both plots contains overflow events from the right-hand side. The hollow histograms show all non-preshowering events, while the filled histograms show the events used for training and testing. The left axis in each plot is the number of showers initiated by photons, while the right axis accounts for those initiated by protons. The histograms have to be interpreted carefully, as the uncertainty of the computed $X_{\mathrm{max}}^{(\mathrm{MC})}$ is very high for deep showers. The histograms do therefore not depict the expected physical behavior, which is shown in Figure 2.3.

some events are estimated to have a very deep X_{max}, which in turn explains the pileup in Figure 7.6. The estimation with $X_{\mathrm{max}}^{(\mathrm{MC})}$ suggests, that the photon-induced showers develop their average particle maximum approximately $200\,\mathrm{g\,cm^{-2}}$ deeper in the atmosphere than proton-induced showers, with many cases where the particle maximum is reached underground. Figure A.5 in the appendix demonstrates how $\Delta X^{(\mathrm{MC})}$ is distributed for different MC energies and MC zenith angles in form of a two-dimensional scatter plot. As expected, showers with high values of $\Delta X^{(\mathrm{MC})}$ are concentrated at high energies and for small zenith angles.

Since the height of the shower origin $(\vec{x}_{\mathrm{o}})_3$ (see Chapter 6) is correlated with the height of the shower maximum, we expect the curvature radius for a photon-induced shower to be smaller than for hadron-initiated showers. Consequently, the algorithm for selecting stations that are part of the shower, which was previously optimized for hadron-initiated

showers, will not be optimal for this scenario. Stations that are triggered too late in comparison to a planar model of the shower front will be rejected as OOT. For a comprehensive reconstruction of photon-initiated events, this may be addressed in future studies by the implementation of an OOT rejection model optimized for photons. This is beyond the scope of this thesis.

In the case of neural networks, it is straightforward to incorporate stations that are rejected as OOT to evaluate potential improvements to the energy reconstruction. This approach was not ultimately employed in the final analysis, as there is currently no research on how the networks would be biased with the inclusion of accidental muons. Since accidental muons are disabled during the simulations and only a single shower is simulated at a time, it can be deduced that as OOT rejected stations in our data set must be part of the actual shower. Therefore, all stations that are rejected as OOT, are rejected falsely. Out of the 46 675 fully-reconstructable, non-preshowering events, 20 468 (43.9 %) contain at least one station that was rejected solely for being OOT. Moreover, given our simulated events, a dataset created without OOT stations yields 314 986 stations, while a data set that includes them amounts to 349 616 stations. Therefore, the training data set can be station-wise expanded by nearly 11 % if OOT stations are included.

To analyze possible performance improvements, a full ensemble of 10 ViT networks is trained with a data set that includes the OOT stations. The comparison is only performed for the ViT architecture, as it was found to be marginally better performing and no architectural dependency is expected for this analysis. Adding the OOT stations leads to a small improvement in the average network performance. This can be seen in Figure 7.7, where bias and resolution of the logarithmic energy prediction are compared for regular ViT networks and those with access to the OOT stations. If evaluated as a function of $\lg(E_{\mathrm{MC}}/\mathrm{eV})$ and $\sin^2\theta_{MC}$, the bias is nearly identical, as can be seen in the top and middle rows of Figure 7.7. The same is true for the resolution, except for a ≈ 0.05 reduction at $E_{\mathrm{MC}} \approx 10^{20.1}\mathrm{eV}$. In the last row of Figure 7.7 we analyze the bias and resolution if plotted as a function of $\Delta X^{(\mathrm{MC})}$. For low values of $\Delta X^{(\mathrm{MC})}$, the shower energy is on average underestimated by approximately 15 % to 20 % for both networks. It is likely that this phenomenon is the result of showers with fluctuations towards X_{max} occurring at higher altitudes, as such showers are expected to produce a lower average signal. The bias reduces with increasing X_{max}, until reaching $\Delta X^{(\mathrm{MC})} \approx 0\,\mathrm{g\,cm^{-2}}$. The same is true for the resolution of the logarithmic predicted energy, which reaches a minimum at $\Delta X^{(\mathrm{MC})} \approx -250\,\mathrm{g\,cm^{-2}}$. Therefore, the shower energy can be estimated most precisely with ML if the showers develop their maximum at $250\,\mathrm{g\,cm^{-2}}$ above the observatory. For showers with an $\Delta X^{(\mathrm{MC})} > 0$ the shower maximum is estimated to be below ground and the networks are underestimating the shower energy by up to 0.25 orders of magnitude. No absolute explanation for this behavior can be given without a more extensive study of how the very deep showers behave. A possible explanation lies in the expansion of the shower transversal to the shower axis being small at ground level for showers with a very deep maximum. This way, the distance between the shower core and the nearest station of the array could introduce a large variability in the energy reconstruction. If a shower core is very far away from all stations, the recorded signal from the stations is probably low, as they only sample the outer edge of the young shower, leading to the large

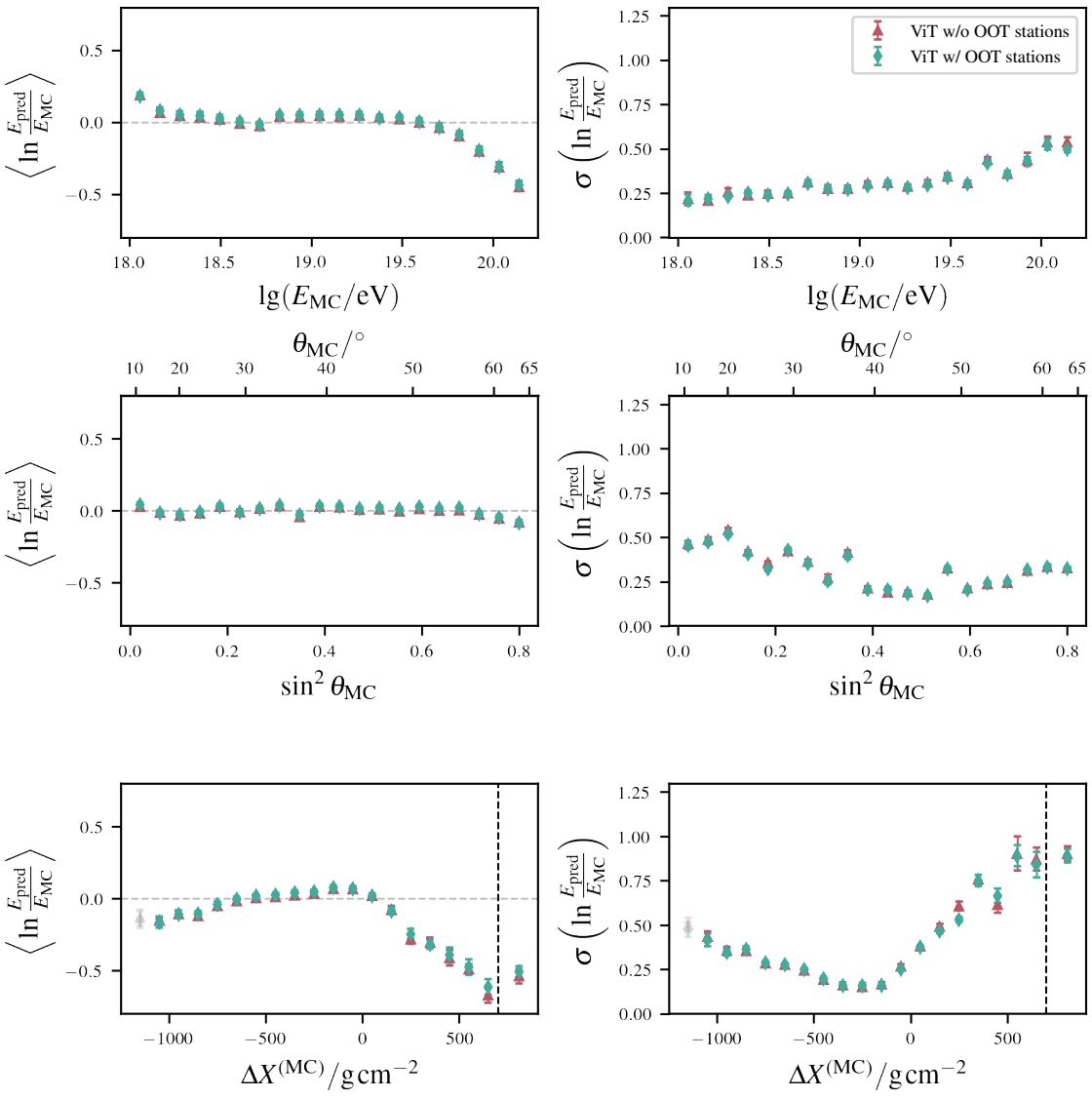


Figure 7.7.: **Bias (left) and resolution (right) of the predicted logarithmic energy as a function of $\lg(E_{\text{MC}}/\text{eV})$, $\sin^2 \theta_{\text{MC}}$, and $\Delta X^{(\text{MC})}$.** The ViT networks with and without access to OOT stations are shown in green and red respectively. All reconstructable, non-preshowering events are included. The data points with their error bars represent the bin-wise mean and standard deviation of the biases and resolutions from the individual networks in the ensemble. The data points situated at the far right in the final row of plots, separated by the dashed lines, represent all overflow events. Bins with less than 20 contributing events are plotted translucently.

observed underestimation of the energy. But for some showers, the core coincides with a station and the network can use this information for a more accurate prediction of the energy. This theory is in line with the worsening resolution for showers with a deeper maximum.

The ViT, that uses the OOT stations shows a slightly better bias for $\Delta X^{(\mathrm{MC})} > 200\,\mathrm{g\,cm^{-2}}$. This small improvement varies between 0.01 and 0.1 per bin. It is expected that the largest improvements are given for very deep showers, as the timing delta between a plane model and a spherical model is the largest for a deep shower maximum. However, the overall improvement is minimal. Therefore, we would not recommend including OOT stations for a network reconstruction, given the risk of introducing noise from stations that are triggered by accidental muons.

8. Comparison of Energy Reconstruction Methods

This thesis presents two different methods for estimating the energy of a photon-induced air shower. The first part of this chapter compares these methods. The second part includes a comparison with the Lookup Table (LUT) approach developed in Reference [1].

8.1. Neural Networks Compared to the Direct Energy Calibration

In Figure 8.1 we compare the bias and resolution of the logarithmic energy predictions of the NNs to the DEC. All subfigures on the left show the bias, while the right side subfigures show the resolution. The depiction in a logarithmic scale is a valid representation for the DEC, as it too is fitted in the logarithmic scale. For the comparison, the best-performing networks from each architecture are selected according to their loss value on the validation data set. By selecting the optimal networks this way, instead of choosing the one that performs best on the test set, we ensure that we are resistant against selecting a network, that performs better on the test set just by chance.

The upper two subfigures show bias and resolution as a function of $\lg(E_{\mathrm{MC}}/\mathrm{eV})$ and only for fully reconstructed events with reconstructed zenith angles between 30° and 60°. The shower angles are reconstructed with the in Section 6.3 derived LDF parameterization for photon-initiated events. It is interesting to see how the DEC behaves for events outside the energy range used for the calibration. The region outside the calibration range is indicated by the gray bands. The bias of the DEC for $E_{\mathrm{MC}} < 10^{19.75}\mathrm{eV}$ is small, with a maximal underestimation of the logarithmic energy by ≈ 0.09, at $E_{\mathrm{MC}} \approx 10^{18.5}\mathrm{eV}$. This is equivalent to an average underestimation by only 0.04 orders of magnitude. It extrapolates well to showers with energies below the calibration interval. Since those small differences are difficult to make out in Figure 8.1, Figure A.4, located in Section A.2 of the appendix, provides the same panels, with a narrower y-range. For energies above the calibration threshold, the DEC underestimates the energy, probably caused by an increasing population of showers with a very deep X_{max}. This is also reflected in the resolution of the logarithmic energy, which worsens up to values of 1.0 for $E_{\mathrm{MC}} > 10^{20}$ eV. If the resolution is calculated for the relative error of the non-logarithmic energy, a spread of 60 % was found for $E_{\mathrm{MC}} > 10^{20}$ eV. That is equal to an average underestimation of the energy by nearly half an order of magnitude. Both neural networks exhibit nearly identical

performance with the applied cuts and show significantly better resolution in all energy intervals, outside and inside the DEC calibration interval. The NNs show yet again the familiar edge effects at the ends of the training energy range.

The second row of subfigures show the bias and resolution as a function of the MC zenith angles. Only events with E_{MC} in the DEC calibration region are selected and intervals with less than 20 contributing events are presented translucently. Moreover, the reconstructed zenith range is limited to values between 17.06° and 65.10°. The evaluation of the function with angles outside this range yields $f_{\mathrm{att}} \leq 0$ which produces non-physical results. Future works may expand on this by demanding that $f_{\mathrm{att}}(\theta) > 0$ for all relevant values of θ during the fitting process. Upon approaching these limits, the DEC prediction significantly overpredicts the energy in a diverging behavior. This is expected as f_{att} closes in on the non-physical region, since $\lim_{f_{\mathrm{att}}(\theta)\to 0}(S_{45}(\theta)) = \infty$. The other reason for the inaccurate prediction is the reconstruction error on the zenith angle. As seen in Figure 6.10, the angular resolution of photon showers does sometimes show a large deviation from the actual shower directions, even with the here derived LDF. Inaccurate zenith angles result in inaccurate predictions of the DEC. The bias of the networks is independent of both effects and remains close to zero for the whole range of $\sin^2\theta_{\mathrm{MC}}$.

As for the resolution, the same divergent behavior for the DEC is visible here. Both networks display a comparable U-shape pattern, albeit to a lesser extent. This suggests, that the energy reconstruction resolution has a general sweet spot at around 40° to 50°. A possible explanation is, that there is just the right amount of atmosphere in front of the detector for the showers to have a consistent relation in shower size and energy. Note that the markers absent in this plot are positioned above the displayed y-range.

The final two plots of Figure 8.1 have the same cuts as before. However, the plots show bias and resolution as a function of the reconstructed zenith angle instead of the MC angle. While angles outside the calibration region are again displayed by a gray band, angles in the non-physical region of f_{att} are marked as a red band. The binning in $\sin^2\theta_{\mathrm{rec}}$ is chosen to match the binning in $\sin^2\theta_{\mathrm{rec}}$. The DEC predictions show similar divergent behavior as for the MC angles. However, the mean bias in the calibration region is slightly reduced, probably due to the absence of effects from inaccurately reconstructed zenith angles. Within the calibration region, the bias of the DEC is close to zero. The NNs exhibit a near-identical behavior when plotted as a function of the reconstructed zenith angles as for the MC zenith angles, which is expected since their predictions are independent of the reconstructed zenith angles.

So far, the resolution and bias were shown while marginalizing along either an energy or a zenith-dependent variable. To see possible correlations, the bias (Figure 8.2(a)) and the resolution (Figure 8.2(b)) of the logarithmic energy predictions are plotted as a function of $\lg(E_{\mathrm{MC}}/\mathrm{eV})$ and $\sin^2\theta_{\mathrm{rec}}$ simultaneously. The binning in both variables is linear and equidistant. The best ViT network (left) is compared to the DEC (right). Bias and resolution are shown on the color axis, where red is an overestimation, blue is an underestimation, and white is a bias-free energy estimation. Similarly, in Figure 8.2(b), white is a perfect resolution, while magenta to red signifies a large spread in the energy predictions. Bins with low statistics (less than 10 events) are presented in gray. Some additional features

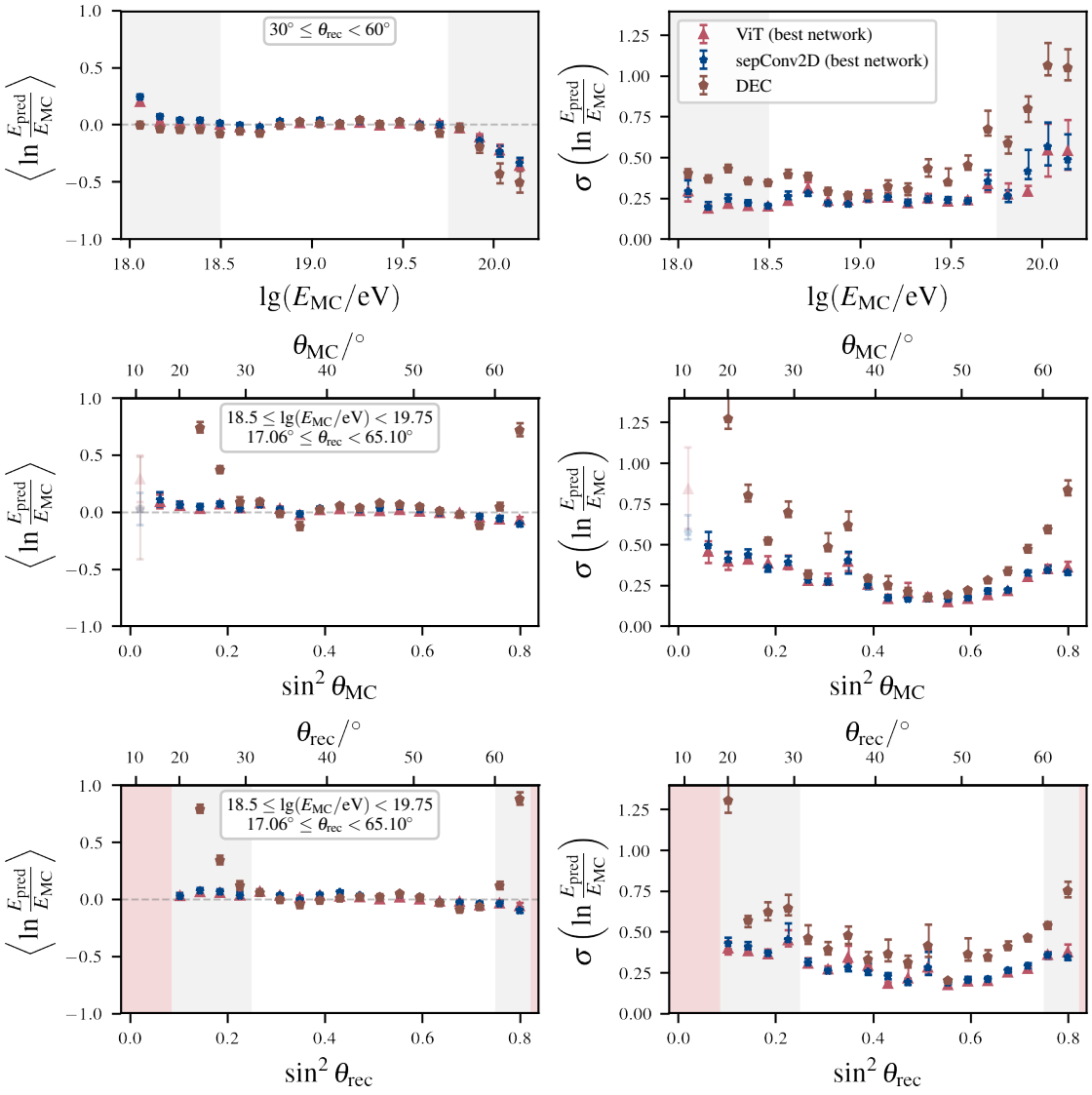
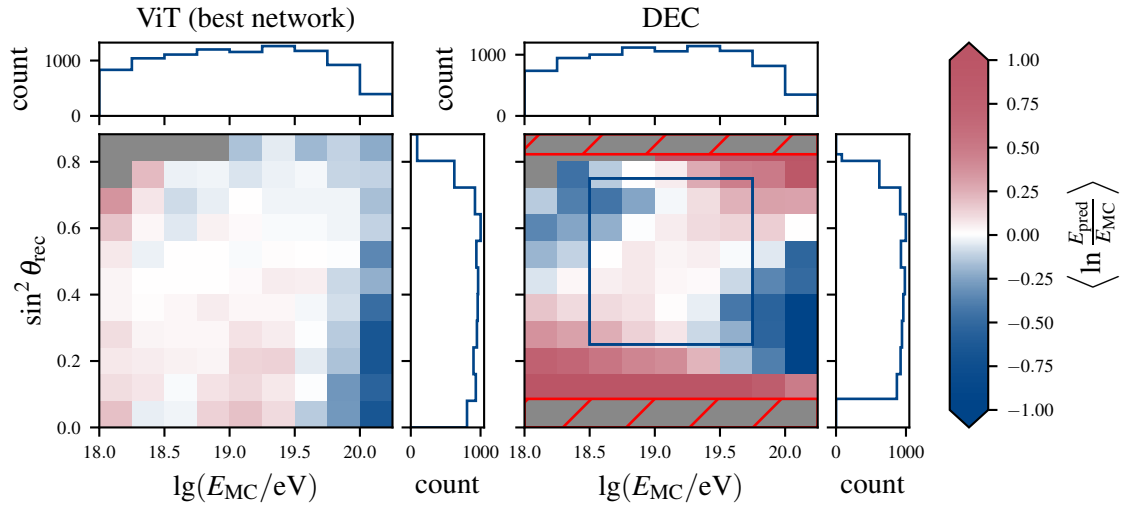
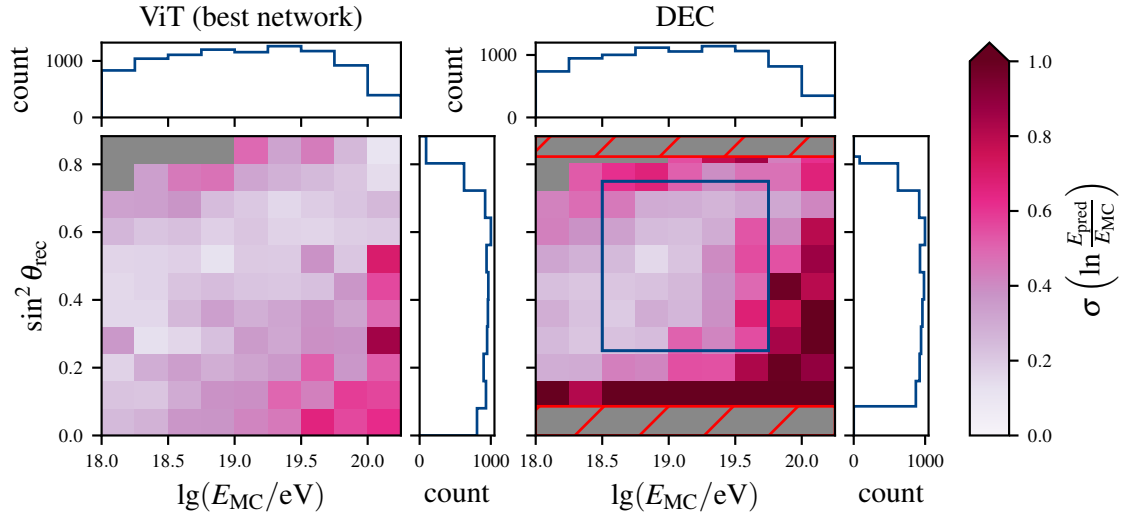


Figure 8.1.: **Bias (left) and resolution (right) of the logarithmic energy as a function of $\lg(E_{MC}/eV)$, $\sin^2 \theta_{MC}$, and $\sin^2 \theta_{rec}$.** The gray bars in the first and last row of plots show the edges of the calibration interval of the DEC method and the red bands in the last plot shows the region where the fitted f_{att} is non-physical. Uncertainties of the bias are calculated as the standard error on the mean. The uncertainties of the resolution represent a two-sided 68.27 % confidence interval, determined via bootstrapping.



(a) bias



(b) resolution

Figure 8.2.: **Bias and resolution of the logarithmic energy for the best ViT network and the DEC depending on $\lg(E_{\text{MC}}/\text{eV})$ and $\sin^2 \theta_{\text{MC}}$.** The histograms show the marginalized amount of events along the respective axis. The blue rectangle in the plots that show the DEC results visualizes the calibration region of the latter. The red hatched region defines the values of θ_{rec} where f_{att} is non-physical given the here derived parameterization.

are visible for the DEC on the right. The blue box indicates the calibration region and the hatched red regions represent zenith angles where the attenuation function is non-physical. In the case of the ViT, all reconstructable non-preshowering events with $\theta_{\mathrm{rec}} < 70^\circ$ are shown, with no restrictions on the energy. For the DEC, events outside the physical region of f_{att} are of course excluded.

For the ViT, the familiar edge effects are evident, with a systematic overestimation of the predicted energy at low MC energies and an underestimation at high MC energies. Edge effects are additionally visible for highly vertical showers and highly inclined showers as an over and underestimation respectively. Events with a very deep show maximum seem to affect the networks, as the bias displays a gradient towards larger underestimations at small zenith angles and MC energies above 10^{20}eV. For showers with $0 \leq \sin^2 \theta_{\mathrm{rec}} < 0.3$ and $E_{\mathrm{MC}} \approx 10^{19}$ eV the energy is slightly overestimated. The reason for this phenomenon is currently not known.

The bias of the DEC paints a more complex picture. It can be described as exhibiting a quadrupole-esque structure, with an overestimation observed in inclined showers at low MC energies and for vertical showers at high MC energies. For inclined showers at high energies and vertical showers at low energies, the DEC method underestimates the shower energies. This phenomenon is not confined to events outside the calibration region; rather, it extends within the calibration interval, albeit to a lesser degree. For the one-dimensional bias and resolution plots, this effect is averaged out.

This behavior is most likely attributable to an increasing population of showers with a very deep X_{max} for high MC energies with zenith angles around 30°. Since the shower size is expected to get smaller for very deep X_{max}, the DEC tries to fit this behavior by adjusting f_{att}. This is consistent with the observed biases and the estimated heights of the shower maximum above (or below) ground (see Figure A.5 in the appendix). This issue may be addressed in future studies by implementing more rigorous cuts or by modifying the DEC, to reflect the behavior of deep showers. As showers with deep maxima are found for vertical showers at high energies, it could prove advantageous to redefine f_{att} with an additional term, that simultaneously depends on θ and S_{1000}. This modification is of course only possible for MC data and is not compatible with a constant intensity assumption, as used in the CIC method. A function like

$$g_{\mathrm{att}}(x, S_{1000}) = f_{\mathrm{att}}(x) + d\, x \lg\left(\frac{S_{1000}}{\mathrm{VEM}}\right), \tag{8.1}$$

could be a useful redefinition. The function g_{att} would then replace f_{att} in the DEC. The parameter x and function $f_{\mathrm{att}}(x)$ would remain unchanged from equation (6.7), while the parameter d would be a new parameter that has to be fitted as well.

The divergent behavior towards small values of f_{att} can be observed, as the reconstructed zenith angles approach the non-physical regions. The effect of very deep showers is reflected by a worse resolution of the DEC for steep showers at high energies.

The energy reconstruction with ML methods yields better results for the bias and resolution and is furthermore applicable to more showers. The reason for this is the great adaptability

of the models, which is needed for the complex footprint of photon-induced showers. Nonetheless, it is unclear if the ML method can be applied to real photon-induced showers as it is not possible to validate that the MC simulations match reality until significant photon candidates are observed. We know that the simulations may not be perfect, as demonstrated by the muon puzzle [80, 81] for hadronic showers. Although the muon puzzle should not affect photon-induced showers, it is impossible to rule out other possible mismatches that could cause unforeseen effects when working with NNs.

The DEC method with the here suggested improvements could prove as a better-understood alternative to the ML models. In contrast to the black box model of the NNs, the steps of the classic reconstruction can be understood and interpreted on a physical level. Although the classic energy reconstruction is only dependent on the integral and not the precise shape of the simulated traces, it too relies on the overall accuracy of the MC models.

8.2. Comparison to other Works

In a previous publication on a photon search above 10^{19} eV [1], a two-dimensional Lookup Table (LUT) is used to estimate the photon energy. The two-dimensional LUT is constructed by taking the weighted average of the base-10 logarithm of the non-preshowering photon MC energies with bins in $\lg(S_{1000}/\mathrm{VEM})$ and $\cos^2\theta_{\mathrm{rec}}^{\mathrm{had}}$. The weights are adjusted so that the simulated spectrum reproduces a reference spectrum that scales according to $\propto E^{-2}$. The variables E_{had} and $\theta_{\mathrm{rec}}^{\mathrm{had}}$ are the reconstructed energy and the reconstructed zenith angle when the photon-induced event is reconstructed as a hadronic event. By deploying this nonlinear ansatz, the LUT eliminates the problem of functionally parameterizing showers with a deep X_{max}. The LUT as driven in Reference [1] can be seen on the left side of Figure 8.3. The right side of the figure shows the ratio of the LUT energy estimator to E_{had}.

Now, we will compare the LUT estimator to our work. The LUT was fitted with events that satisfy the following criteria:

- no preshower,
- 6T5 trigger,
- $\theta_{\mathrm{rec}}^{\mathrm{had}}$ in the range 30° to 60°,
- $E_{\mathrm{had}} > 10^{18.5}$eV.

To this end, the LUT-method will only be compared to our methods with events that fulfill these cuts. This is displayed in Figure 8.4.

As before, we show bias and resolution of the predicted logarithmic energy on the left and on the right respectively. The first row of subfigures shows the predictions of the DEC, the best ViT network, and the LUT as a function of $\lg(E_{\mathrm{MC}}/\mathrm{eV})$. Only the ViT is shown, since the sepConv2D networks do not differ from the ViT significantly. Bins with less than 20 contributing events are again plotted translucently. All three methods overestimate

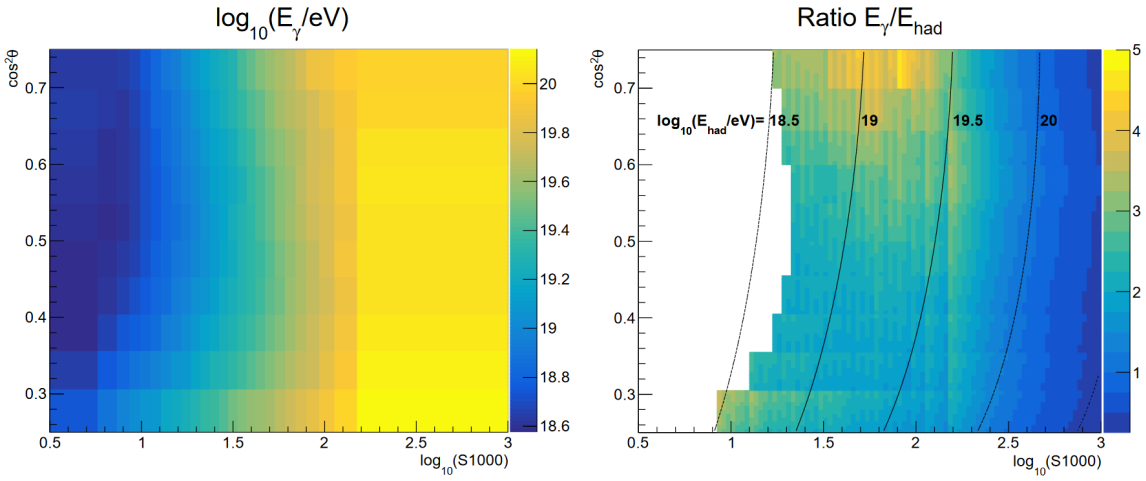


Figure 8.3.: **Lookup table to estimate the photon-induced shower energy and the ratio of the estimated photon energy to the hadronic energy predictor.** The left plot shows the LUT used by Reference [1] to assign the photon energy E_γ as a function of the zenith angle and the shower size. The latter is calculated with the normal Auger LDF, optimized for hadrons. The right plot depicts the fraction of E_γ to the energy E_{had}. E_{had} is the hadronic energy estimator applied to the photon-induced events. Figure from Reference [1].

the energy in the range $E_{\mathrm{MC}} \lessapprox 10^{19}$eV. For the photon search in Reference [1], this is not a problem as the search is only aimed at energies above 10^{19} eV. The underestimation is caused by the final selection criterion of the LUT, which cuts into the distribution of E_{MC} by selecting upwards fluctuations. This is confirmed when looking at the second row of subfigures, where bias and resolution are shown as a function of $\lg(E_{\mathrm{had}}/\mathrm{eV})$. The DEC and the ViT-NN show no significant bias towards low values of $\lg(E_{\mathrm{had}}/\mathrm{eV})$ as both are known to be unbiased for small MC energies. For $E_{\mathrm{had}} > 10^{19.25}$eV, the classic reconstruction starts to overpredict the energy, and the bias grows approximately linearly with $\lg(E_{\mathrm{had}}/\mathrm{eV})$. It is hypothesized, that this one-sided bias could be an effect of the different LDFs, which scale differently at high values of $\lg(E_{\mathrm{had}}/\mathrm{eV})$.

At high MC energies, the LUT method underestimates the energies more than the here-developed approaches. The resolution is slightly better than that of the DEC method in this MC energy region. The LUT method slightly underestimates the energy in nearly all bins of $\sin^2\theta_{\mathrm{rec}}$. This is most likely an artifact of the discrete nature of the LUT, as the bias depends on the spectrum. The reweighting to the reference spectrum that is $\propto E^{-2}$ yields unbiased results for the LUT only, if the given data follow the same reference spectrum. Since our spectrum is $\propto E^{-1}$, this leads to a systematic underestimation.

The resolution of the LUT ansatz as a function of $\sin^2\theta_{\mathrm{rec}}^{\mathrm{had}}$ is mostly comparable to the DEC. Only for $\theta_{\mathrm{rec}} > 55°$, the resolution of the logarithmic energy prediction of the DEC increases to approximately 0.1 larger values than the LUT. Similar behavior is also observed in the bias, where the shower energy is increasingly overestimated, starting at

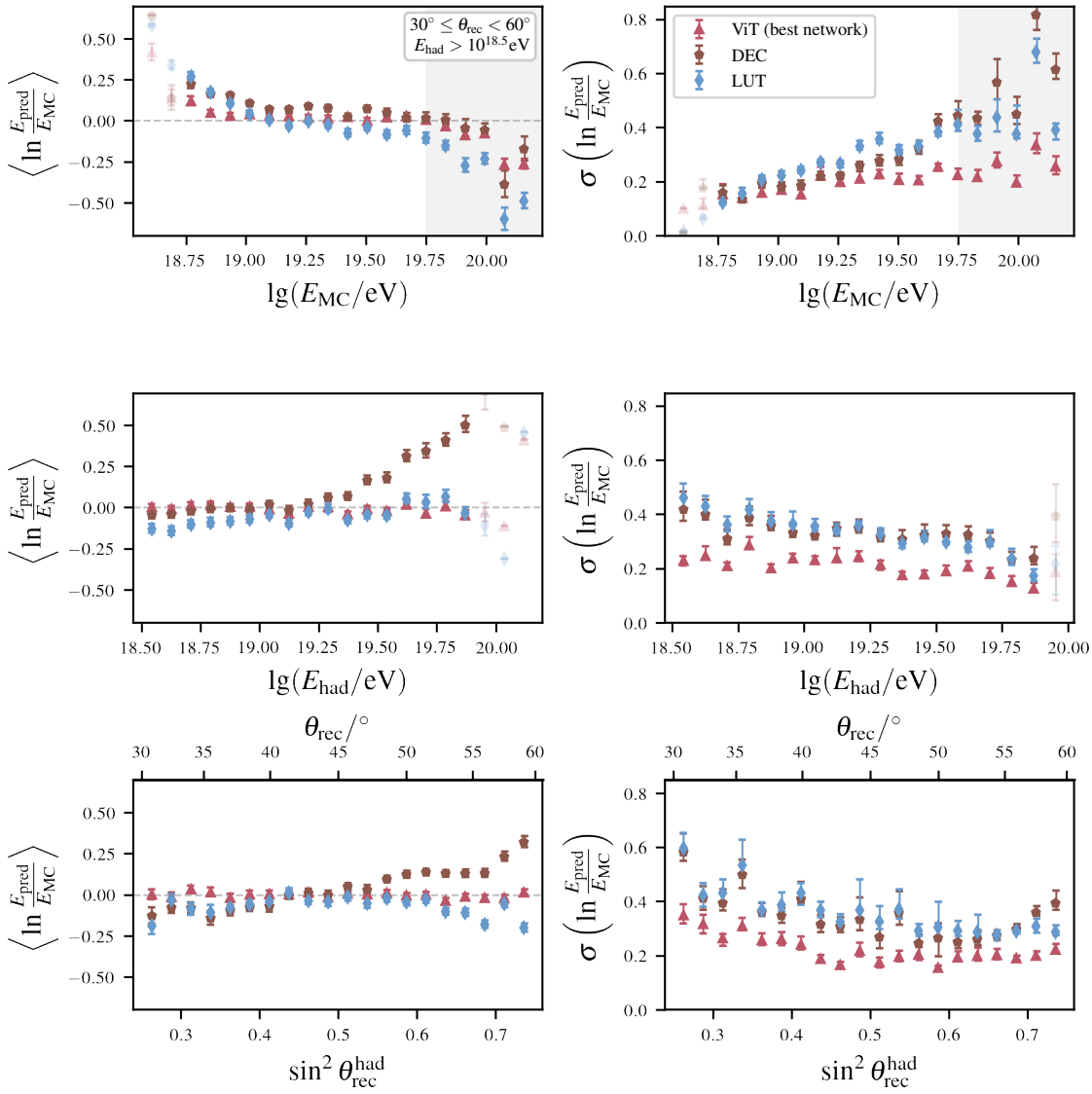


Figure 8.4.: **Comparison of the lookup table with our methods.** Plots on the left side show the bias and plots on the right show the resolution of the logarithmic energy prediction. Three energy estimators are displayed: the best ViT network, the DEC, and the LUT implemented in [1]. All plots use events with reconstructed zenith angles (using the hadron-optimized LDF parameterization) between 30° and 60° and with $E_{had} > 10^{18.5}$eV. Intervals with less than 20 events are displayed translucently. Note the difference in the displayed y-axis range compared to Figure 8.1. Uncertainties of the bias are calculated as the standard error on the mean. The uncertainties of the resolution represent a two-sided 68.27 % confidence interval, determined via bootstrapping.

approximately 45°. As this behavior is not visible in Figure 8.1, where the cut on E_{had} is not present, this is probably again caused by asymmetric selection of events.

Overall, the ViT performs the best, given the event selection prescribed by the LUT. While the DEC and the LUT method behave similarly for the resolution, the LUT shows a better bias, as expected for this ansatz. One big downside of the LUT approach is the discretization of the predicted energies, which is not present in a parametric ansatz. The given LUT with its cuts is only applicable to approximately 35 % of all reconstructable non-preshowering events in our data. Given the goals in Reference [1], this is no limitation. If the LUT ansatz is adopted as a general energy estimator, the range of the LUT can be easily extended. In general, the LUT is an easy-to-implement alternative to the here proposed methods and adapts well to the complicated footprint of photon-initiated showers. However, the discretization of the energy predictions produces a bias, that depends on the spectrum shape of the input.

9. Application to Preshowering Events

This chapter presents an analysis of the performance of the energy estimators developed in this study when applied to preshowering events. Although the DEC and the NNs are not fitted with preshowering events, the probability of an event to be a preshower increases rapidly at approximately $10^{19.75}$eV and the applicability of the models at these energy regions is not guaranteed if preshowers cannot be distinguished from non-preshowering events. In addition to the already established methods, a new ensemble of 10 ViT networks is trained on preshowering and non-preshowering events. The best network is then again selected according to the validation loss and will be compared to the other estimators. The results are presented in Figure 9.1.

Contrary to our initial beliefs, neither the bias nor the resolution of the predicted logarithmic energy worsens with the inclusion of preshowers. This is the case for the ViT, which was trained without any preshowers, and the DEC. In the case of the ML predictor this is even more surprising since it was expected that an energy estimation of the never-seen preshower footprints would not work as expected. As the predictions are robust, even for preshowering events, the applicability range of the normal ViT and the DEC is therefore not limited to energies below $10^{19.75}$eV. With the knowledge, that the effect of preshowers is not as severe as hypothesized it would probably be best practice to repeat the asymmetry parameterization, the LDF fit, and the DEC with the inclusion of preshowers. Unfortunately, this is not feasible within the time constraints of this thesis.

The ViT network that was trained with preshowers shows an even less biased prediction for $E_{\mathrm{MC}} > 10^{19.75}$eV. The reason for this is twofold. Firstly, the network has been subjected to preshowering events at training time, which naturally increases the probability of a successful energy reconstruction for such events. Secondly, with the addition of preshowers, the number of events in the high-energy region increases, which should in turn decrease the severity of edge effects. The sole observed negative consequences of including preshowers during training are a slight increase in the resolution of the logarithmic energy by up to approximately 0.025 in some energy bins and minor shifts towards an overprediction for $E_{\mathrm{MC}} < 10^{19.5}$eV. Consequently, the incorporation of preshowers into the training data set is a sensible approach that should be considered in future works.

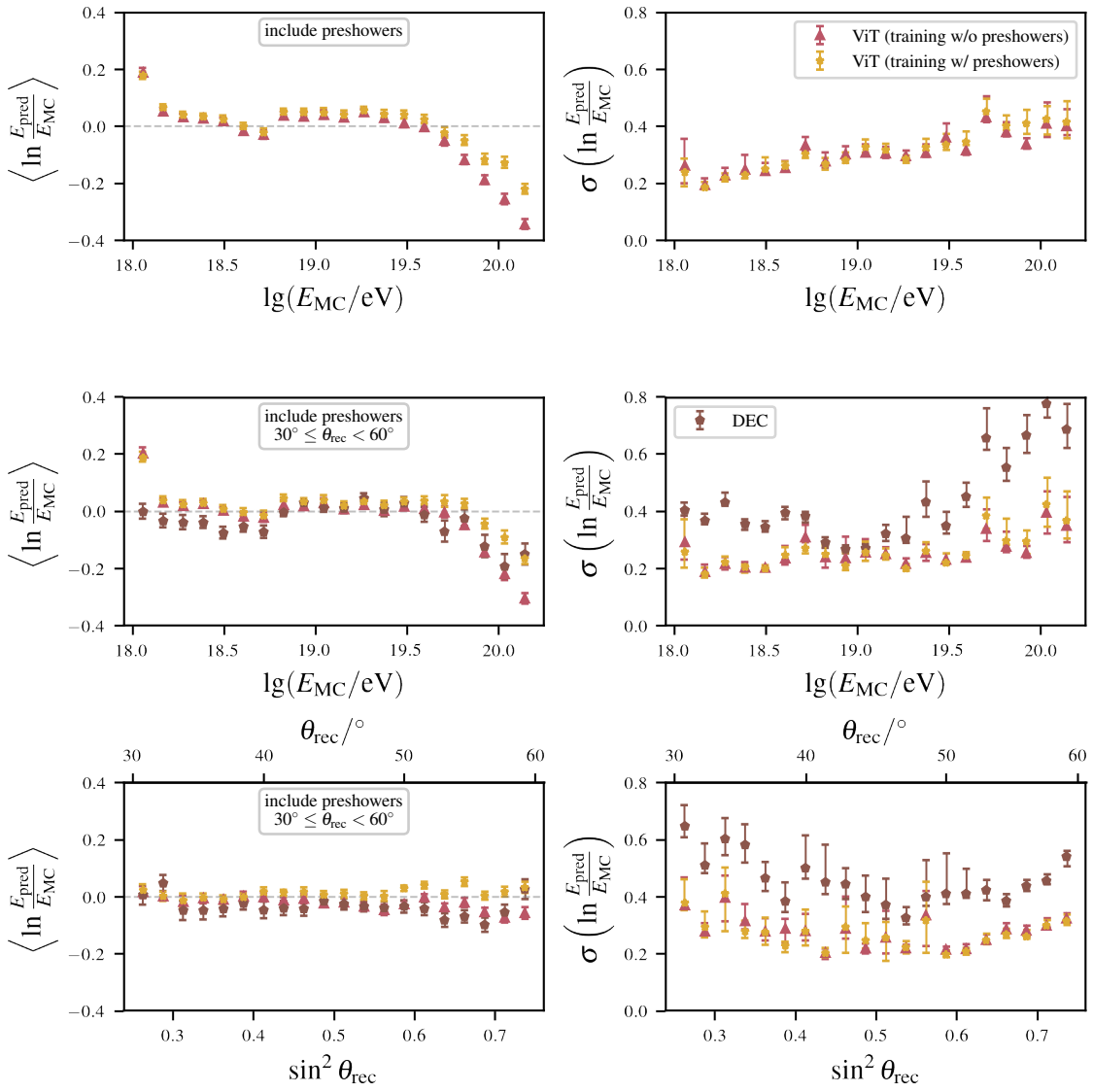


Figure 9.1.: **Bias (left) and resolution (right) of the logarithmic energy prediction for various energy estimators applied to events including preshowers.** The displayed estimators are two identical ViT networks, one trained with preshowers included (yellow) and one without preshowers (red). The third estimator is the unaltered DEC (brown). The first two plots evaluate the estimators for all properly reconstructable events, which includes preshowers. The remaining plots show the subset of events that fulfill $30° \leq \theta_{\text{rec}} < 60°$. Uncertainties of the bias are calculated as the standard error on the mean. The uncertainties of the resolution represent a two-sided 68.27 % confidence interval, determined via bootstrapping.

10. Summary and Outlook

Photons at the Ultra-High-Energy (UHE) frontier may have yet to be discovered, but the Pierre Auger Observatory is one of the best scientific tools for their detection. The observation of a photon with an energy above 10^{18} eV would fundamentally influence our understanding of astroparticle physics models at ultra-high energies. Moreover, it would put strict limits on many beyond the standard model theories. To that end, it is of great interest to reconstruct the exact energy of a hypothetical photon-induced air shower. In this thesis, two methods for this reconstruction were presented.

Reconstructing the primary photon energy is challenging due to the differing signatures of photon-induced showers and hadron-induced showers. The latter have a multitude of hadronic interactions in the atmosphere and produce a large amount of muons during their showering process. In contrast photons and their secondary particles interact mainly electromagnetically, leading to a different shower footprint. Furthermore, photon-induced showers can develop their shower maximum, X_{max}, very late in the atmosphere, with many vertical showers having their X_{max} underground. Another key difference in photon-initiated showers is the existence of preshowers, where the primary photon interacts with the geomagnetic field to produce an e^+e^- pair, which then continues to produce an air shower upon reaching the atmosphere. Preshowers were mostly excluded from this work, but it was shown that the developed models can also successfully reconstruct the energy of preshowering events. To characterize the photon-induced showers, a total of 55 000 air showers with energies between 10^{18} eV and $10^{20.2}$ eV were simulated.

The first method proposed by this work was a classical reconstruction, as it is performed for the hadron-induced showers, but optimized for photon-induced showers. This included a complete parameterization of the Lateral Distribution Function (LDF) as well as fitting parameters for the energy calibration. The LDF for photon-initiated showers improved the 68.27 % percentile of the shower core position error by 12.90 m and that of the shower axis error by 0.294°. Due to showers with deep maxima, preshowers at high energies, and an insufficient detector efficiency at lower energies, the classical estimator was only calibrated with showers that have a reconstructed zenith angle between 30° and 60° and Monte Carlo energies between $10^{18.5}$eV and $10^{19.75}$eV. Although it was not used in the energy reconstruction, the azimuthal signal asymmetry of photon showers with respect to their shower axis has been studied alongside the LDF. It was found that the amplitude of the asymmetry for simulated photon-induced showers is quite significant, due to the strong attenuation of the electromagnetic shower component.

The second presented method relied on Machine Learning (ML) for the energy reconstruction. Two distinct architectures were evaluated. One was based on two-dimenesional

separable convolutions (sepConv2Ds) [90] for the extraction of information between different stations of the Pierre Auger detector array, while the other one was based on the Vision Transformer (ViT) architecture. The optimal performance of the models was ensured by tuning their respective hyperparameters with a Bayesian optimization algorithm.

The classic reconstruction method performed reasonably well within the calibration region, although a zenith and shower size-dependent bias emerged, as the model was not fully capable of capturing the characteristics of the showers. This resulted in a maximum underestimation (overestimation) of the shower energy by approximately 0.19 (0.10) orders of magnitude for certain parameter regions within the calibration region and up to approximately 0.58 (0.16) for energies outside the calibration region. When the relative bias of the energy was calculated, a maximum underestimation (overestimation) of 29.5 % (33.7 %) inside the calibration interval and 92.1 % (43.5 %) for energies outside the calibration interval was observed. When energies outside the calibration were included, the energy and zenith-dependent resolution of the logarithmic energy prediction ranged from 0.15 to 1.02. This is equal to a spread of 0.06 and 0.44 orders of magnitude respectively. All performance metrics were calculated with events that have reconstructed zenith angles between 30° and 60°, which aligns with the calibration interval. For zenith angles outside the calibration interval, the classical predictor overestimated the shower energies.

The Machine Learning (ML) models performed quite well over a large parameter phase space with a maximum relative bias of 2.7 % for underestimations and 18.8 % for overestimations given a range of zenith angles between 30° to 60° and Monte Carlo energies between $10^{18.2}$eV and $10^{19.75}$eV. However, both architectures showed severe edge effects at the borders of the energy range that was used during training, with an overestimation by 0.08 orders of magnitude at low Monte Carlo energies and an underestimation of 0.19 orders of magnitude at high Monte Carlo energies, if averaged over all Monte Carlo zenith angles between 0° and 60°. The energy-dependent resolution of the logarithmic energy predictions for the ML models was found to be lower than for the classic reconstruction with a maximum of approximately 0.2 orders of magnitude. The ViT models have a slight performance advantage over the sepConv2D models, although the different architectures yield very similar results, which is a hint to robust training for both architectures. The edge effects of the ML models can be reduced by extending the training range outside the evaluation range. Compared to the energy estimation with a Lookup Table as used in Reference [1], the ML models showed equal or better performance while being applicable over a larger parameter range.

A preliminary investigation was conducted to conclude if the inclusion of stations that were rejected as out-of-time into the network data sets would impact the predictive power. Only minor improvements were observed in the case of very deep showers.

When applied to preshowering events, all here developed energy estimators demonstrated an improvement in terms of bias and resolution. It is hypothesized, that the LDF fit and the direct energy calibration would benefit from the inclusion of preshowers. It was found that ViT networks trained with preshowers show the most optimal overall performance, without any severe adverse effects on non-preshowering events.

Under the assumption that the simulated photon-induced showers are a close match to hypothetical real showers, the ViT networks trained with preshowers, yield promising results. Future studies should build upon the results achieved here by expanding the training to different atmospheric conditions and testing the impact of accidental muons on the network predictions. It is believed, that the creation of a larger data set, with more events at the low and high energy ends of the training rand would significantly improve the performance of the ML estimators. Future studies may also be able to enhance the energy prediction capabilities of the already existing network ensembles by employing ensemble methods [61, Chapter 7.1]. Given the complex signature of photon-induced showers, ML models present a highly promising alternative to the classical reconstruction method and should be considered in future studies in this area.

Although the classic method for the energy reconstruction performs worse than the ML estimators, it is a well-established approach that may be more resilient to discrepancies between simulations and reality, which makes it a viable energy estimator for photon-induced showers. Moreover, it is believed that modifying the functional form of the direct energy calibration similar to what was suggested in this thesis and incorporating the phenomenology of azimuthal shower asymmetries could significantly improve the performance of the classic reconstruction.

Further improvements in photon energy estimation are expected for both methods with the added functionality provided by the AugerPrime upgrade. The small photomultiplier tubes will allow for the observation of the shower footprint in closer proximity to the shower core. Furthermore, the scintillation detectors, which are more sensitive to the electromagnetic shower component, will probably not only help with the distinction between hadron- and photon-induced showers but will also enhance the energy reconstruction sensitivity of photon-induced showers due to the significant electromagnetic component of the latter.

Acknowledgments

Although written by me, the creation of this thesis would have been impossible without the support from my family, friends, and colleagues, who I now want to thank.

I like to express my deep gratitude to Prof. Dr. Ralph Engel, who served as my first examiner and gave me valuable advice for the content and structure of my thesis. I would also like to extend my gratitude to Priv. Doz. Dr. Roger Wolf, who kindly agreed to be my second examiner.

Furthermore, I am grateful for Dr. Markus Roth, who enabled me to do the research that led to the creation of this thesis and who helped me throughout the year with his expertise and knowledge.

I would like to massively thank the talented Fiona Ellwanger who provided me with know-how and comprehensive explanations, who always had time for me, and who helped me tear down all the walls I ran into.

I thank Steffen Hahn for lending me his knowledge in all things machine learning and aiding me with answers to all the silly little questions I asked.

Additionally, I want to thank Dr. Darko Veberič and Dr. David Schmidt who relentlessly answered all the questions I had and contributed in great parts to shape the analysis and structure of this thesis.

An enormous appreciation for Philipp Meder, Ezequiel Rodriguez, and Emily Pereira Martins, who not only proofread large parts of this thesis but helped me with their expertise in various fields and supported me mentally.

I want to express a heartfelt appreciation to the whole IAP group and all the Auger people who provided me with motivation, fun, and a friendly working environment. Most importantly, big thanks to all the people who brought cake and pretzels to the coffee break.

Finally, I want to thank my family who supported me throughout my whole life as well as my girlfriend and all my friends who supported me emotionally and always believed in me.

Bibliography

[1] P. Abreu et al. "Search for photons above 10^{19} eV with the surface detector of the Pierre Auger Observatory". In: *Journal of Cosmology and Astroparticle Physics* 2023.05 (May 2023), p. 021. DOI: 10.1088/1475-7516/2023/05/021.

[2] V. F. Hess. "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten". In: *Phys. Z.* 13 (1912), pp. 1084–1091.

[3] Z. Cao et al. "Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 gamma-ray Galactic sources". In: *Nature* 594.7861 (May 2021), pp. 33–36. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03498-z.

[4] M. Galaverni and G. Sigl. "Lorentz Violation for Photons and Ultrahigh-Energy Cosmic Rays". In: *Physical Review Letters* 100.2 (Jan. 2008). ISSN: 1079-7114. DOI: 10.1103/physrevlett.100.021102.

[5] H. Li and B.-Q. Ma. "Threshold anomalies of ultra-high energy cosmic photons due to Lorentz invariance violation". In: *Journal of High Energy Astrophysics* 32 (2021), pp. 1–5. ISSN: 2214-4048. DOI: 10.1016/j.jheap.2021.07.001.

[6] V. Berezinsky, M. Kachelriess, and A. Vilenkin. "Ultrahigh-energy cosmic rays without GZK cutoff". In: *Phys. Rev. Lett.* 79 (1997), pp. 4302–4305. DOI: 10.1103/PhysRevLett.79.4302. arXiv: astro-ph/9708217.

[7] "Searches for Ultra-High-Energy Photons at the Pierre Auger Observatory". In: *Universe* 8.11 (Nov. 2022), p. 579. ISSN: 2218-1997. DOI: 10.3390/universe8110579.

[8] M. Fairbairn, T. Rashba, and S. Troitsky. "Photon-axion mixing and ultra-high energy cosmic rays from BL Lac type objects: Shining light through the Universe". In: *Physical Review D* 84.12 (Dec. 2011). ISSN: 1550-2368. DOI: 10.1103/physrevd.84.125019.

[9] C. Meynen. "Search for ultra-high energy photons using the Auger Surface Detector". Diploma thesis. Karlsruhe Institute of Technology, 2014.

[10] A. Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". 2021. DOI: 10.48550/arxiv.2010.11929. arXiv: 2010.11929 [cs.CV].

[11] S. H. Neddermeyer and C. D. Anderson. "Note on the Nature of Cosmic-Ray Particles". In: *Phys. Rev.* 51 (10 May 1937), pp. 884–886. DOI: 10.1103/PhysRev.51.884.

[12] R. L. Workman et al. "Review of Particle Physics - Ch. 30 Cosmic Rays". In: *PTEP* 2022 (2022). DOI: 10.1093/ptep/ptac097.

[13] J. R. Hörandel. "On the knee in the energy spectrum of cosmic rays". In: *Astroparticle Physics* 19.2 (2003), pp. 193–220. ISSN: 0927-6505. DOI: 10.1016/S0927-6505(02)00198-6.

[14] P. Lipari. "Spectral features in the cosmic ray fluxes". In: *Astroparticle Physics* 97 (2018), pp. 197–204. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2017.11.008.

[15] T. Abu-Zayyad et al. "The Knee and the Second Knee of the Cosmic-Ray Energy Spectrum". In: (Mar. 2018). DOI: 10.48550/arxiv.1803.07052. arXiv: 1803.07052 [astro-ph.HE].

[16] G. Brichetto Orquera et al. "The second knee in the cosmic ray spectrum observed with the surface detector of the Pierre Auger Observatory". In: vol. ICRC2023. 2023, p. 398. DOI: 10.22323/1.444.0398.

[17] D. J. Bird et al. "The cosmic-ray energy spectrum observed by the Fly's Eye". In: *The Astrophysical Journal* 424 (Mar. 1994), p. 491. ISSN: 1538-4357. DOI: 10.1086/173906.

[18] A. Aab et al. "Features of the Energy Spectrum of Cosmic Rays above 2.5×10^{18} eV Using the Pierre Auger Observatory". In: *Phys. Rev. Lett.* 125 (12 Sept. 2020), p. 121106. DOI: 10.1103/PhysRevLett.125.121106.

[19] D. J. Bird et al. "Detection of a cosmic ray with measured energy well beyond the expected spectral cutoff due to cosmic microwave radiation". In: *Astrophys. J.* 441 (1995), pp. 144–150. DOI: 10.1086/175344. arXiv: astro-ph/9410067.

[20] P. Abreu et al. "A Search for Photons with Energies Above $2\,10^{17}$eV Using Hybrid Data from the Low-Energy Extensions of the Pierre Auger Observatory". In: *The Astrophysical Journal* 933.2 (July 2022), p. 125. DOI: 10.3847/1538-4357/ac7393.

[21] T. Pierog et al. "First results of fast one-dimensional hybrid simulation of EAS using conex". In: *Nuclear Physics B - Proceedings Supplements* 151.1 (2006). VERY HIGH ENERGY COSMIC RAY INTERACTIONS, pp. 159–162. ISSN: 0920-5632. DOI: 10.1016/j.nuclphysbps.2005.07.029.

[22] J. Álvarez-Muñiz et al. "The Giant Radio Array for Neutrino Detection (GRAND): Science and design". In: *Science China Physics, Mechanics & Astronomy* 63.1 (Aug. 2019). ISSN: 1869-1927. DOI: 10.1007/s11433-018-9385-7.

[23] T. Abu-Zayyad et al. "Upper limit on the flux of photons with energies above 10^{19} eV using the Telescope Array surface detector". In: *Phys. Rev. D* 88 (11 Dec. 2013), p. 112005. DOI: 10.1103/PhysRevD.88.112005.

[24] L. D. Landau and I. Pomeranchuk. "Limits of applicability of the theory of bremsstrahlung electrons and pair production at high-energies". In: *Dokl. Akad. Nauk Ser. Fiz.* 92 (1953), pp. 535–536.

[25] A. B. Migdal. "Bremsstrahlung and Pair Production in Condensed Media at High Energies". In: *Phys. Rev.* 103 (6 Sept. 1956), pp. 1811–1820. DOI: 10.1103/PhysRev.103.1811.

[26] B. McBreen and C. J. Lambert. "Interactions of high-energy ($E > 5 \times 10^{19}$ eV) photons in the Earth's magnetic field". In: *Phys. Rev. D* 24 (9 Nov. 1981), pp. 2536–2538. DOI: 10.1103/PhysRevD.24.2536.

[27] M. Risse and P. Homola. "Search for Ultra-High Energy Photons using Air Showers". In: *Modern Physics Letters A* 22.11 (2007), pp. 749–766. DOI: 10.1142/S0217732307022864.

[28] A. Aab et al. "A SEARCH FOR POINT SOURCES OF EeV PHOTONS". In: *The Astrophysical Journal* 789.2 (June 2014), p. 160. ISSN: 1538-4357. DOI: 10.1088/0004-637x/789/2/160.

[29] A. Aab et al. "A Targeted Search for Point Sources of EeV Photons with the Pierre Auger Observatory". In: *The Astrophysical Journal Letters* 837.2 (Mar. 2017), p. L25. ISSN: 2041-8213. DOI: 10.3847/2041-8213/aa61a5.

[30] G. B. Gelmini, O. E. Kalashev, and D. V. Semikoz. "GZK photons as ultra-high-energy cosmic rays". In: *Journal of Experimental and Theoretical Physics* 106.6 (June 2008), pp. 1061–1082. ISSN: 1090-6509. DOI: 10.1134/s106377610806006x.

[31] C. Bérat et al. "Diffuse Flux of Ultra-high-energy Photons from Cosmic-Ray Interactions in the Disk of the Galaxy and Implications for the Search for Decaying Super-heavy Dark Matter". In: *The Astrophysical Journal* 929.1 (Apr. 2022), p. 55. DOI: 10.3847/1538-4357/ac5cbe.

[32] K. Greisen. "End to the Cosmic-Ray Spectrum?" In: *Phys. Rev. Lett.* 16 (17 Apr. 1966), pp. 748–750. DOI: 10.1103/PhysRevLett.16.748.

[33] Y. A. Fomin et al. "Estimate of the fraction of primary photons in the cosmic-ray flux at energies 10^{17} eV from the EAS-MSU experiment data". In: *Journal of Experimental and Theoretical Physics* 117.6 (Dec. 2013), pp. 1011–1023. ISSN: 1090-6509. DOI: 10.1134/s1063776113140112.

[34] P. Auger et al. "Extensive Cosmic-Ray Showers". In: *Reviews of Modern Physics* 11.3-4 (June 1939), pp. 288–291. ISSN: 0034-6861. DOI: 10.1103/revmodphys.11.288.

[35] V. Novotný et al. "Energy spectrum of cosmic rays measured using the Pierre Auger Observatory". In: *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*. Vol. 395. 2021, p. 324. DOI: 10.22323/1.395.0324.

[36] 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 (2017), pp. 1266–1270. DOI: 10.1126/science.aan4338.

[37] J. Abraham et al. "Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory". In: *Physics Letters B* 685.4 (2010), pp. 239–246. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2010.02.013.

[38] A. Aab et al. "Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV". In: *Phys. Rev. D* 90 (12 Dec. 2014), p. 122005. DOI: 10.1103/PhysRevD.90.122005.

[39] 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 (2010), pp. 227–251. ISSN: 0168-9002. DOI: 10.1016/j.nima.2010.04.023.

[40] C. Meurer and N. S. on behalf of the Pierre Auger Collaboration. "HEAT – a low energy enhancement of the Pierre Auger Observatory". In: *Astrophysics and Space Sciences Transactions* 7.2 (2011), pp. 183–186. DOI: 10.5194/astra-7-183-2011.

[41] E. Varela and (. P. A. Collaboration). "The low-energy extensions of the Pierre Auger Observatory". In: *Journal of Physics: Conference Series* 468.1 (Nov. 2013), p. 012013. DOI: 10.1088/1742-6596/468/1/012013.

[42] 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 (2008), pp. 409–420. ISSN: 0168-9002. DOI: 10.1016/j.nima.2007.12.016.

[43] A. Aab et al. "Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory". In: *Phys. Rev. D* 102 (6 Sept. 2020), p. 062005. DOI: 10.1103/PhysRevD.102.062005.

[44] Pierre Auger Collaboration. "Pierre Auger Observatory Open Data". 2024. URL: https://opendata.auger.org/ (visited on 06/04/2024).

[45] G. Silli et al. "Performance of the 433 m surface array of the Pierre Auger Observatory". In: *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*. Vol. 395. 2021, p. 224. DOI: 10.22323/1.395.0224.

[46] J. Stasielak. "AugerPrime - The upgrade of the Pierre Auger Observatory". In: *International Journal of Modern Physics A* 37.07 (Mar. 2022). ISSN: 1793-656X. DOI: 10.1142/s0217751x22400127.

[47] G. A. Askar'yan. "Excess negative charge of an electron-photon shower and its coherent radio emission". In: *Zh. Eksp. Teor. Fiz.* 41 (1961), pp. 616–618.

[48] G. A. Askar'yan. "Coherent Radio Emission from Cosmic Showers in Air and in Dense Media". In: *JETP* 21 (1965), p. 658.

[49] A. Castellina. "AugerPrime: the Pierre Auger Observatory Upgrade". In: *EPJ Web of Conferences* 210 (2019). Ed. by I. Lhenry-Yvon et al., p. 06002. ISSN: 2100-014X. DOI: 10.1051/epjconf/201921006002.

[50] S. Argirò et al. "The offline software framework of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 580.3 (2007), pp. 1485–1496. ISSN: 0168-9002. DOI: 10.1016/j.nima.2007.07.010.

[51] 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–P10021. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/10/p10021.

[52] X. Bertou et al. "Calibration of the surface array of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 568.2 (Dec. 2006), pp. 839–846. ISSN: 0168-9002. DOI: 10.1016/j.nima.2006.07.066.

[53] 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. 2010), pp. 29–39. ISSN: 0168-9002. DOI: 10.1016/j.nima.2009.11.018.

[54] D. Newton, J. Knapp, and 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: 0927-6505. DOI: 10.1016/j.astropartphys.2006.08.003.

[55] K. Kamata and J. 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.

[56] K. Greisen. “Cosmic Ray Showers”. In: *Annual Review of Nuclear Science* 10.1 (Dec. 1960), pp. 63–108. ISSN: 0066-4243. DOI: 10.1146/annurev.ns.10.120160.000431.

[57] Q. Luce et al. “On the need for unbiasing azimuthal asymmetries in signals measured by surface detector arrays”. In: *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*. Vol. 395. 2021, p. 435. DOI: 10.22323/1.395.0435.

[58] L. Armbrsuter. “Asymmetries of the Lateral Distribution of Particles at the Ground”. MA thesis. Karlsruhe Institute of Technology, 2018.

[59] P. Homola, R. Engel, and H. Wilczyński. “Asymmetry of the angular distribution of Cherenkov photons of extensive air showers induced by the geomagnetic field”. In: *Astroparticle Physics* 60 (2015), pp. 47–53. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2014.05.008.

[60] I. Maris et al. *Comparison of Two Methods to Infer the Cosmic Ray Spectrum at Highest Energies*. GAP-2007_044. URL: https://www.auger.org/gap-notes/download/9-gap-notes-2007/1006-gap2007-044.

[61] I. Goodfellow, Y. Bengio, and A. Courville. “Deep Learning”. MIT Press, 2016. URL: http://www.deeplearningbook.org.

[62] C. C. Aggarwal. “Neural Networks and Deep Learning. A Textbook”. Cham: Springer, 2018, p. 497. ISBN: 978-3-319-94462-3. DOI: 10.1007/978-3-319-94463-0.

[63] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. 2017. DOI: 10.48550/arxiv.1412.6980. arXiv: 1412.6980 [cs.LG].

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

[65] Keras. “Keras Conv2D”. 2024. URL: https://keras.io/api/layers/convolution_layers/convolution2d/ (visited on 04/04/2024).

[66] K. O’Shea and R. Nash. “An Introduction to Convolutional Neural Networks”. 2015. DOI: 10.48550/arxiv.1511.08458. arXiv: 1511.08458 [cs.NE].

[67] OpenAI. “ChatGPT”. 2024. URL: https://openai.com/index/chatgpt/ (visited on 06/04/2024).

[68] A. Vaswani et al. “Attention Is All You Need”. 2023. DOI: 10.48550/arxiv.1706.03762. arXiv: 1706.03762 [cs.CL].

[69] D. Bahdanau, K. Cho, and Y. Bengio. “Neural Machine Translation by Jointly Learning to Align and Translate”. 2016. DOI: 10.48550/arxiv.1409.0473. arXiv: 1409.0473 [cs.CL].

[70] D. Heck et al. *CORSIKA: A Monte Carlo code to simulate extensive air showers.* Tech. rep. 51.02.03; LK 01; Wissenschaftliche Berichte, FZKA-6019 (Februar 98). 1998. DOI: 10.5445/IR/270043064.

[71] T. Pierog et al. “EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider”. In: *Phys. Rev. C* 92 (3 Sept. 2015), p. 034906. DOI: 10.1103/PhysRevC.92.034906.

[72] T. Böhlen et al. “The FLUKA Code: Developments and Challenges for High Energy and Medical Applications”. In: *Nuclear Data Sheets* 120 (2014), pp. 211–214. ISSN: 0090-3752. DOI: https://doi.org/10.1016/j.nds.2014.07.049.

[73] A. Ferrari et al. “FLUKA: A multi-particle transport code (program version 2005)”. CERN Yellow Reports: Monographs. Geneva: CERN, 2005. DOI: 10.5170/CERN-2005-010.

[74] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8.

[75] P. Billoir. “A sampling procedure to regenerate particles in a ground detector from a “thinned” air shower simulation output”. In: *Astroparticle Physics* 30.5 (2008), pp. 270–285. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2008.10.002.

[76] J. Zenglein. “Modellierung und Simulation von kosmischen Luftschauern mit Hilfe der Monte-Carlo-Methode”. MA thesis. Universität Karlsruhe (TH), now Karlsruhe Institute of Technology, 2007.

[77] M. Cranmer. “Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl”. 2023. DOI: 10.48550/arxiv.2305.01582. arXiv: 2305.01582 [astro-ph.IM].

[78] F. James and M. Roos. “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”. In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9.

[79] H. Dembinski et al. “scikit-hep/iminuit”. In: (Dec. 2020). DOI: 10.5281/zenodo.3949207.

[80] A. Aab et al. “Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events”. In: *Phys. Rev. D* 91 (3 Feb. 2015), p. 032003. DOI: 10.1103/PhysRevD.91.032003.

[81] A. Aab et al. “Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory”. In: *Phys. Rev. Lett.* 117 (19 Oct. 2016), p. 192001. DOI: 10.1103/PhysRevLett.117.192001.

[82] P. Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[83] The Pierre Auger collaboration et al. “Deep-learning based reconstruction of the shower maximum Xmax using the water-Cherenkov detectors of the Pierre Auger Observatory”. In: *Journal of Instrumentation* 16.07 (July 2021), P07019. DOI: 10.1088/1748-0221/16/07/P07019.

[84] A. Aab et al. “Extraction of the muon signals recorded with the surface detector of the Pierre Auger Observatory using recurrent neural networks”. In: *Journal of Instrumentation* 16.07 (July 2021), P07016. ISSN: 1748-0221. DOI: 10.1088/1748-0221/16/07/p07016.

[85] F. Ellwanger. “Deep-Learning based Estimation of the Ultra-High Energy Cosmic Ray Spectrum using the Surface Detector of the Pierre Auger Observatory”. MA thesis. Karlsruhe Institute of Technology, 2022.

[86] S. T. Hahn et al. “Boosting the performance of the neural network using symmetry properties for the prediction of the shower maximum using the water Cherenkov Detectors of the Pierre Auger Observatory as an example”. In: *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*. Vol. 395. 2021, p. 239. DOI: 10.22323/1.395.0239.

[87] Pierre Auger Collaboration et al. “Deep-Learning-Based Cosmic-Ray Mass Reconstruction Using the Water-Cherenkov and Scintillation Detectors of AugerPrime”. In: 38th International Cosmic Ray Conference. ICRC 2023. Vol. 444. Proceedings of Science. 2023, Art.–Nr.: 371. DOI: 10.22323/1.444.0371.

[88] M. Abadi et al. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[89] F. Chollet et al. “Keras”. 2015. URL: https://keras.io (visited on 05/29/2024).

[90] “SeparableConv2D”. 2024. URL: https://www.tensorflow.org/api_docs/python/tf/keras/layers/SeparableConv2D (visited on 05/09/2024).

[91] P. J. Huber. “Robust Estimation of a Location Parameter”. In: *The Annals of Mathematical Statistics* 35.1 (1964), pp. 73–101. DOI: 10.1214/aoms/1177703732.

[92] “KerasTuner API”. 2024. URL: https://keras.io/api/keras_tuner/ (visited on 04/11/2024).

[93] T. K. Gaisser and A. M. Hillas. “Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers”. In: *International Cosmic Ray Conference*. Vol. 8. International Cosmic Ray Conference. Jan. 1977, p. 353.

A. Appendix

A.1. Scale of the Relative Bias

It can be useful to display the bias of a network prediction as $\ln(E_{\text{pred}}/E_{\text{MC}})$ instead of the relative bias, $(E_{\text{pred}} - E_{\text{MC}})/E_{\text{MC}}$, since the network is trained with the logarithmic energies. The problem with the latter representation is, that $\langle \ln(E_{\text{pred}}/E_{\text{MC}}) \rangle$ is not the same as $\langle (E_{\text{pred}} - E_{\text{MC}})/E_{\text{MC}} \rangle$, as the mean of the distribution is not invariant under the transformation. Therefore, if the raw network prediction is unbiased, $\langle y_{\text{pred}} - y_{\text{MC}} \rangle$ we will find $\langle \ln(E_{\text{pred}}/E_{\text{MC}}) \rangle = 0$, but not necessarily $\langle (E_{\text{pred}} - E_{\text{MC}})/E_{\text{MC}} \rangle = 0$.

The effect increases, if the spread of predictions is large, which is the case for photons. Figure A.1 shows the difference of the representations. The network in question uses sepConv2D layers in the geometry analyzer and optimized hyperparameters and is trained with non-preshowering events. The evaluation is done for non-preshowering events, within an energy range of $18.8 \leq \lg(E_{\text{MC}}/\text{eV}) < 19.2$ and zenith angles below 60°. The distribution of $\ln(E_{\text{pred}}/E_{\text{MC}})$ shows a near Gaussian spread, as expected for a regression network. For the distribution of $\langle \ln(E_{\text{pred}}/E_{\text{MC}}) \rangle$ an asymmetric spread to the right is visible since overestimations of the logarithmic energies yield larger relative deviations than underestimations. The mean of the relative bias distribution, represented by the colored dashed line, is also shifted to the right.

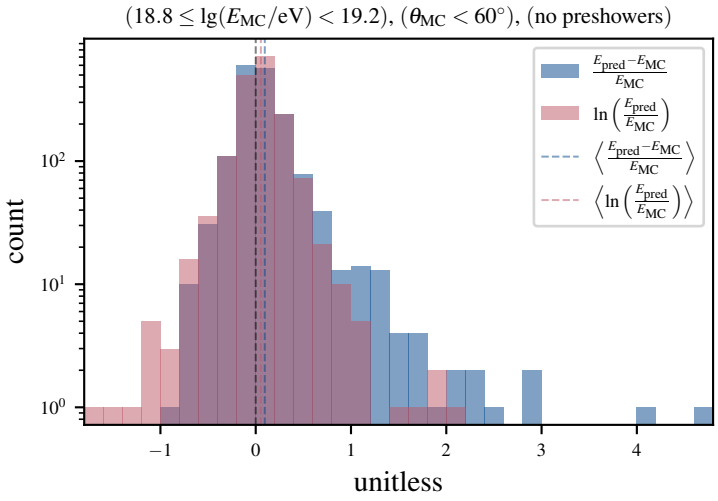


Figure A.1.: **Relative energy bias (blue) and bias of the logarithmic energy (red) of a sepConv2D network.** Only non-preshowering, fully reconstructable events with energies from $10^{18.8}$ eV to $10^{19.2}$ eV and Monte Carlo zenith angels below 60° are evaluated here.

A.2. Supplementary Figures

This section is a collection of supplementary plots.

Figure A.2 shows an exemplary fit of the azimuthal signal asymmetry to equation (4.8). The uncertainty estimations of the signals provided by the uncertainty model show large fluctuations, as indicated by the error bars.

Figure A.3 shows the efficiency $\bar{\varepsilon}$ for binning in the shower energy and the shower zenith angles. The figure shows the total efficiency, which included preshowering events. The slight reduction of the efficiency at $E_{\mathrm{MC}} \approx 10^{20}$ eV and $\sin^2 \theta_{\mathrm{MC}} < 0.4$ is caused by non-preshowering events with a high shower maximum.

Figure A.4 shows the same data as Figure 8.1, only for narrower y-limits. This aids in the discernment of the more nuanced features displayed in the figure.

Figure A.5 shows the distribution of the $\Delta X^{(\mathrm{MC})}$ values of our simulated events, as defined in equation (7.14). The values of $\Delta X^{(\mathrm{MC})}$ are estimated with a fit to the longitudinal shower profile of the MC showers and do not necessarily reflect the physically expected value since the fit is imprecise for deep showers. Please refer to Section 7.6 for more information. The figure demonstrates that $\Delta X^{(\mathrm{MC})}$ increases with the energy, as expected for the LPM effect, and with steeper zenith angles. This figure demonstrates, that most showers with an X_{max} below ground can be avoided by demanding $\theta > 30°$.

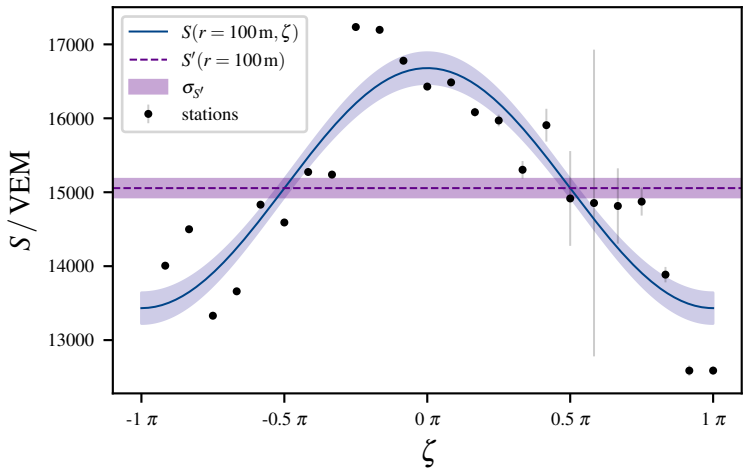


Figure A.2.: **Exemplary fit of the azimuthal asymmetry in showers with high fluctuations in the station uncertainties.** The fit of equation (4.8) (blue) for one dense ring at 100 m distance to the impact point in shower plane coordinates is shown, where the parameters α and S' are free. The estimated value of S' with its uncertainty is plotted too (purple). The gray uncertainties of the station signals are not used in the fit. The event has an energy of 7.1×10^{18} eV and a zenith angle of 52.84°.

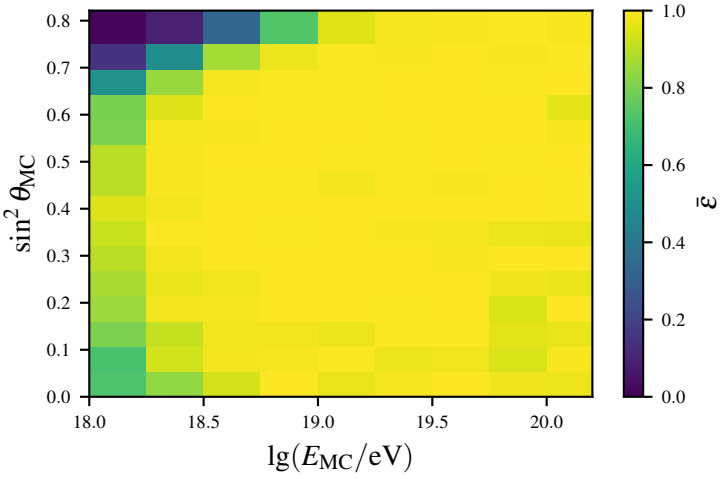


Figure A.3.: **Two-dimensional histogram with the fraction of properly reconstructable events to total events with bins in $\lg(E_{\mathrm{MC}}/\,\mathrm{VEM})$ and $\sin^2 \theta_{\mathrm{MC}}$.** The parameter $\bar{\varepsilon}$ is the efficiency of the Pierre Auger Detector. It is defined as the fraction of events in a bin that pass a T5 and 6T5 trigger and can be fully reconstructed to the total number of events in a bin, including preshowers.

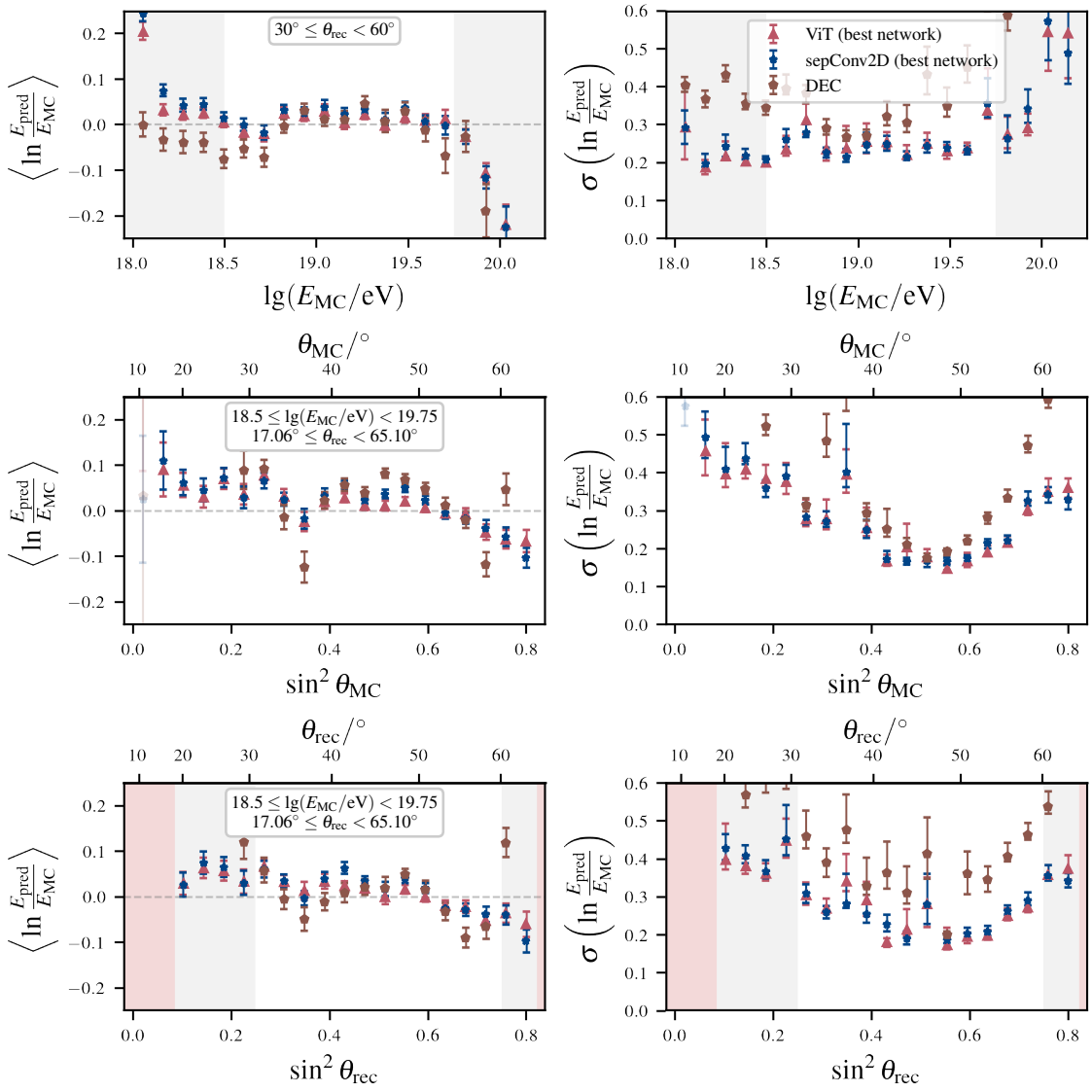


Figure A.4.: **Bias (left) and resolution (right) of the predicted logarithmic energy as a function of $\lg(E_{MC}/eV)$, $\sin^2 \theta_{MC}$, and $\sin^2 \theta_{rec}$.** The gray bars in the first and last row of plots show the edges of the calibration interval of the DEC method and the red bands in the last plot shows the region where the fitted f_{att} is non-physical. Uncertainties of the bias are calculated as the standard error on the mean. The uncertainties of the resolution represent a two-sided 68.27 % confidence interval, determined via bootstrapping.

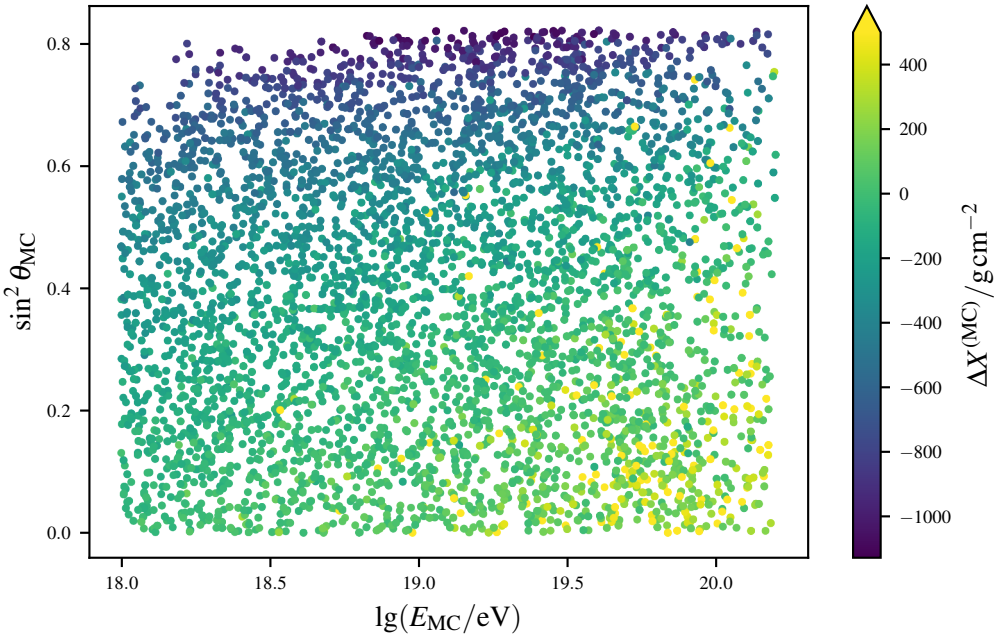


Figure A.5.: **Two-dimensional scatter plot of the $\Delta X^{(\mathrm{MC})}$ distribution.** All non-preshowering events in the evaluation data set that pass a T5 and 6T5 trigger are shown. Showers with $\Delta X^{(\mathrm{MC})} > 0$ have their estimated X_{max} below ground and showers with $\Delta X^{(\mathrm{MC})} < 0$ have their estimated X_{max} above ground.

A.3. Parameter Values

This section serves solely as a collection of the numerical values for fits and parameterization used in this thesis.

Table A.1 shows the coefficients that are used for the parameterization of the shape parameters used in the standard Auger LDF (optimized for hadron-induced events). The parameters are used in equation (4.7).

Table A.2, shows the values that describe the azimuth-dependent signal asymmetry of photon-induced showers (see equation (6.2)), which are determined in this thesis.

The values shown in Table A.3 describe how γ is parameterized as a function of β for the LDF optimization done in this thesis. The values are determined by a fit of equation (6.3).

The values shown in Table A.4 describe how β is parameterized as a function of $\sec\theta$ for the LDF optimization done in this thesis. The values are determined by a fit of equation (6.5).

Table A.5 presents the parameters found when fitting equation (6.7) and equation (6.8) for the DEC of photon-induced events.

Table A.1.: **Coefficients of the hadronic LDF.**

name	value	name	value	name	value
a_0	-3.72	α_0	-1.87	δ_0	0.483
a_1	0.0976	α_1	-0.183	δ_1	0.005
b_0	1.74	β_0	0.490	ϵ_0	-0.272
b_1	-0.242	β_1	-0.065	ζ_0	2.32
c_0	-0.274	γ_0	19.6	η_0	1.95
c_1	0.0349	γ_1	-2.10	η_1	18.01

Table A.2.: **Coefficients of the photon-asymmetry fit.**

name	value
$a_0^{(a)}$	1.072 ± 0.011
$a_1^{(a)}$	-0.206 ± 0.004
$a_2^{(a)}$	0.116 ± 0.011
$a_3^{(a)}$	0.178 ± 0.008
$r_0^{(a)}$	1.337 ± 0.023
$r_1^{(a)}$	-1.069 ± 0.036

Table A.3.: **Parameters from the LDF parameterization: Fitting γ as a function of β.**

name	value
a_γ	-20.93 ± 3.58
b_γ	-7.154 ± 1.072
c_γ	436.3 ± 214.9
d_γ	2.094 ± 0.350

Table A.4.: **Parameters from the LDF parameterization: Fitting β as a function of the zenith angle θ.**

name	value
a_β	-2.885 ± 0.002
b_β	0.4197 ± 0.0012

Table A.5.: **Parameters from the Direct Energy Calibration using photon-induced showers.**

name	value
a	1.196 ± 0.036
b	-7.136 ± 0.091
c	-3.840 ± 0.840
A'	17.740 ± 0.005
B	1.039 ± 0.004