technische universität
dortmund

Hadronic Accelerators in the Universe

First Proton Spectrum measured with the MAGIC Telescopes

Alicia Fattorini
2023

A document submitted in partial fulfillment of the requirements
for the degree of
Doctor rerum naturalium
at
Fakultät Physik, Technische Universität Dortmund

Supervised by
Prof. Dr. Dr. Wolfgang Rhode and Prof. Dr. Carsten Westphal

Er glaubt heute sehr viel weniger als damals. Aber er weiß ein bisschen mehr. Und mitunter, besonders an sonnigen Nachmittagen, hält er das für keinen ganz schlechten Tausch.
Erich Kästner - Eine Auswahl, 1956

Abstract

The search for the origin of charged cosmic rays remains one of the greatest challenges in astrophysics. Extremely accelerated particles propagate through the universe carrying the secrets of the most energetic cosmic phenomena. While neutral particles are not deflected by magnetic fields and point back to their sources, charged cosmic rays arrive on Earth as a diffuse flux, making it nearly impossible to identify their origin. The MAGIC telescopes, primarily designed to detect high-energetic gamma rays, also have the potential to study charged cosmic rays.

This work presents the analysis chain to produce a proton spectrum from data measured with the MAGIC telescopes. The analysis chain includes data preparation, machine learning algorithms for particle reconstruction, and unfolding techniques which consider remaining background contributions. New simulations of air showers induced by charged cosmic rays are used in this analysis and tested accordingly.

This work illustrates the potential of IACTs for the research of charged cosmic rays and provides the first proton spectrum of MAGIC, which constitutes a valuable addition to previous measurements by other cosmic-ray experiments.

Kurzfassung

Die Suche nach dem Ursprung der geladenen kosmischen Strahlung ist nach wie vor eine der größten Herausforderungen der Astrophysik. Extrem beschleunigte Teilchen propagieren durch das Universum und tragen die Geheimnisse der höchstenergetischen kosmischen Phänomene in sich. Während neutrale Teilchen nicht von Magnetfeldern abgelenkt werden und zu ihren Quellen zurückweisen, erreicht geladene kosmische Strahlung die Erde als diffuser Teilchenstrom, was es nahezu unmöglich macht, ihren Ursprung zu bestimmen. Die MAGIC-Teleskope, die in erster Linie für die Untersuchung von hochenergetischer Gammastrahlung konzipiert sind, haben auch das Potenzial, geladene kosmische Strahlung zu untersuchen.

In dieser Arbeit wird die Analysekette zur Erstellung eines Protonenspektrums aus den mit den MAGIC-Teleskopen gemessenen Daten entwickelt. Die Analysekette umfasst die Datenaufbereitung, Algorithmen für maschinelles Lernen zur Teilchenrekonstruktion, und Entfaltungstechniken unter Berücksichtigung von verbliebenen Untergrundbeiträgen. Neue Simulationen der durch geladene kosmische Strahlung induzierten Luftschauer werden in dieser Analyse verwendet und entsprechend getestet.

Diese Arbeit veranschaulicht das Potenzial von IACTs für die Forschung im Bereich der geladenen kosmischen Strahlung und liefert das erste Protonenspektrum von MAGIC, welches eine wertvolle Ergänzung zu den bisherigen Messungen anderer Experimente für kosmische Strahlung bildet.

Contents

Introduction

1

The search for the origin and acceleration mechanisms of cosmic rays has fascinated astrophysicists for more than a century. These high-energy particles travel through our universe and carry the secrets of the most energetic processes occurring at cosmic distances. Today it is known that the high-energy cosmic messengers include neutral particles, such as gamma rays and neutrinos, but also charged particles, like electrons and atomic nuclei, including protons, helium, and heavier nuclei.

Neutral particles are not deflected in magnetic fields; therefore, they point directly back to their source of origin. Up to now, several thousand [22] gamma-ray sources are known while identifying neutrino sources remains a more difficult task: the detection of neutrinos is challenging since they only interact weakly and therefore, huge detector volumes are necessary. Nowadays, neutrino experiments, such as the IceCube detector [1] with a volume of $1\,\mathrm{km}^3$, are in operation, dedicated to the search for cosmic neutrino sources.

In particular, the origin of charged cosmic rays is still a mystery: various experiments have detected a high hadronic particle flux over an enormous energy range for decades, yet no source could be assigned with sufficient certainty. Charged cosmic rays are deflected by magnetic fields in the cosmos and reach the Earth as a diffuse particle flux, which does not directly point back to its origin. Although the charged cosmic particles change their direction on the way to Earth, they offer a unique window into the most energetic events in our universe.

Since the 1920s, detectors have been launched in balloons to high altitudes to measure cosmic rays, the mysterious *Höhenstrahlung* [67]. With the conquest of space, the first satellite-based detector was sent into orbit in the late 1960s [85]. At the same time, research on cosmic-ray-induced air showers developed, and water-filled tanks were built on Earth to measure Cherenkov radiation produced by charged air shower particles [83]. Short time later, Imaging Air Cherenkov Telescopes (IACTs) were built to detect gamma rays indirectly. In addition to Cherenkov light from gamma-induced air showers, IACTs also measure light from hadron-induced air showers as unavoidable background. Initially, the signal-background separation of gamma- and hadron-induced air showers had to be done manually. Nowadays, machine learning algorithms do this work, making the reconstruction fast, robust, and reproducible.

The Major Atmospheric Gamma-Ray Imaging Cherenkov (MAGIC) experiment has been operating for over 20 years, since 2012, as a two-telescopes-system and has achieved groundbreaking results in high-energy gamma-ray physics like the detection of the most promising candidate for high-energetic cosmic neutrinos [2] and the first detection of a Gamma-Ray Burst (GRB) in the TeV energy range [7].

Although MAGIC is designed to detect gamma rays, it has great potential to contribute to the study of charged cosmic rays: most air showers are hadron-induced and thus provide impressive statistics. Due to the complexity of hadron-induced showers compared to gamma-induced showers, a detailed analysis is only now possible with new computational resources. This work aims to verify that IACTs like MAGIC can produce a spectrum of charged cosmic rays: the proton spectrum.

This thesis begins with an introduction to the challenges of identifying cosmic-ray sources and discusses the most prominent candidates in chapter 2. An explanation of the detection mechanisms follows in chapter 3. This chapter includes an outline of the history of astrophysics, starting with the discovery of cosmic rays in 1912 and leading to the MAGIC telescopes and their detection mechanisms.

In chapter 4, the data preparation is described and performed, focussing on the parameterization of the camera images in the context of cosmic-ray analyses. In addition to the measurements, protons, helium, and iron nuclei simulations are produced for this analysis. The simulations for the MAGIC experiment are discussed in detail in chapter 5. After processing the simulations the same way as the measured data, a data-simulation comparison ensures that the simulations represent the measured data well and rules out a mismatch due to systematic uncertainties in the development of cosmic-ray showers, such as the muon puzzle describes in section 5.2.

In chapter 6, machine learning algorithms are trained to identify protons and to estimate their energies. The particle identification is performed in two steps: first, a random forest is trained to separate iron nuclei from the lighter helium nuclei and protons; in the second step, another random forest separates the lighter particles from each other, namely protons from helium. A third random forest is used to estimate the proton energy. Additionally, the performance of the machine learning algorithms is evaluated.

Chapter 7 deals with the detector response. The detector properties like its efficiency and energy resolution, are summarized under the term Instrument Response Functions (IRFs) and calculated with the help of simulations. The detector properties are required for the unfolding in the following chapter 8. In unfolding, the true particle flux is calculated from the observed counts, avoiding the problem that the simple inversion of the detector response is ill-conditioned by minimizing the likelihood of approaching the true spectrum. In this work, the background is taken into account in the unfolding. The background consists of the helium and iron nuclei, incorrectly identified as protons.

This work performs the first unfolding of the proton spectrum considering the remaining background calculated with data from the MAGIC experiment.
Additionally, the spectrum is unfolded in different ranges of zenith distance. Due to the assumption that the proton flux is independent of the observed sky direction, this calculation serves as a test for the analysis by checking whether the results are consistent.
Finally, the spectrum is compared with published data of previous experiments.
The summary of the results and an outlook in chapter 9 complete this thesis.

Theory of Hadronic Accelerators 2

The search for the origin of cosmic rays is still one of the biggest questions in modern astroparticle physics. Gravity is most likely the fundamental accelerator of particles in the universe so that they can reach enormous kinetic energies up to 10^{20} eV [65, 144]. Gravitational energy is partially transformed into kinetic energy when matter collapses, as in a supernova or in the accretion from surrounding matter of supermassive black holes in active galactic nuclei. In following interactions in these extreme environments, particles can be further accelerated. To date, it is not fully understood what sources and mechanisms lead to the shape of the cosmic-ray spectrum as measured and observed on Earth. In modern astroparticle physics, different experiments cooperate and share their results because each experiment specializes in certain types of particles in limited energy ranges. The observed particle types are charged cosmic rays, photons, and neutrinos; additional cosmic messengers are gravitational waves.

Charged cosmic rays consist of protons and heavier nuclei of helium, carbon, oxygen, and iron. They also include leptons like electrons and positrons. Magnetic fields deflect charged particles, so it is challenging to study their origin - they do not take a straight path to Earth and hence do not point back to their source.

Photons have no charge and are therefore not deflected by magnetic fields, but they can be absorbed by interstellar matter. At very high energies above several 100 GeV, photons begin to interact with the Extragalactic Background Light (EBL) [62, 63]: the EBL absorption $\gamma_{\mathrm{VHE}}\,\gamma_{\mathrm{EBL}} \to e^+\,e^-$ leads to the annihilation of very-high-energy photons.

Neutrinos only interact weakly and can cross cosmic distances almost unaffected. This makes them ideal information carriers but also challenging to detect.

Gravitational waves are only observable in certain events, such as the merger of two black holes or neutron stars. The measurement of gravitational waves is a very young field of experimental astrophysics; the first detection of gravitational waves was achieved by the LIGO/Virgo collaboration in 2015 [4] and promises exciting discoveries in the future.

Various experiments (radio, microwave, infrared, optical, ultraviolet, X-ray, and gamma-ray instruments) have detected numerous photon sources, and depending on the wavelength, excellent directional resolution has been achieved.

2.1 Cosmic-Ray Spectra

Because of the deflection of charged particles, reconstructing the origin of cosmic rays is difficult. Until today no source of cosmic rays could be identified. Over wide energy ranges, the spectrum is well known and can be decomposed into the spectra of the individual elements. However, the exact acceleration mechanisms behind the spectra can only be speculated up to now. Magnetic fields deflect the charged particles that reach the Earth in a diffuse flux. The composition of elements in the energy range above 30 TeV is listed in table 2.1. The diffuse fluxes of the cosmic rays used for this analysis are shown in figure 2.1. At lower energies, the data points of different measurements differ strongly, most likely due to solar modulation. The magnetic field of the Sun is variable, and depending on the solar phase, cosmic particles are subject to more or less strong magnetic fields. This affects low-energy particles more than high-energy ones because the Lorentz force acting on moving particles in a magnetic field is larger not only with a larger magnetic field but also with a smaller kinetic energy of the particles.

Table 2.1: Composition of cosmic rays above 30 TeV [139, 140]. The elements are divided into groups: low, medium, high, very high, indicating their mass. The assumption is that elements of similar mass numbers interact similarly with air molecules when entering the Earth's atmosphere.

group	elements	Z	$E > 30\,\mathrm{TeV}$
low	H - He	1 - 2	64 %
medium	C - O	6 - 8	11 %
high	Ne - S	10 - 16	12 %
very high	Cl - Fe	17 - 26	13 %

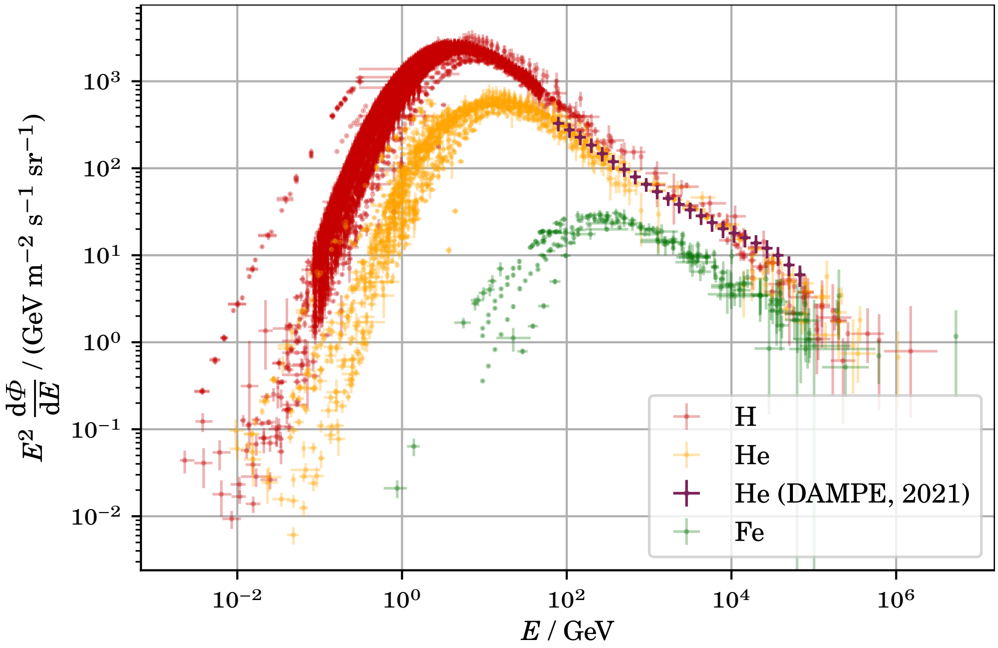

Figure 2.1: Cosmic-ray spectra of measurements with different experiments during the last decades: Data of proton, helium and iron nuclei fluxes per nucleon. The experiments contributing to the spectra are shown in figure 5.1, 5.2 and 5.3. A full description of the experiments is available in appendix C.

2.2 Source Candidates for Hadronic Acceleration

In general, scientists assume that galactic sources located in the Milky Way can only be responsible for cosmic rays up to energies of a few PeV. Because of the large occurrence, Supernova Remnants (SNRs) are favored candidates for most cosmic-ray sources in this energy range. The acceleration of cosmic rays up to almost ZeV energies must be caused by more powerful events than those known in the Milky Way. As extragalactic sources, Active Galactic Nuclei (AGN) are prime candidates. Other exciting sources could be GRBs: short, point-like, and extremely strong gamma-ray flashes. Many have already been detected, but the process behind GRBs is still unclear. Nevertheless, it can be

Table 2.2: Hot candidates for the acceleration of cosmic particles in the universe and the estimated maximum energy the particles can reach in their environment [36, p. 595]

.

source	magnetic field	radius	maximum energy
SNR	30 µG	1 pc	3×10^{16} eV
AGN	300 µG	10^4 pc	10^{21} eV
GRB	1 GG	10^{-3} AU	2×10^{20} eV

assumed that these events can also strongly accelerate charged cosmic rays. A table of the favored candidates for the origin of cosmic rays is given in table 2.2. The energy range covered by this work is between several hundred GeV and several hundred TeV, so all accelerators listed here are possible sources.

2.2.1 Supernova Remnants

Supernova Remnants (SNRs) are assumed responsible for a large fraction of cosmic rays in the TeV range. About 90 % [36, p. 595] of the galactic sources discovered up to now are indeed identified as SNRs: the leftovers of massive stars after their violent death. Depending on their spectral lines, Supernovae (SNe) are divided into different types. If these contain hydrogen lines, they belong to type II SNe, otherwise to type I. The criteria were first created by Minkowski [97] and later specified further. They are summarized in figure 2.2. The different types can be divided into two families that describe the initial event of the SN.

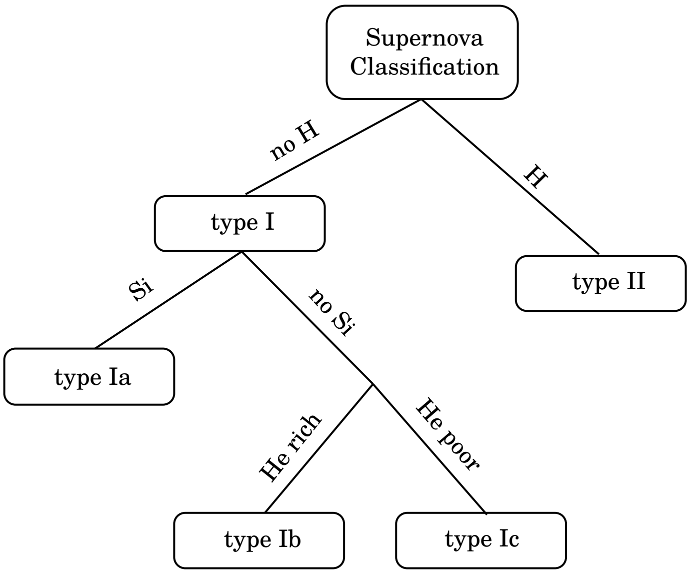

Figure 2.2: Types of SNe, classified according to the measured light curves and absorption lines, which infer their chemical element composition.

Core-collapse SNe (type II, Ib and Ic): The life of a massive star ends in a SN, an explosion. When a star has exhausted its hydrogen supply, its radiation pressure drops. The equilibrium between gravity and radiation pressure, which keeps a star stable, fails, and its core contracts under gravity. Due to the contraction of the matter, the density and temperature increase enormously. The new conditions allow three helium nuclei to fuse into carbon, and the released energy can counteract the collapse. Once the helium supply is burned, the pattern repeats: decrease of radiation pressure, core contraction, increase of temperature and density, fusion of heavier elements until they are exhausted. Depending on the progenitor's mass, this pattern can continue until silicon fuses to iron, the heaviest element, where nuclear fusion releases more energy than it requires. When all the fuel is burned, the nucleus collapses and ejects its shell: an enormous amount of gravitational energy is released as neutrinos. If the remaining mass of the collapsed nucleus has a mass larger than $1.4\,M_{\odot}$, a neutron star is formed; at masses larger than $3\,M_{\odot}$ to $5\,M_{\odot}$, a black hole remains [36, p. 597].

Thermonuclear SNe (type Ia): A less massive star can find its end in a SN under certain circumstances. Typically it collapses into a white dwarf after a complete fuel burn in its preliminary final stage. If matter is accreted, for example, from a dying companion star in a binary system that inflates to a red giant, the white dwarf rapidly gains mass. Due to the increasing self-gravity, it collapses, and finally, the incipient carbon burning results in a SN [36, p. 597].

The most famous SNR is the Crab Nebula. It is the leftover of an observed SN in 1054. In gamma-ray astronomy, it is used as a standard candle (a calibration source) because, on the one hand, the Crab Nebula is a very strong and well-researched gamma-ray source, and on the other hand, the intensity has remained nearly stable over the years. This makes it the ideal object to show the performance of new gamma-ray detectors or new analyses and to compare experiments amongst each other. An image of the Crab Nebula is shown in figure 2.3, an overlay of images from multiple telescopes operating in different energy ranges. In the center of the nebula is a pulsar, a rapidly rotating neutron star encircled by an enormous magnetic field capable of emitting electromagnetic radiation at its poles. Like a rotating lighthouse, it flashes with great temporal reliability. While the SNR emits a nearly constant gamma-ray signal, the slightly less energetic gamma-ray pulses from the neutron star in its center reach the Earth every 33 ms [56, p. 269].

2.2.2 Active Galactic Nuclei

Super Massive Black Holes (SMBHs) with masses from $10^6\,M_{\odot}$ to more than $10^{10}\,M_{\odot}$ are located in the centers of galaxies. In about 1 % of the cases, the black hole is active, thus accretes matter and exhibits strong emission [36, p. 603]. Such active SMBHs in the center of galaxies, their host galaxies, are called Active Galactic Nucleus (AGN). In figure 2.4, an AGN is sketched. In the center is the SMBH, surrounded by the accretion disk, which rotates around the black hole. The infalling matter turns into two collimated jets of highly relativistic particles, ejected perpendicular to the accretion disk in opposite directions. The jet pointing away from the observer is called a counter

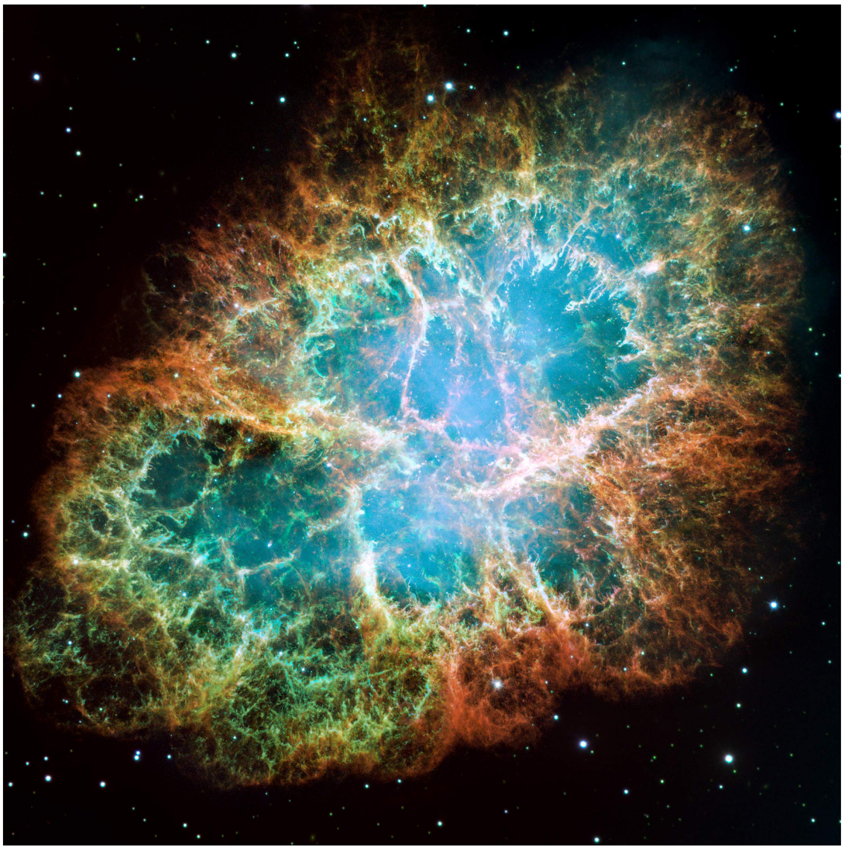

Figure 2.3: Image of the Crab Nebula taken by the Hubble Space Telescope [71] funded and built by the National Aeronautics and Space Administration (NASA) [101] and European Space Agency (ESA) [128]. In the nebula's center is the Crab pulsar, a rotating neutron star that emits radiation in jets at its poles.

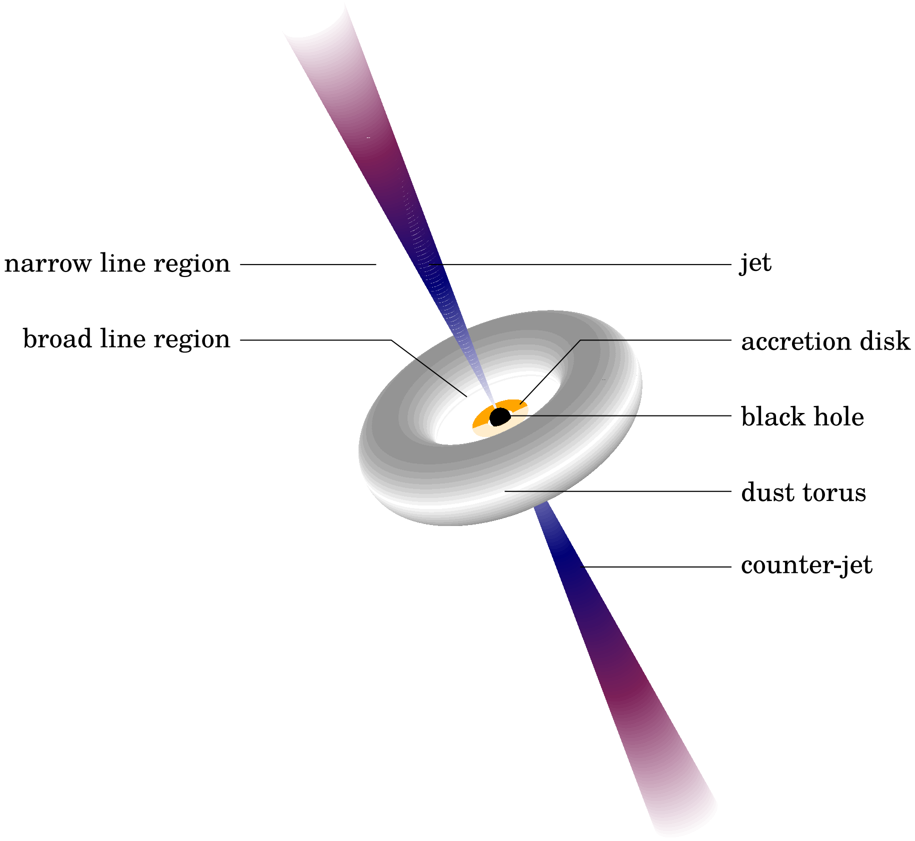

Figure 2.4: Illustration of an AGN. The accretion disk rotates around the SMBH in its center and supplies matter. The collimated jets at the poles emit extremely accelerated matter perpendicular to the accretion disk. The dust torus surrounds the SMBH and supplies the accretion disk with matter. The region near the accretion disk is called the broad line region, and the region around the main jet is called the narrow line region. These regions are not empty; particles interact with the accelerated jet particles.

jet to distinguish it from the jet that faces the observer. Extreme conditions prevail in the environment of AGN: the energy of emitted photons ranges from radio to gamma, and the luminosity can reach up to 10^{40} W [36, p. 604]; this is about 10,000 times the luminosity of a typical galaxy with a non-active black hole in its center. If a jet of an AGN points exactly towards Earth, the object is called blazar. The measured photons from blazars are extremely energetic (in the GeV to TeV range), making AGN also exciting candidates for the origin of charged cosmic rays.

The synchrotron self-Compton model can explain some of the emitted gamma radiation: It assumes a population of high-energy electrons deflected in the prevailing magnetic fields and emit synchrotron radiation. These low-energy synchrotron photons can, in turn, be accelerated by inverse Compton scattering from these same electrons.

$$\gamma + e^- \to \gamma + e^-$$

In this purely leptonic model, no presence or emission of protons or other hadronic matter is necessary. Nevertheless, the synchrotron self-Compton model is not sufficient to explain the shape of the measured blazar spectra [56, p. 291ff].

A second hadronic model provides an explanation: neutral, relativistic pions π^0 in the jet decay into a pair of photons [102].

$$\pi^0 \to \gamma + \gamma$$

These pions must be produced in hadronic interaction processes. In these processes, charged pions $\pi^\pm$ can also be produced.

$$\begin{aligned} p + Y &\to Y' + \pi^0 \\ p + X &\to X' + \pi^\pm \end{aligned}$$

Therefore, charged pions are also predicted in the hadronic model in addition to neutral pions. Charged pions decay into muons and neutrinos.

$$\begin{aligned} \pi^+ &\to \mu^+ + \nu_\mu \\ \pi^- &\to \mu^- + \bar{\nu}_\mu \end{aligned}$$

Since the hadronic model predicts neutrinos in addition to high-energetic photons, the identification of blazars as neutrino sources would support the theory that hadronic particles are accelerated in highly relativistic jets of AGN and serve as a source of cosmic rays[87]. In 2018, a high-energy and likely cosmic neutrino was measured with the IceCube detector [1], and its direction was reconstructed. Spatially and temporally coincident with this neutrino, a gamma-ray flare from the blazar TXS0506+056 was observed by different experiments [2]. This event solidifies the assumption that blazars are sources of cosmic rays like protons.

The Instrument's Setup

3

Earth's atmosphere allows the survival and evolution of life on our planet: it protects terrestrial life from the high-energy radiation in our universe. While the atmosphere is transparent to visible light and radio waves, ionizing radiation like high-energy photons or cosmic protons interact with the molecules of the atmosphere and do, fortunately, not reach the Earth. Nevertheless, the atmosphere's opacity for ionizing radiation makes it difficult for astronomers to explore its origin. Until the early 20th century, the existence of charged cosmic rays was unknown. In 1912, during a balloon flight from Bohemia to Brandenburg, Victor Hess measured, against all expectations, the increasing ionization of the air with higher altitude [67]. A decrease in ionization was expected since terrestrial causes were assumed. The opposite phenomenon could only be explained by extraterrestrial radiation sources: Cosmic rays had been discovered.

Nowadays, two methods help to overcome the atmosphere's opacity to measure cosmic rays: on the one hand, satellites leave the atmosphere behind to measure cosmic particles directly; on the other hand, ground-based detectors measure the secondary radiation of cosmic rays on Earth. Both methods have their advantages and disadvantages. While satellites allow the direct measurement of particles, they are costly, and their detector sizes are limited. Ground-based detectors can be built much larger and operate in arrays. However, reconstructing the primary particles from the measurement requires a great degree of understanding of the particle interactions in the atmosphere and computational effort to simulate these processes.

Today, a multitude of astrophysical experiments on Earth and in space exist. The detectors are specialized for certain particle types and energy ranges. Experiments like the *Fermi*-LAT [19], *Swift*-BAT [23], and *DAMPE* [31] are launched on satellites surrounding the Earth. Famous ground-based experiments are gamma-ray telescopes such as MAGIC [14], H. E. S. S. [10], VERITAS [70] and gamma-ray detectors like HAWC [6], as well as cosmic-ray detector systems like the Pierre Auger Experiment [129] and LHAASO [30]. A full description of the experiments is available in appendix C.

While detectors onboard satellites measure radiation directly, ground-based experiments measure their interaction products propagating in the atmosphere either with water tanks or with IACTs, telescopes with a reflecting dish collecting light into a very sensitive camera. The exact processes in the atmosphere are explained in section 3.1.1.

3.1 Imaging Air Cherenkov Telescopes

Fred Whipple and his research group built the first successfully operating Imaging Air Cherenkov Telescope (IACT) on Mount Hopkins in southern Arizona 1968. The Whipple telescope reflected light with a 10 m parabolic mirror into a camera of 37 Photo Multiplier Tubes (PMTs). A few years later, the first gamma-ray sources were detected: the galactic SNR Crab Nebula [137] and the extragalactic AGN Markarian 421 [110]. The High Energy Gamma-Ray Astronomy (HEGRA) telescope array [17] is the predecessor of the MAGIC telescopes at the Observatorio del Roque de los Muchachos (ORM) on La Palma, Spain, and a pioneer in the field of Cherenkov astronomy by operating several telescopes in a stereoscopic mode. The HEGRA experiment was in operation from 1987 to 2002: it consisted of various detector types like scintillation counters and Airshower Observation By angle Integrating Cherenkov Counters (AIROBICC), and in 1992 it started to operate the first of multiple IACTs for the detection of gamma rays. The HEGRA experiment operated in a broad energy range from 1 TeV to several PeV. During the operation of the experiment, the HEGRA collaboration has achieved many successes, such as the detection of the highest energy photons at that time (16 TeV), originating from an extragalactic object, the blazar Markarian 501. In 2002, the HEGRA telescopes were dismantled and replaced by the MAGIC I telescope, the first telescope of the new generation of IACTs on La Palma.
However, to understand how an IACT works, knowing what processes occur when high-energy cosmic particles enter the Earth's atmosphere is essential. These processes are discussed in the following subsections.

3.1.1 Air Showers

When a very-high-energy proton or a heavier nucleus enters the Earth's atmosphere, this primary particle interacts with the air molecules. This collision creates many new particles: secondary hadrons, typically pions. These pions can further interact with more air molecules, creating more secondary hadrons. The shower component resulting from these secondary hadrons is called hadronic shower component [112, p. 26]. A simplified model of a hadron-induced air shower is shown in figure 3.1.

$$\pi + N \rightarrow \pi + X + \ldots$$

Alternatively, pions can decay. A charged pion $\pi^{\pm}$ has a mean lifetime of $\tau = (2.6033 \pm 0.0005) \times 10^{-8}$ s and decays most likely ($\Gamma_j/\Gamma = (99.987\,70 \pm 0.000\,04)\,\%$) into a muon $\mu^{\pm}$ and a muon neutrino ν_{μ} [141]. These processes initiate the muonic component [112,

p. 28] of hadron-induced air showers, see figure 3.1.

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$
$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$$

A neutral pion π^0 has a mean lifetime of $\tau = (8.43 \pm 0.13) \times 10^{-17}$ s and decays most likely ($\Gamma_j/\Gamma = (98.823 \pm 0.034)\,\%$) into a pair of photons [141].

$$\pi^0 \rightarrow \gamma + \gamma$$

In the vicinity of atomic nuclei, a high-energy photon γ can generate an electron e^- and its antiparticle, the positron e^+. This process is called pair production, and the photon's energy is converted mainly into the mass of the electron-positron pair and their kinetic energy.

$$\gamma \rightarrow e^- + e^+ \qquad \text{(pair production)}$$

In turn, the electron and the positron emit photons, the so-called bremsstrahlung, which is produced when charged particles are deflected in the electromagnetic fields of the atoms in the atmosphere. Hence, a cascade of electrons, positrons, and photons develops, the so-called electromagnetic shower component [112, p. 23], see figure 3.1.

$$e^\pm \rightarrow e^\pm + \gamma \qquad \text{(bremsstrahlung)}$$

A cascade of particles develops, and the initial energy of the photon is divided among many particles until the energy of the individual particles is too small to trigger further processes. With primary particles of higher energy, more secondary particles are produced in the first collisions. Furthermore, the number of interaction generations also increases. In addition to pions, the collision of very-high-energy hadrons with air molecules can produce other secondary particles: kaons and baryon-antibaryon pairs. These particles can, in turn, interact with other air molecules or decay. For example, a charged kaon $K^\pm$ has an average lifetime of $\tau = (1.2380 \pm 0.0020) \times 10^{-8}$ s and most likely decays through one of the following decay channels [141].

$$K^+ \rightarrow \mu^+ + \nu_\mu \qquad \Gamma_j/\Gamma = (63.56 \pm 0.11)\,\%$$
$$K^- \rightarrow \mu^- + \bar{\nu}_\mu \qquad \Gamma_j/\Gamma = (63.56 \pm 0.11)\,\%$$
$$K^\pm \rightarrow \pi^\pm + \pi^0 \qquad \Gamma_j/\Gamma = (20.67 \pm 0.08)\,\%$$

When high-energy gamma rays enter the atmosphere, they also induce air showers like charged cosmic rays. By pair production, an electron-positron pair is created from

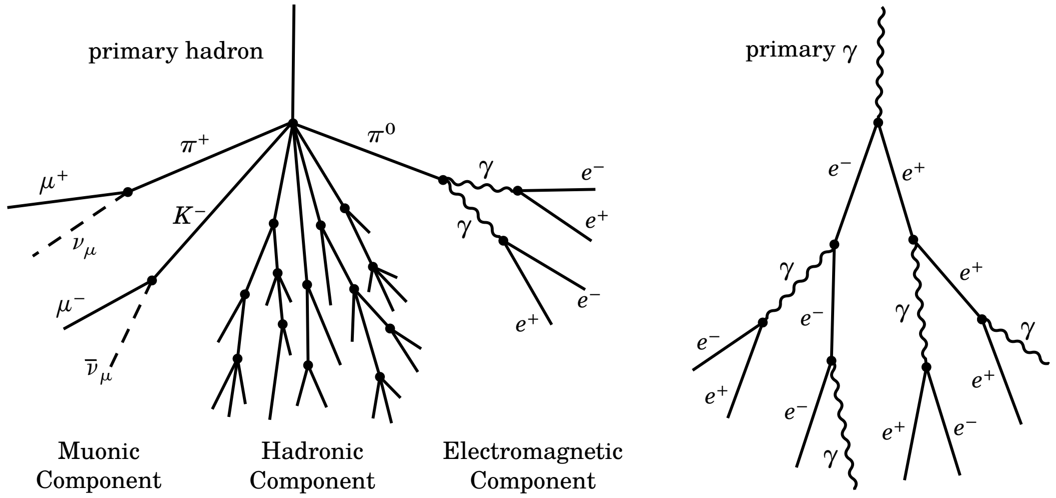

Figure 3.1: Simplified model of an air shower produced by a hadron (left) and a high-energetic photon (right) when entering Earth's atmosphere. While the gamma particle induces an electromagnetic shower, the collision of a hadron with an air molecule produces a more complex shower with different components: an electromagnetic, a hadronic, and a muonic one.

the primary photon in the vicinity of the air atoms. The repetition of bremsstrahlung and pair production occurs, as described for the electromagnetic component of hadron-induced air showers, and an electromagnetic shower is formed. The simplified model of an electromagnetic air shower induced by a primary high-energetic photon is shown in figure 3.1.

3.1.2 Cherenkov Radiation

The primary particle inducing the air shower travels with relativistic energy through a nearly perfect vacuum before entering the atmosphere. When the primary particle interacts with atmospheric atoms, newly created particles achieve kinetic energies close to the speed of light in a vacuum. The velocity of light in a given medium depends on the density and, therefore, the refractive index n of the medium, and while the constant c_0, the speed of light in vacuum, cannot be exceeded, the maximum velocity $c = {}^{c_0}/_{n}$ of light in a given medium can. When a charged particle propagates through a dielectric medium such as air at a velocity v greater than the speed of light in that medium c, the Cherenkov effect occurs: Atoms in the dielectric medium become polarized by the passing charged particle, and when they return to their original state, the atoms emit photons with energies in the optical range. As the charged particle propagates through the medium faster than light, a cone-shaped shock wave of electromagnetic radiation is formed behind the charged particle, as shown in figure 3.2.

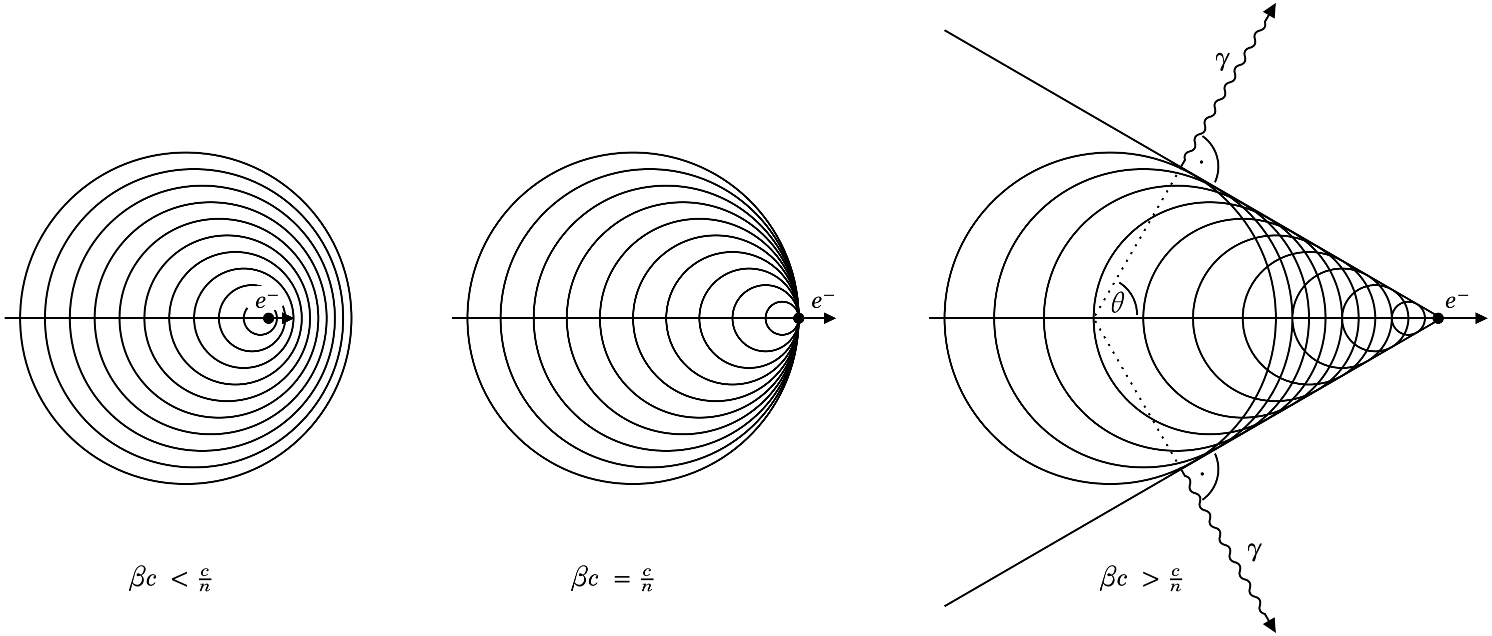

Figure 3.2: Cherenkov effect: a charged particle propagates through a dielectric medium. In the process, the atoms become polarized, and when they return to their original state, they emit photons. The charged particle has the velocity $v = \beta c$, and the speed of light in the medium is $c = {}^{c_0}/_{n}$. The three possible cases are sketched here: if $v < c$, the waves superpose and cancel. If $v = c$, a wavefront is formed in front of the particle. If $v > c$, a cone-shaped shock wave is formed due to constructive interference. The charged particle here is an electron, but it can also be a heavier charged particle like a muon.

The angle θ of light emission depends on the relativistic velocity $v = \beta c_0$ of the charged particle and the refractive index n of the medium.

$$\cos(\theta) = \frac{c}{v} = \frac{1}{\beta n} \tag{3.1}$$

The cone is called Mach cone and is known for objects with ultrasonic speed. The higher the kinetic energy of the charged particle, the greater the light-emission angle θ, and its maximal value is determined by the refractive index n of the medium.

3.2 The MAGIC System

The Major Atmospheric Gamma-Ray Imaging Cherenkov (MAGIC) telescopes are two gamma-ray telescopes located at an altitude of 2200 m above sea level on the Roque de los Muchachos on the Canary Island of La Palma. The first telescope was inaugurated in 2004, and the second telescope and an upgrade of the first telescope were carried out in 2011 and 2012 [14], [15]. Since 2012, the two telescopes have operated in stereo mode, which means they observe the same position in the sky and trigger only an event if both telescopes collect sufficient light from the air shower. A photo of the two telescopes is available in figure 3.3. Each of the telescopes has a dish of 17 m in diameter. The telescope dish consists of several individual mirrors that can be aligned individually. Thus, the mirrors can be focused on and defocused away from the camera. The telescopes are particularly sensitive to gamma rays between about 20 GeV and several 100 TeV. The energy of the primary particles can be determined with a resolution of about 16 %. The telescope's field of view is about 3.5°, and the angular resolution is 0.1°. The structure of the telescopes consists of carbon fiber tubes, as these are exceptionally light yet stable and robust. Thanks to the light structure, the telescopes can be moved

Figure 3.3: The MAGIC telescope system: MAGIC I (left) and MAGIC II (right). In between is the counting house: the computing, electronics, and operation center. The Light Detection and Ranging (LIDAR) tower (white) is mounted on the terrace of the counting house to protect the LIDAR from sunlight, rain, and wind.

very quickly, reaching a maximum speed of 7° per second, and it is possible to rotate the telescopes for 180° within 27 s in total. This maneuverability is essential since one of the main goals of MAGIC is to measure transient objects like GRBs to explore their properties and origins. Since neutrino events of the Ice Cube detector [1] are also published via the Gamma-ray Coordinates Network (GCN) [61], these transient events are also observed by the MAGIC telescopes. The agility of the MAGIC telescopes is also a great advantage.

3.2.1 The Telescope Dish

The two MAGIC telescopes have a dish with a diameter of 17 m each. Both telescope dishes consist of individually movable mirrors. This construction allows the mirrors to be refocused toward the camera at any time. Since the gravitational forces act on the camera depending on the zenith position of the telescopes, a software named Active Mirror Control (AMC) focuses the mirrors after significant repositionings, such as changing the target during the night of observation.

The quality of the measurement depends on the reflectivity of the mirrors and their focus on the camera. The reflectivity of the mirrors usually decreases over time, as the surfaces are permanently exposed to weather conditions such as rain, hail, and sand storms due to the site's proximity to the Sahara [60]. The reflectivity decreases if the dust settles on the mirror surfaces over an extended dry period. Rain can wash it off again and cause the reflectivity to increase suddenly. The focus quality can be evaluated by the Point Spread Function (PSF) of an optical point source, typically a bright star. The measurement of the PSF of such a point source is performed at the beginning of each observation night. On a longer time scale, the focus quality would decrease. Therefore new calibration parameters for the AMC are determined approximately once a year and counteract this effect.

3.2.2 The Camera

The camera is mounted on a carbon fiber arch and faces the dish. A lid protects the camera from sunlight and humidity during the day. The lid can also be closed at night, for example, when car headlights or another strong light source shine into the camera or weather conditions worsen, like upcoming storms or rain. Behind the lid, the camera is additionally protected by an acrylic glass window.

The cameras themselves are each about 1 m in diameter and consist of 1039 PMTs. The incoming light is guided to the PMTs by hexagonal to circular light guides, the so-called Winston Cones. The design prevents gaps between the PMTs and the resulting loss of Cherenkov photons.

The PMTs are exceptionally well designed to detect weak light signals. An incoming photon causes a photoelectron to be released in the cathode at the tube's entrance, which triggers an electron cascade in downstream dynodes. Since the number of electrons potentially increases with the number of dynodes, even a few incoming photons are sufficient to generate a measurable current at the end of the tube. Due to their sensitivity, strong light can significantly damage PMTs, and they should never be powered on in daylight with the camera lid open.
The signals from the PMTs are converted into optical signals using Vertical-cavity Surface-emitting Lasers (VCSELs) and transmitted via optical fibers to the electronics room in the counting house. The optical fibers are about 160 m long, and compared to electrical wires, they have the advantage that no electromagnetic interference occurs between the cables.

3.2.3 Trigger and Readout System

Once the optical signals from the individual pixels reach the electronics room in the counting house, photodiodes convert them to electrical signals at the receiver board. The trigger decides whether to store an incoming signal. An overview of the trigger levels of the MAGIC system is shown in figure 3.4. MAGIC has specific trigger criteria: if a strong pulse is detected in several, especially neighboring pixels, the camera likely captured a shower that caused the signal pattern. The camera information must pass through several trigger levels. Signals pass the level-0 trigger if pulses in single pixels exceed a

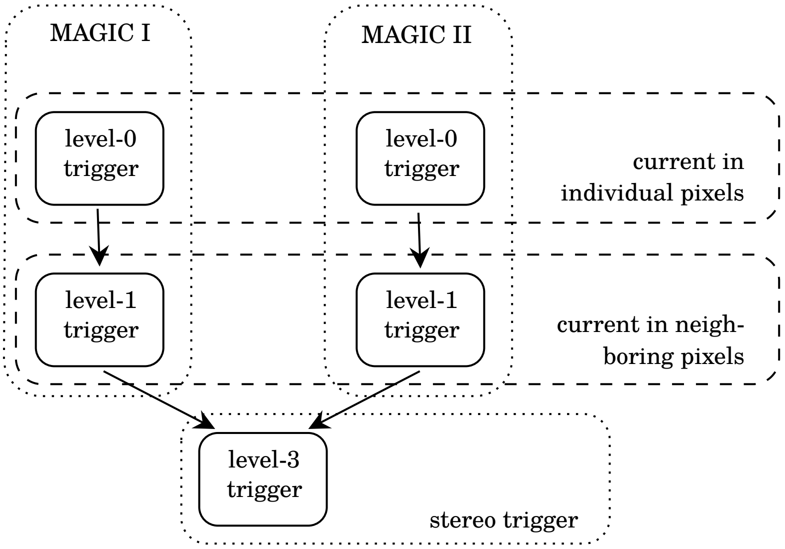

Figure 3.4: Trigger level of the MAGIC telescopes. The level-0 and level-1 triggers must be fulfilled for both telescopes within a short time window so that the level-3 trigger is activated.

certain discriminator threshold. After that, the signal reaches the level-1 trigger stage. Here such events trigger that contain pulses in neighboring pixels that exceed in sum a second discriminator threshold within a short time window. The last stage is the level-3 trigger: while the previous triggers must be passed through independently for both telescopes, this stage uses stereoscopic information: if the images of both telescopes pass the level-1 trigger within a time compatible with the spatial distance between the telescopes, then it is probable that both telescopes detected an atmospheric shower and the light pulses of all pixels in both MAGIC I and MAGIC II cameras around the trigger time are stored.

The discriminator thresholds of the individual trigger stages depend on the observation conditions: in fully dark night skies, the thresholds can be set lower, but with strong moonlight, the threshold must be set higher; otherwise, false triggering due to scattering moonlight would dominate the measurements.

Besides the standard trigger, MAGIC also has the *sum trigger* [57], a trigger bundles the camera pixels into cells, for which it requires certain thresholds. With this trigger, the energy threshold of MAGIC can be slightly reduced, which is especially advantageous for pulsar observations.

The readout system of MAGIC was upgraded in 2012 to a compact Domino Ring Sampler 4 (DRS4) system. Cherenkov flashes typically last between 2 and 20 ns, and the new system is able to sample signals with 2 Gsamples/s with a short deadtime of 26 µs and negligible channel-to-channel cross-talk [14].

3.3 Analysis Overview

The analysis consists of three parts: the data preparation, the data reconstruction, and the inclusion of the detector properties to draw physical results from the measurement. Figure 3.5 gives an overview of the analysis, including all steps.

The first part, the data preparation in chapter 4, is about generating information about the Cherenkov showers from the raw telescope data. This is done by calibrating the measured signals, cleaning the images, and extracting parameters from the measured images that characterize the shower. This process is not called parameterization and can be done for the images of both telescopes separately (mono parameterization) but also with the images of both telescopes together (stereo parameterization). These steps are done with the analysis tool of the MAGIC collaboration `mars`. The software `mars` was developed to analyze gamma-induced showers but can be used for the parameterization because here, no distinction of the primary particles takes place yet, and gamma-induced and hadron-induced showers are treated the same. In addition, new parameters are generated. This process is called feature generation. New features are generated from

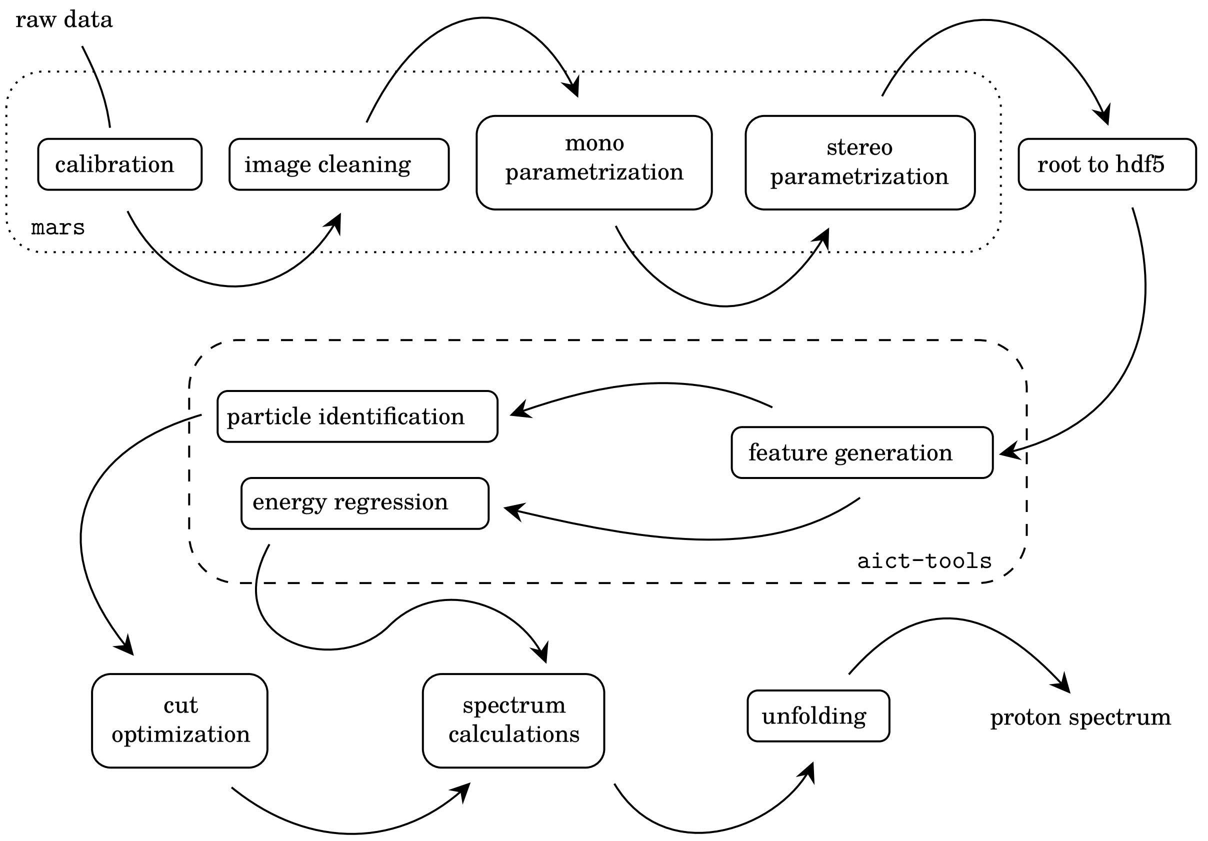

Figure 3.5: Overview of the analysis in this thesis. The analysis chain is sketched from data collection to the proton spectrum. The boxes show which steps are performed with the analysis tools Magic Analysis and Reconstruction Software (`mars`) (dodded) and `aict-tools` (dashed). New software was written in this work for the remaining analysis steps and further tests.

the set of parameters to emphasize the shower characteristics more and reconstruct the events better.

The reconstruction takes place in the second step in chapter 6: from the features, the particle type, and the energy of the primary particles are estimated with the help of machine learning algorithms. For this purpose, three random forests are trained: one for separating iron, one for separating helium, and one for estimating the proton energy. The reconstruction, as well as the model testing, is done by the tool `aict-tools`. Afterward, the best cuts are searched and evaluated.

The third part is about the detector properties and calculating the final spectrum in the chapters 7 and 8. In order to infer the true proton spectrum from the measurement, the detector must be understood in detail. The detector properties are calculated with the help of simulations. An unfolding with consideration of the background finally calculates the proton spectrum.

Data Preparation

4

This chapter focuses on data preparation in order to process the *raw data*, the data stored by the experiment, to the level that allows analyzing it with machine learning algorithms. When the events trigger, the DRS4 stores the time series of the light in each pixel. This is an enormous amount of data, and it is difficult for a random forest to process because it is not able to combine single pixels as logical images. Accordingly, the data volume is first reduced without losing essential information, and the camera images are parameterized in a way that the machine learning algorithm can be taught to reconstruct shower-inducing events using these parameters. The parameterization of shower images and the generation of shower features is explained in section 4.1. In section 4.2, the observation data and simulations of this analysis are initiated and prepared for the testing of the newly produced simulations in chapter 5 and for machine learning following in chapter 6. Preparing the observation data and the simulations includes quality checks and discarding poorly reconstructable events.

4.1 Method: Data Preparation with `mars`

The software `mars` [98] is the standard analysis program of the MAGIC collaboration. Its primary goal is separating gamma-like events from the background, such as protons and heavier particles, and reconstructing their energy and direction to calculate the spectra of potential gamma-ray sources.

The standard software `mars` can be used for processing pixel-wise waveforms to merge them into camera images, clean the shower images from electrical noise and scattered background light, and parameterize them since the aim is to describe the showers with parameters that can be applied to machine learning algorithms.

The raw data consists of the waveforms of the individual camera pixels, meaning the measured photoelectrons in a 30 ns time window around the trigger moment. This raw data must be further processed so that individual attributes of the events can be extracted in order to feed them into machine learning algorithms.

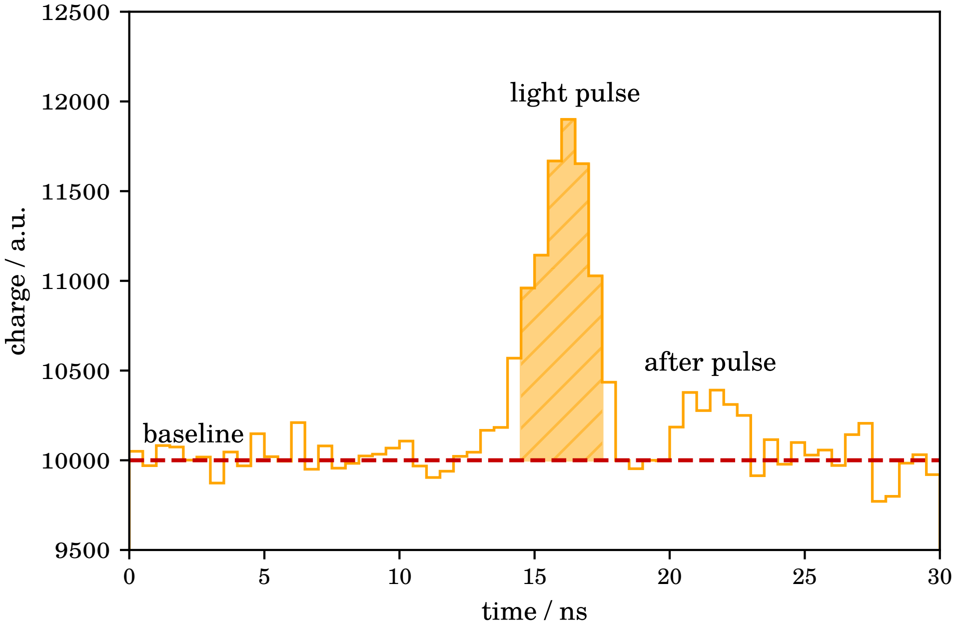

Figure 4.1: Sketch of a typical waveform in a single pixel when recording an air shower. The light pulse which images the Cherenkov flash is followed by the after pulse, a well-known characteristic of PMTs. The baseline describes the background light and is subtracted from the total signal of the light curve. The filled area corresponds to the charge in the pixel triggered by the shower.

4.1.1 From Waveform to Charge and Arrival Time: Calibration

With the calibration, the arrival time of the event and its exposed charge is extracted from the waveform from each camera pixel. The sketch of a waveform in a single pixel when detecting an air shower is shown in figure 4.1. This exemplary waveform describes a strong signal, while most pixels in the camera measure no or very little Cherenkov light of a typical air shower. The charge and arrival time in single pixels are calculated as follows.

Charge In addition to the light from the event shower, the background light reflects into the camera. The background light can be moonlight or light from sunrise or sunset, it can come from the cities or stars, and clouds can reflect light into the camera. Electronic noise also occurs in the PMTs, current generated due to the operation of the PMTs. Pedestal measurements are made before and during the observation to eliminate the electronic and background noise: they are measurements that are not triggered by events but by random triggers. From the waveforms of pedestal measurements, a baseline is calculated. The baseline is subtracted from the event waveforms; ideally, only the charge from the Cherenkov showers remains.

The sliding window method is used to obtain the charge of the event in each pixel:

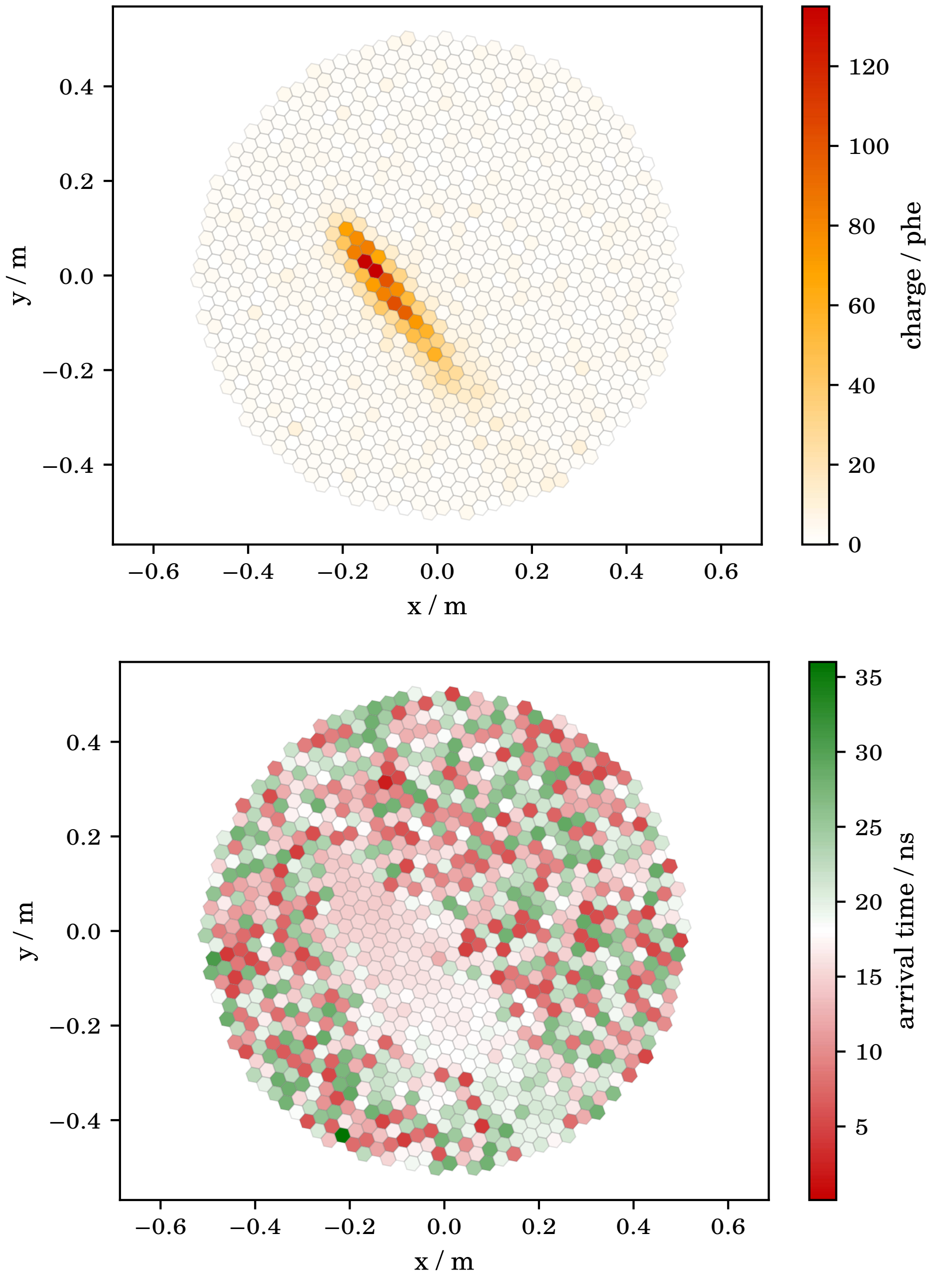

Figure 4.2: The values of the pixels extracted with the calibration from the waveforms: charge (top) and arrival time (bottom) of a real event measured with the MAGIC II camera.

a fixed time range of 3 ns (= 6 time slices) slides over the light curve, and the time range is taken where the integrated charge is the largest.
Up to this step, the charge was measured in readout counts. The calibration converts readout counts into photoelectrons. Therefore, calibration measurements are taken before and during the observations. A calibration box in the center of the telescope dish sends short light pulses with a constant light intensity into the camera, and these pulses are recorded with a special trigger. Since the intensity of the calibration events is known, the transformation from readout counts to photoelectrons can be calculated. PMTs differ among themselves; they age differently. The individual camera pixels vary by 10 % in photoelectrons while homogeneously illuminated during calibration. The operating current is used to counteract the differences in the individual pixels: this process is called flat fielding. Calibration events are taken at a frequency of 50 Hz during observations to ensure constant flat fielding results.

Arrival Time The arrival time is calculated with the resulting time range from the sliding window method. The time bins within this window are weighted with the signal in these time bins. Its average corresponds to the final arrival time of the pixel.

Figure 4.2 shows the camera image of a measurement of a particle shower. Per pixel, one value for the charge and one for the arrival time are stored and displayed here.

4.1.2 Image Cleaning

As chapter 3.2.3 explains, the PMT current is stored per pixel. Besides Cherenkov photons, electronic noise or the night sky background can also generate current in the pixels. Such pixels that show a current, which does not stem from Cherenkov photons, can lead to biases in the parameterization. Therefore, image cleaning is performed beforehand by identifying pixels likely to be from the air shower and discarding those containing only random currents. In the cleaning procedure, all pixels survive, which, together with their N neighboring pixels, have exceeded a certain signal threshold within a fixed time interval of about 1 ns. The number of neighboring pixels N can be between two and four, and depending on the number, the signal and time thresholds are stricter or looser. In the next step, signal thresholds are defined for the individual pixels. For core pixels (pixels surrounded by neighboring pixels), this threshold corresponds typically to 6 phe and for neighboring pixels to 3.5 phe.

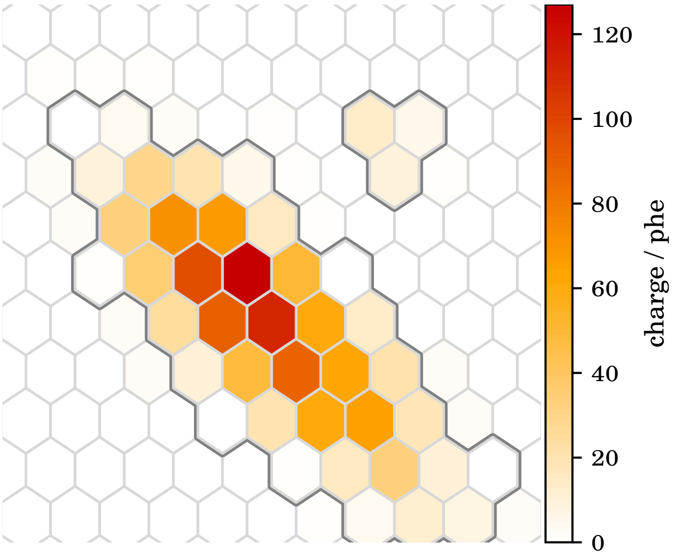

Figure 4.3: Typical shower image in camera coordinates. The pixels that survive the image cleaning are outlined in gray. An exemplary image cleaning was performed, in which all pixels containing 3 phe or more are kept. Next to the main shower, an island has also been recognized.

4.1.3 Image Parameterization

The machine learning methods used here are not able to use the information of the pixels directly to reconstruct the particle's type and energy correctly. Parameterization describes the shower characteristics and aims the reduction of data volume by extracting relevant information about the shower.

The pattern of the shower images provides information about the type of primary particle and its energy. For example, there is a direct correlation between the shower pixels' total charge and the primary particle's energy: the greater the primary particle's energy, the more Cherenkov light is produced in Earth's atmosphere, which the camera can collect. The parameterization of shower images was first presented in 1985 by Hillas [68]. Initially developed for Whipple Observatory data, see section 3.1, the parameterization is still used today to describe images from the present generation of Cherenkov telescopes. The basis of the parameterization is the definition of a shower axis calculated with a Principal Component Analysis (PCA). The first component of the PCA is called `length`, and the second component is called `width`. Over the years, many attributes have been added, and depending on the experiment and analysis, specific attributes are more important than others. The important attributes `width` and `length`, as well as the center of gravity `cog`, are shown in figure 4.4. The main attributes considered for this analysis are

`cog`: center of gravity is the weighted mean of the signal position in the camera.

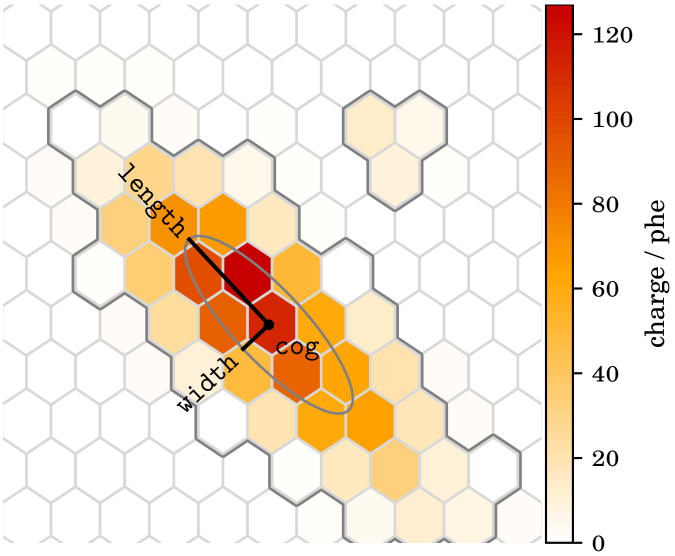

Figure 4.4: Hillas parameters `width` and `length` of a typical shower image with the corresponding `cog`. The pixels that survive the image cleaning are outlined in gray.

`size`: number of photoelectrons of all shower pixels. A strong indicator for the primary particle's energy.

`length`: first component of the PCA.

`width`: second component of the PCA.

`leakage1/2`: fraction of pixel charge in the camera edges (outermost camera ring/two outermost camera rings).

`concentration`: ratio of photoelectrons in the brightest pixel to the photoelectrons in all shower pixels.

`rms_time`: arrival time spread of all pixels assigned to the shower. Defined by $t - RMS = \sqrt{\sum_{i=1}^{k}(t_i - t_{\mathrm{mean}})^2}$.

`p1_grad`: gradient of arrival time along major axis obtained by a linear fit of $y = mx + b$ with m as the gradient.

`num_islands`: number of pixel islands including the main island. A strong indicator for fragmentation of the shower and, therefore, particle type.

`m3_longitudinal`: third longitudinal momentum of the shower image along the image's major axis.

`m3_transversal`: third transversal momentum of the shower image along the image's minor axis.

`asymmetry`: distance between the pixel with the highest charge and the center of gravity.

Additional features, such as the `size_main_island`, are generated from the shower images. Two shower images exist for all events, one recorded by each MAGIC telescope. Therefore, all the parameters are extracted for two images, and the suffixes `_1` and `_2` distinguish them. A complete list of all features used in this analysis can be found in appendix A.3.

4.1.4 Stereo Parameterization

Since MAGIC observes in stereo mode, each event leads to two shower images. From a pair of shower images, a geometrical reconstruction of the shower is possible and further attributes, the stereo parameters, can be calculated. The main stereo parameters of the analysis are explained here:

`impact`: the distance to the core impact point. The core impact point is where the primary particle would hit the ground if it did not interact in the atmosphere. The impact point can be geometrically reconstructed from the intersection of the projected shower axis in each telescope camera. The parameters `impact_m1` and `impact_m2` are the corresponding distances from the respective `cog` and the core impact point. An illustration of the geometrical reconstruction of the `impact` is shown in figure 4.5.

`max_height`: the distance from the ground to the spot where the shower is the brightest. The spot is projected into the camera as the `cog`. The intersection of the following lines can reconstruct it: the lines of sight of the `cog` of each telescope and the reconstructed shower axis. Since these three lines rarely cross at a point in the sky, `max_height` is defined as the height at which the triangle spanned by the three lines and parallel to the ground has the smallest perimeter. An illustration of the geometrical reconstruction of the `max_height` is shown in figure 4.6.

`cherenkov_radius`: radius of the Cherenkov light pool at the ground, that a typical shower electron with the energy of 86 MeV (critical energy of air) produces at the reconstructed `max_height`. This approximates the radius of the Cherenkov light pool of the event.

`cherenkov_density`: density of the Cherenkov light pool at the ground produced by an electron with the critical energy of the air at the height of the `max_height`.

`cos_bs_angle`: the cosine of the angle between the shower axis and the geomagnetic field.

The full list of stereo parameters can be found in appendix A.3.

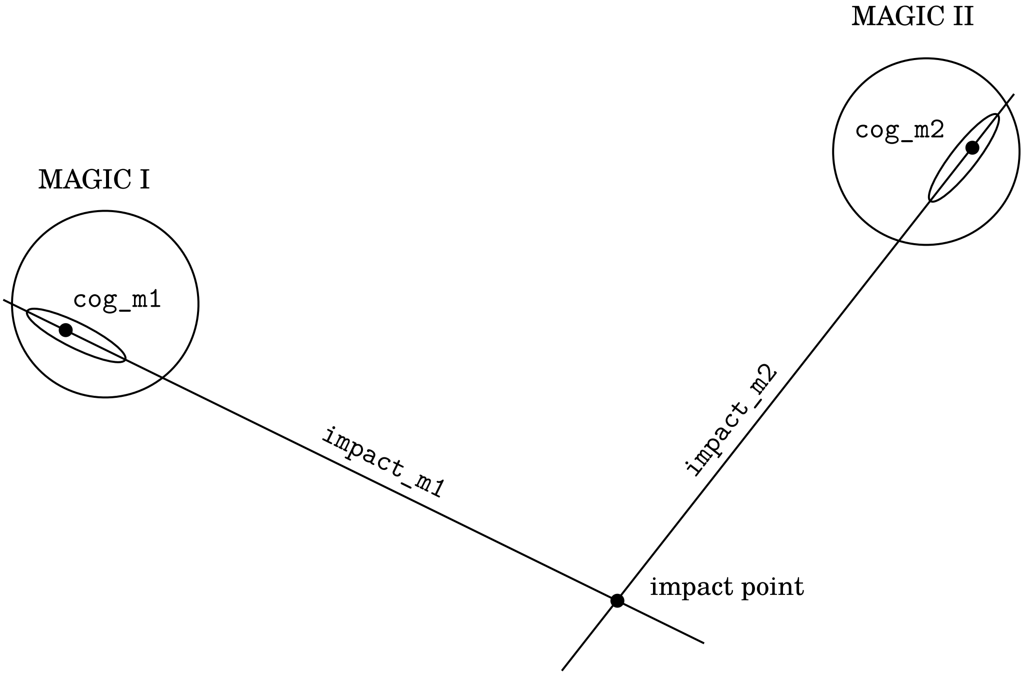

Figure 4.5: Illustration of the parameters `impact_m1` and `impact_m2`. The core impact point is the intersection of the main axes of the shower images of MAGIC I and MAGIC II. The `impact` for each telescope is the distance from the `cog` to the core impact point.

4.1.5 Feature Selection

Some features contain more information about the primary energy or the particle type than others. For the classification of cosmic-ray nuclei, the attributes that correlate with the mass number of the primary particle are extremely important. First, it is assumed that the distributions of `max_height` of heavy nuclei are more spread out than those of light nuclei. Figure 4.7 plots the distribution of `max_height` of the different particle type simulations. The distributions are normalized so that their peaks can be easily compared. The predicted differences in the peaks can be seen in the figure.

Another important shower characteristic for separating different particle types is the attribute `num_islands` since heavier nuclei with larger mass numbers are more likely to form many subshowers in the atmosphere. At the same time, protons tend to form fewer subshowers. Figure 4.8 plots the number of subshowers measured during the event. The differences in the `num_island` distributions of the three particle types can be noticed.

Another feature that illustrates the differences of the air showers of the used particle types could be the `concentration`: in figure 4.9, the distributions of the particle types are shown, and differences in the simulation sets are recognizable.

An essential feature for primary energy reconstruction is the `size` of the shower images.

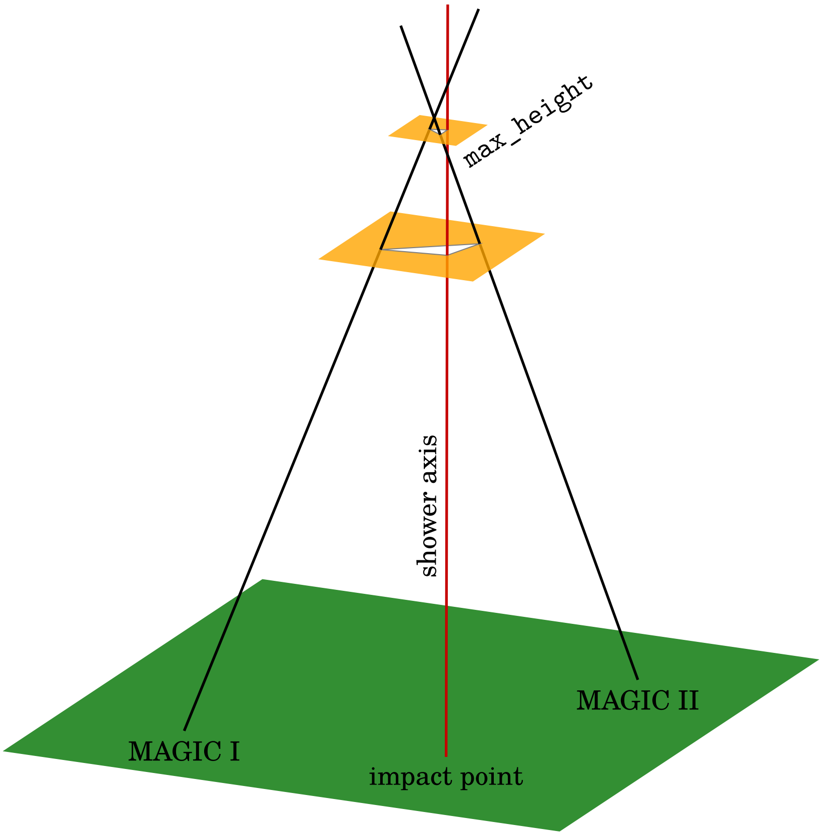

Figure 4.6: Calculation of the parameter `max_height`, the height of the brightest spot of the air shower. The lines of sight from MAGIC I and MAGIC II to their `cog` and the reconstructed shower axis span a triangle parallel to the ground. The height of the triangle's minimal perimeter is the estimated height of the shower maximum, the `max_height`.

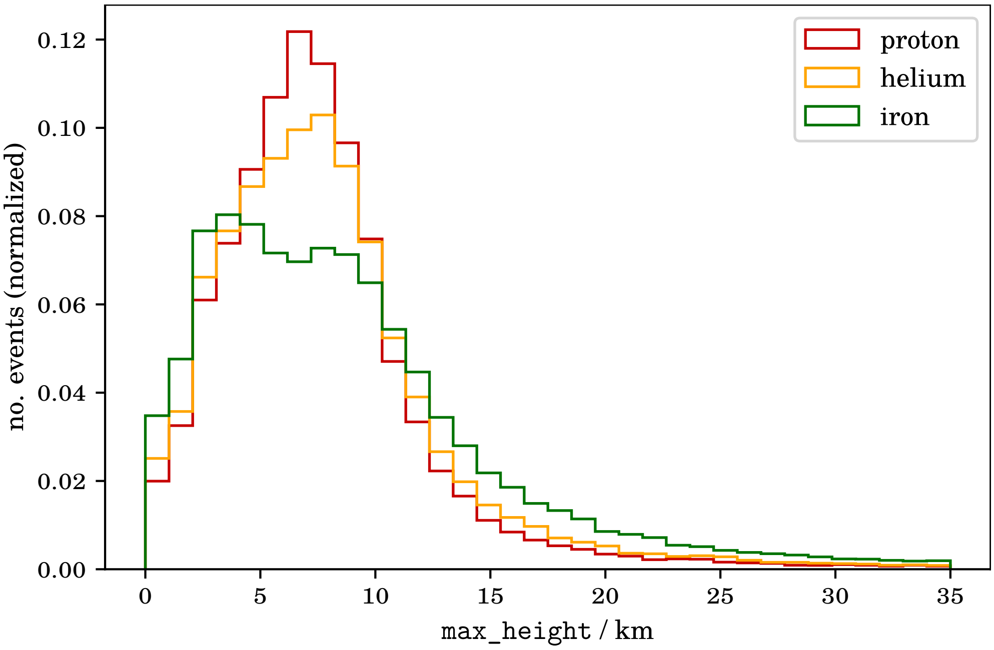

Figure 4.7: Distribution of the geometrically reconstructed `max_height` for the three particle types obtained from simulations. While protons show a clear peak, the distribution smears with higher mass numbers. Hence, the `max_height` could be useful for the particle classifier.

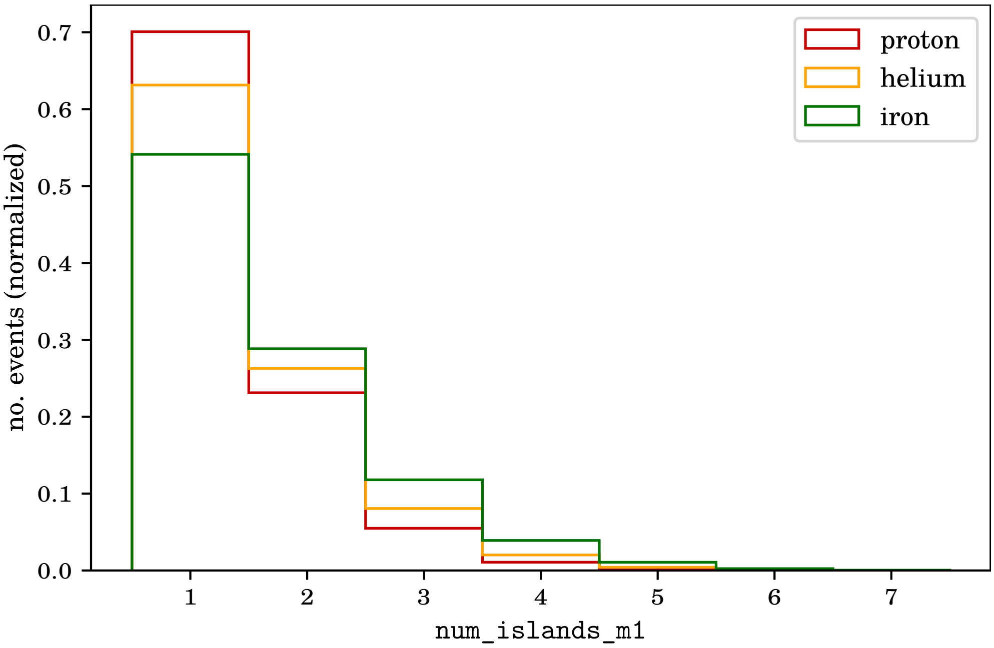

Figure 4.8: Distribution of `num_islands` for the three particle types obtained from simulations. A nucleus with many nucleons is likely to induce more subshowers than a single proton. This relationship between mass number and `num_islands` may be an important feature for the particle classifier.

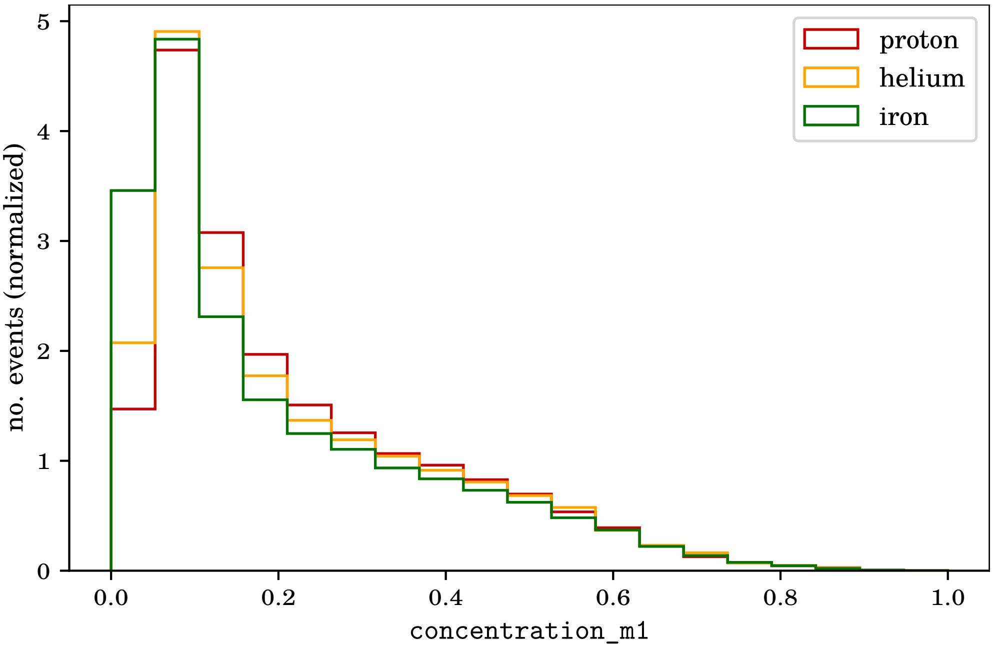

Figure 4.9: Distribution of the parameters `concentration` for the three types of particles obtained from simulations. The `concentration` of the showers is smaller at higher mass numbers since the showers spread out more.

A clear correlation between the true energy of the simulated events and the `size` in the camera is evident. If a shower is not completely inside the camera, the energy reconstruction is more difficult because important information is missing. The correlation between energy and `size` is shown in figure 4.10. Since the number of particles potentially decreases with energy, it makes sense to use logarithmic energy values and `size` for the analysis. Thus, there are enough events in each energy range during the training. Besides the statistical reasons, it is also more convenient for the machine learning algorithm to provide a smaller range of values for the target observable. Differences of `size` distributions of the three types of particles are visible: the heavier the nucleus, the smaller the fraction of Cherenkov light reaching the camera. This is because the energy of the heavy nuclei is divided among more subshowers and scattered further away from the shower core. This means that a shower image with a $\log_{10}(\texttt{size})$ of 2.5 could result from a proton with moderate primary energy or an iron nucleus with much higher primary energy. For this reason, particle separation is particularly important.

4.1.6 Feature Generation

In some cases, generating additional features for the random forest is helpful. In this analysis, the attributes should provide the information needed to reconstruct and

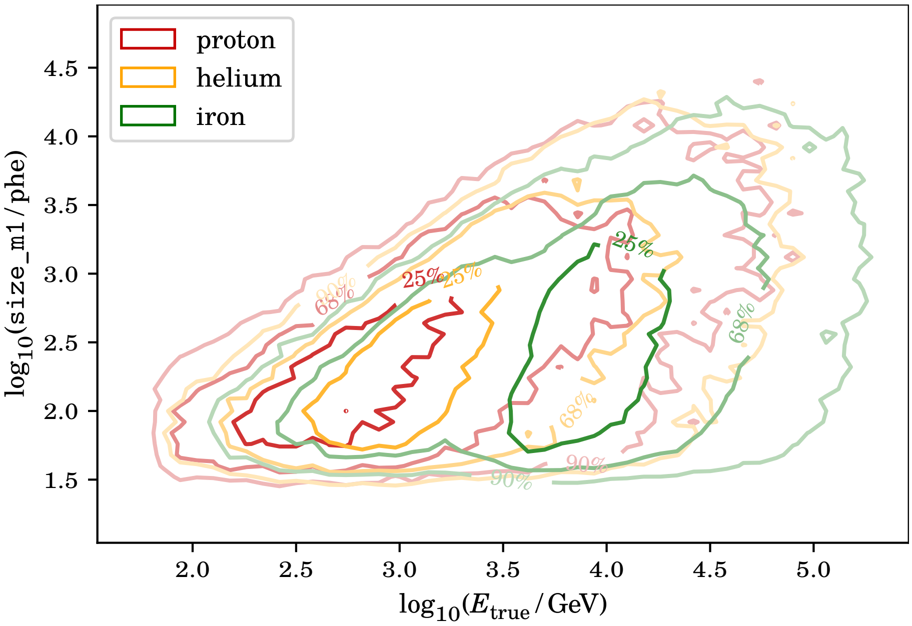

Figure 4.10: Correlation between the `size` of the shower images and the primary energy of the three simulated particle types. The heavier the nucleus, the smaller the fraction of Cherenkov light reaching the camera.

distinguish hadronic primary particles in the most applicable form. As described in section 4.1.5, the distribution of the energy and consequently the parameter `size` approximate a power law. If the parameter `size` were divided into equal-sized bins, some bins would have very few entries, leading to problems with the estimator. With the logarithm of `size`, the events are more evenly distributed among the bins. Other features are generated to combine information logically: the shower shape is characteristic of the particle type, so the ratio of the `length` and `width` can lead to better identification of the primary particle type. Some of the generated features are:

`log_size`: $\log_{10}(\texttt{size})$

`width_length_ratio`: `width/length`

`abs_p1_grad`: $\sqrt{\texttt{p1\_grad} \cdot \texttt{p1\_grad}}$

`area`: $\texttt{width} \cdot \texttt{length} \cdot \pi$

`impact_per_cherenkov_rad`: `impact/cherenkov_radius`

A complete list of the generated features can be found in appendix A.3.

4.2 Data for the Analysis

This chapter outlines the data selection and preparation for this analysis. The considered selection of data and the quality cuts in data and simulations are motivated.

4.2.1 Observation Data

In principle, any data taken with the telescope during standard observations can be used for the analysis since the assumption is that cosmic rays arrive on Earth approximately homogeneously in the field of view of the telescopes. By selecting observations of targets with little or no gamma-ray emission, the already small fraction of gamma rays in the analysis is minimized and thus negligible. Only observations in excellent weather conditions are used, which means that the transmission of the atmosphere is high and the light pollution by, for example, the moon or the galactic plane is low. In addition, data where technical problems occurred, is excluded. The period ST.03.07 is the time between 2016-04-30 and 2017-06-03. In spring 2016, after some rainfalls, the reflectivity of the mirrors and, correspondingly, the optical PSF of the telescopes had improved significantly compared to the previous months. In late summer 2017, both dropped again as dust and sand of the calima sand storm from the Sahara reached La Palma and gradually polluted the mirrors. Due to the stable properties of the telescope during this period, a single set of simulations can be used to analyze the data from this time. Furthermore, the zenith distance distributions of the measurements and simulations are essential for the data selection. The data must be selected in the comparable zenith distance range to the simulations because this affects the shower development: at low zenith, the path through the atmosphere is larger than at high ones. Up to now, the complex simulations cover a range from 5° to 35°. In general, MAGIC observes sources with zenith distances up to 90°. The result is a dataset of observations collected over 166 h considering all the above quality criteria. The criteria for the measurement conditions of the data are shown in table 4.1. A complete list of the observation data used in this analysis can be found in appendix B.

Table 4.1: Conditions for observation data selection. Due to the enormous amount of measurement data in the Monte Carlo period, relatively strict cuts in the quality of the measurements can be made, and measurements in good weather conditions (high atmosphere transmission) can be selected.

	condition
period	ST.03.07
zenith distance	5 - 35°
transmission	> 90 %

4.2.2 Simulation Data

For this particular analysis, hadronic showers are simulated, as they have not been required in any MAGIC analysis before. A total of 124 737 500 camera images of proton showers are produced. Of these showers, only 105 165 events trigger according to the criteria of the observations. In addition, helium- and iron-induced shower are simulated for the background investigations. The size of these datasets is similar to the proton dataset. The settings for the simulations are listed in table 4.2 and correspond to the conditions of the data: good reflectivity of the mirrors, no opacity of the atmosphere for Cherenkov photons, observations in low zenith distances in the range from 5° to 35°. The energy ranges in which the particles are simulated depend on the one hand on the aimed analysis range, taking computing power and time into account, and on the other hand on the energy threshold at which the telescope can detect the particle types at all. The heavier the particle and the higher its energy, the more computing time is required to simulate the interactions in the atmosphere.

Table 4.2: Settings for the simulation of protons, helium and iron nuclei. In total, 421 491 400 events of different energies are produced for this analysis.

	proton	helium	iron
simulated events	124 737 500	124 003 800	172 750 100
triggered events	105 165	157 653	308 536
triggered event ratio	0.08 %	0.13 %	0.18 %
energy E_{min} / GeV	70	140	400
energy E_{max} / TeV	500	2000	6000
spectral index γ	−2	−2	−2
view cone Ω / °	4	4	4
scatter radius r_{max} / m	1500	1500	1500

The proton-induced showers are simulated between 70 GeV and 500 TeV for the primary energy. Figure 4.11 shows the energy distribution of the simulated and triggered events. The percentage of triggered protons is between 0.0043 % at the lower and 5.5 % at the upper energy limit. Helium nuclei are simulated between 140 GeV and 2 PeV. The energy distributions of simulated and triggered helium events are visualized in figure 4.12. The percentage of triggered helium nuclei ranges from 0.002 % at the lower to 7.4 % at the upper energy limit. The simulations of iron nuclei start at 400 GeV and go up to 6 PeV. The energy distributions of the simulated as well as the triggered primary particles are shown in figure 4.13. While in the first two energy bins, no iron event was triggered, and also in the third one, the percentage of triggered events remains very small (4.9×10^{-6} %), the fraction of 11.1 % at high energies is comparatively high. The view cone of 4° and the scatter radius of 1500 m ensure that the events are simulated

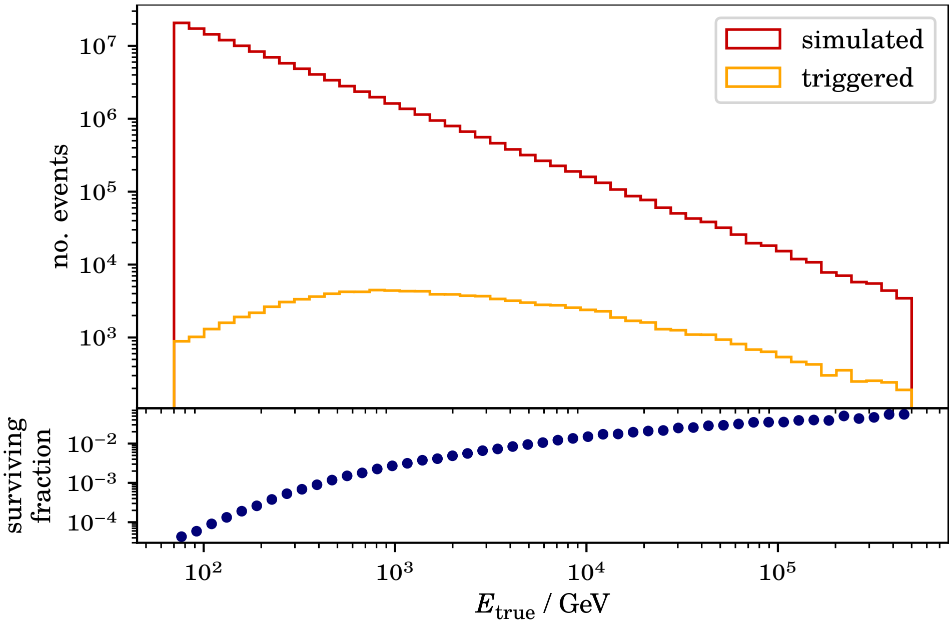

Figure 4.11: Fraction of simulated protons whose shower images collected in the camera fulfill the trigger conditions of the MAGIC telescopes. The surviving fraction indicates the telescope efficiency for protons.

in the field of view of the telescopes and that the Cherenkov light has a high probability of hitting the telescopes to trigger. The simulations of all particle types have an energy distribution following a power law with a spectral index of $\gamma = -2$. The simulated spectral index deviates intentionally from the real spectral index of the cosmic rays, on the one hand, so that the random forest is not biased by the number of events in the certain energy range, and on the other hand, to trigger enough high-energy events in the simulation to cover this range with sufficient statistics for the analysis.

4.2.3 Quality Cuts

The telescope records all the events that have passed the trigger. However, the images of the events are of different quality: some are easier to analyze than others. The quality cuts aim to exclude these events from the analysis to prevent unnecessary computing time and not worsen the analysis results. For example, if an air shower is not entirely collected in the camera, important information is missing, and the estimation of its properties is less confident. At low energies, the Cherenkov light in the resulting showers is faint, and the telescope is less efficient, resulting in only a fraction of all events being detected, creating a large uncertainty in the analysis results in the low energy range. The following cuts in table 4.3 are made to ensure the data quality in the analysis. The same quality cuts are made for the simulations, as the simulations must

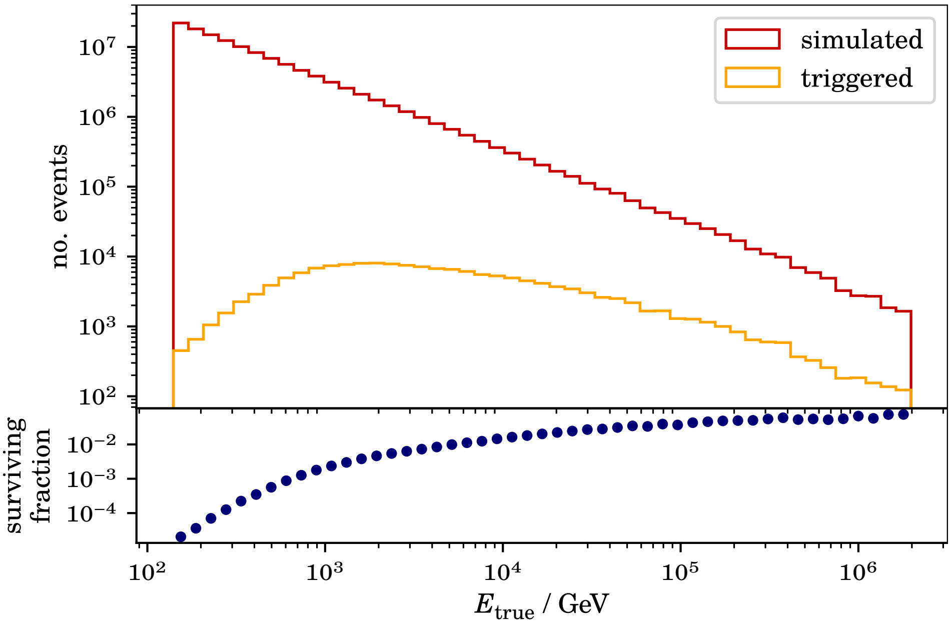

Figure 4.12: Fraction of simulated helium nuclei whose shower images collected in the camera fulfill the trigger conditions of the MAGIC telescopes. The surviving fraction indicates the telescope efficiency for helium nuclei.

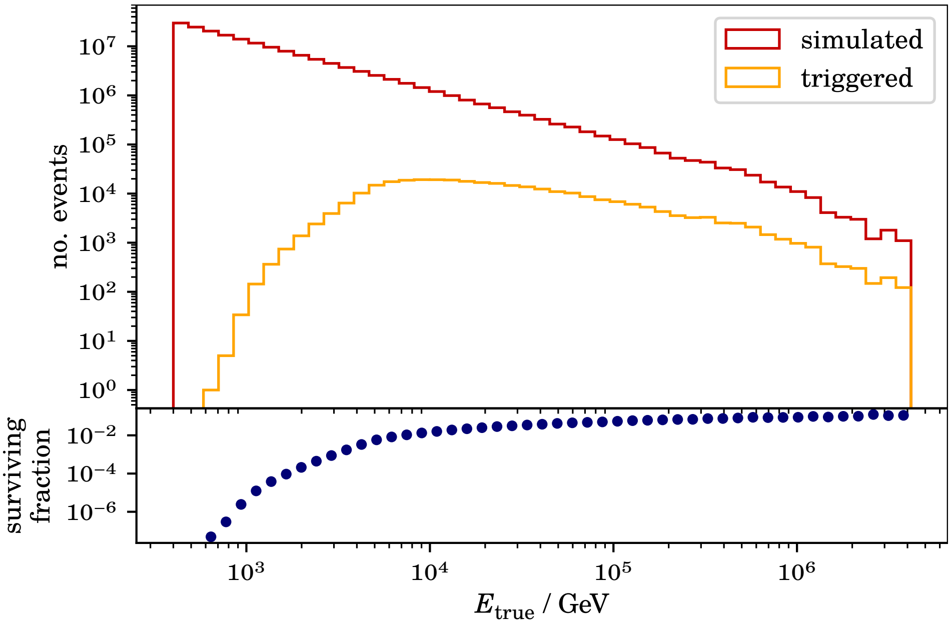

Figure 4.13: Fraction of simulated iron nuclei whose shower images collected in the camera fulfill the trigger conditions of the MAGIC telescopes. The surviving fraction indicates the telescope efficiency for iron nuclei.

Table 4.3: Quality cuts for analysis. The cuts are applied to both measurement data and the simulations. The cuts are used to analyze only those events that have produced complete, well-reconstructable shower images in the camera. If the parameter `valid` is 1, it indicates problem-free data processing up to this level.

parameter		cut
`size_m1`, `size_m2`	>	250 phe
`cog_rad_m1`, `cog_rad_m2`	<	340 mm
`cherenkov_radius`	>	0 cm
`cherenkov_density`	>	0 a.u.
`valid`	==	1

go through the same analysis processes as the data. The plots in figure 4.14 show the `size` distributions of simulations before and after the cuts. The `size` cut eliminates those events in which only little Cherenkov light has reached the camera. The plots in figure 4.15 show the distributions of simulations before and after the cut concerning the `cog_rad`. The `cog_rad` is the distance from the center of gravity of the shower image to the center of the camera. If the shower is closer to the edge of the camera, the value is larger. The cut eliminates events with a large proportion outside the camera radius. The last two plots in figure 4.16 show the distributions of the `cherenkov_density` and the `cherenkov_radius` for the three simulated particles before and after the cuts. The cuts in the Cherenkov parameters are made to remove all non-physical values below zero because when the calculation fails, `mars` assigns the value = -1 to the event.

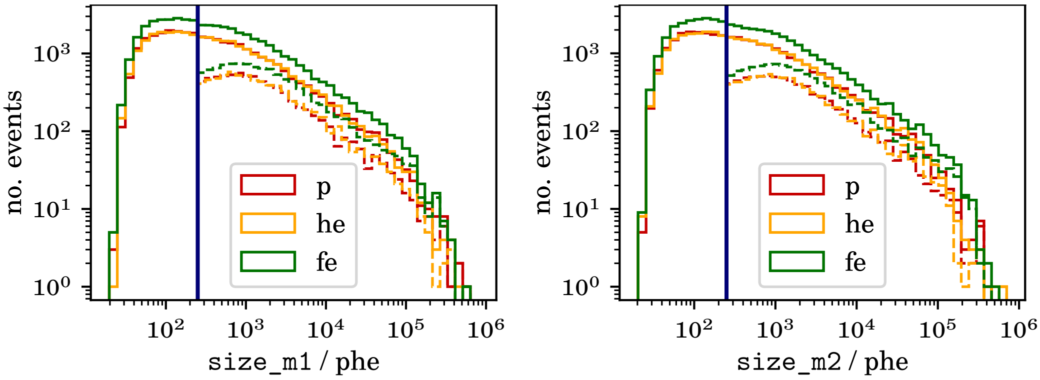

Figure 4.14: The features `size_m1` and `size_m2` before (solid lines) and after (dashed lines) the quality cuts.

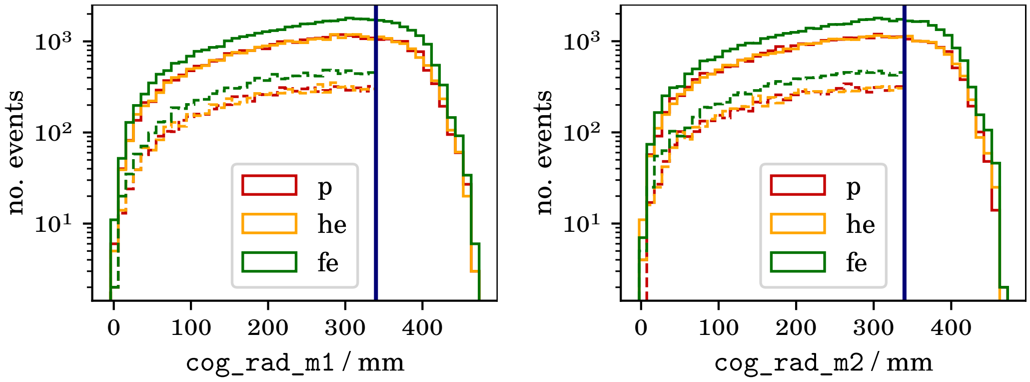

Figure 4.15: The features `cog_rad_m1` and `cog_rad_m2` before (solid lines) and after (dashed lines) the quality cuts.

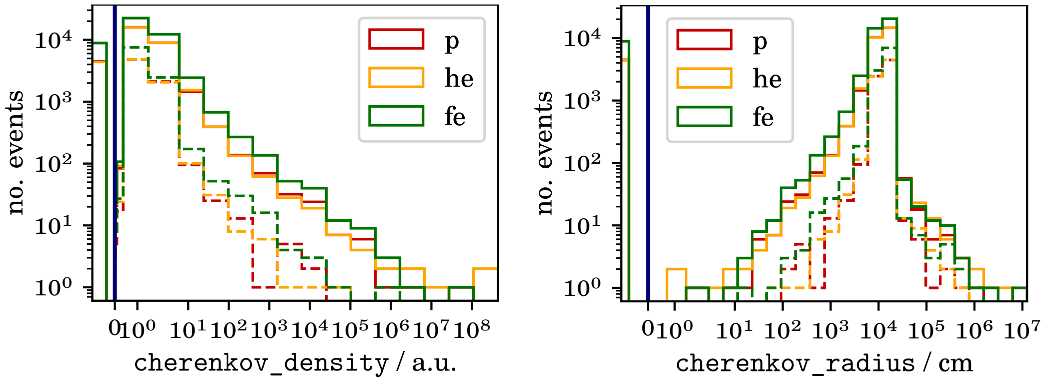

Figure 4.16: The features `cherenkov_density` and `cherenkov_radius` before (solid lines) and after (dashed lines) the quality cuts.

Simulation of Air Showers 5

In order to reconstruct a primary particle's properties from the measurement, datasets of events are needed for which the particle types and energies are known, so-called labeled data. Since these cannot be generated under laboratory conditions, simulations of the air showers help out: all known and relevant interactions of the primary and secondary particles with the atmospheric molecules and electromagnetic fields must be considered and simulated. This chapter discusses the simulation status and current issues in understanding air showers induced by charged cosmic rays. Furthermore, this chapter includes the weighting of simulations. Generally, the energy distribution of simulated events follows a generic power law with a spectral index of $\gamma = -2$. The assumption about the true spectral index and the measurement time is included in the weighting so that the simulations approximate a realistic measurement with corresponding measurement time. The true spectra of the different types of particles are estimated with previous measurements from other experiments. The weighting time is the observation time of the measured data analyzed in this work. The weighting is followed by testing the simulations by comparing them to the measured data from this analysis.

5.1 Monte Carlo Simulations

As described in section 3.1.1, the primary particle induces an air shower when entering Earth's atmosphere. Charged secondary particles in the shower emit Cherenkov light, guided via the known mirror geometry into the camera. In the analysis of IACT data, the simulation of these processes is necessary: the primary particle inducing the air shower, the shower propagation with the emitted Cherenkov light, and the detection by the telescopes.

Finally, a simulated shower image is obtained, which matches the shower image from a measurement in the relevant properties. The simulation is repeated for different particle types and energies to provide sufficient statistics for the analysis. It is called Monte Carlo simulation, a method based on repeated random sampling from a distribution to obtain numerical results.

The simulation of a hadron-induced electromagnetic particle shower is more computationally time-consuming than the simulation of gamma-induced ones. With an increasing mass number, the simulations of the hadronic showers become more complex. For the standard analysis of MAGIC, where gamma-ray sources are studied, the simulation of gamma-induced electromagnetic showers is sufficient. For the particular research in this thesis, in which protons are to be separated from heavier particles, both protons and the most contributing heavy particles, helium, and iron nuclei, are simulated. The simulation software of MAGIC bases on Cosmic Ray Simulations for Kascade (`corsika`) [66], a software developed to simulate air showers for the Karlsruhe Shower Core and Array Detector (KASCADE) [54] experiment in the 1980s. After some modifications and extensions, MAGIC simulates its data with the adapted and customized software MAGIC Monte Carlo Software (`mmcs`). The simulation of the telescope response due to the mirror geometry and reflection is done within the standard software `mars`.

5.2 The Muon Puzzle

A current challenge in understanding the physics of hadronic air showers is called muon puzzle. At very high energies above 10^{15} eV, cosmic rays can only be measured indirectly with ground-based detectors due to the limited surface of satellite-based detectors. Well-known high-energy cosmic-ray detectors are Pierre Auger Experiment [129] in western Argentina and IceTop [3], the surface component of the antarctic IceCube Neutrino Observatory [1]. A full description of the experiments is available in appendix C. An essential feature for these experiments is the number of muons N_μ produced in pion or kaon decays (see section 3.1.1) in a shower induced by cosmic rays since the number of muons strongly correlates with the mass A of the primary particle [112, p. 29].

$$N_\mu \propto A^{1-\alpha_\mu} \tag{5.1}$$

The exponent α_μ depends on the muon energy E_μ: the higher the muon energies, the steeper the muon spectrum due to decreasing energy loss through ionization and decay with increasing muon energies. Today, several models describe the processes and predict the number of muons in cosmic-ray showers.

A hitherto unsolved problem is the muon puzzle: the number of measured muons is higher than the number of expected muons – across different models. In 2018, a collaboration of eight experiments published an energy-dependent discrepancy between prediction and measurement with a significance of $8\,\sigma$ [37]. That means the interaction processes, including muon production in air showers, still need to be fully understood. This leads to problems of realistically simulating air showers for analyzing and interpreting the measurements. The basis of all models for predicting light mesons and muons produced in meson decays are measurements from particle accelerators. Since all models systematically underestimate the number of muons, the assumption is that the cause of the discrepancy between prediction and measurement is a fundamental problem in measuring the production processes of pions and kaons in collider experiments [13]. The difficulties arise because light mesons from collisions in particle accelerators tend to have a small transverse momentum. In general, it is more convenient to reconstruct those particles which strongly depart from the beam pipe. Also, the muons keep their small transverse momentum, so it is complicated to calculate the vertices of meson production and meson decay in the forward direction of the collider. One reasonable assumption is that the cause of the muon discrepancy lies in a false estimation of the π^0 fraction, the fraction of neutral pions produced in an air shower induced by cosmic rays. The neutral pions, in turn, generate electromagnetic subshowers when decaying into photons.

Even though the muon puzzle seems to be a problem at high energies so far: for MAGIC standard analyses simulations of hadron-induced showers have not been required in the past. The simulations were made especially for the proton analysis presented in this thesis. Verifying that data and simulations agree in the energy range relevant to the analysis is necessary.

5.3 Reweighting of Simulations

At multiple points of this analysis, the simulations must be weighted to represent the real measurement. This means the weighting, on the one hand, by the observation time of the measured data, and on the other hand, with a realistic assumption of the spectral index γ since it has been set to a generic index $\gamma = -2$ previously.

Balloon-based detectors and other experiments have already measured the cosmic-ray flux over a wide energy range. In the following, the true spectral index is estimated by collating published data from these experiments and performing a linear regression on the logarithm of the data points. Future reweighting of the simulated spectrum in the analysis will be based on the result of this regression. A complete list of experiments with an additional brief description is provided in appendix C.

5.3.1 Fit to Data from Previous Experiments

In the following, the assumption is made that the collected data of all experiments is statistically distributed around the true flux so that a fit to all the data approximates the true spectra of the particles.

First, the simulations must be tested to verify that they represent the measured data well. For this purpose, the simulations are reweighted to the respective assumed spectra and added up to form the total cosmic-ray spectrum as measured. Under the hypothesis that with protons, helium, and iron nuclei, a large part of the particle spectrum in the TeV energy range is covered, the simulations are compared with the measured data. Histograms of the most important features are generated for both datasets to exclude a data-simulation mismatch.

Second, the proton simulations are needed to calculate the instrument response functions (IRFs). The efficiency of the experiment depends on the energies of the primary particles, as explained in chapter 7. To calculate the energy-dependent efficiency using the simulations, the spectral index of the simulation distribution should be close to reality. The simulations are produced with a spectral index of $\gamma = -2$ and therefore need to be reweighted to calculate the IRFs. Third, the analysis requires an assumption of the true flux of the background particles. In the unfolding, chapter 8, the background is taken into account, and thus included. For this, the simulations of the helium and iron nuclei have to be reweighted so that the spectrum approximates the true spectrum as closely as possible.

The data required to determine the true flux is taken from the open source data of the particle data group [125] for all three particle types proton, helium and iron nuclei, measured with different detectors. The experiments contributing data to the proton spectrum are listed along with the period of data taking and the data energy range in figure 5.1. The lists of helium experiments and iron experiments are shown in figures 5.2 and 5.3. The newest data of the helium nuclei spectrum is extracted from the publication of the *DArk Matter Particle Explorer* (*DAMPE*) collaboration in Alemanno et al. [16]. More detailed information about the energy ranges are given in the tables A.2.1 to A.2.3 in appendix A.2.

The weighting factors for the simulations are obtained by fitting a function that describes the data well in the energy range relevant to this analysis. In this approach, power law functions describe the spectra, assuming they are smooth and continuous. However, since different experiments provide the data, some of which are in conflict, an alternative approach, such as interpolating the data, does not produce reliable results. Especially in lower energies, the fluxes vary because of different solar periods. The power law function that models the flux of cosmic rays in the GeV and TeV range has the spectral index γ and the flux normalization Φ_0.

$$\Phi = \Phi_0 \left(\frac{E}{\mathrm{GeV}} \right)^{\gamma} \tag{5.2}$$

The datasets contain statistical and, in some cases, systematic uncertainties. In order to take into account the uncertainties σ_f of the data points f in the fit, meaning to assign those with small uncertainties more influence on the result, the fit was performed with

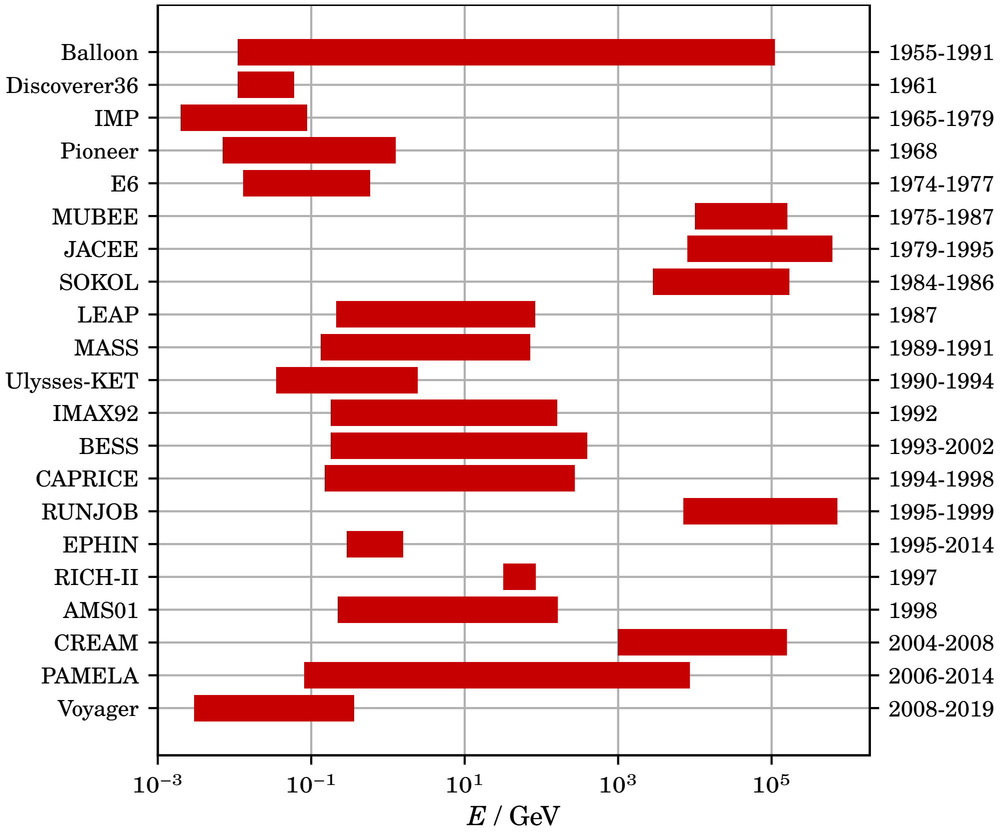

Figure 5.1: Measurements of cosmic protons with different experiments during the last decades. Each experiment is sensitive to specific energies, and they can cover a wide range together. All experiments in this list are balloon- or satellite-based. A full description of the experiments is available in appendix C.

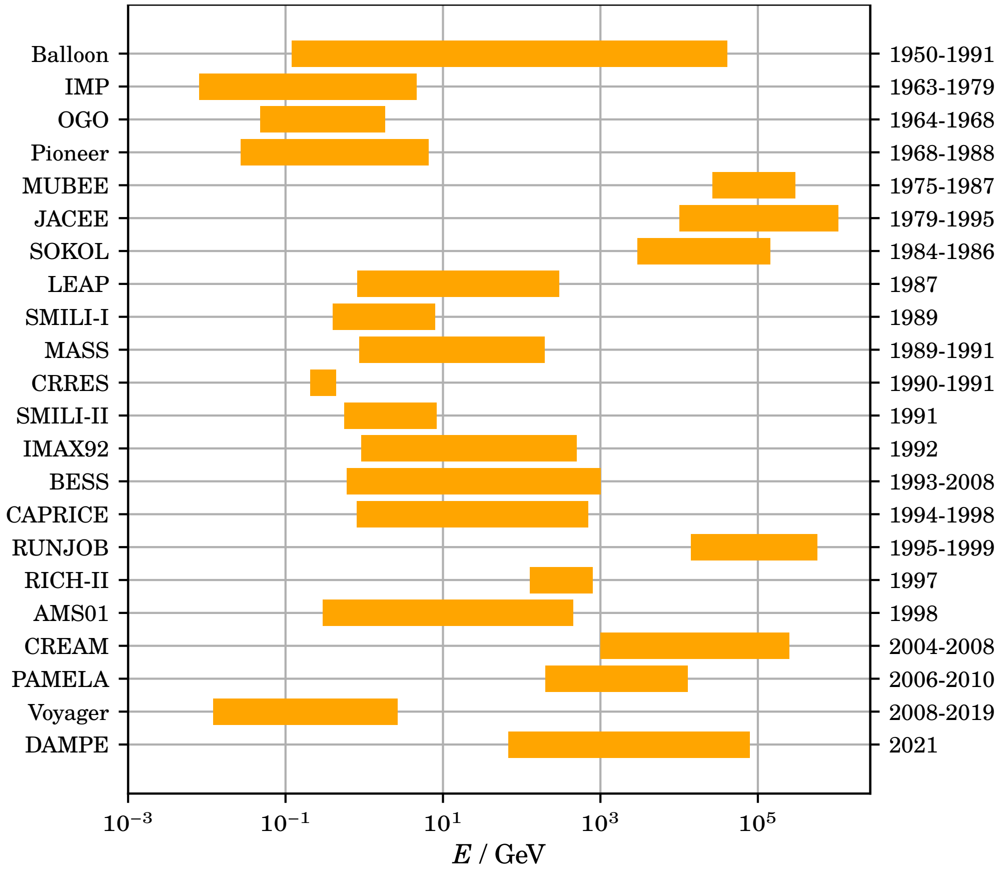

Figure 5.2: Measurements of cosmic helium nuclei with different experiments during the last decades. Each experiment is sensitive to specific energies, and they can cover a wide range together. All experiments in this list are balloon- or satellite-based. A full description of the experiments is available in appendix C.

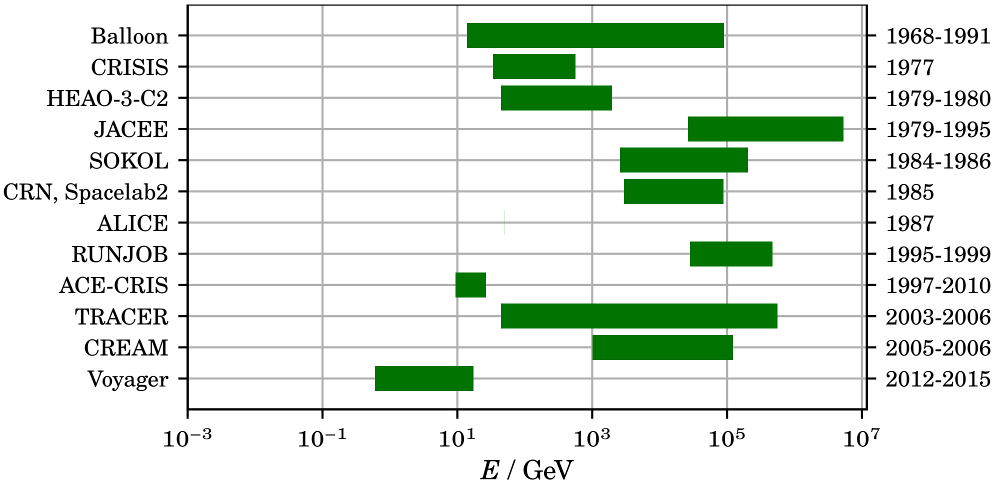

Figure 5.3: Measurements of cosmic iron nuclei with different experiments during the last decades. Each experiment is sensitive to specific energies, and they can cover a wide range together. All experiments in this list are balloon- or satellite-based. A full description of the experiments is available in appendix C.

an inverse-variance weighting with the weights w for all data points.

$$w = \frac{1}{\sigma_f^2} \tag{5.3}$$

The fit is above a threshold energy of 20 GeV per nucleon. For numerical reasons, the logarithm of the flux is used for the fit. Thus, the flux is described as

$$\log(\Phi) = \gamma \cdot \log\left(\frac{E}{\mathrm{GeV}}\right) + \log(\Phi_0) \tag{5.4}$$

and a linear regression $y = m{\cdot}x{+}b$ can be made. The data from the different experiments, especially the most recent *DAMPE* measurements of helium nuclei, and the fit results for the different particle types above the threshold energy were already shown in chapter 2 in figure 2.1. The common representation of cosmic spectra according to the Particle Data Group (PDG) is the particle flux versus energy per nucleon. In this visualization, the spectra are optically separated from each other so that their individual shape can be inspected. The particle spectra normalized to the number of nucleons are shown in figure 5.4. The plots of the data separated by particle types and with visual subdivision of the experiments can be found in appendix A.1 in figures A.1.1 to A.1.3. The results of the fit parameters are shown in table 5.1 below.

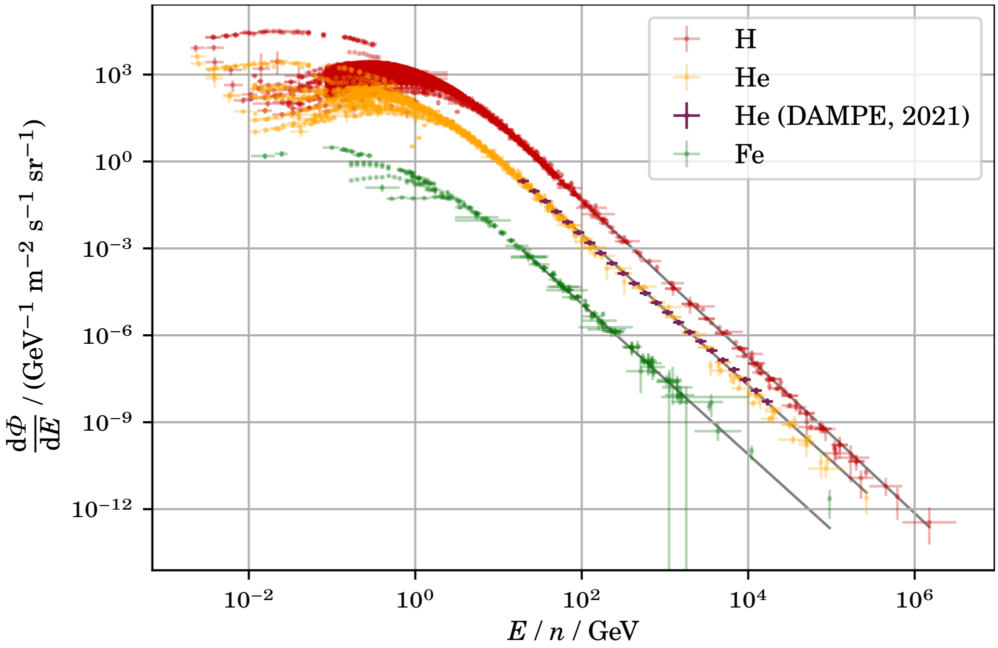

Figure 5.4: Cosmic-ray spectra of measurements with different experiments during the last decades: Data of proton, helium and iron nuclei fluxes per nucleon. A power law is fitted to the data above 20 GeV per nucleon. The experiments contributing to the spectra are shown in figure 5.1, 5.2 and 5.3. A full description of the experiments is available in appendix C.

5.3.2 Calculation of Weights

The simulations for the three particle types are generated with a spectral index of $\gamma = -2$ as described in section 4.2.2. They are weighted to the true cosmic-ray spectra for the data-simulation comparison, and with the observation time t_{eff} so that the simulation reflects the distributions of events that were measured during an observation of the effective observation time t_{eff} and can then be compared with the measured data.
The simulations are weighted with the results for the parameters of the different particle spectra from section 5.3.1 section using the `pyirf` package [104]. The weighting factors w_i for each event with the true energy E_i are calculated with the simulated spectra Φ_{sim} and the target spectra Φ_{target}, the respective true spectra fpr the particles.

$$w_i = \frac{\Phi_{\text{target}}(E_i)}{\Phi_{\text{sim}}(E_i)} \tag{5.5}$$

The target spectra $\Phi_{\text{target}}(E)$ are calculated with the fit results in table 5.1.

$$\Phi_{\text{target}}(E) = \Phi_0 \left(\frac{E}{\text{GeV}}\right)^{\gamma} \tag{5.6}$$

Table 5.1: Results of linear regression to the logarithm of the spectrum from equation (5.2) to the data in figure 2.1 (or figure 5.4) for protons, helium and iron nuclei. The energy is given per nucleon.

	E_{min} / GeV	f_0 / $(\mathrm{GeV\,s\,m^2\,sr})^{-1}$	γ
protons	20	10 909.2	−2.696
helium	80	4589.5	−2.605
iron	1120	1202.6	−2.598

The simulated flux Φ_{sim} is calculated from the settings of the Monte Carlo production: number of simulated events N_{sim} including possible shower reuse, limits of the simulated energy range E_{min} and E_{max}, the maximal simulated impact parameter r_{max}, the simulated spectral index γ_{sim} and, in case of a diffuse flux like for cosmic rays, the opening angle of the view cone Ω_{sim}. With the effective observation time t_{eff} of the observation data, the simulated events are weighted so that the distributions of data and simulations represent the same measurement.

$$\Phi_{0,\mathrm{sim}} = \frac{N_{\mathrm{sim}}\,(\gamma_{\mathrm{sim}}+1)\,\mathrm{GeV}^{\gamma_{\mathrm{sim}}}}{t_{\mathrm{eff}}\,\pi\,r_{\mathrm{max}}^2\,\Omega_{\mathrm{sim}}\left(E_{\mathrm{max}}^{\gamma_{\mathrm{sim}}+1} - E_{\mathrm{min}}^{\gamma_{\mathrm{sim}}+1}\right)} \tag{5.7}$$

5.4 Data-Simulation Comparison

As described in section 5.1, simulating data is a complex task. However, simulation tools have become more advanced over time, and more details have been taken into account, bringing the simulation closer to real measurements. Nevertheless, during the simulation, some assumptions and simplifications are made that cause the simulation dataset to differ from the measured dataset. For the analysis based on machine learning algorithms in chapter 6, it is important that these differences are not reflected in the attributes used in the analysis. If a random forest focuses on those attributes that perform poorly in the data-simulation comparison, applying the random forest to the measured data compromises the reconstruction and might result in bad performance. In order to verify that the simulations sufficiently reflect the dataset, a data-simulation comparison is essential.

The following section provides two distinct data-simulation comparisons: a visual comparison covering the entire parameter ranges in section 5.4.1 and a quantitative comparison that searches for those attributes with a larger mismatch in section 5.4.2.

5.4.1 Comparison with Weighted Simulations

Data and simulations must match the attributes that play a major role in the analysis. The `size`, the number of photoelectrons in the pixels assigned to the shower, correlates directly with the Cherenkov light of the shower and thus with the energy of the primary particles. The data-simulation comparison is shown in figure 5.5. The relative deviation is also shown there. With few statistics, mostly in the simulations, the relative deviation can become large. The difference in the last few bins of the size distribution can be explained by the fact that the Monte Carlo simulations are simulated only up to a certain energy. However, with several hundred hours of measurement time, particles with higher energies will likely be detected. The higher the energy of the particles, the larger the amount of Cherenkov light in the air shower and the resulting `size` parameter of the detected event. Accordingly, there is a cut-off in the `size` distribution of Monte Carlo simulations, which is not present in the `size` distribution of the measured dataset.

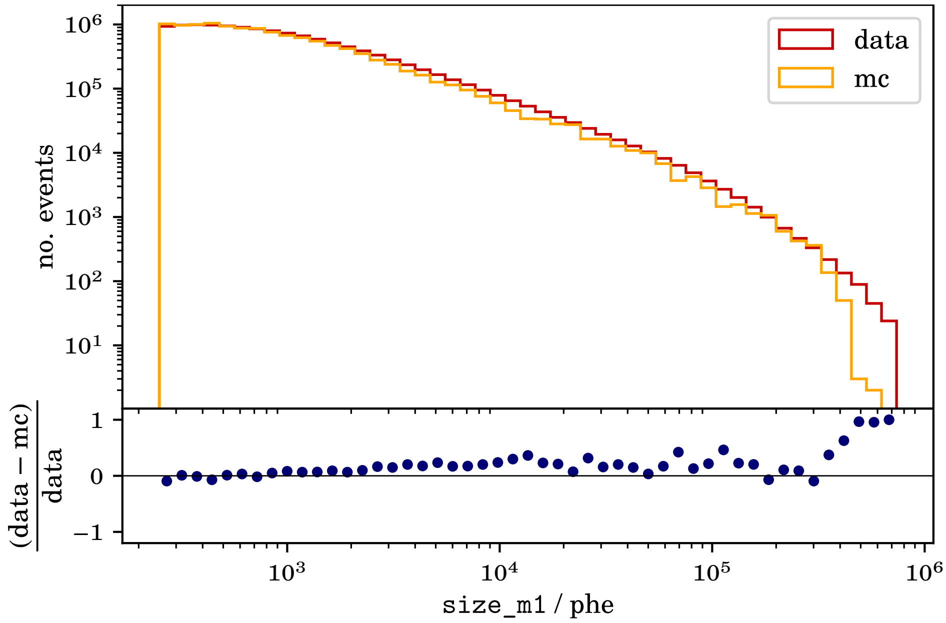

Figure 5.5: Data-simulation comparison for the parameter `size`. At very high values, a mismatch occurs because the simulations have an upper energy threshold resulting in a cut-off in high `size` values in contrast to the measured data.

The parameters `width` and `length` are essential for characterizing the shower images. The plots for the data-simulation comparison can be found in figure 5.6 and figure 5.7. They show good agreement over a large central range. There is a stronger relative deviation between data and simulations for the smallest and largest values, which lower statistics in these ranges can explain. The distributions of all features used in this work can be found in appendix A.4.1 to A.4.25.

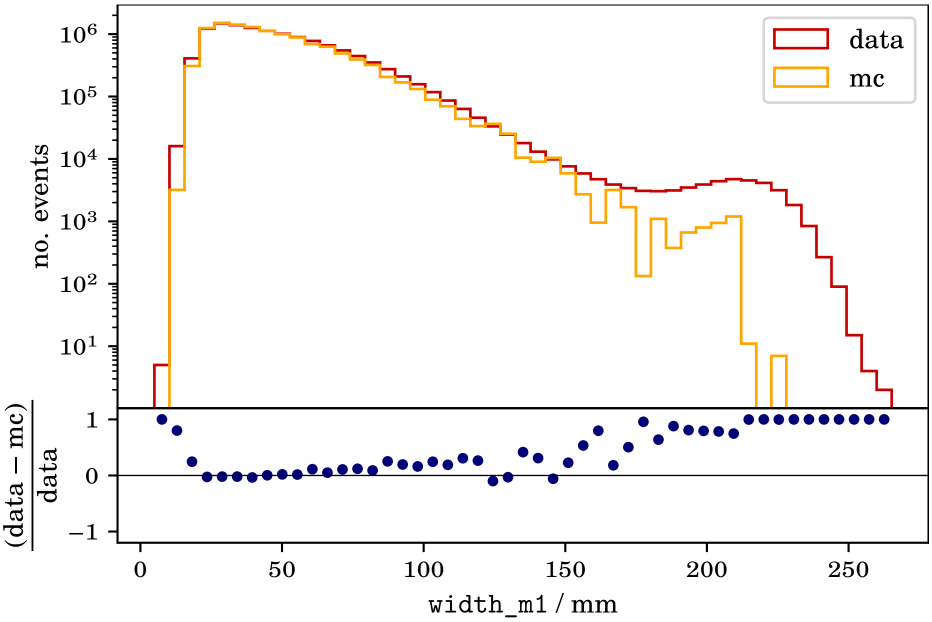

Figure 5.6: Data-simulation comparison for the parameter `width`. Bins with many entries match well, while those with fewer entries have statistics-related mismatches between data and simulations.

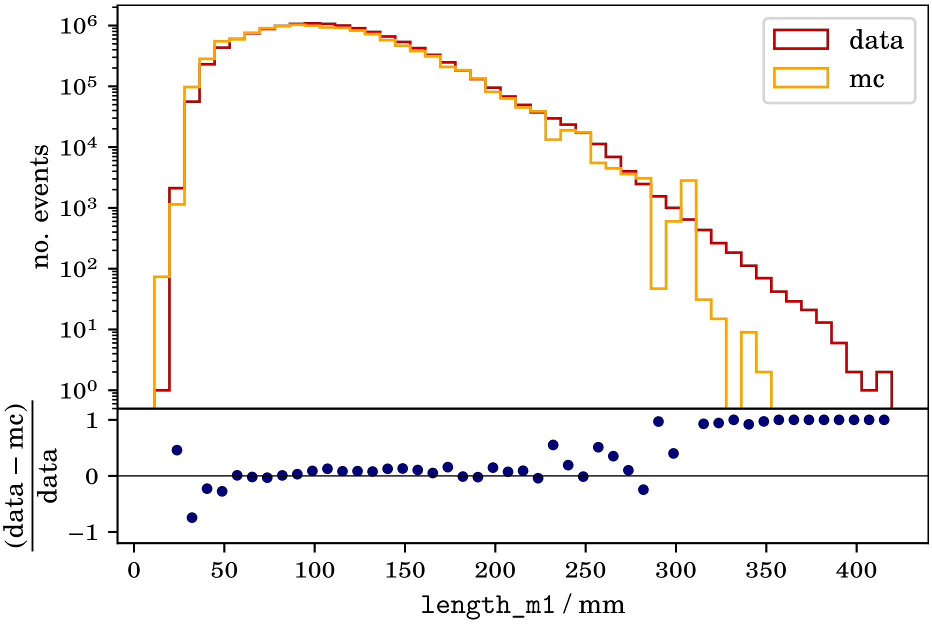

Figure 5.7: Data-simulation comparison for the parameter `length`. Bins with many entries match well, while those with fewer entries have statistics-related mismatches between data and simulations.

5.4.2 Separability Test with Machine Learning Algorithms

A method to quantify the quality of the simulations and to find the attributes in which the data-simulation mismatch is largest is to train a machine learning classifier. Machine learning algorithms are described in section 6.1. Here, a random forest is trained to separate the simulations (*signal*) from the measured data (*background*).

The performance of the random forest is used to determine the separability of the simulations from the data. If the simulations are close to reality, separating them from the data is difficult, and they are considered suitable for the main analysis. The validation of machine learning algorithms is further explained in chapter 6.2.

Additionally, the test of the feature importance reveals how well the attributes separate the datasets: those with high feature importance have a significant data-simulation mismatch of the data, whereas the simulations match well with the measurements in attributes with lower feature importance. The main analysis must consider these results by excluding the attributes with a large data-simulation mismatch. The random forest is trained with all simulations and the data used for the analysis. In figure 5.8, the Receiver Operating Characteristic (ROC) curve is shown. With an Area Under the Curve (AUC) of 8.1, the random forest performs moderately, which means it can partially

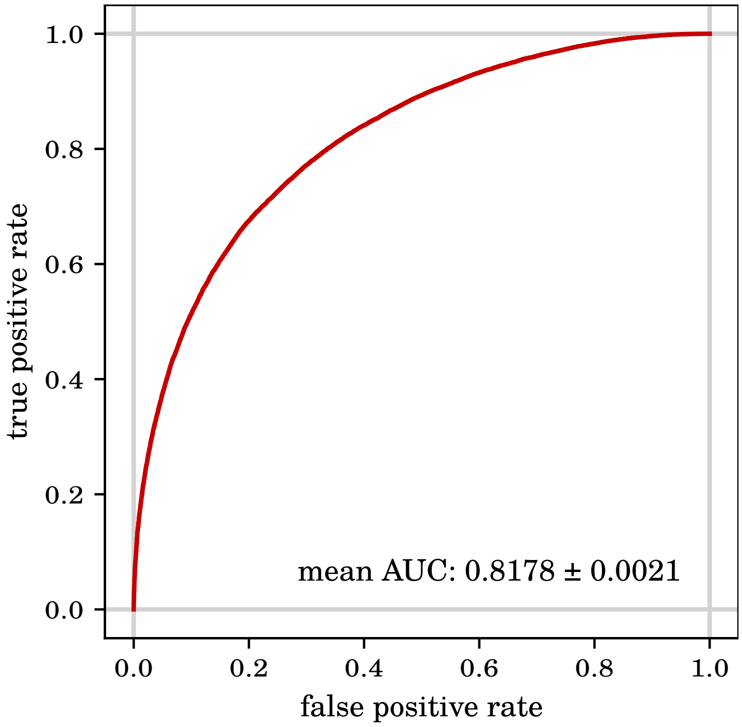

Figure 5.8: The ROC curve of the classifier to test the separability of data and simulations.

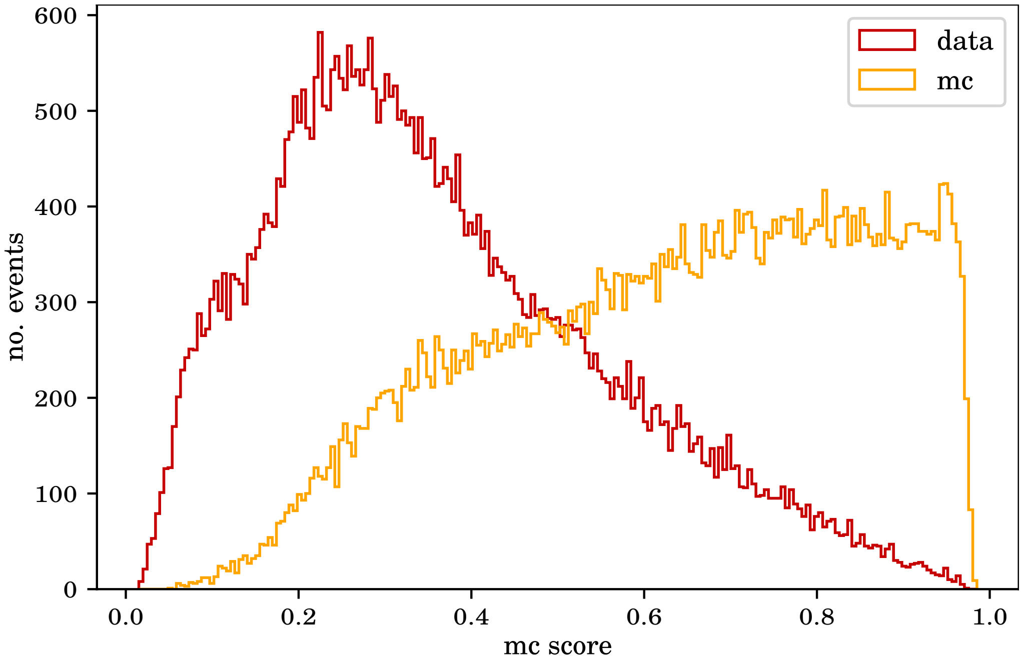

Figure 5.9: The classifier's score distributions of data and simulations to test the separability of data and simulations.

separate the data and simulations. Figure 5.9 shows the classification score for the data and the simulations. In figure 5.10, both precision and recall are given for different cuts. It is not possible to separate the measurement data from the simulations by cuts in the estimator and obtain pure datasets. Of particular interest is the feature importance of the individual attributes. These provide information on the extent to which the attributes have contributed to the separation of the datasets. Figure 5.11 shows the feature importance for the individual attributes. First of all, none of the attributes stands out against the other attributes. This means that the simulations do not differ significantly from the measurements in any attribute. The four most important features for separating data and simulations are the spectral `concentration_m1`, the Root Mean Square (RMS) of the arrival times in the individual camera pixels `rms_time_m1`, the Cherenkov radius `cherenkov_radius` and the `cos_bs_angle`. In the main analysis, attention is paid accordingly to these attributes and their influence on the results.

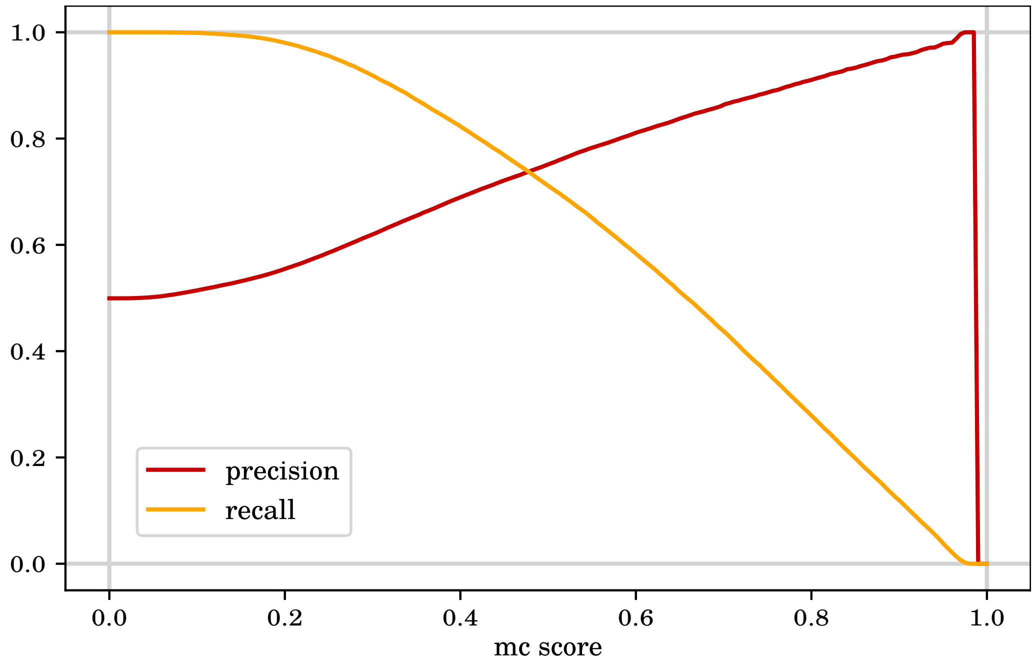

Figure 5.10: Precision and recall of the classifier to test the separability of data and simulations.

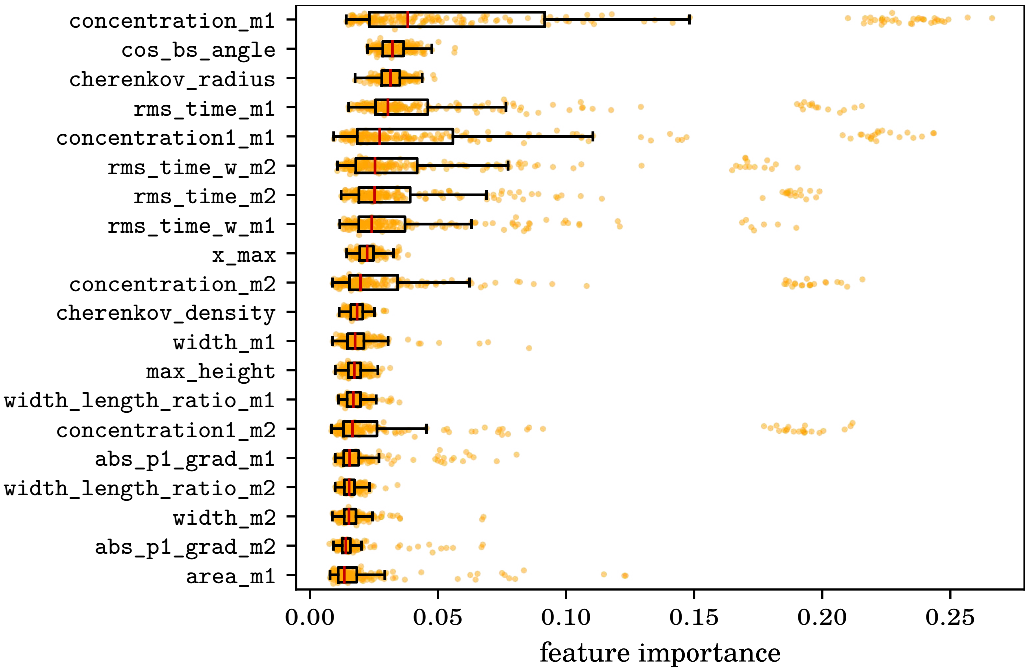

Figure 5.11: The 20 most important features for separating data and simulations. No attribute stands out in particular, which means that no feature has excessive data-simulation mismatch compared to the other features.

Reconstruction with Machine Learning Algorithms

6

The next part of the analysis chain is the extraction of physical parameters from the attributes of the shower images. In this analysis, the focus is on identifying the type of primary particles and their energy. The analysis distinguishes between protons as *signal* and helium and iron nuclei as *background*. The energy of the protons is to be estimated. Machine learning algorithms are explained in this chapter, and then models are trained to estimate the physical properties of the primary particles.

Since the MAGIC telescopes trigger events at a rate around 250 Hz to 300 Hz [15], a hundred million images are evaluated in this analysis of 164 observation hours. The challenge is to analyze enormous amounts of data to reconstruct the properties of the primary particles from the attributes of the camera images. Machine learning algorithms are the appropriate tools to process huge amounts of data.

6.1 Method: Data Reconstruction with `aict-tools`

In astroparticle physics, machine learning has established itself as the most reliable tool for analyzing large datasets in recent decades. This analysis uses the software `aict-tools` [103], developed in Dortmund. The tool is written in the programming language `python` and mainly integrates the package `scikit-learn` [108] for the machine learning algorithms.

Analysts had to refer to manual methods before machine learning algorithms took over the task of predicting particle properties on large datasets. It was common to study the shower images and determine typical patterns to gain information about the properties of the primary particles. Then, cuts were made in the parameter space to divide the datasets into their shower properties and assign certain primary particle predictions. With the establishment of machine learning algorithms, the performance of the estimation has increased, and human bias has been reduced.

Nowadays, there are various methods of machine learning. Random forests are used in this analysis, a classic machine-learning approach for event-based analyses. It is a supervised learning method, meaning that the algorithm is trained on labeled data:

data of which the true values for the target parameters are known. The method is robust and well-tested, moreover very well applicable to IACT data. In this analysis, a random forest is used for the first time to separate proton events from heavier nuclei for the MAGIC experiment.

6.1.1 Binary Decision Trees

A random forest consists of several decision trees. A decision tree has the task of estimating what a certain property is, for example, the energy or the type of the primary particle, based on the event's features, the attributes. For training a random forest, labeled data is needed. This differs from the observation data in that the primary particle's true properties, the target parameters, are already known. These datasets are simulated as explained in section 5.1.

Given a labeled dataset, it can be split based on a feature, with the premise being to split that dataset as best as possible into two subsets. The validation of cuts is explained in section 6.2 below. Cuts on other features further split these subsets of data, see figure 6.1. The structure resembles the branching of a tree, hence its name decision tree. The cutting points in the decision trees are called nodes. At each node, the feature that best splits the dataset is taken from a randomly chosen subset of features. This ensures that a second decision tree makes different decisions given the same training dataset. If a dataset is not subdivided further, it ends up in a node called a leaf. Each leaf contains an almost pure subset, mainly signal or background.

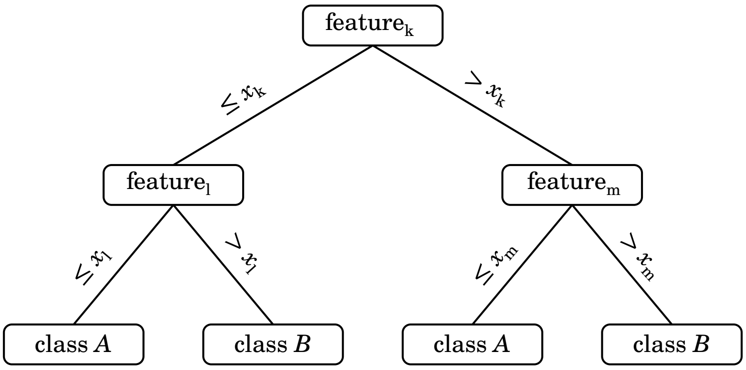

Figure 6.1: Illustration of a decision tree with three features and four leaves. In reality, decision trees have more layers and more leaves.

6.1.2 Random Forests

A single decision tree may be able to reconstruct the training dataset as best as possible, but the reconstruction performance on another dataset might suffer: this problem is

called overtraining. An entire set of decision trees, a random forest, is generated to prevent overtraining. Even if trained with the same datasets, these decision trees differ because each is trained on a randomly chosen feature subset.
The prediction of a random forest classifier, see section 6.1.3 is a score between 0 and 1: depending on the fractions of decision trees assigning an event to class A and to class B. The analyst decides the score threshold for the final classes. In the case of random forest regressors, see section 6.1.4, the estimates of the target parameters are also averaged from the results of the individual decision trees. Here, the random forest gives a concrete value, as in this analysis, the estimated energy.

6.1.3 Classifiers

If the random forest separates the events into discrete classes, it is called a classifier. This analysis trains classifiers to estimate the particle type with the target classes proton, helium, and iron nuclei. The class most represented in the leaf during training determines the class label.
The splitting performance of the classifier node is calculated using the Gini-Index [117]. The Gini-Index indicates which feature is best suited for separation and which split gives the purest subsets. The Gini-Index I_{Gini} is defined by

$$I_{\text{Gini}} = 1 - \sum_{i=1}^{J} p_i^2 \tag{6.1}$$

with the probability p_i that the event is correctly assigned to class i out of $J = 2$ classes. Training the classifiers requires a simulated dataset that includes proton, helium, and iron nuclei. The classification takes place in two steps: a first classifier is trained to separate the heavy nuclei (iron nuclei) from the light nuclei (proton and helium nuclei). A second classifier is trained to separate the lighter nuclei: helium from proton events. This random forest is trained using only helium and proton simulations since the assumption is made that the estimator reconstructs the remaining iron nuclei more likely as helium than as proton due to their nuclear properties discussed in section 6.3.

6.1.4 Regressors

A regressor estimates continuous values from a given range instead of classes. In this analysis, the prediction of the proton energy is such a regression problem. The random forest is trained on a dataset with proton simulations with known energy values.
The estimated energy in the leaf of the decision tree is averaged from the energies of the training data in that leaf. The best splits in the nodes are determined with the minimization of the loss function. The criterion in this work is the minimization of the

mean squared error I_{MSE} as

$$I_{\mathrm{MSE}} = \sum_{x_j \in X} (y_j - \bar{y})^2 \tag{6.2}$$

which reduces the variance for all features x_j in the subsample X of a node to find the feature with the best split of the data into subsets.

The goal of the analysis is to obtain a proton spectrum, so the focus is on estimating the proton energy. Whether the energy of the helium and iron nuclei, erroneously reconstructed as protons, is correctly estimated is secondary. Therefore, the energy regressor is trained only with proton simulations and applied to all data identified as proton by the classifier in the previous step. The additional analysis step, the unfolding in chapter 8, considers that the proton set is not pure after classification and corrects this error.

6.2 Method: Validation of Random Forests

Essential for the evaluation of the trained models is a validation of the random forests. As mentioned in section 6.1.2, the result of a classifier, the score, is between 0 and 1, depending on whether the event is more likely signal or background. Therefore, the analyst decides which score threshold to classify an event as signal or background. Depending on the analysis and the task, a pure signal dataset is required, or it is essential to obtain many events in the final dataset for statistical reasons. Various criteria can determine the quality of the separation. The most common quality criteria for binary classifiers are purity, efficiency, and accuracy. Figure 6.2 shows a scheme of binary classification. The dataset for validation consists of simulations. Besides the reconstructed parameters, the truths of the parameters are also known. The cut divides the dataset into positive events, classified as signal events, and negative events, classified as background. Correctly classified events are connoted true positive and true negative. If a signal event is incorrectly sorted into the background dataset due to a low score, it is called a false negative; if the score of a background event is higher than the cut, it is a false positive.

Various metrics are commonly used to optimize the classification threshold. They are calculated with the amount of true and false negatives or positives depending on the possible threshold and quantify the quality of datasets obtaining by the different thresholds.

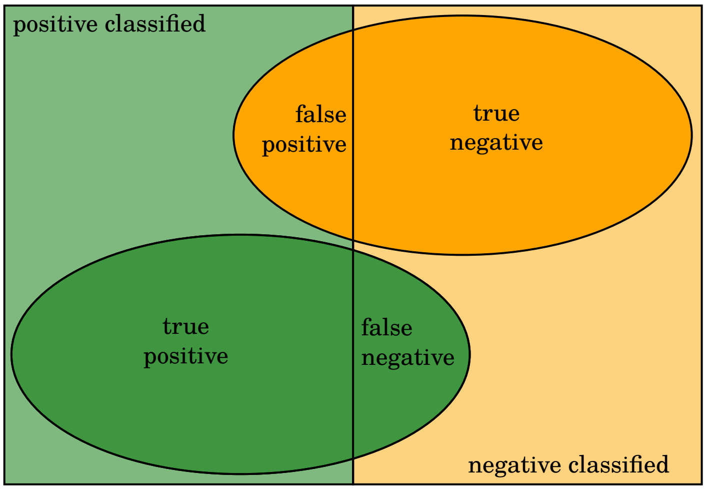

Figure 6.2: Scheme of the cut dividing the dataset into positive (left) and negative (right) classified events. The green circle includes all signal events classified correctly (true) and incorrectly (false). The yellow circle represents the background events, also correctly (true) and incorrectly (false) classified here.

Purity is the fraction of correctly classified signal events out of all positive events. Also known as **precision** or **positive predictive value**. This metric is preferred when the signal dataset must be as pure as possible for further analysis.

$$\text{purity} = \frac{\boxed{\boxed{\text{TP}}}}{\boxed{\boxed{\text{TP}}} + \boxed{\boxed{\text{FP}}}} \tag{6.3}$$

True Positive Rate (TPR) is the fraction of correctly classified signal events out of all signal events. Also known as **efficiency**, **sensitivity**, or **recall**.

$$\text{TPR} = \frac{\boxed{\boxed{\text{TP}}}}{\boxed{\boxed{\text{TP}}} + \boxed{\boxed{\text{FN}}}} \tag{6.4}$$

False Positive Rate (FPR) is the fraction of background events incorrectly classified as signal events out of all background events.

$$\text{FPR} = \frac{\boxed{\boxed{\text{FP}}}}{\boxed{\boxed{\text{FP}}} + \boxed{\boxed{\text{TN}}}} \tag{6.5}$$

Accuracy is the fraction of correctly classified events out of all events.

$$\text{accuracy} = \frac{\boxed{\mathrm{TP}} + \boxed{\mathrm{TN}}}{\boxed{\mathrm{TP}} + \boxed{\mathrm{FN}} + \boxed{\mathrm{FP}} + \boxed{\mathrm{TN}}} \tag{6.6}$$

For the validation with purity and accuracy, it is important that the signal and background datasets are about the same size or that they mirror the real signal background ratio.

6.2.1 ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve is commonly used to visualize the separating power of a classifier. It shows the TPR against the FPR at different score cuts. The TPR is the efficiency in equation (6.4), meaning the fraction of all signal events that were correctly identified. The FPR in equation (6.5) is the fraction of all background events that were incorrectly classified as signal events. By varying the score cuts to cover efficiencies from zero to one and plotting the efficiencies against the corresponding FPR, the ROC curve is obtained.
A typical ROC curve is shown in figure 6.3. The Area Under the Curve (AUC) can reach a value between 0.5 and 1 and gives information about the separating power of the classifier: A perfect classifier separates two populations perfectly and the AUC is one. A classifier that cannot find a way to distinguish two populations, randomly guesses the classes and the result in the ROC curve is an AUC of 0.5.

6.2.2 Feature Importance

Feature importance is a standard metric for validating random forests: it indicates the relevance of a given feature for dividing datasets into distinct subclasses. It is based on the metrics from the sections 6.1.3 and 6.1.4, the Gini-Index I_{Gini} in equation (6.1) for classifiers, and the mean squared error I_{MSE} in equation (6.2) for regressors. The metric is weighted by the position of the node n in the tree, the factor $\Delta(x_n)$, since large amounts of data are split at the first nodes, giving the best separation. In contrast, only small amounts of data are split at the last nodes, giving little importance to the overall separation. The following equation from Louppe [84] shows the calculation of

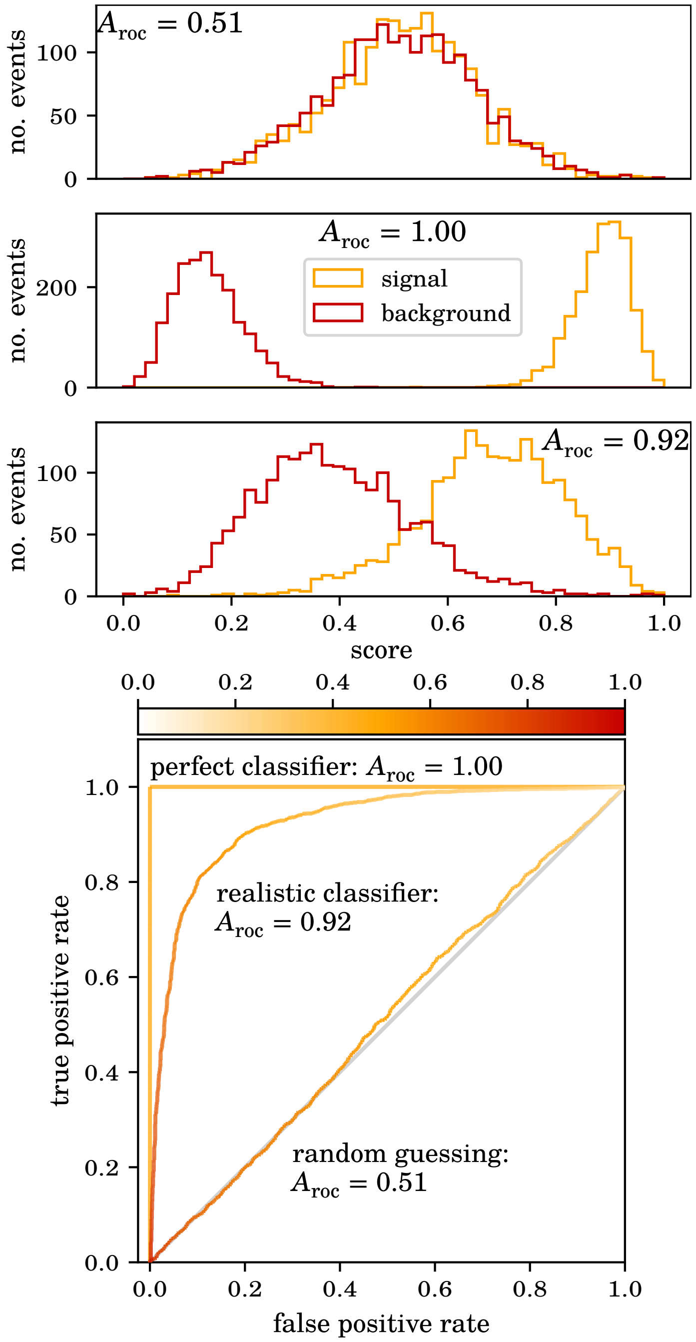

Figure 6.3: Distributions of a classifier with little separation strength ($A_{roc} = 0.51$, top), perfect separation ($A_{roc} = 1.00$, center) and a realistic classifier ($A_{roc} = 0.92$, bottom). The ROC curves of the distributions are shown below. The gray line shows the ROC curve for a classifier for which the two populations are indistinguishable.

the feature importance f of a feature X_i in a random forest.

$$f(X_i) = \frac{1}{M}\sum_{m=1}^{M} f(X_i, m) \tag{6.7}$$

$$= \frac{1}{M}\sum_{m=1}^{M}\sum_{\mathrm{n}\in N_m} \mathbb{1}(i_n = i)\,[p(n)\Delta x_n] \tag{6.8}$$

If the feature X_i in the node n from all nodes N of the decision tree m contributes to the separation ($i_n = i$), the validation value $p(n)$ is weighted by the factor Δx_n. It is summed over all nodes in which the feature X_i is used in a decision tree. This obtained value $f(X_i, m)$ is averaged over all decision trees M in the random forest. The feature importance calculated this way is also called Mean Decrease Impurity (MDI) importance.

6.2.3 Bias and Resolution

The energy dispersion is the deviation between true and estimated energies. The dispersion of the reconstruction is quantified in this analysis by two metrics: the bias and the resolution. Both values are based on the relative error δE of the events.

$$\delta E = \frac{E_{\mathrm{est}} - E_{\mathrm{true}}}{E_{\mathrm{true}}} \tag{6.9}$$

The bias is defined as the median of the relative error.

$$\mathrm{bias} = \mathrm{median}(\delta E) \tag{6.10}$$

The resolution is half the distance between the upper and lower 1σ-quantile of the normal distribution of the events, meaning the quantile containing 68.2 % of the events.

$$\mathrm{resolution} = \frac{Q_{84.1}(\delta E) - Q_{15.9}(\delta E)}{2} \tag{6.11}$$

The bias in the different energy bins is corrected by unfolding in chapter 8. The energy resolution gives the uncertainty of the results of a random forest regressor.

6.2.4 Cross-Validation

Cross-validation is a common method to evaluate how reliably the trained models can be applied to an independent dataset. In a k-fold cross-validation, the training dataset is divided into k subsets. While training the random forest with $k-1$ subsets, the remaining subset is used to validate the model. This means that the model is applied to the validation subset and evaluated with corresponding metrics, typically precision in equation (6.3) and recall in equation (6.4) for the classifiers, and bias in equation (6.10) and resolution in equation (6.11) for the regressors. In a subsequent iteration, another subset is retained for validation, and the random forest is trained on the remaining $k-1$ subsets. The training and evaluation procedure is repeated k times so that all subsets have been used once for the validation. Figure 6.4 schematically shows 4-fold cross-validation. The k-fold cross-validation produces k estimators for the performance of the model and hence robust error estimates. The advantage of this method is that the datasets can be used for training and validation. Thus, no separate validation dataset must be simulated beforehand to obtain the error estimators. Nevertheless, the k-fold cross-validation has to be performed k times, which increases the computation time for this analysis step by a factor k, and enough simulations are necessary to obtain enough statistics in each sub-dataset.

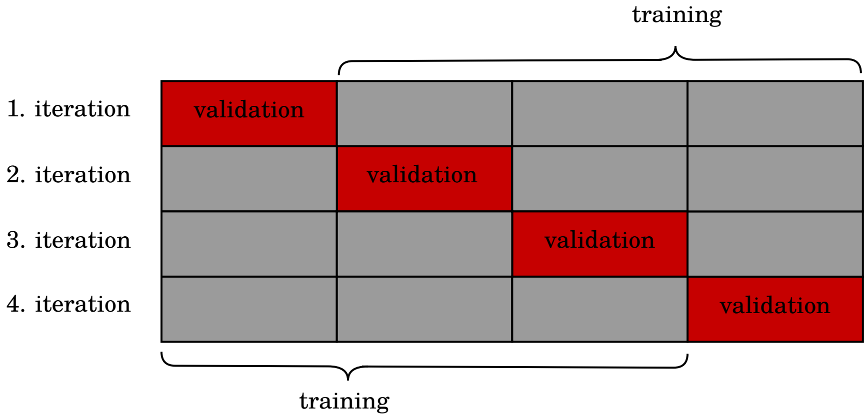

Figure 6.4: Schematic representation of a 4-fold cross-validation. Each iteration uses a different one of the four subsets to validate the model, while the remaining three subsets are used to train the random forest.

6.3 Classification of Cosmic Rays

In the classification, shower images produced by primary protons must be separated from the ones of heavy particles, namely helium and iron nuclei. This analysis is done with a two-step classification. In the first step, a random forest is trained, fed with iron simulations as *signal* and helium nuclei and protons as *background*. It learns to separate iron particles from the rest because iron nuclei consist of 56 nucleons ($A = 56$) and, thus, of significantly more nucleons than protons with a mass number $A = 1$ and helium nuclei with a mass number $A = 4$. The shower images produced by iron nuclei interacting in Earth's atmosphere are predicted to significantly differ from the images produced by the light primary particles, protons and helium nuclei.

In the second step, a helium classifier is trained. Here, helium simulations are used as *signal*, and proton simulations as *background*. This classifier learns to distinguish between proton and helium: two light particle types that are theoretically challenging to separate. In addition to protons and helium events, iron events can be misclassified by the first classifier. The assumption is that the iron classifier can identify many iron nuclei, and those incorrectly classified as light nuclei more closely resemble helium nuclei than protons in their properties. Thus, there is a high chance that surviving iron nuclei are selected from the helium classifier. Section 6.5.1 will discuss whether an iron classifier is necessary for this analysis or whether a helium classifier is sufficient to separate iron from the sample.
For the training of the random forests, 40 % of the simulated data is used. The iron and helium classifiers are trained with the same set of proton and helium simulations. The iron classifier additionally makes use of the set of iron simulations.

6.3.1 Iron Classifier

As mentioned before, for the training of the first random forest, iron simulations are used as *signal*, and protons and helium nuclei as *background* events. The random forest, which classifies iron and separates it from protons and helium nuclei, is evaluated with cross-validation. Figure 6.5 displays the ROC curve for the iron classifier. The AUC of

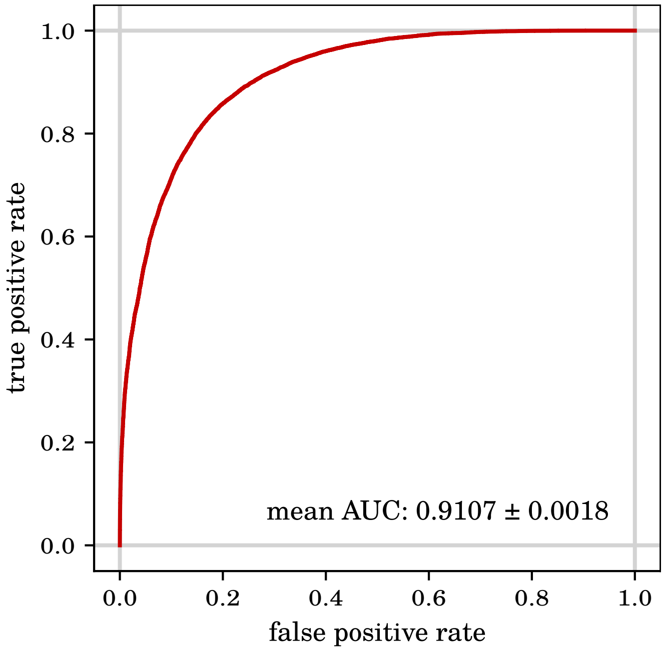

Figure 6.5: ROC curve of the iron classifier. The mean AUC is calculated from the cross-validation.

0.91 shows that iron can be separated well from the rest. Figure 6.6 shows the score of the respective iron and *background* events. This figure confirms the conclusion of the ROC curve: the iron dataset tends to a score towards zero, while the *background* dataset tends to a score towards one. Hence the datasets are well separated based on the score. Figure 6.7 shows precision and recall as a function of the score cut. Here, relatively high values can be obtained with different cuts, which also indicates the good quality of the classifier. Figure 6.8 shows the feature importance of the individual attributes: here, it can be seen that such attributes, which describe the shower geometry, contributed most to the separation of the events. The `width_length_ratio` and the `width` seem particularly characteristic features for the iron showers. Also, the arrival time dispersion in the shower image characterized in `rms_time` and `rms_time_w` is an important attribute to separate the samples.

The comparison of the results with the list of the 20 most important features for the separation of data and simulations from figure 5.11 rules out that the iron classifier favors such attributes with a larger data-simulation mismatch.

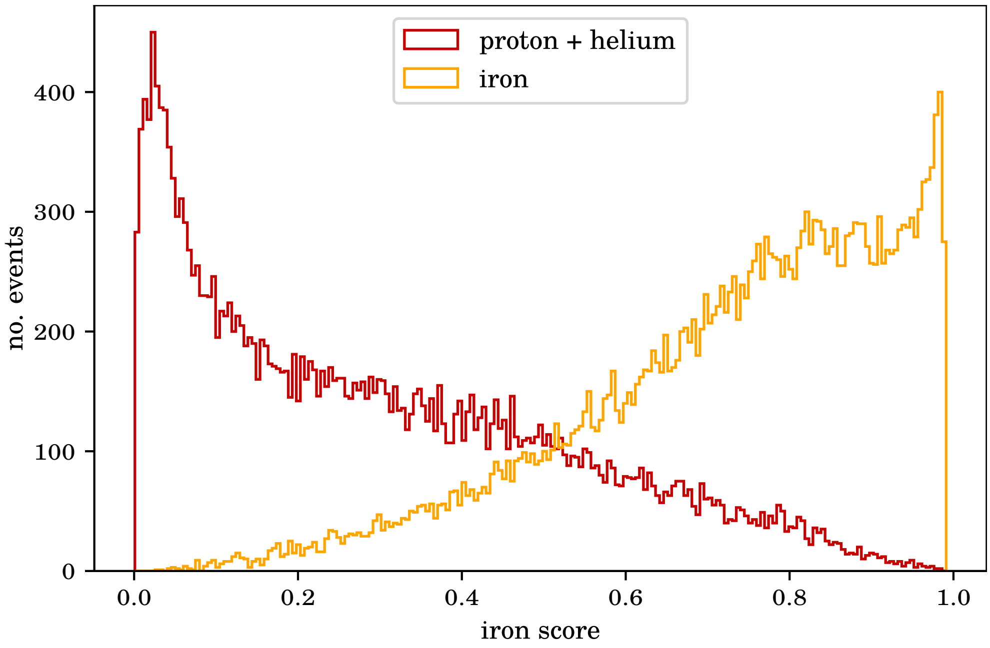

Figure 6.6: The score distributions of the iron nuclei (yellow) and the lighter particles, protons, and helium nuclei (red) estimated with the iron classifier. The distributions can be efficiently separated by a well-chosen cut in the score.

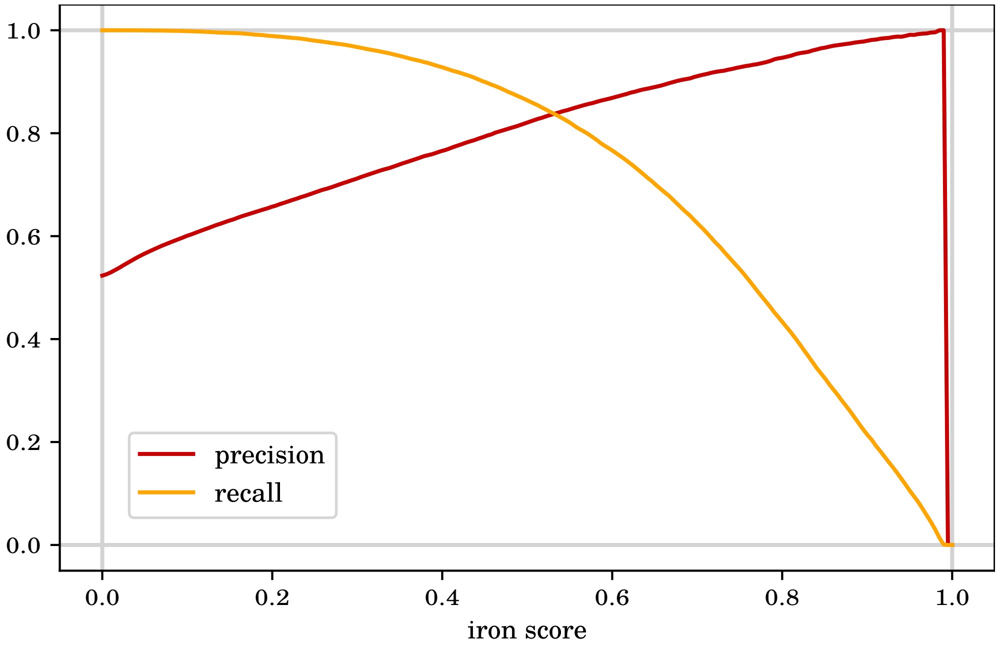

Figure 6.7: Precision and recall of the iron classifier calculated from the separated simulation data at varying score cuts. Depending on the requirements of the analysis, high precision or high recall may be desired The iron score cuts are discussed in section 6.5. High values in precision and recall indicate the good quality of the classifier.

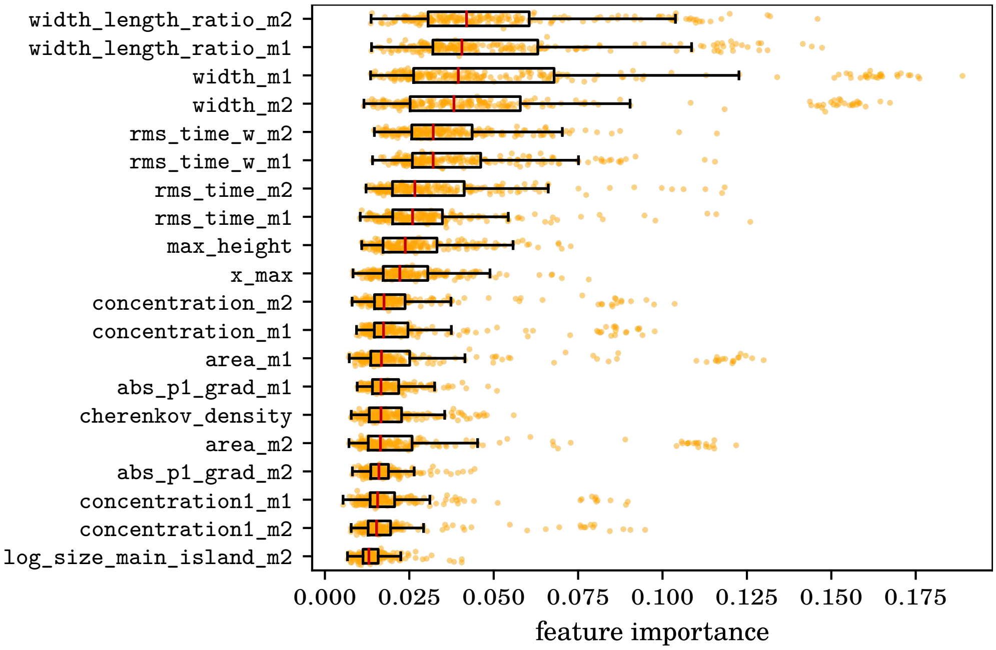

Figure 6.8: The 20 most important features of the iron classifier. The most important features describe the topology of the shower (`width`), the arrival time distributions (`rms_time`), and the height of the shower (`max_height` and `x_max`).

6.3.2 Helium Classifier

For the helium classifier, protons and helium simulations are used: the random forest is trained on helium events as *signal* and protons as *background*. The results of the cross-validation are shown in the following. Already in the ROC curve in figure 6.9, an AUC of 0.72 indicates that the separation of helium and protons is much more difficult than the separation of iron from the lighter particles. This result is also reflected in figure 6.10: the helium and proton distributions overlap strongly in the score, so a cut does not result in a clear separation of the two samples. Figure 6.11 shows precision and recall and again demonstrates that different cuts in the score do not produce particularly strong separation results. Figure 6.12 shows the feature importance of the various attributes; it can be seen that no attribute contributes particularly strongly to the separation of the data sts. The attributes contributing most to the separation of helium nuclei and protons are the geometric attributes `width` and `length`. Additionally, the attribute `x_max` plays a larger role than in the iron classifier.

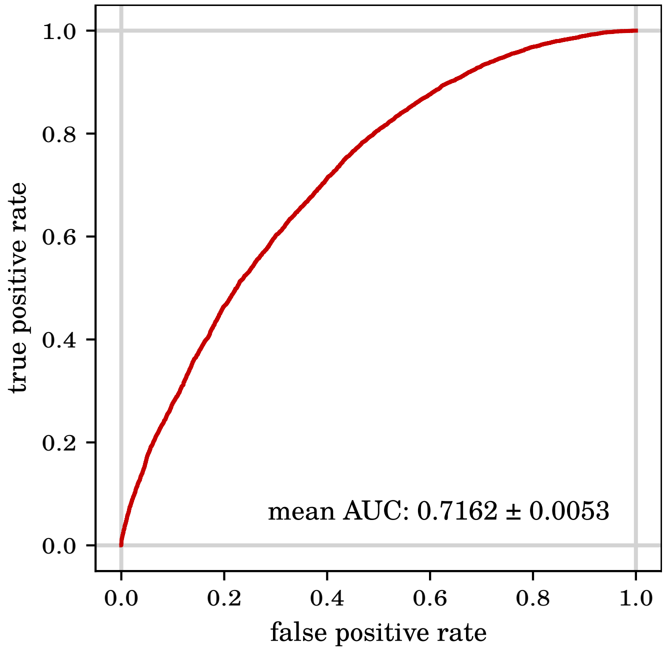

Figure 6.9: ROC curve of the helium classifier. The mean AUC is calculated from the cross-validation.

The comparison of the results with the list of the 20 most important features for the separation of data and simulations from figure 5.11 shows that the helium classifier does not favor such attributes in which the simulations poorly represent the measured data.

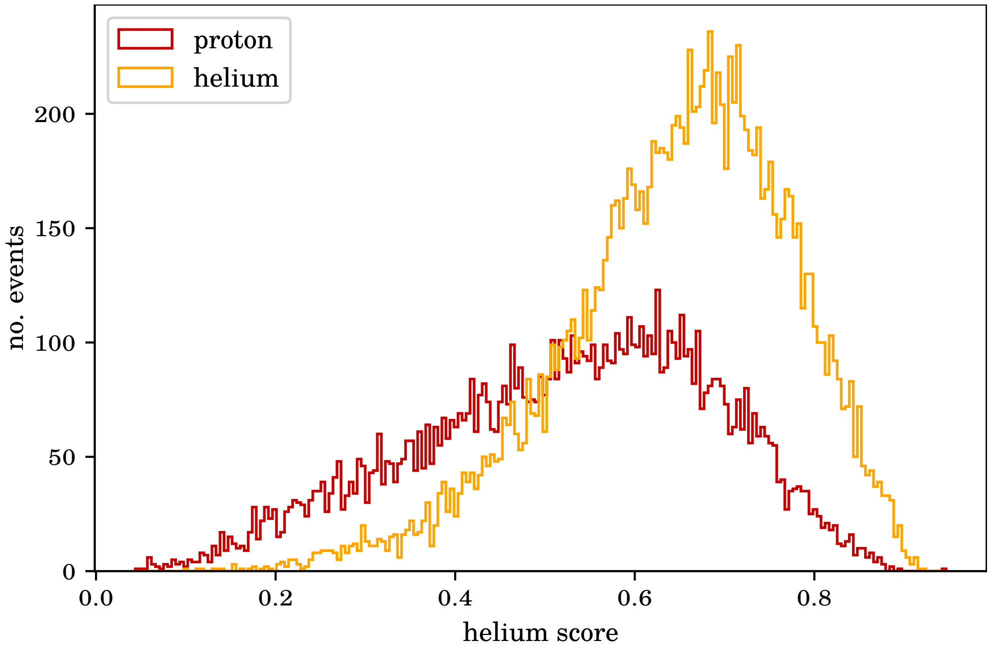

Figure 6.10: The score distributions of helium nuclei (yellow) and protons (red) estimated with the helium classifier. The distributions overlap, so the score cut must be well-chosen.

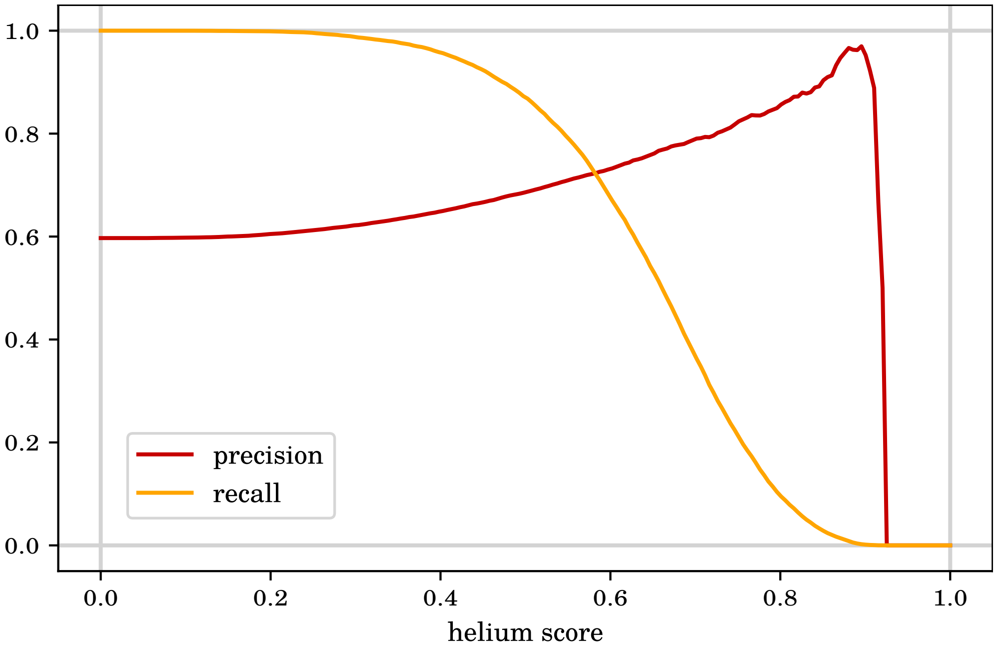

Figure 6.11: Precision and recall of the helium classifier calculated from the separated simulation data at varying score cuts. Depending on the requirements of the analysis, high precision or high recall may be desired. The helium score cuts are discussed in section 6.5.

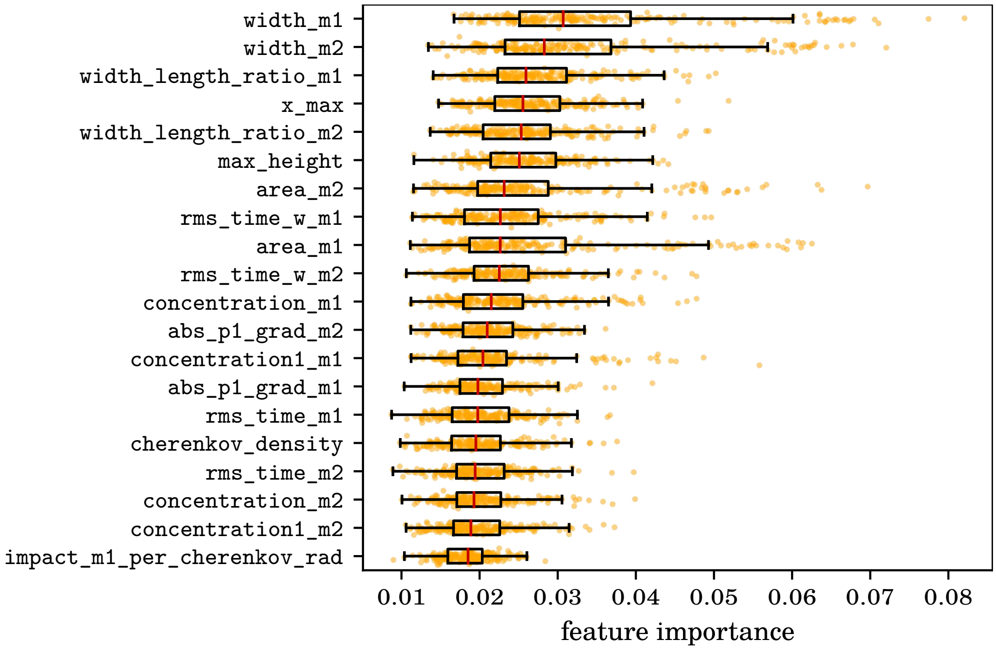

Figure 6.12: The 20 most important features of the helium classifier. The most important features describe the topology of the shower (`width`), the height of the shower (`max_height` and `x_max`), and the arrival time distributions in the images (`rms_time`).

6.4 Energy Reconstruction

For the energy reconstruction, a random forest is trained with the proton sample, which has already been used for training the random forests for particle identification and contains 40 % of all simulated proton events, the same as used for the training of the classifier in section 6.3. The priority of the analysis is a sufficiently well-estimated proton energy; whether the energy of the helium and iron nuclei performs well is secondary since it is used only secondarily in the analysis. The random forest is applied to all events, both the protons and the helium and iron simulations.

The energy dispersion for the protons is shown in figure 6.13. The limited energy resolution of the detector leads to a dispersion of the true primary's energy which the here-shown migration matrix can express. Further descriptions of the energy dispersion are given in section 7.1.2. At low energies, the energies tend to be overestimated, while at high energies, they tend to be underestimated. This underestimation at high energies and overestimation at low energies is a typical characteristic of machine learning models. The models can only predict values in the range for which they have been trained: at the boundaries, the predictions smear inward, while in the middle of the range, they can smear in either direction. These systematic deviations can be further quantified and eliminated by unfolding described in chapter 8. Figure 6.14 shows the bias and

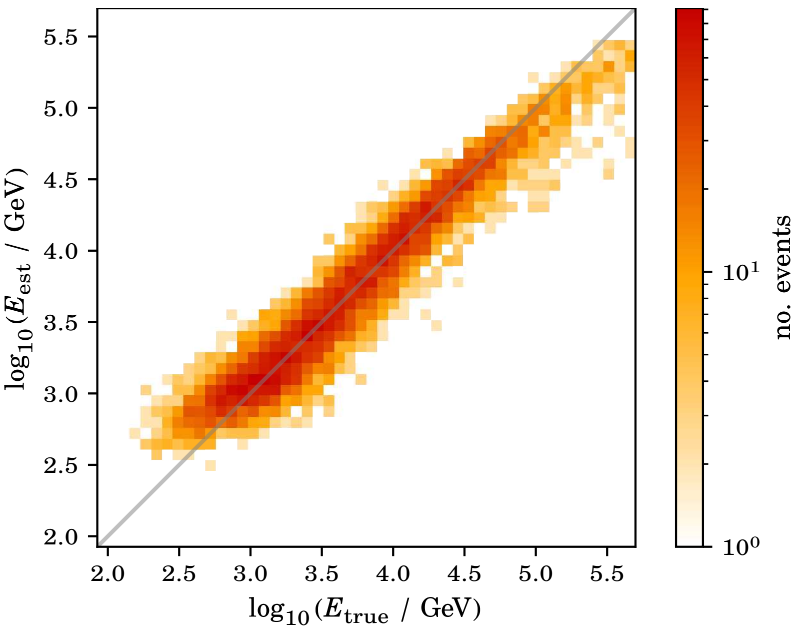

Figure 6.13: Migration from true primary energy of protons to reconstructed energy. The characteristic of the regressor is the overestimation at low energies and the underestimation at higher ones.

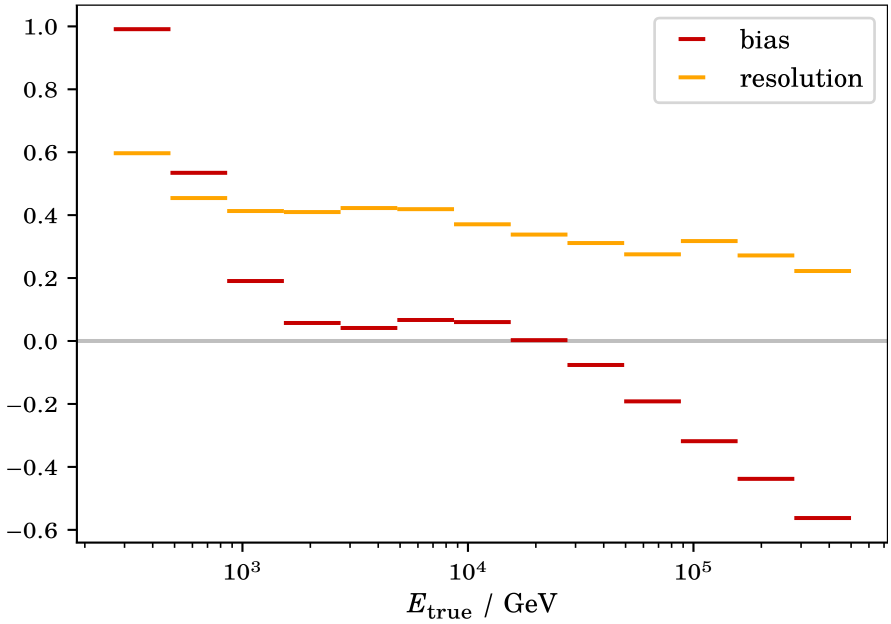

Figure 6.14: Bias and resolution of the proton energy regressor. The bias shows the random forest overestimating low energies and underestimating high energies. The resolution is constant with an improving tendency towards increasing energy.

the resolution of the energy estimator. Also, the bias shows that the deviation of true and estimated energy is stronger at high and at low energies; at medium energies from 1 TeV to 80 TeV, the bias goes to zero. The resolution is higher at low energies and then decreases with increasing energy. This means that the uncertainty on the estimator is lower at high energies than at lower energies. Figure 6.15 shows the 20 most important features of the random forest. Especially time attributes like `rms_time` and the showers `size` are important for estimating the energy.

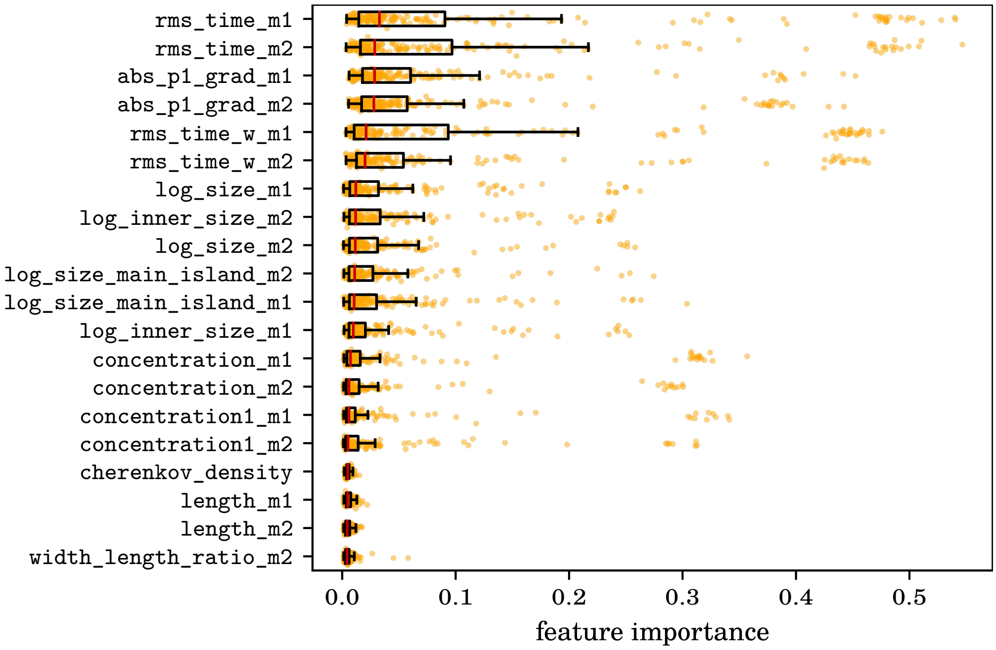

Figure 6.15: The 20 most important features of the proton energy regressor. The most important features describe features related to the arrival times of the image (`p1_grad` and `rms_time`) and the detected Cherenkov light of the shower (`size`).

The comparison of the results with the list of the 20 most important features for the separation of data and simulations from figure 5.11 rules out that the energy regressor favors such attributes with a larger data-simulation mismatch.

The tools in this thesis can also be used to train an energy regressor for helium and iron nuclei. The performances for the two estimators are in appendix A.6 and show that machine learning algorithms are also suitable for this purpose. Therefore, the analysis in this work could be performed analogously to the protons with the other particle types to calculate a corresponding helium or iron spectrum. However, the respective results depend on the assumptions about the background, the remaining particles. Thus, the calculation of all spectra with this method would have to consider the dependencies amongst the spectra.

6.5 Energy-dependent Cuts in Helium and Iron Score

The random forests have calculated a score for each event, indicating how many decision trees conclude that the primary particle is iron or helium or belongs to the corresponding background. The analyst must define a cut on the score to assign a class to each event. If the analysis requires a sample with high purity, using only those events identified as protons with high confidence makes sense. The disadvantage of such a strict cut is that many events, including protons, are rejected, and the sample for the analysis becomes smaller.

While enormous amounts of measured data are available for this analysis, the statistics of the simulations are limited by the amount of required computation time. The cuts are chosen to obtain many protons for the analysis so that the fraction of protons in the sample becomes sufficiently large. A loose cut will also carry more helium and iron nuclei. This must be considered in the later analysis and subtracted in the unfolding in chapter 8.

This analysis chooses the efficiency from equation (6.4) as the evaluation criterion for the score cuts. When training the machine learning algorithms, for practical reasons, the nucleus that is filtered out of the dataset is defined as the signal. Iron is labeled as signal in the first random forest and helium in the second. Accordingly, protons are in both cases in the background dataset. The criterion used in this analysis is the efficiency with which protons are classified, called background efficiency in the following.

The validation dataset is used for the calculation of the background efficiency and the decision of the cuts. This dataset corresponds to about 30 % of the simulated data of each particle type. While for the iron classifier, the cut can be chosen slightly looser because here the separation is clearer, for the helium score, the cut has to be chosen more carefully. The background efficiency for the iron cut is 0.7, meaning that 70 % of the background (protons and helium nuclei) must be correctly reconstructed. The cut in the helium score is chosen with a background efficiency of 0.8 in order to keep 80 % of the protons that already survived the iron cut.

In order to achieve optimal cuts, the validation sample is weighted to the energy spectra of the particle types. Thus a realistic representation can be made: with increasing energy, the iron fraction among all particle types increases; at low energies, it is relatively small; at high energies, it becomes comparatively large. Once the spectra are weighted, the samples are divided into logarithmic energy bins, and the cuts are optimized for the respective energy ranges. It should be noted here that the reconstructed energy is used since only the reconstructed energy, but not the true energy of the data to which the cuts are to be applied, is known, and the estimate sufficiently represents the dependencies of the energy.

The weighted validation samples of protons, helium, and iron nuclei are shown in figures 6.16 to 6.18 with the corresponding cut in each energy bin to visualize the cut on individual particle types. Events pass this analysis step if they have a score below the cuts (black) in the figures. Figure 6.16 shows the energy-dependent iron and helium scores of the iron simulations and the selected cuts. The events in the figure are weighted to the expected iron spectrum. Iron particles are filtered out to a large extent (89.5 %) with the cuts in the iron score. Iron events also receive high helium scores in the second random forest since they are classified here as helium rather than iron. After the second cut, only 3.9 % of the iron particles survive.

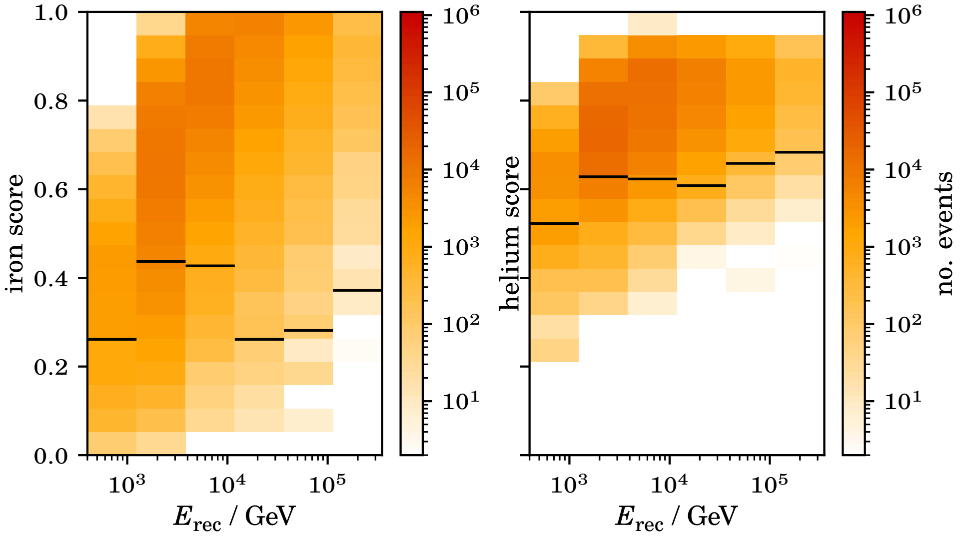

Figure 6.16: Energy-dependent cuts in classifier score for the iron (left) and helium (right) classifier. The cuts are calculated at 70 % background efficiency for the iron classifier and 80 % background efficiency for the helium classifier. The z-axis shows the distribution of the scores of all iron validation simulations calculated by the iron and helium classifiers.

Figure 6.17 shows the scores of the simulated helium events as a function of the reconstructed energy and the chosen score cuts. With 43.8 %, already a large fraction of the helium simulations does not survive the cuts in the iron score, and the second cut in helium score filters out even more events, resulting in only 31.1 % of helium events being misclassified as protons.

The scores of the simulated protons as a function of the reconstructed energy and the cuts in iron and helium score can be found in figure 6.18. 24.4 % of the protons do not survive the cut in the iron score. In the cut in the helium score, 15.3 % are filtered from the proton sample because the reconstruction was not clear enough or incorrect. Of the protons, a total of 60.3 % survive the score cuts, while 68.9 % of the helium nuclei and 96.1 % of the iron nuclei are filtered out by the iron and helium score cuts.

After displaying the cuts with the distributions of the different simulated particle types,

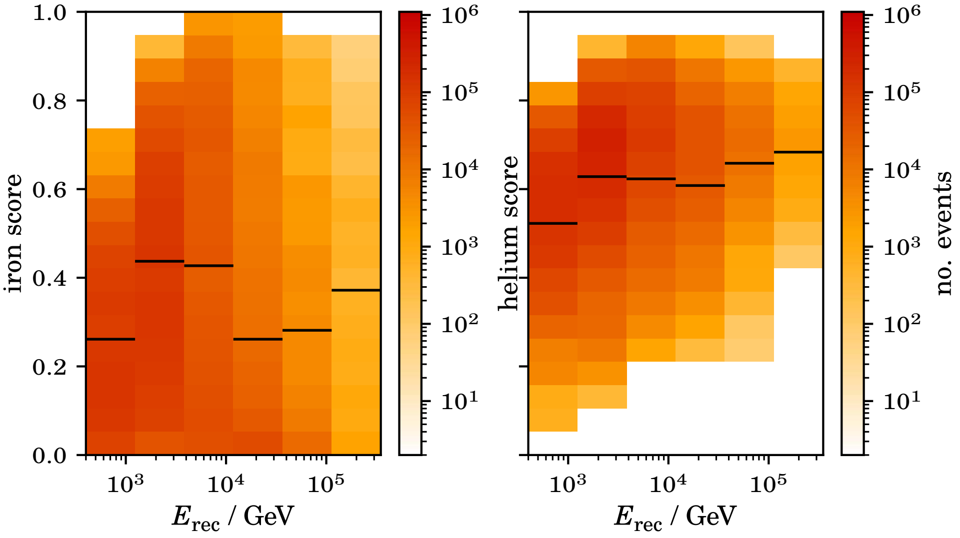

Figure 6.17: Energy-dependent cuts in classifier score for the iron (left) and helium (right) classifier. The cuts are calculated at 70 % background efficiency for the iron classifier and 80 % background efficiency for the helium classifier. The z-axis shows the distribution of the scores of all helium validation simulations calculated by the iron and helium classifiers.

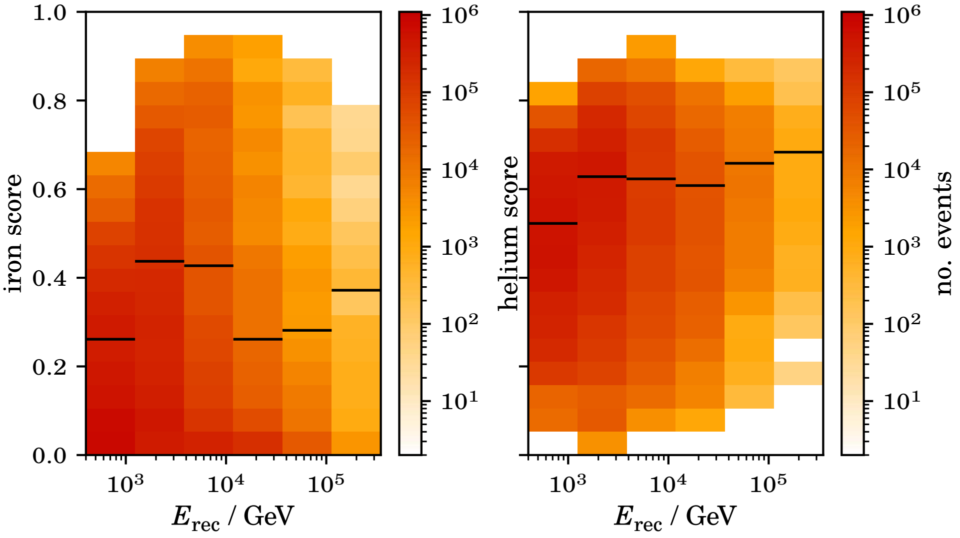

Figure 6.18: Energy-dependent cuts in classifier score for the iron (left) and helium (right) classifier. The cuts are calculated at 70 % background efficiency for the iron classifier and 80 % background efficiency for the helium classifier. The z-axis shows the distribution of the scores of all proton validation simulations calculated by the iron and helium classifiers.

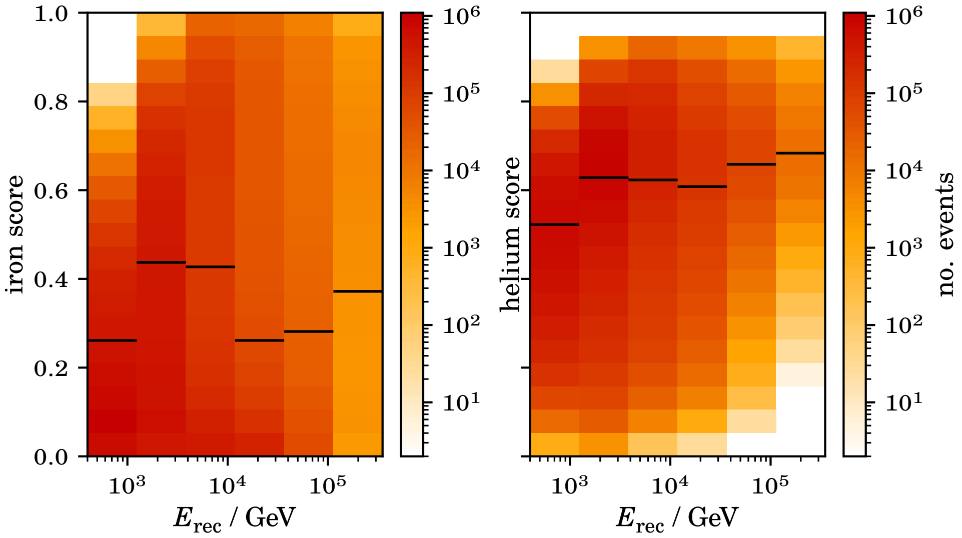

Figure 6.19: Energy-dependent cuts in classifier score for the iron (left) and helium (right) classifier. The cuts are calculated at 70 % background efficiency for the iron classifier and 80 % background efficiency for the helium classifier. The z-axis shows the distribution of the scores of the observation data calculated by the iron and helium classifiers.

the influence of the cuts on the real data is also visualized: figure 6.19 shows the iron and helium scores of the observation data as well as the energy-dependent cuts for the scores. With a relatively strict cut in the iron score already 36 % of all measured events are discarded, 15 % do not survive the more moderate cut in the helium score. In total, 49 % of the observation data is classified as protons. Earlier measurements predict that the group of light nuclei consisting of protons and helium nuclei accounts for about 64 % of the total cosmic-ray flux above an energy of 30 TeV as shown in table 2.1. Although the energy limits do not match exactly, the result of the classifiers with 64 % of all observed events classified as light particles, protons and helium nuclei, is exceedingly consistent with expectations.

Table 6.1 shows the fractions of the reconstructed particle types depending on the true particle type. The columns show the percentages calculated with the weighted number of events of the simulated proton, helium, and iron validation samples, as well as the observation data and different simulated subset combinations. The rows indicate the estimated particle type after the energy-dependent cuts. With a background efficiency of 0.7, the cut in the iron score is set such that 30 % of the background events, here protons and helium nuclei (p + he) do not pass while 70 % survive the cut.

Moreover, the fraction of the summed simulations match the fractions of the observation data. From this, the weighted particles behave similarly in the cuts as the data and underline the data-simulation comparison from section 5.4.

Table 6.1: Fractions of reconstructed events of different datasets. The simulated and weighted proton, helium, and iron sample, as well as the observation data, are listed. The rows show the reconstructed particle types proton, helium, and iron.

	proton sample	helium sample	p + he sample	iron sample	all particles	observation data
iron	24.4 %	43.8 %	29.9 %	89.5 %	30.8 %	36 %
helium	15.3 %	25.1 %	18.1 %	6.6 %	17.9 %	15 %
proton	60.3 %	31.1 %	52.0 %	3.9 %	51.3 %	49 %

In addition to the fractions in table 6.1, the total event numbers of the subsets obtained by the score cut for the particle classification are shown in table 6.2. The numbers correspond to the real event numbers from the measurement, as well as the numbers of weighted simulation events.

Table 6.2: Counts of reconstructed events of different datasets. The simulated and **weighted** proton, helium, and iron sample, as well as the observation data, are listed. The rows show the reconstructed particle types proton, helium, and iron.

	proton sample	helium sample	p + he sample	iron sample	all particles	observation data
iron	2 073 393	1 471 918	3 545 311	167 165	3 712 476	5 623 417
helium	1 302 660	843 451	2 146 111	12 240	2 158 351	2 377 647
proton	5 123 082	1 047 608	6 170 690	7371	6 178 061	7 804 203

The appendix A.5 contains the total number of unweighted events in table A.5.1 and the corresponding percentages in table A.5.2.

Figure 6.20 and figure 6.21 summarize the information about the cut decisions. It shows the scores of the different particle types and the corresponding values for different cut criteria in each of the six energy bins of this analysis.

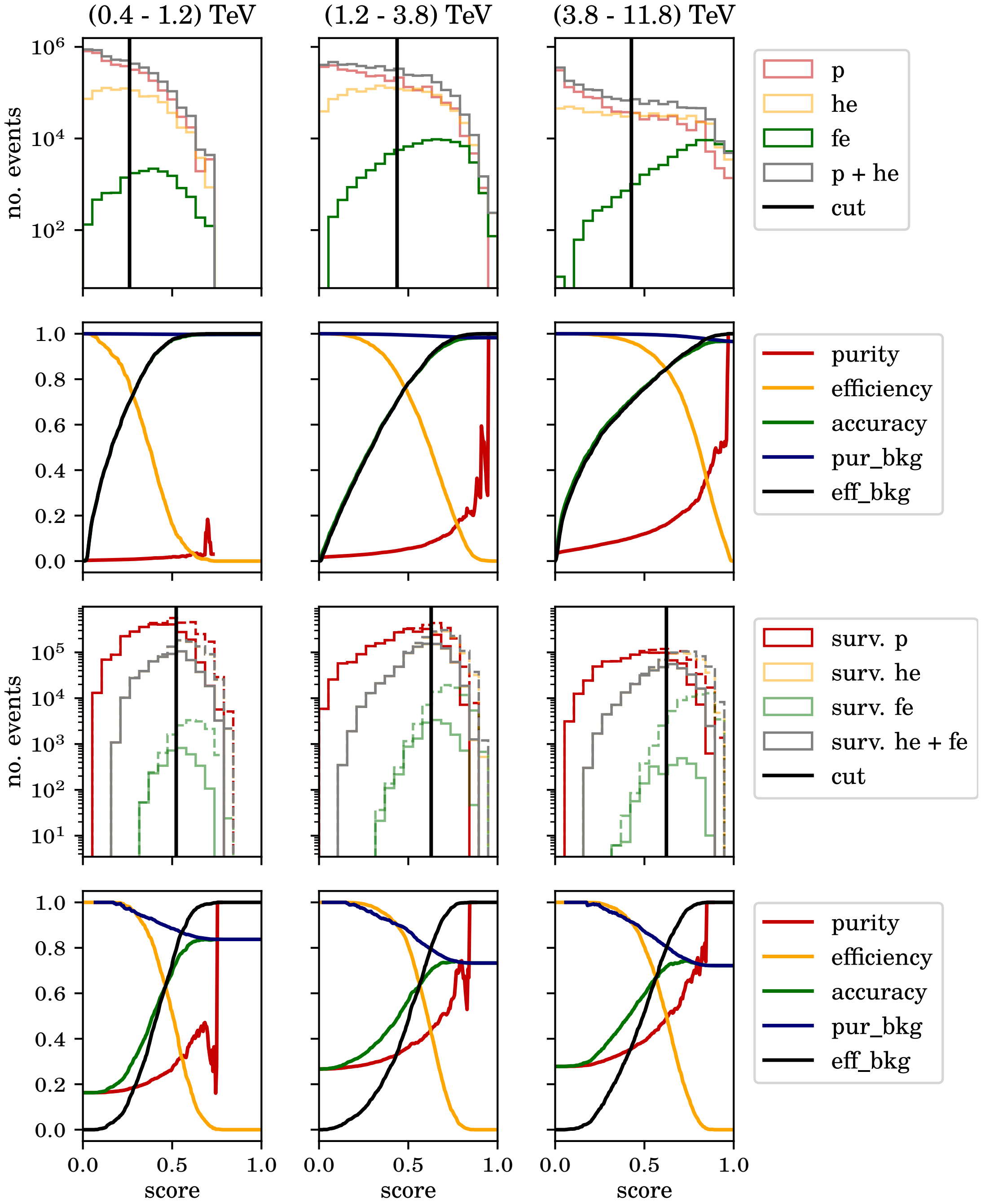

Figure 6.20: Selected cuts (black) for the iron score (top) and the helium score (bottom) with the simulated validation samples. The rest (gray) corresponds to all proton and helium simulations (top) and all helium and iron simulations that survived the iron cut (bottom). The dashed lines indicate the populations before the cuts in the iron score. The simulations are weighted to approximate the real spectra. Here the first three energy bins are shown.

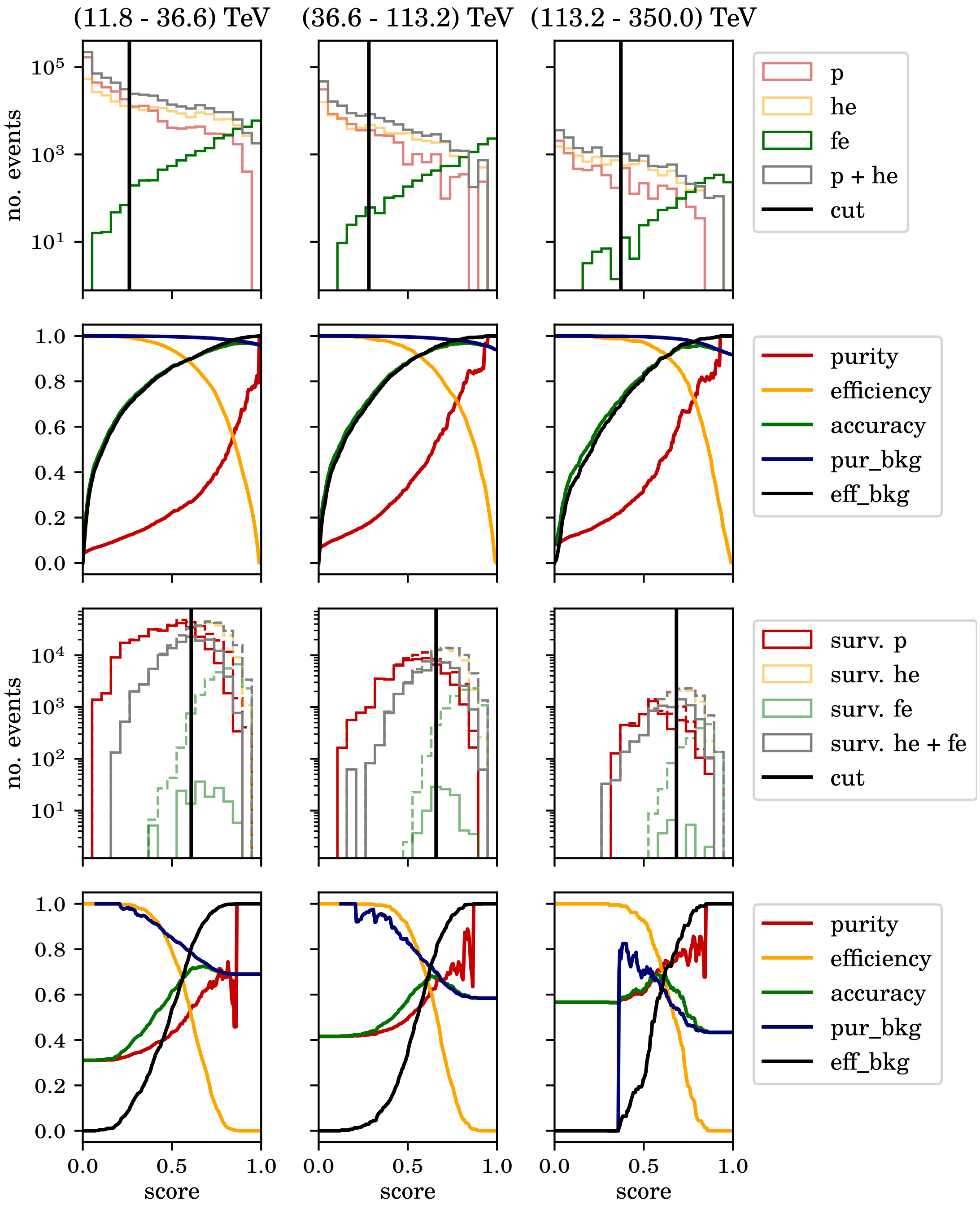

Figure 6.21: Selected cuts (black) for the iron score (top) and the helium score (bottom) with the simulated validation samples. The rest (gray) corresponds to all proton and helium simulations (top) and all helium and iron simulations that survived the iron cut (bottom). The dashed lines indicate the populations before the cuts in the iron score. The simulations are weighted to approximate the real spectra. Here the last three energy bins are shown.

6.5.1 Is an Iron Score necessary?

At this point, the question arises whether the cut in the iron score is necessary or whether the helium cut is sufficient to separate enough background from the protons. For illustration, the scores of the different particle types are shown in figure 6.22 in the corresponding energy bins. The chosen score cuts are indicated with black lines. The 90 % and the 68 % quantiles indicate the ranges in which there is a corresponding percentage of events. In the first two energy bins, the cut in the iron score seems to detect additional iron events, but at higher energies, a cut in the helium score already throws out many iron events. The question of whether a cut in the iron score would be sufficient is based on the assumption that an iron-induced shower in its signature is more similar to a helium-induced shower than to a proton-induced shower since the mass number of an iron nucleus (Z=56) is closer to that of a helium nucleus (Z=4) than to that of a proton (Z=1). As discussed in section 4.1.5, the `max_height`, thus the shower height and also the number of islands in the shower image, the attribute `num_islands` is an indicator of the mass of the primary particle, since these should correlate. Accordingly, it is reasonable to assume that an iron shower is preferably classified as helium and not as proton by a proton-helium classifier. This tendency for iron events with a high helium score to also have a high iron score can be seen in the figures.

In principle, it makes sense to use iron simulations to classify the particles since they are just available. However, in addition to the most prominent particles, protons, helium, and iron nuclei, a smaller fraction of other elements, such as oxygen and carbon, with next-higher scores, also hit the atmosphere. Since the tests confirm the assumption that particles with higher mass numbers tend to get a high helium score in the proton-helium classifier, it can be assumed that other heavier particles are also more likely to be classified as helium. Accordingly, the proton-helium classifier can reasonably classify the entire hadronic background in the measured data.

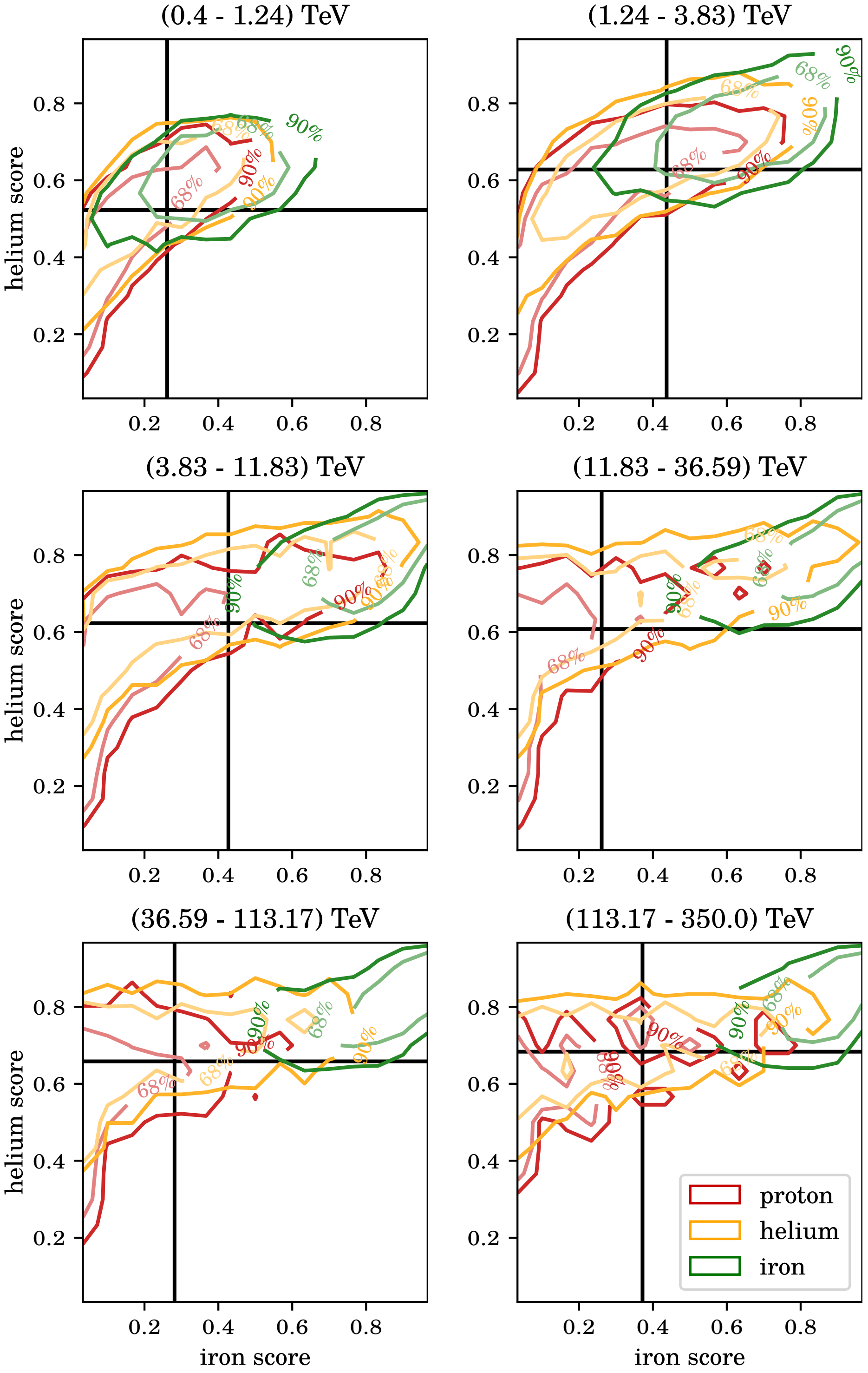

Figure 6.22: Iron and helium scores of the proton, helium, and iron samples given in the 90 % and the 68 % quantiles and the energy-dependent iron and helium score cuts. A strong correlation between helium and iron scores is visible in the iron sample (green). The helium random forest tends to classify iron as helium rather than proton.

Detector Properties

7

So far, the energy and the particle type are reconstructed for each event by applying the decision trees to the datasets and choosing cuts in the case of classification: this results in a dataset of events classified as protons, and the estimated energy of each event. The main goal is to calculate the true proton flux as a function of energy: from the particles detected by the instrument and identified as protons during the analysis, calculate how many particles in which energy range originally hit the Earth in the observed area of the sky during the time of the observation. Understanding the detector and the analysis is essential to obtain the physical result from the measurement.

First, the telescope has a certain efficiency: the proportion of detected protons to all incoming protons. Further protons are eliminated in the analysis due to quality cuts or misidentification. This leads to only a certain fraction of all cosmic rays in a given area being measured and analyzed. Thus, both telescope efficiency and the efficiency in the analysis must be studied and understood to correct the measured number of particles in the observed area.

Second, a detector usually has a certain dead time: the time after triggering is necessary to store the event. During this time, no further event can trigger the electronics. Accordingly, the detector is not in operation during this time, and the dead time after each triggered event must be subtracted from the total observation time to calculate the effective measurement time.

Third, the detector has a limited energy resolution. The energy of the primary particle is estimated from the shower images. A shower image cannot be assigned exactly to a true energy value because the detector is not perfect: the energy estimation produces results deviating from the truth, visualized in the migration matrix based on simulations. Additionally, there is a bias in the energies, a characteristic of the energy estimator. The bias can be corrected once it is quantified. This chapter explains and calculates the Instrument Response Functions (IRFs) for this analysis.

7.1 Method: Instrument Response Functions

The discrete *Fredholm* integral equation [99, p. 282] describes the expectation of the observed distribution in bins of the reconstructed energy E_{rec} as the vector $\hat{g}_j$ with the elements $j = 1, 2, ..., N_j$ given the true distribution in bins of the true energy E_{true} as the vector f_i with the elements $i = 1, 2, ..., N_i$.

$$\hat{g}_j = \sum_{i=1}^{N_i} A_{ji} \cdot f_i' + b_j \qquad \text{with} \qquad f_i' = \varepsilon_i \cdot f_i \tag{7.1}$$

First, it must be considered that only a fraction of the incoming true number of protons f_i trigger the experiment and survive the cuts. Thus the proton efficiency ε_i influences the final result. The deviation of the energy due to the measurement and reconstruction methodology is described by the energy dispersion matrix A_{ji}, so in order to be able to determine the reconstructed energy from the true energy, the acceptance-corrected proton distribution f_i' must be folded with the matrix. Thus we obtain a realistic description of the protons, as the telescope would measure them. Since it must be assumed that the score cuts do not produce a pure proton sample and some particles such as iron and helium are misclassified as protons, a background b_j must also be included in the calculation.

7.1.1 Efficiency and Effective Area

The efficiency is the fraction of all events detected by the instrument and identified as protons in the analysis. The efficiency depends on the primary particle's energy since shower images of lower energies produce less Cherenkov light, making detection more difficult. As the energy of the primary particle increases, the number of Cherenkov photons in the shower also does, and triggering the event is, therefore, more likely.

$$\varepsilon(E) = \frac{N_{\text{identified}}(E)}{N_{\text{all}}(E)} \tag{7.2}$$

The ratio of detected and well-classified protons $N_{\text{identified}}$ and all incoming protons N_{all} depending on the energy E corresponds to the efficiency ε.

If the instrument had an efficiency of $\varepsilon = 1$, it would be a perfect detector. The effective area of the instrument is provided to illustrate the efficiency. The effective area is the detector area of a perfect instrument measuring the same number of protons. The simulated validation set is the base for the efficiency estimation: the simulated shower images have undergone the same trigger procedures and cuts as measured data. The efficiency is calculated in the same energy bins as the final energy spectrum.

$$\begin{aligned} A_{\text{eff}}(\Delta E) &= \varepsilon(\Delta E) \cdot A_{\text{simulated}}(\Delta E) \qquad (7.3) \\ &= \frac{N_{\text{identified}}(\Delta E)}{N_{\text{simulated}}(\Delta E)} \cdot A_{\text{simulated}}(\Delta E) \qquad (7.4) \end{aligned}$$

The effective area $A_{\text{eff},i}$ is the area that an ideal detector would have to detect the same number of particles $N_{\text{identified},i}$ as measured by the real detector.

7.1.2 Energy Dispersion

The energy dispersion is the deviation between true and estimated energies. The dispersion is a statistical problem of a real, non-perfect detector and can be described by the probability density function for the measurement of the value E_{true}. Once the dispersion is known, the energy spectrum can be statistically reconstructed. Energy dispersion is studied with simulations. For this purpose, the energy estimator is applied to the validation set of protons. Thus, each event receives a reconstruction E_{est} in addition to the simulated true energy E_{true}. Filling a two-dimensional histogram with these energies yields what is known as the migration or dispersion matrix, as shown in Figure 6.13.

7.1.3 Background

In reality, the background is composed of all particles different from the proton and yet produces similar showers in the telescope camera. Most of the background likely consists of heavier particles like helium and iron. Gamma rays can also produce similar showers, but the flux is negligible, as discussed in chapter 2. In principle, random electronic noise or a bright flash in the camera can also trigger and contribute to the background. Anything that triggers the telescopes and is classified as a proton-induced shower in the analysis, but is not a proton event, must be counted as background $b(E_{\text{rec}})$: misclassified helium $he_{\text{mis}}(E_{\text{rec}})$ and misclassified iron $fe_{\text{mis}}(E_{\text{rec}})$ are considered as background $b(E_{\text{rec}})$ since they are assumed to be the largest component and therefore considered as simulations in this analysis. The background $b(E_{\text{rec}})$ is composed by the simulations.

$$b(E_{\text{rec}}) = he_{\text{mis}}(E_{\text{rec}}) + fe_{\text{mis}}(E_{\text{rec}}). \tag{7.5}$$

The background is weighted with the spectra from section 5.3.1 and the effective time t_{eff} from the measurement to provide a realistic representation. For testing purposes, $g(E_{\text{rec}})$ is generated from simulations of protons, helium, and iron nuclei classified as protons.

$$g(E_{\text{rec}}) = p(E_{\text{rec}}) + he_{\text{mis}}(E_{\text{rec}}) + fe_{\text{mis}}(E_{\text{rec}}) \tag{7.6}$$

The simulations for $g(E_{\text{rec}})$ are weighted as well, so the simulation sets and therefore $g(E_{\text{rec}})$ resembles the observation data.

7.2 Method: Flux Calculation

The diffuse proton flux is a quantity that is independent of the detector and its characteristics as well as the measurement conditions. It is defined by

$$\Phi = \frac{\mathrm{d}^2 N}{\mathrm{d}A\mathrm{d}t} \tag{7.7}$$

and describes the number of events N measured per detector area A and time t. The particle flux depends on the particle energy and follows a power law; see equation (5.2). Accordingly, the particle flux depends on the width of the selected energy range. This can be avoided by using the differential energy spectrum

$$\frac{\mathrm{d}\Phi}{\mathrm{d}E} = \frac{\mathrm{d}^3 N}{\mathrm{d}E\mathrm{d}A\mathrm{d}t} \tag{7.8}$$

which takes the energy range dE of the spectrum into account. In the case of a diffuse particle flux Φ_{diffuse}, the size of the observed part of the sky must be known: the solid angle Ω. With the solid angle Ω, the final equation for the differential flux results.

$$\frac{\mathrm{d}\Phi_{\text{diffuse}}}{\mathrm{d}E} = \frac{\mathrm{d}^4 N}{\mathrm{d}\Omega\mathrm{d}E\mathrm{d}A\mathrm{d}t} \tag{7.9}$$

This equation applies to a perfect detector with an efficiency of 1 and continuous operation between the start and end of the measurement. Furthermore, the equation assumes that the true energy of the particles is known. The fact that the detector and the reconstruction of the particles' properties is not perfect is discussed in the following. Furthermore, corrective methods are explained.

7.2.1 Dead Time and Effective Time

The storing of an event follows each triggering. Due to the hardware, no further event can be triggered during this process. The length of the processing time depends on the detector electronics and is called dead time. The observation time must be corrected since the detector is not operating for the dead time after each trigger event. The effective time t_{eff} depends on the number of triggered events $N_{\text{triggered}}$, as well as the trigger rate $(1/\bar{\Delta t}_{\text{triggered}})$ and the dead time t_{dead}.

$$
\begin{aligned}
t_{\text{eff}} &= t_{\text{observed}} - N_{\text{triggered}} \cdot t_{\text{dead}} && (7.10)\\
&= \sum_i (\Delta t_{\text{triggered},i}) - N_{\text{triggered}} \cdot t_{\text{dead}} && (7.11)\\
&= N_{\text{triggered}} \cdot \frac{1}{N_{\text{triggered}}} \sum_i (\Delta t_{\text{triggered},i}) - N_{\text{triggered}} \cdot t_{\text{dead}} && (7.12)\\
&= N_{\text{triggered}} \cdot [\frac{1}{N_{\text{triggered}}} \sum_i (\Delta t_{\text{triggered},i}) - t_{\text{dead}}] && (7.13)\\
&= N_{\text{triggered}} \cdot [\text{mean}(\Delta t_{\text{triggered}}) - t_{\text{dead}}] && (7.14)
\end{aligned}
$$

$$
t_{\text{eff}} = N_{\text{triggered}} \cdot [(\bar{\Delta t}_{\text{triggered}}) - t_{\text{dead}}] \qquad (7.15)
$$

The dead time of MAGIC amounts 26 µs since the hardware update in 2012 [15]. To calculate the effective time of the measurement, the mean time interval between two successive events $\bar{\Delta t}_{\text{triggered}}$ is calculated and, after subtracting the dead time $t_{\text{dead}} = 26\,\mu\text{s}$, multiplied by the number of triggered events $N_{\text{triggered}}$.

7.2.2 Steradian

In this analysis, the diffuse proton flux is calculated within the telescope's field of view. The assumption is that the proton flux is homogeneously distributed within the field of view. The flux is given divided by the solid angle to give the physical flux independent of the field of view of the telescope. Provided that the sky is a spherical cap, the solid angle Ω for a given view cone α is

$$
\Omega = 2\pi \left(1 - \cos\left(\frac{\alpha}{2}\right)\right) \text{sr} \qquad (7.16)
$$

with the unit sr (steradian). The MAGIC telescopes, with their field of view of 3.5°, cover a section of sky with a solid angle of $\Omega \approx 0.003\,\text{sr}$.

7.3 Calculation of IRFs

The IRFs are obtained from the simulations of proton showers. At different measurement inclinations, the so-called zenith distances, shower particles have to penetrate the atmosphere for different distances. The air conditions influence the shower evolution and the fraction of Cherenkov light collected in the camera. Therefore, the IRFs are assumed to be different for different zenith ranges.
Generally, it is sufficient if the data and simulations cover a similar zenith range to

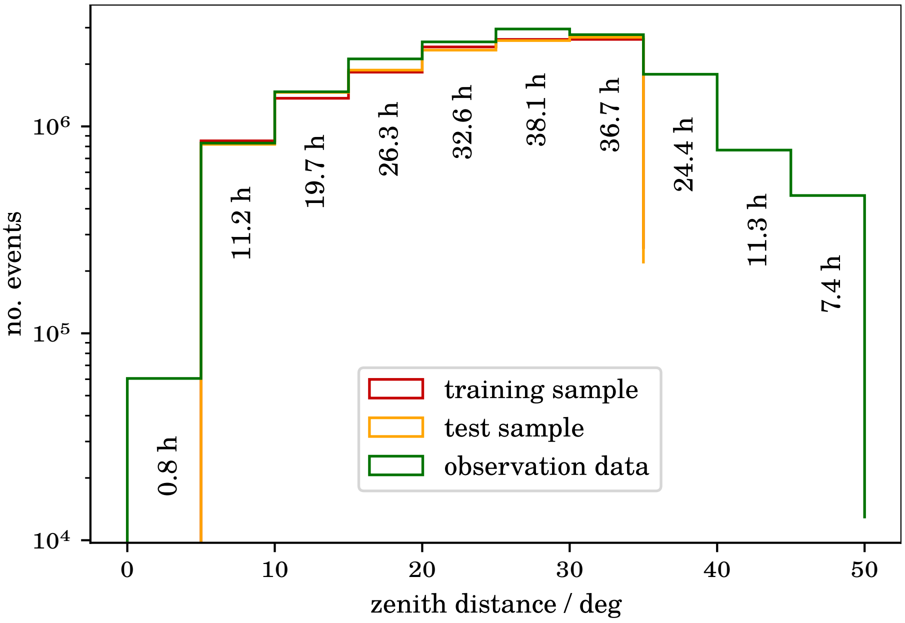

Figure 7.1: Distribution of zenith distances of the events of the observation data (green) and the simulations (red and yellow). The simulations are provided from 5° to 35° zenith distance. The times in the bars are the effective times t_{eff} of the observation data in the corresponding zenith ranges. The simulations are weighted according to the assumed spectra of the particle types and the effective time of the observation data.

comparable ratios. Figure 7.1 shows the distribution of zenith distances with the corresponding effective times of the observation data. Due to the significantly larger statistics of measurements compared to the simulations, the measured data were selected to match the available simulations. The table 7.1 shows the observation time t_{obs} and the effective time t_{eff} calculated from t_{obs} according to equation (7.15) for the different zenith distance ranges. In the range from 5° to 35° the effective time is 164.6 h. Overall, it is clear that the analysis statistics are limited due to the simulations, not the data. Due to the limited statistics of the simulations, the analysis is performed in six energy bins plus two extra bins, the overflow and the underflow bin. The surviving events are shown after triggers and all necessary cuts in table 7.2 binned in true energy.

The same table is shown again for the binning in reconstructed energy in table 7.3. It can be seen that enough events are available from the observational data. At the same time, the statistics are limited in the simulations, and finer binning could result in too few statistics being left for unfolding in the individual bins. The detector efficiency is calculated according to equation (7.2) and thus from the ratio of simulated protons and ultimately triggered and identified protons. The distributions of the simulated protons and the survived protons are shown in figure 7.2. The resulting energy-dependent efficiency can be seen in figure 7.3. The efficiency depends on true energy E_{true} since

Table 7.1: The observation times t_{obs} in the different zenith distances and the calculated effective times t_{eff}. Simulations are provided at zenith distances from 5° to 35°, so the effective time of the measurement is $t_{\text{eff}} = 164.591\,\text{h}$. It is relevant for the reweighting of the simulations. For the rest of the analysis, the effective times of the corresponding zenith ranges of the data must be considered.

zd / deg	t_{obs} / h	t_{eff} / h
5 - 10	11.288	11.173
10 - 15	19.944	19.740
15 - 20	26.632	26.332
20 - 25	32.904	32.567
25 - 30	38.481	38.072
30 - 35	37.108	36.707
5 - 35	166.357	164.591

Table 7.2: Number of events in the chosen bins in true energy E_{true} for the proton train and test samples.

E_{rec} / GeV			train sample	test sample
0	-	400	116	105
400	-	1237.1	702	699
1237.1	-	3825.9	1223	1236
3825.9	-	11 832.2	1039	977
11 832.2	-	36 593.1	666	661
36 593.1	-	113 170.5	299	339
113 170.5	-	350 000	101	114
350 000	-	inf	15	15

Table 7.3: Number of events in the chosen bins in reconstructed energy E_{rec} for the proton train and test samples as well as for the observed data.

E_{rec} / GeV			train sample	test sample	data sample
0	-	400	8	3	29 765
400	-	237.1	741	724	2 389 637
237.1	-	3825.9	1278	1327	2 373 589
3825.9	-	11 832.2	1010	933	884 467
11 832.2	-	36 593.1	739	737	362 410
36 593.1	-	113 170.5	321	339	92 721
113 170.5	-	350 000	64	83	8190
350 000	-	inf	0	0	0

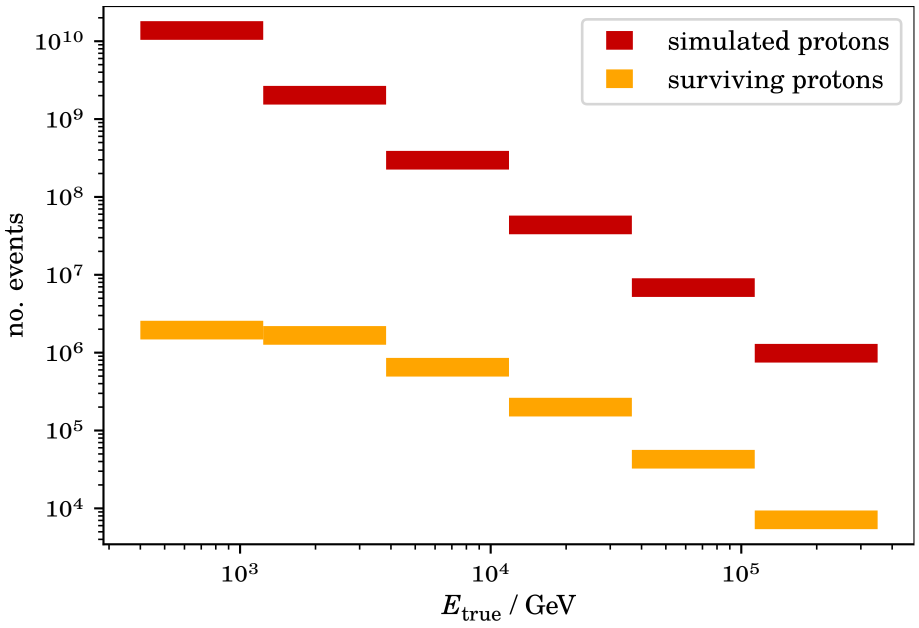

Figure 7.2: All simulated protons (red) and of these the protons that fulfill all trigger conditions and survive the quality and score cuts of the analysis (yellow) as a function of simulated energy E_{true}. The proton simulations are weighted according to the assumed proton spectrum and the effective time of the observation data.

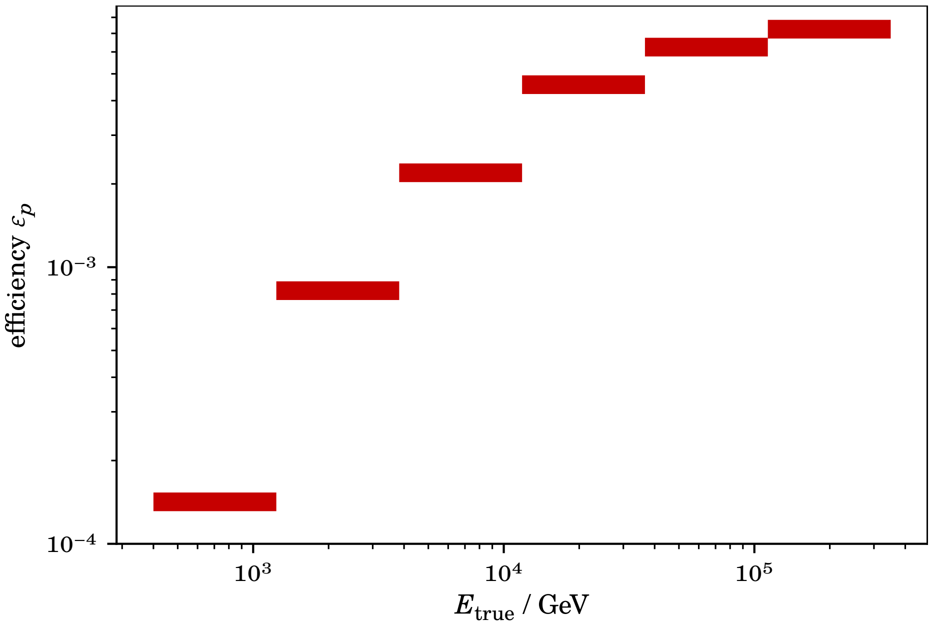

Figure 7.3: Proton efficiency ε_p of the complete analysis as a function of the simulated energy E_{true}, describing the fraction of all simulated protons that trigger and subsequently survive the analysis cuts. The proton simulations are weighted according to the assumed proton spectrum and the effective time of the observation data.

the non-triggered events cannot be reconstructed and thus be given in reconstructed energy. The unfolding in chapter 8 calculates the spectrum in bins of true energy from the distribution in bins of reconstructed energy; in this context, the efficiency can be considered a function of the true energy. Another method would be to calculate the efficiency by dividing the survived protons in reconstructed energy bins by all simulated protons in true energy bins - this method is called *poor mans unfolding* because it skips a proper unfolding and is only useful when a first look at the data is needed. This work performs a proper unfolding in chapter 8.

In addition to the efficiency, the energy dispersion also belongs to the IRFs and must be described with simulations. For this purpose, the proton simulations are used again: In the chosen energy binning, E_{rec} and E_{true} show the energy dispersion in a two-dimensional histogram. The dispersion matrix mathematically describes the IRF due to the energy resolution and is shown in figure 7.4.

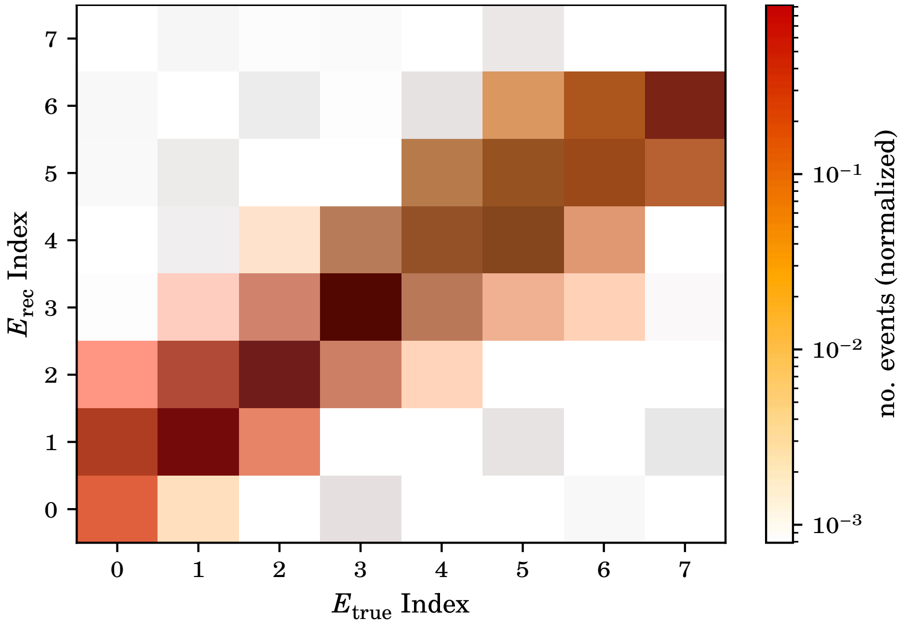

Figure 7.4: Migration matrix of the energy estimator trained with a pure proton simulation set. The migration matrix was created from the so-called validation set, a dataset different from the training and test sets.

The weighted simulation sets can now be compared with the data for testing purposes. According to equation (7.1), the simulated proton distribution is multiplied by the IRFs, and the background is added. The result for the training and test datasets is shown in figure 7.5 and compared with the observation data.

Figure 7.5 shows that the weights are well estimated and that the IRFs appear reasonable. Statistical fluctuations and uncertainties in the spectral indices of the weights can explain some discrepancies. So, the fact that further elements are not available in this comparison leads to discrepancies: the observation data contains more particle types like nuclei of carbon, oxygen, and other elements. Table 2.1 demonstrates that elements other than the here simulated ones comprise a considerable part of the all-particle spectrum. Although a significant fraction is assumed to have been sorted out with the helium classifier, some events will have survived the score cut and may be responsible for minor discrepancies.

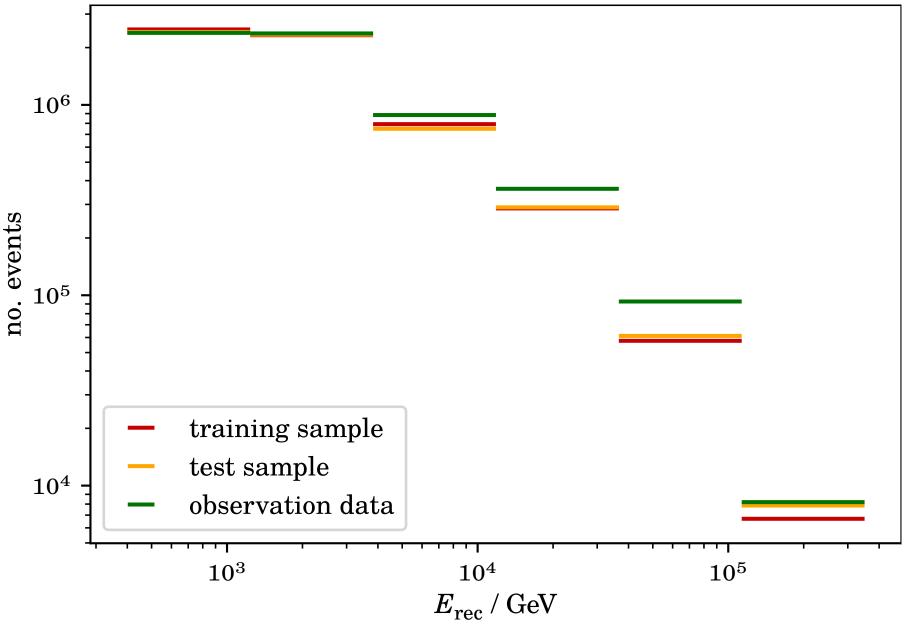

Figure 7.5: Comparison of simulations (red and yellow), weighted by the effective time of the observational data and the assumed spectra for the given particle types, and observational data (green).

Proton Spectrum

8

The challenge is to determine the true proton spectrum from the measured data. The dataset consists of all particles that triggered the telescope and survived the quality and score cuts, thus were identified as protons. An energy E_{rec} was reconstructed for each event in the sample.

Understanding what the measurement process of the experiment looks like is essential for estimating the true number of protons per energy bin from the measurement.

8.1 Method: Inverse Problems

The Fredholm integral in equation (7.1) from the previous chapter describes the mathematical correlation between the binned true proton distribution f_i and the expected values for the binned measurement $\hat{g}_j$ depending on the instrument response $\boldsymbol{A}$. For simplifying reasons, the abbreviated notation for vectors can also be used.

$$\hat{\boldsymbol{g}} = \boldsymbol{A} \cdot \boldsymbol{f}' + \boldsymbol{b} \tag{8.1}$$

If a specific prediction $f_{\mathrm{th},i}$ exists and is to be tested, the equation (7.1) can be used to make a prediction about the measurement $g_{\mathrm{th},j}$ and compare it to the actual measurement $\hat{g}_j$. In the case that no prediction is available or the result is to be calculated without a model assumption in order to remain unbiased and result-open, one can also reconstruct f_i from $\hat{g}_j$.

The most straightforward way to infer the true proton spectrum from the measurement is to invert the equation to

$$\hat{\boldsymbol{f}} = [\boldsymbol{A}^{+} \cdot (\boldsymbol{g} - \boldsymbol{b})] \cdot \frac{1}{\boldsymbol{\varepsilon}} \tag{8.2}$$

with the pseudo-inverse matrix $\boldsymbol{A}^{+}$ of the energy response matrix. This process describes statistically and mathematically an ill-conditioned problem. This naive approach method leads in practice to highly oscillating, poorly useful solutions. Instead of naive inversion, a different technique known as Poisson likelihood unfolding is common.

8.1.1 Poisson Likelihood Unfolding

Since the MAGIC analysis is event-based and the bin contents of the observables correspond to count rates, it can be assumed that the bin contents follow a Poisson distribution.

$$f(g_j|\lambda_j) = \frac{e^{-\lambda_j}\lambda_j^{g_j}}{g_j!} \tag{8.3}$$

According to Fredholm's integral equation (7.1), the expected value λ_j for the single bin contents j is given by

$$\lambda_j = \sum_i A_{ji} f_i' + b_j \tag{8.4}$$

with which the Poisson likelihood equation is calculated.

$$\mathscr{L} = \prod_j \frac{\left(\sum_i A_{ji} f_i' + b_j\right)^{g_j}}{g_j!} \cdot e^{-\left(\sum_i A_{ji} f_i' + b_j\right)} \tag{8.5}$$

$$= \prod_j \frac{(\boldsymbol{A} \cdot \boldsymbol{f}' + \boldsymbol{b})_j^{g_j}}{g_j!} \cdot e^{-(\boldsymbol{A} \cdot \boldsymbol{f}' + \boldsymbol{b})_j} \tag{8.6}$$

The variable f_i is now chosen to maximize the likelihood and, thus, is an estimate of the true bin entries. In practice, a minimizer is used for this; accordingly, the likelihood is modified in a way the minimizer can properly handle it and calculates an $\boldsymbol{f}'$ that minimizes the negative log-likelihood

$$-\log \mathscr{L} = \sum_j \left[(\boldsymbol{A} \cdot \boldsymbol{f}' + \boldsymbol{b})_j - g_j \log \left((\boldsymbol{A} \cdot \boldsymbol{f}' + \boldsymbol{b} + \xi)_j \right) \right] \tag{8.7}$$

where ξ is a small value above zero. In the case of empty bins, adding ξ prevents calculating the logarithm of zero.

8.1.2 Regularization

The problem that inverse problems might have fluctuating solutions can be solved with the help of regularization. The regularization process injects an additional term into the loss function from equation (8.7) that includes prior knowledge about the result. A valid criterion in astroparticle physics is that the resulting logarithmic spectrum of $\boldsymbol{f}$ must be flat over a certain energy range.

In this work, the Tikhonov regularization [130, p. 100] is used, which is expressed by the penalty term

$$\mathscr{L}_{\text{tikh}}(\boldsymbol{f}) = \tau \cdot \sum_i \|\boldsymbol{C} \cdot \log(\boldsymbol{f} + \xi)\|_i^2 \tag{8.8}$$

with the requirement of flatness of the acceptance-corrected logarithmic spectrum $\log(\boldsymbol{f}) = \log({}^{f'}\!/_{\varepsilon})$ over the analyzed energy range. The discrete second derivative matrix is used for this purpose since it describes the slope variation. The norm of $C \cdot f$ is squared so that the solution is positive and small values are favored in minimization over a strong negative curvature. The derivative matrix is given by

$$\boldsymbol{C}_{n\times n} = \begin{pmatrix} 1 & -1 & 0 & & \cdots & & 0 \\ -1 & 2 & -1 & 0 & \cdots & & \\ 0 & -1 & 2 & & & \vdots & \vdots \\ & 0 & & \ddots & & 0 & \\ \vdots & \vdots & & & 2 & -1 & 0 \\ & & \cdots & 0 & -1 & 2 & -1 \\ 0 & & \cdots & & 0 & -1 & 1 \end{pmatrix}$$

and the regularization strength τ is the penalty factor. This factor τ should be well chosen: at $\tau \to 0$, the regularization does not influence the final result for $\boldsymbol{f}$, while at $\tau \to \infty$ a linear solution for $\boldsymbol{f}$ is forced. The final likelihood for the minimizer is composed of the Poisson likelihood from equation (8.7) and the regularization term from equation (8.8).

$$\begin{aligned} \mathscr{L}_{\text{final}} &= -\log \mathscr{L}_{\text{poisson}} + \mathscr{L}_{\text{tikh}} \\ &= \sum_j \left[(\boldsymbol{A} \cdot \boldsymbol{f}' + \boldsymbol{b})_j - g_j \log\left((\boldsymbol{A} \cdot \boldsymbol{f}' + \boldsymbol{b} + \xi)_j \right) \right] + \tau \cdot \sum_i \|\boldsymbol{C} \cdot \log(\boldsymbol{f} + \xi)\|_i^2 \end{aligned} \tag{8.9}$$

In this work, the `migrad` minimizer from the `iminuit` python package [38, 74] is used to find a solution for $\boldsymbol{f}'$, and acceptance-corrected $\boldsymbol{f}$.

8.1.3 Regularization strength

There are multiple ways to determine the optimal regularization strength τ. An intuitive approach would be to perform a grid search in a reasonable range of τ and compare the regularized distribution f_{reg} with the true distribution f_{true}. Using a bin-by-bin test such as the χ^2-test [107] or the Kolmogorov–Smirnov (KS) test [76, 121], or a cross-bin metric like the Wasserstein metric [111], a conclusion can then be made about the best regularization strength.

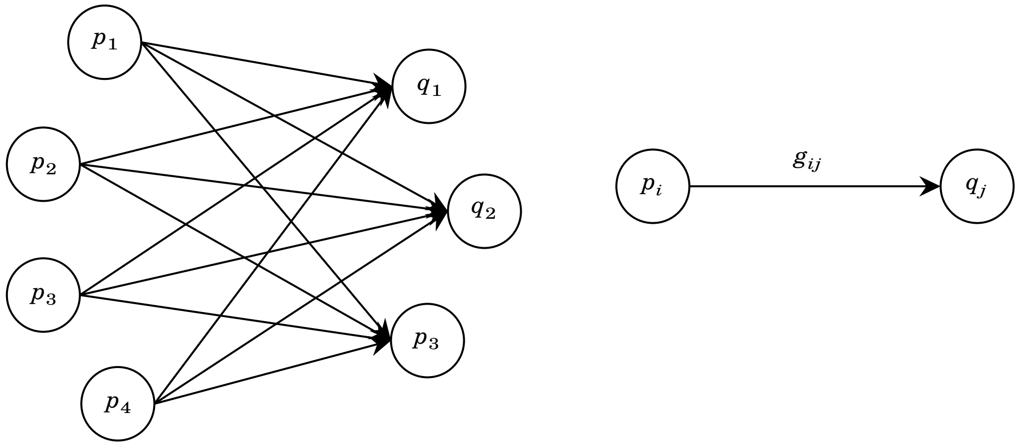

Figure 8.1: Illustration of the earth mover distance: the amount of earth g_{ij} is transported from pile p_i to pile q_j at the positions u_i and v_j. Adapted from [111].

The Wasserstein distance is also called earth mover distance [82] because the following scenario can describe the metric: a set of earth piles shall be shifted, resulting in another set of piles. The value g_{ij} describes the amount transported from pile p_i to pile q_j over the ground distance d_{ij} between the piles' positions u_i and v_j. The illustrative representation of the shift of the piles is shown in figure 8.1. The Wasserstein distance can be described as follows.

$$W(\hat{\boldsymbol{g}}, \boldsymbol{g}) = \min \frac{\sum_{i,j=1}^{N_{p,q}} d_{ij} g_{ij}}{\sum_{i,j=1}^{N_{p,q}} g_{ij}} \tag{8.10}$$

In principle, the minimal work required to get from an initial to the target situation is described. With this criterion, the chosen regularization strength is a fixed value transferred for all datasets.

A more reasonable approach would be to determine the regularization strength for each dataset individually. Such an approach provides the criterion of the global correlation coefficient ρ_{global} [99, p. 298]. It is given by

$$\rho_{\text{global}} = \frac{1}{N_{\text{true}}} \sum_{j=1}^{N_{\text{true}}} \sqrt{1 - \left((V_f)_{jj} \cdot (V_f^{-1})_{jj}\right)^{-1}} \tag{8.11}$$

where V_f is the covariance matrix of f, the inverse Hessian matrix of the likelihood minimum. When minimizing the global correlation coefficient, the unfolded bins of the result are as uncorrelated as possible. The reason is that the correlation of the unfolded bins arises from the estimation from the observed bins. Minimizing the bin correlation has proven to be the best method with the most physically plausible results.

8.2 Unfolding of the Data

Unfolding is first performed on the test dataset to prove the applicability of the analysis. The test dataset consists of 30 % of all available simulations and differs from the dataset used to determine the score cuts and the IRFs in chapter 6.

The test dataset consists of weighted simulations of protons, helium, and iron nuclei, whose particle type and energy were estimated with the random forests from chapter 6 and classified with the score cuts. Thus, it includes all particles classified as protons - including helium and iron nuclei misclassified as protons. The misclassified particles are considered background following equation (7.5).

Using the migration matrix and the proton efficiency from section 7.3, a Poisson likelihood with regularization term, equation (8.9), is then established and minimized to obtain the unfolded bin entries of the proton spectrum. The unfolded solution of the test dataset is shown in figure 8.2, once without regularization $\tau = 0$ and once with the optimized regularization according to the criterion of minimal bin correlation. The mean global correlation coefficients ρ_{global} for different regularization strengths τ can be seen in figure 8.3. Despite a large dispersion of the data points, a minimum is visible at a regularization strength around $10^2 < \tau < 10^3$, and the selected minimum is at a

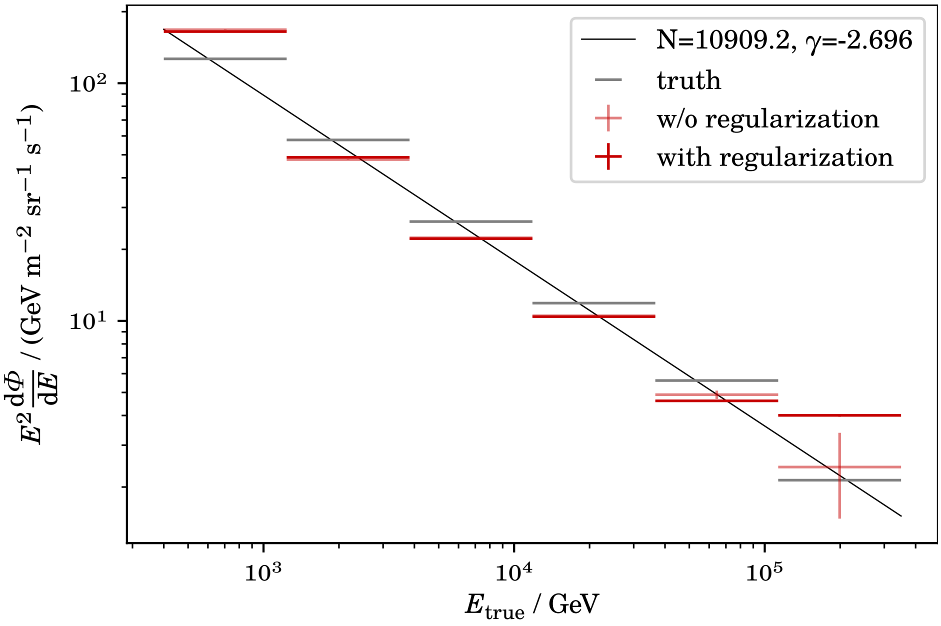

Figure 8.2: Proton spectrum obtained from the test sample consisting of protons, helium, and iron events. All particle types were classified, and their energy was estimated. The data points are unfolded once with and once without regularization. As a reference spectrum, the fit of the proton spectrum from section 5.3.1 is shown in black. In the first four bins, the regularization effect is difficult to see: the data points with and without regularization are almost on top of each other.

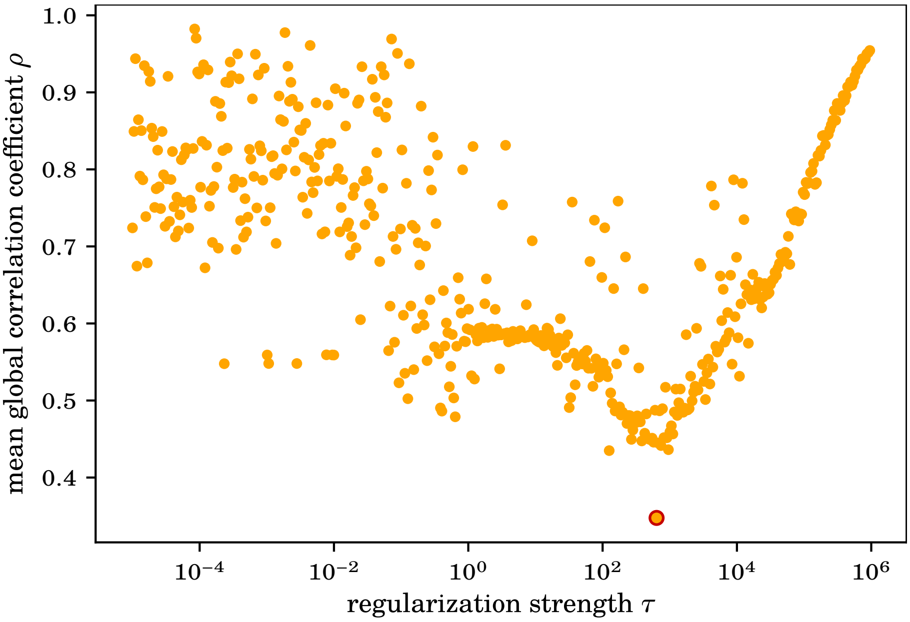

Figure 8.3: Mean global correlation coefficients of the simulated test sample unfolded with 500 regularization strengths τ. The red circle marks the chosen regularization strength at the minimal mean global correlation coefficient ρ.

regularization strength of $\tau = 282$.

The earth mover distance tests how much the solution deviates from the truth. It is plotted in figure 8.4 for the 500 different regularization strengths. It can be seen that the earth mover distance is little affected by the regularization up to $\tau \approx 10^3$, but increases rapidly with greater values. This agrees with the results from the figure 8.3 since also the mean global correlation coefficient ρ_{global} reveals a rapid deterioration of the solution with stronger regularization. Due to only small differences in the quality of the results at a smaller regularization strength than $\tau \approx 10^3$, it can be discussed whether a regularization is necessary in this analysis.

The same unfolding is then performed with the observation data: The resulting plot is shown in figure 8.5. Again, the regularization strength τ is determined using the mean global correlation coefficient ρ_{global}, and the corresponding values depending on the regularization strength τ can be seen in figure 8.6. The regularization strength $\tau = 2381$ can reach the minimal correlation.

8.2.1 Statistical and Systematical Error

The errors of a Poisson likelihood unfolding can be calculated with the Hessian matrix H that the minimizer returns [27]. At the minimum, the inverse Hessian matrix is a valid approximation for the covariance matrix $\boldsymbol{G}$, which indicates the covariance between the

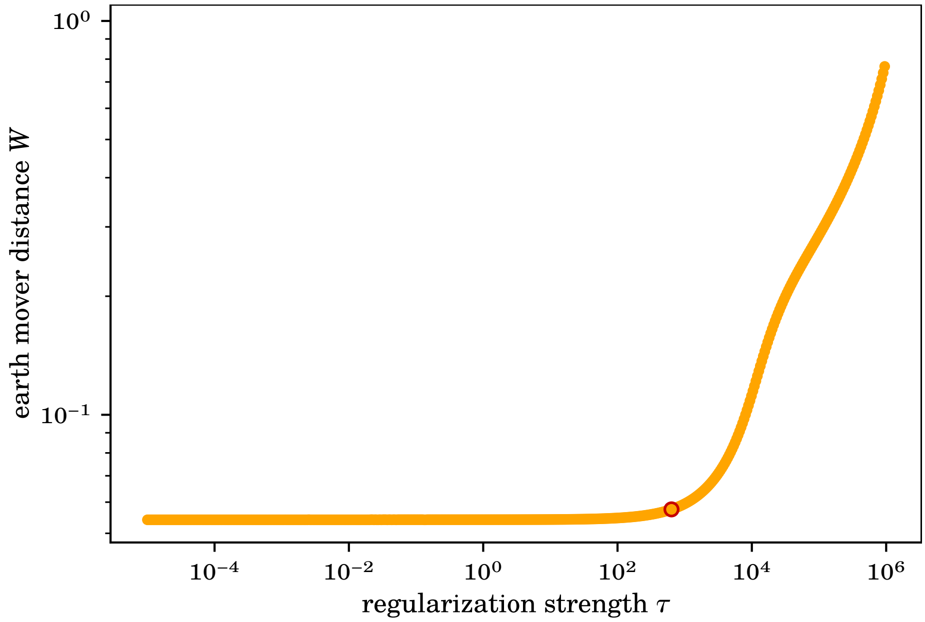

Figure 8.4: Earth mover distances of the simulated test sample unfolded with 500 regularization strengths τ. The red circle marks the chosen regularization strength at the minimal mean global correlation coefficient ρ.

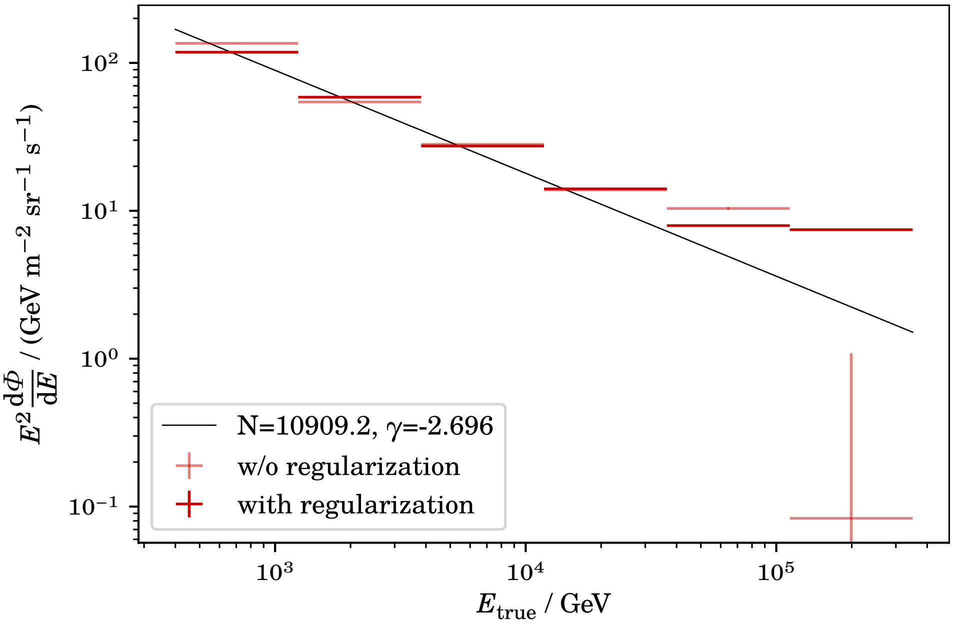

Figure 8.5: Proton spectrum obtained from the observation data sample. The events were classified, and their energy was estimated. The data points are unfolded once with and once without regularization. As a reference spectrum, the fit of the proton spectrum from section 5.3.1 is shown in black.

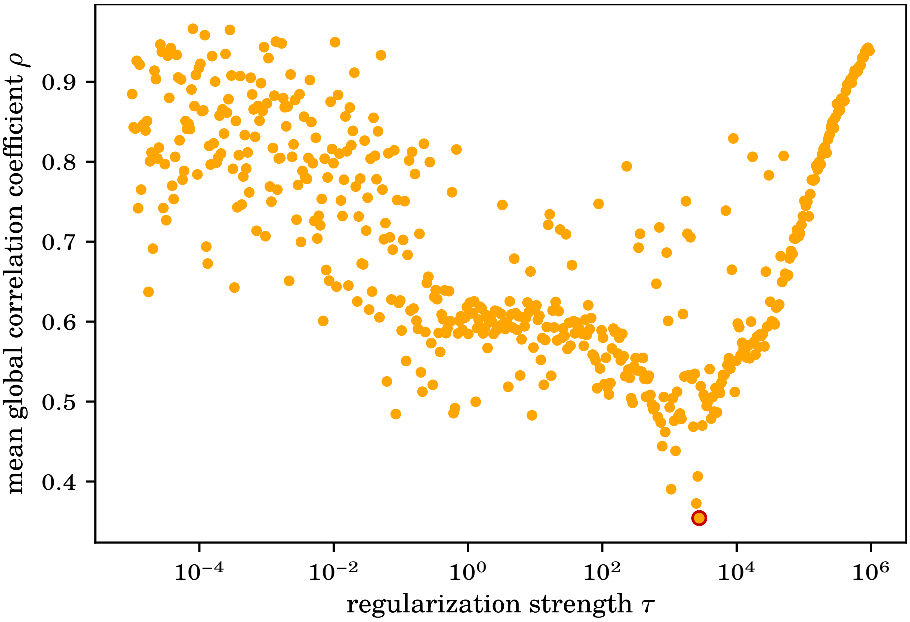

Figure 8.6: Mean global correlation coefficients of the data sample unfolded with 500 regularization strengths τ.

bin contents of the unfolded result.

$$\boldsymbol{G} = \boldsymbol{H}^{-1} \tag{8.12}$$

The diagonal of the covariance matrix shows the variance of the bin contents, while the root of the variance corresponds to the standard deviation σ.

$$\sigma(f_i) = \sqrt{\boldsymbol{H}_{ii}^{-1}} \tag{8.13}$$

Alternatively, a Bayesian approach can be used: a Markov Chain Monte Carlo (MCMC) algorithm repeatedly samples the posterior likelihood from the unfolding. The statistical error corresponds to the 1σ-quantile containing 68.2 % of the MCMC results. Here, the statistical error is estimated with the Hessian matrix since the results are more conservative than those from the MCMC approach.

However, in this analysis, the systematic error likely has a much more significant influence on the final error than the statistical error. The systematic error includes the error of prior assumptions, such as the assumptions about interaction cross sections in the simulation of air showers. For the spectra of gamma-ray sources, MAGIC assumes an error of 30 % by default. Naively, this assumption can be adopted for the proton spectrum. Future calculations of the systematic error of the proton spectrum are highly recommended.

8.3 Zenith Independency Check

This section studies whether the zenith distance affects the measurement of the proton flux on Earth. A zenith-dependent analysis is performed in the following. Therefore, the data is divided into subsets of certain zenith ranges and unfolded.

The expectation is that the flux is independent of the zenith angle of the measurement of the telescopes since a diffuse proton flux is predicted to be isotropically distributed over the sky without any hotspots or trends. Accordingly, the zenith-dependent analysis is simultaneously a cross-check of different subsamples.

For zenith-dependent unfolding, the test simulations are divided into bins of 5°. However, the determination of the IRFs is a challenge. Calculating the efficiency requires the number of all protons simulated; all simulated events are stored in event lists with the particle type and the primary's energy. In contrast, the zenith distances of these simulated events are missing since these are not stored to keep the data volume manageable. Therefore, there is no way to determine the number of simulated events per zenith range.

There are two possibilities to handle the problem. The first possibility considers the detector response being zenith-independent. This assumption is naive and serves more as a test: if the unfolded spectra show systematical behavior, the IRFs can not be assumed to be zenith-independent. The advantage of this method is that the IRFs are calculated with many simulations. Thus every energy bin contains enough events, especially in the two-dimensional histogram of the energy migration matrix.

The acceptance and the energy dispersion are estimated from the simulated protons of all zenith ranges, resulting in one migration matrix and one efficiency for all zenith bins. Thus it assumes that the energy dispersion and the acceptance remain approximately constant over the range in this analysis. The unfolded spectra for the zenith ranges are in figure 8.7 for the simulations and in figure 8.8 for the observation data.

The unfolded data points in figure 8.7, which are created with the simulation test data set, show a slight systematic effect: the smaller the zenith distance, the stronger the underestimation of the proton flux, the larger the zenith distance the stronger the overestimation. Unfolded data points from observations in figure 8.8 also show this systematic behavior even more strongly. In addition, the unfolded spectrum has more fluctuations and uncertainties, especially at small zenith distances. Overall, the zenith dependency of the results is visible, most likely because they are based on the naive assumption of the zenith independency of IRFs.

The second way to deal with the problem of missing information about the zenith distances of the simulated primary spectrum assumes the IRFs to be zenith-dependent and is based on a simple trick: the cut in the zenith areas is treated as a detector or analysis effect and is included in the acceptance. The acceptance is thus equivalent to one of a

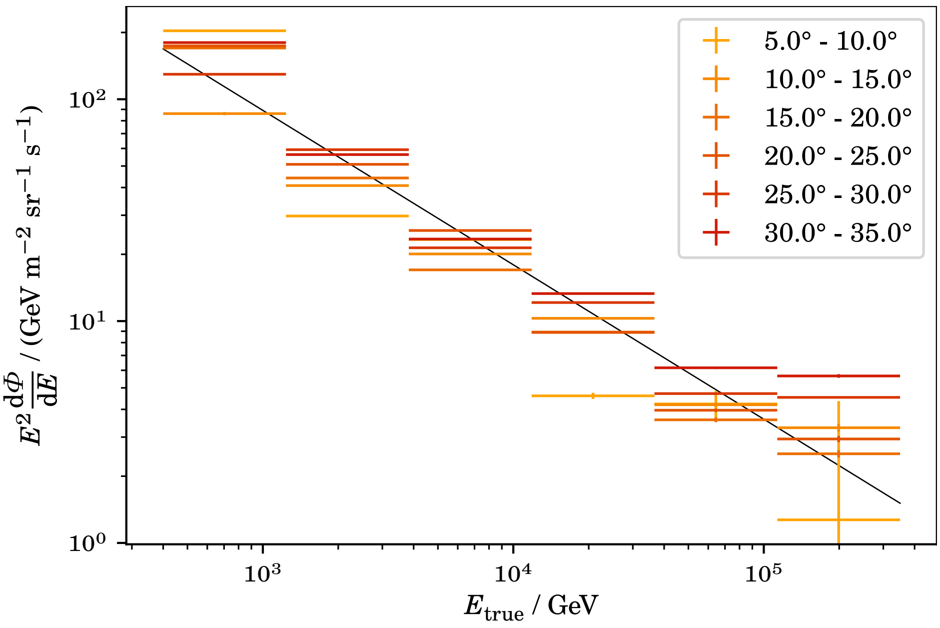

Figure 8.7: Unfolded spectra obtained from simulations for zenith ranges of 5° each. The energy dispersion and the acceptance are based on proton simulations of zenith distances from 5° to 35° with the assumption that they stay constant over the zenith range.

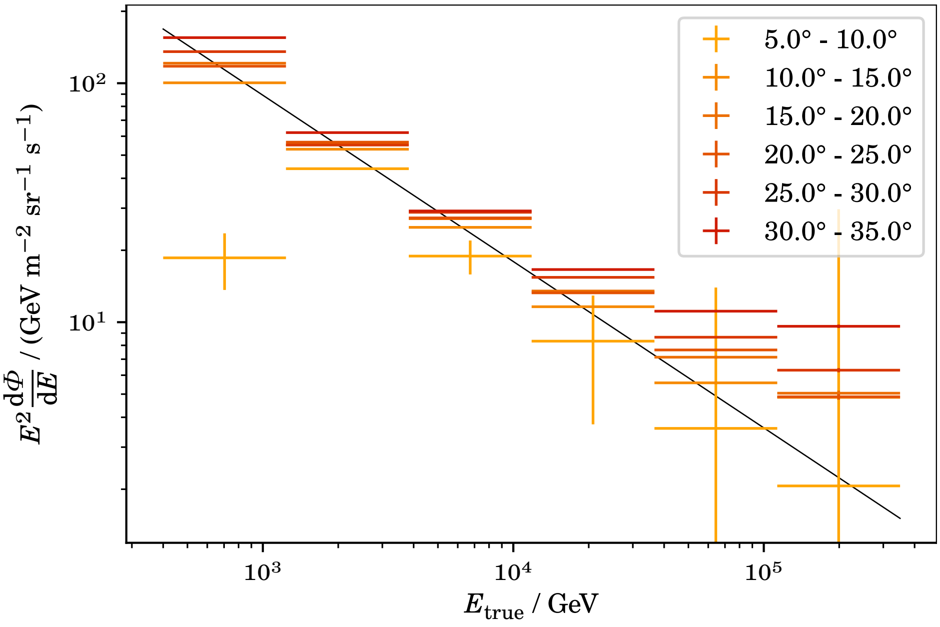

Figure 8.8: Unfolded spectra obtained from observation data for zenith ranges of 5° each. The energy dispersion and the acceptance are based on proton simulations of zenith distances from 5° to 35° with the assumption that they stay constant over the zenith range.

telescope that has operated over a large zenith range but has triggered only those events that occurred in the 5° range. Thus, the effective time in this method is the effective time of the measurement over the entire zenith range. This way, the zenith dependency of the acceptances is included. In this method, the migration matrix is calculated from the proton simulations of the corresponding zenith ranges, and lower statistics in the binned matrix are tolerated. The results of the second method are shown in figure 8.9 for the simulations and figure 8.10 for the observations. The unfolded data points in figure 8.9, created with the simulation test dataset, fluctuate around the assumed true spectrum without any visible systematical behavior. The unfolded data points from observations in figure 8.10 are consistent amongst each other and do not show any systematical behavior either. They match well with the results of the whole dataset in section 8.2.

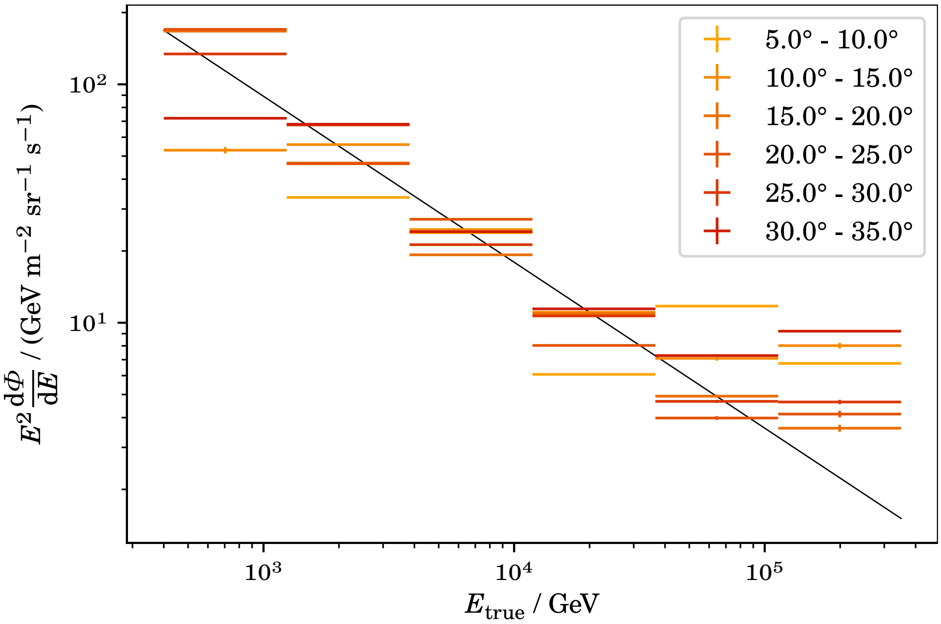

Figure 8.9: Unfolded spectra obtained from simulations for zenith ranges of 5° each. One energy matrix and zenith-dependent acceptance calculations underlay the data points.

While the first method turns out to be unsuitable for the unfolding, and the results confirm the zenith dependency of the detector response, the second method, which considers the zenith dependency of the IRFs, shows that the proton flux does not depend on the zenith distance of the observation.

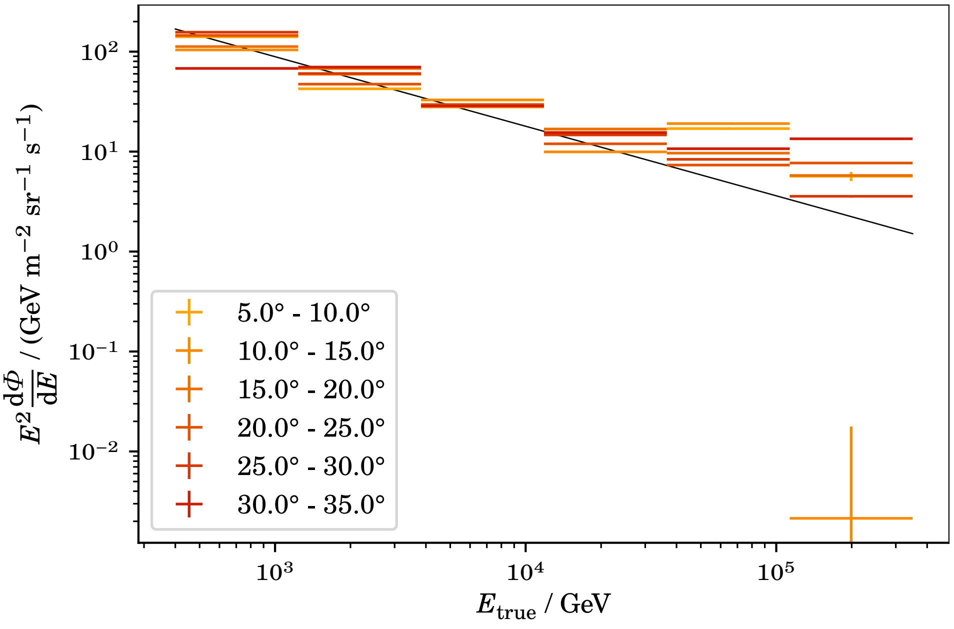

Figure 8.10: Unfolded spectra obtained from observation data for zenith ranges of 5° each. One energy matrix and zenith-dependent acceptance calculations underlay the data points.

8.4 Comparison to Previous Measurements

The proton spectrum obtained from the data measured with the MAGIC experiment is now compared with the spectra from previous experiments. For this purpose, the data points of all experiments, including the MAGIC data, are shown in figure 8.11. The unfolded spectrum of this work lies perfectly over the proton spectrum from the fit and is consistent with the measured data from the other experiments. It shows that MAGIC, with its indirect measurement method and originally designed to measure gamma radiation, is a suitable instrument for obtaining a proton spectrum.

The data points of the other experiments were calculated in much smaller energy bins than in this work. This is because this analysis is performed with an experiment that differs in many ways from all other experiments: the MAGIC experiment performs an indirect measurement, which means that the protons are not measured directly but indirectly via the air shower signatures. The separation of the different particle types is much more complex in this case: machine learning algorithms must be used to classify the particle types. Simulations that reproduce the complex shower structures are necessary for training. This requires a tremendous amount of resources and time, and at the time of this analysis, only the available simulations could be created. The limiting factor is the statistics of the simulations - which prohibits finer binning in the energy.

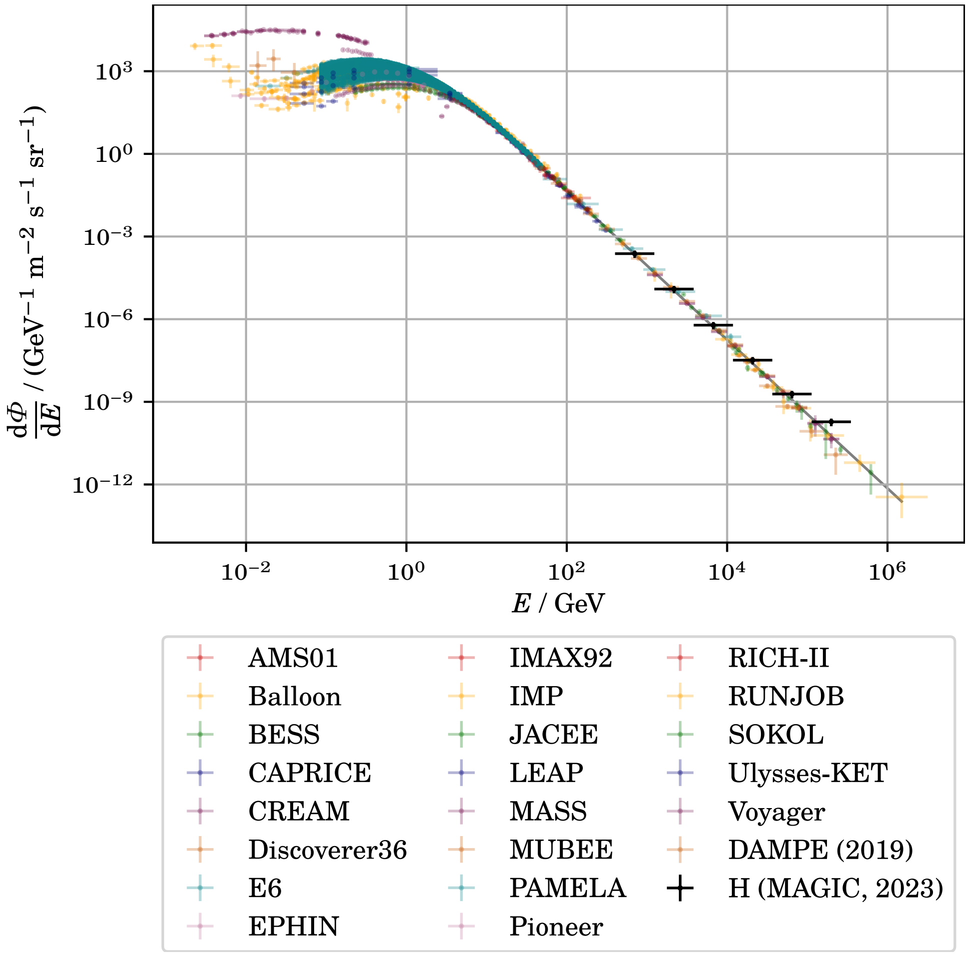

Figure 8.11: Proton spectrum of measurements with different experiments during the last decades including this work: the first proton spectrum of the MAGIC experiment (black).

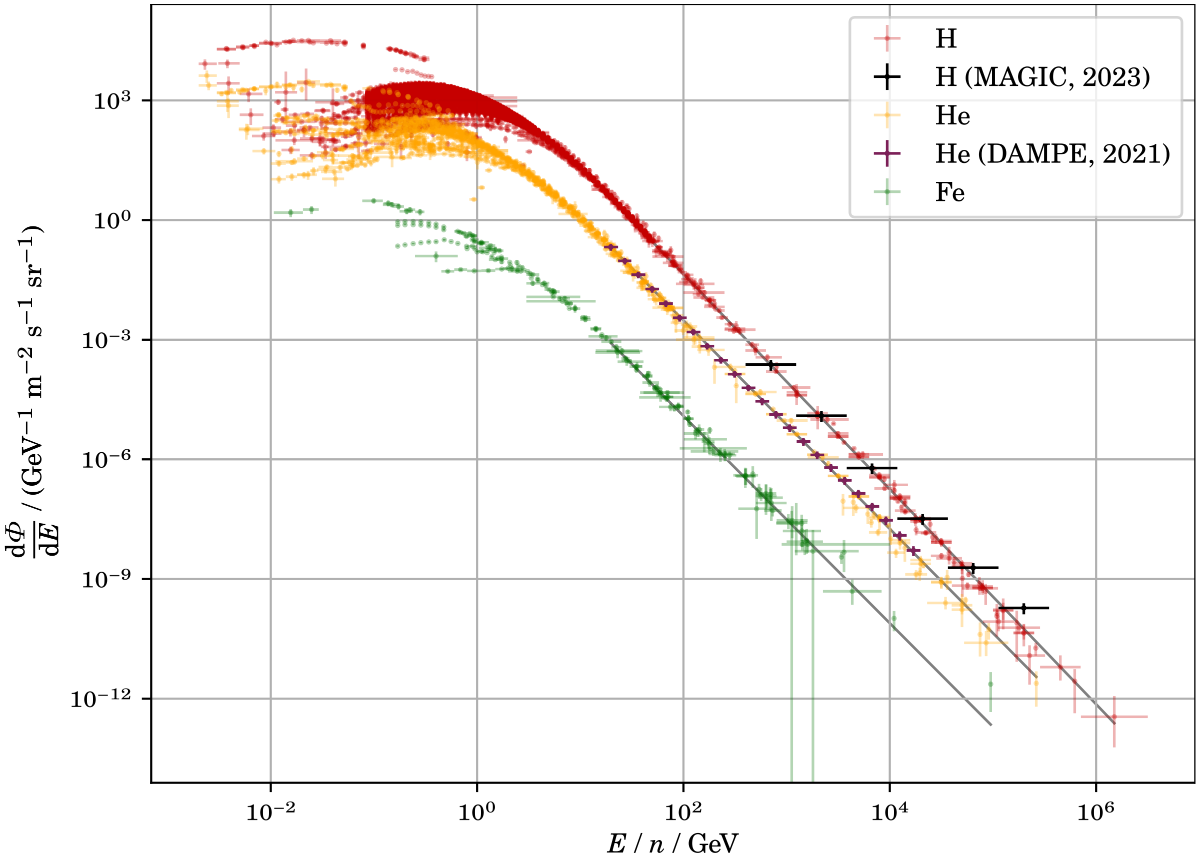

Figure 8.12: Cosmic-ray spectra of measurements with different experiments during the last decades including this work: the first proton spectrum of the MAGIC experiment (black).

Overall, the values calculated in this work fit into the graph. Here, only a deviation of 30 % is assumed for the systematic error estimation, as discussed in section 8.2.1 and typically done in the standard analysis of MAGIC. However, the simulation of hadronic showers is much more complex and unexplored than the simulation of gamma-induced showers. Accordingly, the systematic deviation will probably be larger than the assumed 30 %. Unlike in gamma-ray analyses, no measurement can be made for the background estimation. In gamma-ray analyses, a point in the sky is chosen where no gamma-ray source is assumed to estimate background from measurements. In the case of protons, the background, meaning all particles that are not protons, cannot be determined in any separate measurement. Therefore, the background estimation must be done with simulations. Here, too, new uncertainties arise which have to be investigated.

For illustration, the MAGIC proton spectrum points are added to the graphs with the three elements, protons, helium, and iron nuclei. The spectra of the three elements are shown in figure 8.12.

The MAGIC data points also fit well with the previous cosmic-ray measurements in this representation. Overall, the proton spectrum in this work confirms previous measurements and shows that MAGIC is a suitable cosmic-ray instrument, representative also for the future IACT generation, the Cherenkov Telescope Array (CTA).

Conclusion and Outlook

9

In this work, the spectrum of cosmic protons was calculated for the first time using data from the MAGIC experiment. The question of whether an experiment like MAGIC, which is specialized to observe gamma rays and records hadronic showers as background, can estimate a proton spectrum was answered: yes, in fact, MAGIC can produce a competitive proton spectrum.

This work sets up the entire analysis chain for calculating a proton spectrum from data measured by IACTs.

New proton, helium, and iron simulations were created for this analysis. They were compared to MAGIC observations containing mostly non-gamma-induced air showers. A data-simulation comparison was performed to check whether these new simulations agree with the observed data. First, the feature distributions of data and simulations were graphically presented, and the relative deviations of the bin contents were calculated. This visual comparison per attribute made it possible to check which values the data and simulations match best. Subsequently, a machine learning algorithm was used to test how well data and simulations could be separated. It was found that none of the features stood out and played a particular role in separability.

In the next part of the analysis, machine learning algorithms were used to classify the particles and reconstruct the energy. The particle classification took place in two steps: in the first step, the iron nuclei were separated from the rest, and in the second step, helium nuclei were separated from protons. The performance of the iron random forest proved to be very efficient, allowing many iron nuclei to be filtered from the datasets without excessively discarding protons. The helium random forest performed slightly weaker, as expected. This can be explained by the similarity in the shower structure of light nuclei. Subsequently, energy-dependent classification thresholds were chosen based on weighted simulations. The criterion for the cut choice was the efficiency of the surviving protons in the dataset. It was shown that events with a high iron score also have a high helium score at the same time. It can be concluded that the helium classifier tends to filter out other heavier nuclei as well. In this analysis, only simulations of proton helium and iron nuclei were available. However, it can be assumed that the helium classifier could also separate other particles, such as oxygen and carbon, from protons.

The energy estimator was trained with simulations of proton events and has shown good performance in the validation. Additional regressors for helium and iron were trained and tested to show the possibility of unfolding their spectra in the future.

Next, an unfolding with consideration of the background was performed. For this purpose, the instrument response functions were first calculated from the proton simulations. In the unfolding, not only the efficiency and energy dispersion were considered, but also the background of helium and iron nuclei misclassified as protons. The background of surviving helium and iron nuclei was estimated using simulations weighted with the spectra measured by previous experiments. The unfolding took place with regularization. The regularization strength was determined using the mean global correlation coefficient. Unfolding was tested with a test dataset that confirmed the unfolded and regularized data points. An additional zenith-dependent study confirmed the result from the unfolding before and proved the zenith independency of the proton flux.

The proton spectrum of this work aligns with previous data from other experiments. From an astrophysical point of view, this work confirms the current knowledge about the proton flux and fits into the measurements of many previous experiments. Thus, it was shown that the MAGIC experiment is capable of performing a thorough analysis of cosmic protons. More simulations need to be created to generate more data points for the spectrum. The limit of this work was not the statistics in data but simulations. The aim of this work was to set up an analysis chain for proton spectrum calculations and a limited simulation dataset was used because of the enormous computational effort. Especially the simulation of iron nuclei requires an immense amount of resources. With newly available resources, additional proton, helium, and iron simulations could be produced and included in this analysis in future work. Further elements like oxygen and carbon could further improve the analysis.

At the same time, it was proven that the simulations of protons, helium, and iron nuclei themselves were successful. Simulations of hadronic showers have been challenging because not all interaction cross sections are known with small uncertainties. The muon puzzle is one of the largest uncertainties in understanding the interaction processes in hadronic showers. The number of measured muons in air showers is significantly higher than that of predicted muons. One assumption is that the contribution of neutral pions in air showers is wrongly estimated, which is responsible for the electromagnetic component in air showers. This work has shown no evidence of a discrepancy between measurement and simulation. Therefore there is no evidence that the electromagnetic shower component in hadronic showers in the energy range of several hundred GeV to several hundred TeV contributes to the muon puzzle.

However, it is exciting that it is possible to use IACTs to study cosmic rays. New possibilities will open up soon with the next generation of Cherenkov telescopes, the CTA.

In addition to higher sensitivity and a wider energy range, the array will also provide the possibility to detect sufficient muon events from the muonic shower component to include them in the analysis. On the one hand, this will significantly improve the classification of the primary particles, and on the other hand, it will be a great help in the search for the solution of the muon puzzle itself.

This work impressively showed that IACTs like MAGIC provide data that can be reconstructed with machine learning algorithms and unfolded, including the hadronic background, to produce the proton spectrum. Analogously, this analysis can be used to calculate the spectra of heavier elements as well.

Bibliography

1. M. Aartsen et al. (IceCube Collaboration). “The IceCube Neutrino Observatory: instrumentation and online systems”. *Journal of Instrumentation* 12:03, 2017, P03012–P03012.

   DOI: 10.1088/1748-0221/12/03/p03012

2. M. Aartsen et al. (IceCube, *Fermi*-LAT, MAGIC, AGILE, ASAS-SN, HAWC, H. E. S. S., *INTEGRAL*, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, *Swift*/NuSTAR, VERITAS, VLA Collaboration). “Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A”. *Science* 361:6398, 2018.

   DOI: 10.1126/science.aat1378

3. R. Abbasi et al. “IceTop: The surface component of IceCube”. *Nuclear Instruments and Methods* 700, 2013, pages 188–220.

   DOI: 10.1016/j.nima.2012.10.067

4. B. P. Abbott et al. (LIGO and Virgo Collaboration). “Observation of Gravitational Waves from a Binary Black Hole Merger”. *Physical Review Letters* 116, 6 2016, page 061102.

   DOI: 10.1103/PhysRevLett.116.061102

5. K. Abe et al. “Measurements of Cosmic-Ray Proton and Helium Spectra from the BESS-Polar Long-duration Balloon Flights over Antarctica”. *The Astrophysical Journal* 822:2, 65, 2016, page 65.

   DOI: 10.3847/0004-637X/822/2/65. ARXIV: 1506.01267 [astro-ph.HE]

6. A. U. Abeysekara et al. “Daily Monitoring of TeV Gamma-Ray Emission from Mrk 421, Mrk 501, and the Crab Nebula with HAWC”. *The Astrophysical Journal* 841:2, 2017, page 100.

   DOI: 10.3847/1538-4357/aa729e

7. V. A. Acciari et al. (MAGIC Collaboration). “Teraelectronvolt emission from the γ-ray burst GRB 190114C”, 2020.

   DOI: 10.48550/ARXIV.2006.07249

8. O. Adriani et al. “Measurements of cosmic-ray proton and helium spectra with the PAMELA calorimeter”. *Advances in Space Research* 51:2, 2013, pages 219–226.
   DOI: `10.1016/j.asr.2012.09.029`
9. O. Adriani et al. “Time Dependence of the Proton Flux Measured by PAMELA during the 2006 July-2009 December Solar Minimum”. *The Astrophysical Journal* 765:2, 91, 2013, page 91.
   DOI: `10.1088/0004-637X/765/2/91`. ARXIV: `1301.4108 [astro-ph.HE]`
10. F. Aharonian et al. “A deep spectromorphological study of the gamma-ray emission surrounding the young massive stellar cluster Westerlund 1”. *Astronomy and Astrophysics* 666, 2022, A124.
    DOI: `10.1051/0004-6361/202244323`
11. H. S. Ahn et al. “Energy Spectra of Cosmic-ray Nuclei at High Energies”. *The Astrophysical Journal* 707:1, 2009, pages 593–603.
    DOI: `10.1088/0004-637X/707/1/593`. ARXIV: `0911.1889 [astro-ph.HE]`
12. H. Aizu et al. “Heavy Nuclei in the Primary Cosmic Radiation at Prince Albert, Canada. II”. *Physical Review* 121:4, 1961, pages 1206–1218.
    DOI: `10.1103/PhysRev.121.1206`
13. J. Albrecht et al. “The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider”. *Astrophysics and Space Science* 367:3, 2022.
    DOI: `10.1007/s10509-022-04054-5`
14. J. Aleksić et al. “The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system”. *Astroparticle Physics* 72, 2016, pages 61–75.
    DOI: `10.1016/j.astropartphys.2015.04.004`
15. J. Aleksić et al. “The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula”. *Astroparticle Physics* 72, 2016, pages 76–94. ISSN: 0927-6505.
    DOI: `10.1016/j.astropartphys.2015.02.005`
16. F. Alemanno et al. “Measurement of the Cosmic Ray Helium Energy Spectrum from 70 GeV to 80 TeV with the DAMPE Space Mission”. *Physical Review Letters* 126:20, 2021.
    DOI: `10.1103/physrevlett.126.201102`
17. F. Arqueros et al. (HEGRA Collaboration). “Energy spectrum and chemical composition of cosmic rays between 0.3-PeV and 10-PeV determined from the Cherenkov light and charged particle distributions in air showers”. *Astronomy and Astro-*

physics 359, 2000, pages 682–694.
ARXIV: astro-ph/9908202

18. K. Asakimori et al. “Cosmic-Ray Proton and Helium Spectra: Results from the JACEE Experiment”. *The Astrophysical Journal* 502:1, 1998, pages 278–283.
DOI: 10.1086/305882

19. W. B. Atwood et al. “The Large Area Telescope on the *Fermi Gamma-Ray Space Telescope* mission”. *The Astrophysical Journal* 697:2, 2009, pages 1071–1102.
DOI: 10.1088/0004-637x/697/2/1071

20. G. D. Badhwar et al. “Measurements of the Low-Energy Cosmic Radiation during the Summer of 1966”. *Physical Review* 163:5, 1967, pages 1327–1342.
DOI: 10.1103/PhysRev.163.1327

21. V. K. Balasubrahmanyan et al. “The Multiply Charged Primary Cosmic Radiation at Solar Minimum 1965”. *Journal of Geophysical Research* 71, 1966, page 1771.
DOI: 10.1029/JZ071i007p01771

22. J. Ballet et al. (The Fermi-LAT Collaboration). *Fermi Large Area Telescope Fourth Source Catalog Data Release 4 (4FGL-DR4)*. 2023.
DOI: 10.48550/ARXIV.2307.12546

23. S. D. Barthelmy et al. “The Burst Alert Telescope (BAT) on the SWIFT Midex Mission”. *Space Science Reviews* 120:3-4, 2005, pages 143–164.
DOI: 10.1007/s11214-005-5096-3

24. J. J. Beatty, M. Garcia-Munoz, and J. A. Simpson. “The cosmic-ray spectra of H-1, H-2, and He-4 as a test of the origin of the hydrogen superfluxes at solar minimum modulation”. *The Astrophysical Journal* 294, 1985, pages 455–462.
DOI: 10.1086/163311

25. J. J. Beatty et al. “The Cosmic-Ray 3He/ 4He Ratio from 100 to 1600 MeV AMU -1”. *The Astrophysical Journal* 413, 1993, page 268.
DOI: 10.1086/172994

26. R. Bellotti et al. “Balloon measurements of cosmic ray muon spectra in the atmosphere along with those of primary protons and helium nuclei over midlatitude”. *Physical Review D* 60:5, 052002, 1999, page 052002.
DOI: 10.1103/PhysRevD.60.052002. ARXIV: hep-ex/9905012 [hep-ex]

27. V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Vieweg+Teubner Verlag, 1998.
DOI: 10.1007/978-3-663-05690-4

28. M. Boezio et al. “The cosmic-ray proton and helium spectra measured with the CAPRICE98 balloon experiment”. *Astroparticle Physics* 19:5, 2003, pages 583–604.
DOI: `10.1016/S0927-6505(02)00267-0`. ARXIV: `astro-ph/0212253 [astro-ph]`

29. J. Buckley et al. “A New Measurement of the Flux of the Light Cosmic-Ray Nuclei at High Energies”. *The Astrophysical Journal* 429, 1994, page 736.
DOI: `10.1086/174357`

30. Z. Cao et al. *The Large High Altitude Air Shower Observatory (LHAASO) Science Book (2021 Edition)*. 2022.
ARXIV: `1905.02773 [astro-ph.HE]`

31. J. Chang et al. “The DArk Matter Particle Explorer mission”. *Astroparticle Physics* 95, 2017, pages 6–24.
DOI: `10.1016/j.astropartphys.2017.08.005`

32. E. G. Clayton, T. G. Guzik, and J. P. Wefel. “CRRES Measurements of Energetic Helium During the 1990-1991 Solar Maximum”. *Solar Physics* 195:1, 2000, pages 175–194.
DOI: `10.1023/A:1005251630568`

33. G. M. Comstock, C. Y. Fan, and J. A. Simpson. “Energy Spectra and Abundances of the Cosmic-Ray Nuclei Helium to Iron from the Ogo-I Satellite Experiment”. *The Astrophysical Journal* 155, 1969, page 609.
DOI: `10.1086/149895`

34. G. M. Courtier and A. D. Lenney. “The flux of the cosmic ray hydrogen and helium nuclei at Kiruna, Sweden”. *Planetary and Space Science* 14:6, 1966, pages 503–518.
DOI: `10.1016/0032-0633(66)90006-7`

35. A. C. Cummings et al. “Galactic Cosmic Rays in the Local Interstellar Medium: Voyager 1 Observations and Model Results”. *The Astrophysical Journal* 831:1, 18, 2016, page 18.
DOI: `10.3847/0004-637X/831/1/18`

36. A. De Angelis and M. Pimenta. *Introduction to Particle and Astroparticle Physics. Multimessenger Astronomy and its Particle Physics Foundations*. Undergraduate Lecture Notes in Physics. Springer, 2018.
DOI: `10.1007/978-3-319-78181-5`

37. H. P. Dembinski et al. (EAS-MSU, IceCube, KASCADE-Grande, NEVOD-DECOR, Pierre Auger, SUGAR, Telescope Array, Yakutsk EAS Array Collaboration). “Report on Tests and Measurements of Hadronic Interaction Properties with Air Showers”. *EPJ Web of Conferences* 210, 2019. Ed. by I. Lhenry-Yvon et al.,

page 02004.
DOI: 10.1051/epjconf/201921002004. ARXIV: 1902.08124 [astro-ph.HE]

38. H. Dembinski et al. *scikit-hep / iminuit*. 2023.
DOI: 10.5281/ZENODO.8191656

39. V. A. Derbina et al. “Cosmic-Ray Spectra and Composition in the Energy Range of 10-1000 TeV per Particle Obtained by the RUNJOB Experiment”. *The Astrophysical Journal* 628:1, 2005, pages L41–L44.
DOI: 10.1086/432715

40. J. H. Derrickson et al. “A measurement of the absolute energy spectra of galactic cosmic rays during the 1976-77 solar minimum”. *International Journal of Radiation Applications and Instrumentation* 20:3, 1992, pages 415–421

41. E. Diehl et al. “The energy spectrum of cosmic-ray protons and helium near 100 GeV”. *Astroparticle Physics* 18:5, 2003, pages 487–500.
DOI: 10.1016/S0927-6505(02)00157-3

42. N. Durgaprasad, C. E. Fichtel, and D. E. Guss. “Solar Modulation of Cosmic Rays and Its Relationship to Proton and Helium Fluxes, Interstellar Travel, and Interstellar Secondary Production”. *Journal of Geophysical Research* 72, 1967, page 2765.
DOI: 10.1029/JZ072i011p02765

43. M. A. Duvernois et al. “The Isotopic Composition of Galactic Cosmic-Ray Elements from Carbon to Silicon: The Combined Release and Radiation Effects Satellite Investigation”. *The Astrophysical Journal* 466, 1996, page 457.
DOI: 10.1086/177524

44. J. J. Engelmann et al. “Charge composition and energy spectra of cosmic-ray nuclei for elements from Be to Ni - Results from HEAO-3-C2.” *Astronomy and Astrophysics* 233, 1990, pages 96–111

45. A. Engler et al. “Primary cosmic ray α-particles — II”. *Il Nuovo Cimento* 20:6, 1961, pages 1157–1165.
DOI: 10.1007/BF02732525

46. A. Engler et al. “Primary cosmic-ray α-particles — I”. *Il Nuovo Cimento* 19:6, 1961, pages 1090–1099.
DOI: 10.1007/BF02731385

47. J. A. Esposito et al. “The ALICE instrument and the measured cosmic ray elemental abundances”. *Astroparticle Physics* 1:1, 1992, pages 33–45.
DOI: 10.1016/0927-6505(92)90007-M

48. F. Foster and B. E. Schrautemeier. "Energy spectrum and geomagnetic cut-off of primary cosmic-ray α-particles near 41 °N mag". *Nuovo Cimento A Serie* 47:2, 1967, pages 189–194.
    DOI: 10.1007/BF02818342
49. P. H. Fowler et al. "The low energy end of the cosmic ray spectrum of alpha-particles". *Philosophical Magazine* 2:14, 1957, pages 157–175.
    DOI: 10.1080/14786435708243805
50. P. S. Freier, E. P. Ney, and P. H. Fowler. "Cosmic Rays and the Sunspot Cycle: Primary α-Particle Intensity at Sunspot Maximum". *Nature* 181:4619, 1958, pages 1319–1321.
    DOI: 10.1038/1811319a0
51. P. S. Freier and C. J. Waddington. "Intensity of 80- to 200-Mev Protons over Fort Churchill on August 26, 1960". *Journal of Geophysical Research* 70:9, 1965, pages 2111–2117.
    DOI: 10.1029/JZ070i009p02111
52. P. S. Freier and C. J. Waddington. "Singly and doubly charged particles in the primary cosmic radiation". *Journal of Geophysical Research* 73:13, 1968, pages 4261–4271.
    DOI: 10.1029/JA073i013p04261
53. P. S. Freir and C. J. Waddington. "Electron, Hydrogen Nuclei, and Helium Nuclei Observed in Primary Cosmic Radiation during 1963". *Journal of Geophysical Research* 70:23, 1965, pages 5753–5768.
    DOI: 10.1029/JZ070i023p05753
54. D. Fuhrmann et al. (KASCADE-Grande Collaboration). *KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays*. 2013.
    ARXIV: 1308.2098 [astro-ph.HE]
55. F. Gahbauer et al. "A New Measurement of the Intensities of the Heavy Primary Cosmic-Ray Nuclei around 1 TeV amu^{-1}". *The Astrophysical Journal* 607:1, 2004, pages 333–341.
    DOI: 10.1086/383304
56. T. K. Gaisser, R. Engel, and E. Resconi. *Cosmic Rays and Particle Physics*. Cambridge University Press, 2016.
    DOI: 10.1017/cbo9781139192194
57. J. R. García et al. *Status of the new Sum-Trigger system for the MAGIC telescopes*. 2014.
    ARXIV: 1404.4219 [astro-ph.IM]

58. M. Garcia-Munoz, G. M. Mason, and J. A. Simpson. “The anomalous He-4 component in the cosmic-ray spectrum at below approximately 50 MeV per nucleon during 1972-1974”. *The Astrophysical Journal* 202, 1975, page 265.
    DOI: `10.1086/153973`
59. T. L. Garrard, E. C. Stone, and R. E. Vogt. “Solar Modulation of Cosmic-Ray Protons and He Nuclei”. In: *International Cosmic Ray Conference*. Vol. 2. International Cosmic Ray Conference. 1973, page 732
60. M. Gaug et al. *Atmospheric Monitoring for the MAGIC Telescopes*. 2014.
    ARXIV: 1403.5083 [astro-ph.IM]
61. *GCN: NASA’s Time-Domain and Multimessenger Alert System*. `https://gcn.nasa.gov/`.
    Note: Accessed: 2023-06-21
62. N. M. Gerasimova, A. I. Nikishov, and I. L. Rosenthal. “Interation of Nuclei and Photons of High Energies with a Thermal Radiations in the Universe”. *Journal of the Physical Society of Japan Supplement* 17, 1962, page 175
63. R. C. Gilmore, F. Prada, and J. Primack. “Modelling gamma-ray burst observations by Fermi and MAGIC including attenuation due to diffuse background light”. *Monthly Notices of the Royal Astronomical Society* 402:1, 2009, pages 565–574.
    DOI: `10.1111/j.1365-2966.2009.15909.x`
64. G. Gloeckler. “Solar Modulation of the Low-Energy Galactic Helium Spectrum as Observed on the Imp 1 Satellite”. *Journal of Geophysical Research* 70:21, 1965, pages 5333–5343.
    DOI: `10.1029/JZ070i021p05333`
65. K. Greisen. “End to the Cosmic-Ray Spectrum?” *Physical Review Letters* 16:17, 1966, pages 748–750.
    DOI: `10.1103/physrevlett.16.748`
66. D. Heck et al. “CORSIKA: A Monte Carlo code to simulate extensive air showers”. en, 1998.
    DOI: `10.5445/IR/270043064`
67. V. Hess. “Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten”. *Physikalische Zeitschrift* 13, 1912, pages 1084–1091
68. A. M. Hillas. “Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei”. In: *19th International Cosmic Ray Conference (ICRC19)*. Vol. 3. International Cosmic Ray Conference. 1985, page 445

69. J. Hinton. “The status of the HESS project”. *New Astronomy Reviews* 48:5-6, 2004, pages 331–337.
    DOI: `10.1016/j.newar.2003.12.004`
70. J. Holder et al. “VERITAS: Status and Highlights”, 2011.
    DOI: `https://doi.org/10.48550/arxiv.1111.1225`
71. *Hubble Space Telescope*. `https://hubblesite.org/home`.
    Note: Accessed: 2023-06-21
72. M. Ichimura et al. “Observation of heavy cosmic-ray primaries over the wide energy range from ∼100 GeV/particle to ∼100 TeV/particle: Is the celebrated “knee” actually so prominent?” *Physical Review D* 48:5, 1993, pages 1949–1975.
    DOI: `10.1103/PhysRevD.48.1949`
73. I. P. Ivanenko et al. “Energy Spectra of Cosmic Rays above 2 TeV as Measured by the ’SOKOL’ Apparatus”. In: *23rd International Cosmic Ray Conference (ICRC23), Volume 2*. Vol. 2. International Cosmic Ray Conference. 1993, page 17
74. F. James and M. Roos. “Minuit - a system for function minimization and analysis of the parameter errors and correlations”. *Computer Physics Communications* 10:6, 1975, pages 343–367.
    DOI: `10.1016/0010-4655(75)90039-9`
75. E. Kamioka et al. “Azimuthally controlled observation of heavy cosmic-ray primaries by means of the balloon-borne emulsion chamber”. *Astroparticle Physics* 6:2, 1997, pages 155–167.
    DOI: `10.1016/S0927-6505(96)00051-5`
76. A. Kolmogorov-Smirnov, A. N. Kolmogorov, and M. Kolmogorov. “Sulla determinazione empirica di una legge di distribuzione”. In: 1933
77. H. A. Krimm et al. “The *Swift* /BAT Hard X-Ray Transient Monitor”. *The Astrophysical Journal Supplement Series* 209:1, 2013, page 14.
    DOI: `10.1088/0067-0049/209/1/14`
78. P. Kühl, R. Gómez-Herrero, and B. Heber. “Annual Cosmic Ray Spectra from 250 MeV up to 1.6 GeV from 1995 - 2014 Measured with the Electron Proton Helium Instrument onboard SOHO”. *Solar Physics* 291:3, 2016, pages 965–974.
    DOI: `10.1007/s11207-016-0879-0`. ARXIV: `1603.00676 [physics.space-ph]`
79. K. A. Lave et al. “Galactic Cosmic-Ray Energy Spectra and Composition during the 2009-2010 Solar Minimum Period”. *The Astrophysical Journal* 770:2, 117, 2013, page 117.
    DOI: `10.1088/0004-637X/770/2/117`

80. J. A. Lezniak and W. R. Webber. “Solar modulation of cosmic ray protons, helium nuclei, and electrons: A comparison of experiment with theory”. *Journal of Geophysical Research* 76:7, 1971, page 1605.
    DOI: 10.1029/JA076i007p01605

81. J. A. Lezniak and W. R. Webber. “The charge composition and energy spectra of cosmic-ray nuclei from 3000 MeV per nucleon to 50 GeV per nucleon.” *The Astrophysical Journal* 223, 1978, pages 676–696.
    DOI: 10.1086/156301

82. H. Ling and K. Okada. “An Efficient Earth Mover’s Distance Algorithm for Robust Histogram Comparison”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29:5, 2007, pages 840–853.
    DOI: 10.1109/TPAMI.2007.1058

83. J. Linsley and L. Scarsi. “Arrival Times of Air Shower Particles at Large Distances from the Axis”. *Physical Review* 128:5, 1962, pages 2384–2392.
    DOI: 10.1103/physrev.128.2384

84. G. Louppe. *Understanding Random Forests: From Theory to Practice*. 2015.
    ARXIV: 1407.7502 [stat.ML]

85. G. H. Ludwig. “Cosmic-Ray Instrumentation in the First U. S. Earth Satellite”. *Review of Scientific Instruments* 30:4, 1959, pages 223–229.
    DOI: 10.1063/1.1716522

86. R. C. Maehl et al. “Energy Spectra of Cosmic Ray Nuclei: $4 \leq Z \leq 26$ and $0.3 \leq E \leq 2$ GeV amu^{-1}”. *Astrophysics and Space Science* 47:1, 1977, pages 163–184.
    DOI: 10.1007/BF00651365

87. K. Mannheim, R. J. Protheroe, and J. P. Rachen. “Cosmic ray bound for models of extragalactic neutrino production”. *Physical Review D* 63:2, 2000.
    DOI: 10.1103/physrevd.63.023003

88. J. Marquardt and B. Heber. “Galactic cosmic ray hydrogen spectra and radial gradients in the inner heliosphere measured by the HELIOS Experiment 6”. *Astronomy and Astrophysics* 625, A153, 2019, A153.
    DOI: 10.1051/0004-6361/201935413. ARXIV: 1905.01052 [physics.space-ph]

89. M. Martucci et al. “Proton Fluxes Measured by the PAMELA Experiment from the Minimum to the Maximum Solar Activity for Solar Cycle 24”. *The Astrophysical Journal Letters* 854:1, L2, 2018, page L2.
    DOI: 10.3847/2041-8213/aaa9b2. ARXIV: 1801.07112 [physics.space-ph]

90. G. M. Mason. “Interstellar Propagation of Galactic Cosmic-Ray Nuclei $2 \leq Z \leq 8$ in the Energy Range 10-1000 MeV Per Nucleon”. *The Astrophysical Journal* 171, 1972, page 139.
DOI: `10.1086/151266`
91. F. B. McDonald and W. R. Webber. “Changes in the Low-Rigidity Primary Cosmic Radiation during the Large Forbush Decrease of May 12, 1959”. *Journal of Geophysical Research* 65, 1960, page 767.
DOI: `10.1029/JZ065i002p00767`
92. F. B. McDonald. “Primary Cosmic-Ray Intensity near Solar Maximum”. *Physical Review* 116:2, 1959, pages 462–463.
DOI: `10.1103/PhysRev.116.462`
93. F. B. McDonald and W. R. Webber. “Proton Component of the Primary Cosmic Radiation”. *Physical Review* 115:1, 1959, pages 194–205.
DOI: `10.1103/PhysRev.115.194`
94. W. Menn et al. “The Absolute Flux of Protons and Helium at the Top of the Atmosphere Using IMAX”. *The Astrophysical Journal* 533:1, 2000, pages 281–297.
DOI: `10.1086/308645`
95. P. Meyer and R. Vogt. “Primary Cosmic Ray and Solar Protons. II”. *Physical Review* 129:5, 1963, pages 2275–2279.
DOI: `10.1103/PhysRev.129.2275`
96. G. Minagawa. “The abundances and energy spectra of cosmic ray iron and nickel at energies from 1 to 10 GeV per AMU”. *The Astrophysical Journal* 248, 1981, pages 847–855.
DOI: `10.1086/159210`
97. R. Minkowski. “Spectra of Supernovae”. *Publications of the Astronomical Society of the Pacific* 53, 1941, page 224
98. A. Moralejo et al. *MARS, the MAGIC Analysis and Reconstruction Software*. 2009.
DOI: `10.48550/ARXIV.0907.0943`
99. K. Morik and W. Rhode, eds. *Discovery in Physics*. De Gruyter, 2022.
DOI: `10.1515/9783110785968`
100. D. Mueller et al. “Energy Spectra and Composition of Primary Cosmic Rays”. *The Astrophysical Journal* 374, 1991, page 356.
DOI: `10.1086/170125`

101. *NASA: National Aeronautics and Space Administration*. `https://www.nasa.gov/`.

Note: Accessed: 2023-08-10

102. L. Nellen, K. Mannheim, and P. L. Biermann. “Neutrino production through hadronic cascades in AGN accretion disks”. *Physical Review D* 47, 1993, pages 5270–5274.

DOI: `10.1103/PhysRevD.47.5270`. ARXIV: `hep-ph/9211257`

103. M. Nöthe, K. Brügge, and J. Buß. *fact-project / aict-tools: v0.27.1 – 2021-05-05*. 2021.

DOI: `10.5281/ZENODO.3338081`

104. M. Nöthe et al. *cta-observatory / pyirf: v0.8.0 – 2022-11-25*. Version v0.8.0. 2022.

DOI: `10.5281/zenodo.7362244`

105. A. Obermeier et al. “Energy Spectra of Primary and Secondary Cosmic-Ray Nuclei Measured with TRACER”. *The Astrophysical Journal* 742:1, 14, 2011, page 14.

DOI: `10.1088/0004-637X/742/1/14`. ARXIV: `1108.4838 [astro-ph.HE]`

106. J. F. Ormes and W. R. Webber. “Proton and helium nuclei cosmic-ray spectra and modulations between 100 and 2000 Mev/nucleon”. *Journal of Geophysical Research* 73:13, 1968, pages 4231–4245.

DOI: `10.1029/JA073i013p04231`

107. K. Pearson. “X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50:302, 1900, pages 157–175.

DOI: `10.1080/14786440009463897`

108. F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”, 2012.

DOI: `10.48550/ARXIV.1201.0490`

109. C.-L. Project et al. *Observations of the Crab Nebula and Pulsar with the Large-Sized Telescope Prototype of the Cherenkov Telescope Array*. 2023.

DOI: `10.48550/ARXIV.2306.12960`

110. M. Punch et al. “Detection of TeV photons from the active galaxy Markarian 421”. *Nature* 358:6386, 1992, pages 477–478.

DOI: `10.1038/358477a0`

111. A. Ramdas, N. Garcia, and M. Cuturi. *On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests*. 2015.

ARXIV: `1509.02237 [math.ST]`

112. M. Rao and B. Sreekantan. *Extensive Air Showers*. World Scientific, 1998. ISBN: 9789810228880

113. C. Rastoin et al. "Time and space variations of the Galactic cosmic ray electron spectrum in the 3-D heliosphere explored by Ulysses." *Astronomy and Astrophysics* 307, 1996, pages 981–995

114. M. J. Ryan, J. F. Ormes, and V. K. Balasubrahmanyan. "Cosmic-Ray Proton and Helium Spectra above 50 GeV". *Physical Review Letters* 28:15, 1972, pages 985–988.
    DOI: 10.1103/PhysRevLett.28.985

115. T. A. Rygg, J. J. Ogallagher, and J. A. Earl. "Modulation of cosmic ray protons and helium nuclei near solar maximum". *Journal of Geophysical Research* 79:28, 1974, page 4127.
    DOI: 10.1029/JA079i028p04127

116. T. A. Rygg and J. A. Earl. "Balloon measurements of cosmic ray protons and helium over half a solar cycle 1965-1969". *Journal of Geophysical Research* 76:31, 1971, page 7445.
    DOI: 10.1029/JA076i031p07445

117. I. Saud, M. Khan, and A. Khan. "Diversity of order α". *Journal of Theoretical Biology* 107:4, 1984, pages 521–530.
    DOI: 10.1016/s0022-5193(84)80130-7

118. W. R. Scarlett, P. S. Freier, and C. J. Waddington. "The Charge and Energy Spectra of Heavy Cosmic-Ray Nuclei". *Astrophysics and Space Science* 59:2, 1978, pages 301–311.
    DOI: 10.1007/BF01023921

119. E. S. Seo et al. "Measurement of Cosmic-Ray Proton and Helium Spectra during the 1987 Solar Minimum". *The Astrophysical Journal* 378, 1991, page 763.
    DOI: 10.1086/170477

120. Y. Shikaze et al. "Measurements of 0.2 20 GeV/n cosmic-ray proton and helium spectra from 1997 through 2002 with the BESS spectrometer". *Astroparticle Physics* 28:1, 2007, pages 154–167.
    DOI: 10.1016/j.astropartphys.2007.05.001. ARXIV: astro-ph/0611388 [astro-ph]

121. N. Smirnov. "Table for Estimating the Goodness of Fit of Empirical Distributions". *The Annals of Mathematical Statistics* 19:2, 1948, pages 279–281.
    DOI: 10.1214/aoms/1177730256

122. E. C. Stone et al. "Voyager 1 Observes Low-Energy Galactic Cosmic Rays in a Region Depleted of Heliospheric Ions". *Science* 341:6142, 2013, pages 150–153.
    DOI: 10.1126/science.1236408

123. E. C. Stone. “A Measurement of the Primary Proton Flux from 10 to 130 Million Electron Volts”. *Journal of Geophysical Research* 69:19, 1964, pages 3939–3945.
doi: `10.1029/JZ069i019p03939`

124. E. C. Stone et al. “Cosmic ray measurements from Voyager 2 as it crossed into interstellar space”. *Nature Astronomy* 3, 2019, pages 1013–1018.
doi: `10.1038/s41550-019-0928-3`

125. M. Tanabashi et al. (Particle Data Group Collaboration). “Review of Particle Physics”. *Physical Review D* 98, 3 2018, page 030001.
doi: `10.1103/PhysRevD.98.030001`

126. B. J. Teegarden, F. B. McDonald, and V. K. Balasubrahmanyan. “Spectra and charge composition of the low energy galactic cosmic radiation from $Z = 2$ to 14”. In: *International Cosmic Ray Conference*. Vol. 1. International Cosmic Ray Conference. 1970, page 345

127. The CTA Consortium. *Science with the Cherenkov Telescope Array*. World Scientific, 2018.
doi: `10.1142/10986`

128. *The European Space Agency*. `https://www.esa.int/`.
Note: Accessed: 2023-06-21

129. The Pierre Auger Collaboration. “The Pierre Auger Cosmic Ray Observatory”. *Nuclear Instruments and Methods* 798, 2015, pages 172–213.
doi: `10.1016/j.nima.2015.06.058`

130. A. N. Tikhonov et al. *Numerical Methods for the Solution of Ill-Posed Problems*. Springer Netherlands, 1995.
doi: `10.1007/978-94-015-8480-7`

131. W. R. Webber, R. L. Golden, and S. A. Stephens. “Cosmic ray proton & helium spectra from 5 - 200 GV measured with a magnetic spectrometer.” In: *International Cosmic Ray Conference*. Vol. 1. International Cosmic Ray Conference. 1987, pages 325–328

132. W. R. Webber and P. R. Higbie. “Galactic propagation of cosmic ray nuclei in a model with an increasing diffusion coefficient at low rigidities: A comparison of the new interstellar spectra with Voyager data in the outer heliosphere”. *Journal of Geophysical Research (Space Physics)* 114:A2, 2009, A02103.
doi: `10.1029/2008JA013689`

133. W. R. Webber and F. B. McDonald. “Cerenkov Scintillation Counter Measurements of the Intensity and Modulation of Low Rigidity Cosmic Rays and Features of the Geomagnetic Cutoff Rigidity”. *Journal of Geophysical Research* 69:15, 1964, pages 3097–3114.

DOI: 10.1029/JZ069i015p03097

134. W. R. Webber and F. B. McDonald. “The Cosmic-Ray Oxygen and Helium Spectra Measured at Pioneer 10 over the Time of the 1987 Modulation Minimum, and Implications for the He/O Source Ratio”. *The Astrophysical Journal* 435, 1994, page 464.

DOI: 10.1086/174828

135. W. R. Webber and S. M. Yushak. “A measurement of the energy spectra and relative abundance of the cosmic-ray H and He isotopes over a broad energy range”. *The Astrophysical Journal* 275, 1983, pages 391–404.

DOI: 10.1086/161541

136. W. R. Webber et al. “A Measurement of the Cosmic-Ray 2H and 3He Spectra and 2H/ 4He and 3He/ 4He Ratios in 1989”. *The Astrophysical Journal* 380, 1991, page 230.

DOI: 10.1086/170578

137. T. C. Weekes et al. “Observation of TeV gamma rays from the Crab nebula using the atmospheric Cerenkov imaging technique”. *The Astrophysical Journal* 342, 1989, page 379.

DOI: 10.1086/167599

138. J. P. Wefel et al. “Measurements of Cosmic Ray Helium During the 1991 Solar Maximum”. In: *International Cosmic Ray Conference*. Vol. 2. International Cosmic Ray Conference. 1995, page 630

139. B. Wiebel-Sooth. “Measurement of the allparticle enregy spectrum and chemical composition of cosmic rays with the HEGRA detector”. PhD thesis. Universität Wuppertal, 1998

140. B. Wiebel-Sooth, P. L. Biermann, and H. Meyer. *Cosmic Rays VII. Individual element spectra: prediction and data*. 1997.

DOI: 10.48550/ARXIV.ASTRO-PH/9709253

141. R. L. Workman et al. (Particle Data Group Collaboration). “Review of Particle Physics”. *PTEP* 2022, 2022, page 083C01.

DOI: 10.1093/ptep/ptac097

142. Y. S. Yoon et al. “Proton and Helium Spectra from the CREAM-III Flight”. *The Astrophysical Journal* 839:1, 5, 2017, page 5.

DOI: 10.3847/1538-4357/aa68e4. ARXIV: 1704.02512 [astro-ph.HE]

143. J. S. Young et al. “The elemental and isotopic composition of cosmic rays - Silicon to nickel”. *The Astrophysical Journal* 246, 1981, pages 1014–1030. doi: 10.1086/158997

144. G. T. Zatsepin and V. A. Kuz’min. “Upper Limit of the Spectrum of Cosmic Rays”. *Soviet Journal of Experimental and Theoretical Physics Letters* 4, 1966, page 78

145. V. Zatsepin et al. “Energy Spectra of Primary Protons and Other Nuclei in Energy Region 10-100 TeV/nucleus”. In: vol. 2. 1993, page 13

Glossary

AGN Active Galactic Nucleus. 7, 9, 11, 12, 16, 174

AIROBICC Airshower Observation By angle Integrating Cherenkov Counters. 16, 169

AMC Active Mirror Control. 21

AUC Area Under the Curve. 56, 64, 68, 71

`corsika` Cosmic Ray Simulations for Kascade. 46

CTA Cherenkov Telescope Array. 109, 112, 168

DAMPE *DArk Matter Particle Explorer*. 15, 48, 51, 139, 168

DRS4 Domino Ring Sampler 4. 23, 27

EBL Extragalactic Background Light. 5

ESA European Space Agency. 10

***Fermi*-LAT** *Fermi* Large Area Telescope. 15, 169

FPR False Positive Rate. 63, 64

GCN Gamma-ray Coordinates Network. 21, 169, 173

GRB Gamma-Ray Burst. 2, 7, 21, 169, 173

HAWC High Altitude Water Cherenkov. 15, 169

HEGRA High Energy Gamma-Ray Astronomy. 16, 169

H. E. S. S. High Energy Stereoscopic System. 15, 170

IACT Imaging Air Cherenkov Telescope. 1, 2, 15, 16, 45, 60, 109, 111–113, 169–171, 173, 174

IRF Instrument Response Function. 2, 89, 93, 94, 99, 103, 105

KASCADE Karlsruhe Shower Core and Array Detector. 46, 171

KS Kolmogorov–Smirnov. 97

LHAASO Large High Altitude Air Shower Observatory. 15, 171

LIDAR Light Detection and Ranging. 20

LIGO Laser Interferometer Gravitational-Wave Observatory. 5, 171

MAGIC Major Atmospheric Gamma-Ray Imaging Cherenkov. 2, 3, 15, 16, 20–23, 27, 29, 33–35, 39, 40, 46, 47, 59, 60, 96, 102, 106–109, 111–113, 171

`mars` Magic Analysis and Reconstruction Software. 23, 24, 27, 43, 46

MCMC Markov Chain Monte Carlo. 102

MDI Mean Decrease Impurity. 66

`mmcs` MAGIC Monte Carlo Software. 46

NASA National Aeronautics and Space Administration. 10, 169, 170

ORM Observatorio del Roque de los Muchachos. 16, 168, 169

PCA Principal Component Analysis. 31, 32

PDG Particle Data Group. 51

PMT Photo Multiplier Tube. 16, 21, 22, 28, 30, 169, 170, 174

PSF Point Spread Function. 21, 39

RMS Root Mean Square. 57

ROC Receiver Operating Characteristic. 56, 64, 65, 68, 69, 71

SMBH Super Massive Black Hole. 9, 11

SN Supernova. 8, 9

SNR Supernova Remnant. 7–9, 16, 174

***Swift*-BAT** *Swift* Burst Alert Telescope. 15, 173

TPR True Positive Rate. 63, 64

VCSEL Vertical-cavity Surface-emitting Laser. 22

VERITAS Very Energetic Radiation Imaging Telescope Array System. 15, 173

Appendix

A Analysis Appendix

A.1 Cosmic-Ray Spectra of other Experiments

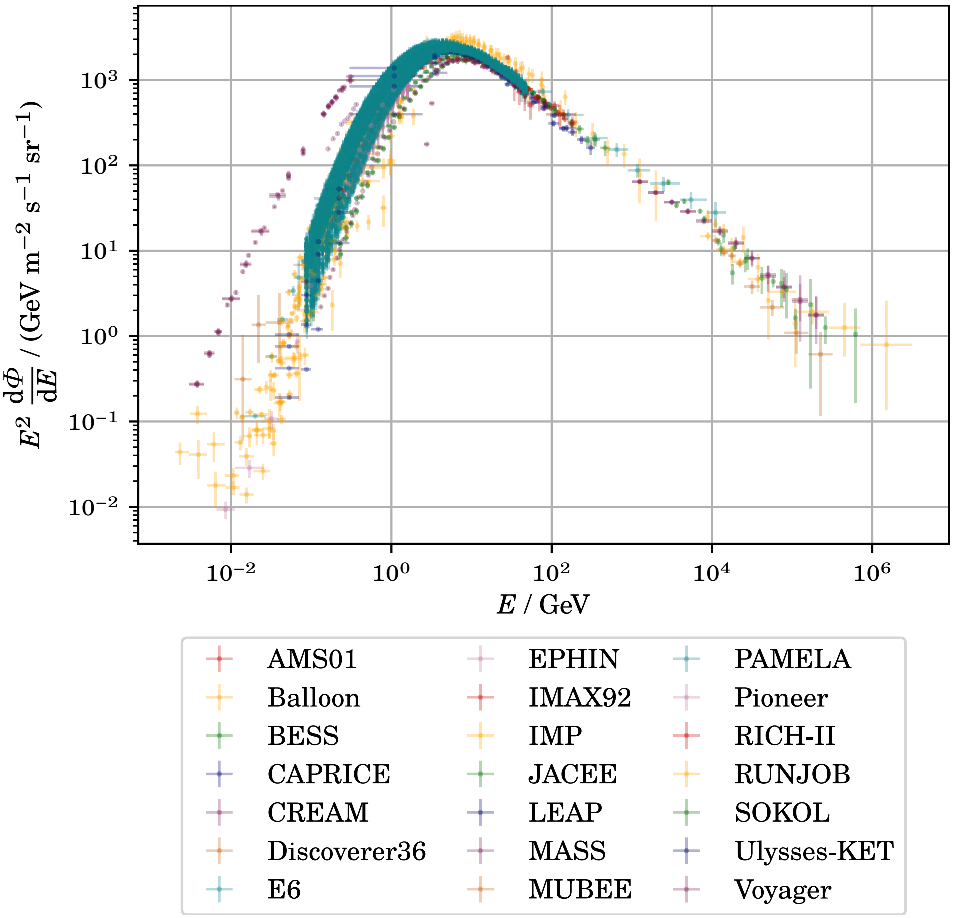

Figure A.1.1: Proton spectrum of measurements with different experiments during the last decades. The experiments contributing to the spectrum are shown in figure 5.1, more detailed in table A.2.1. A full description of the experiments is available in appendix C.

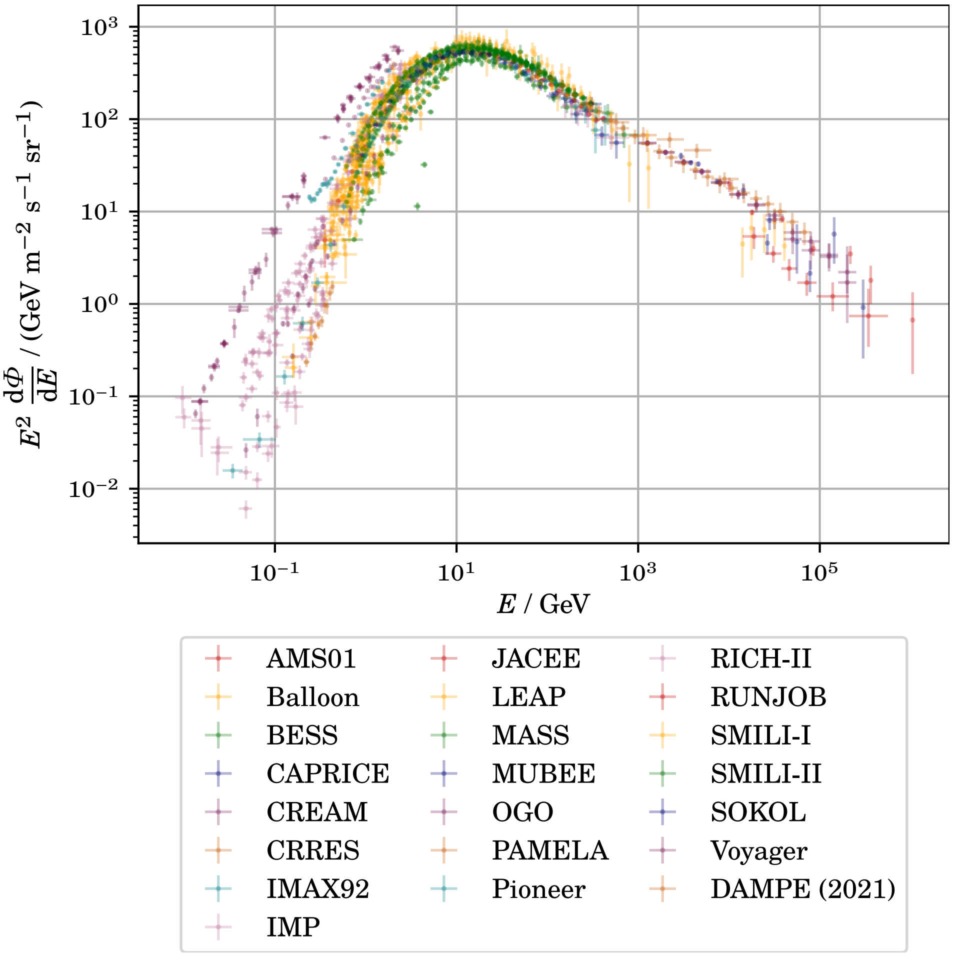

Figure A.1.2: Helium spectrum of measurements with different experiments during the last decades. The experiments contributing to the spectrum are shown in figure 5.2, more detailed in table A.2.2. A full description of the experiments is available in appendix C.

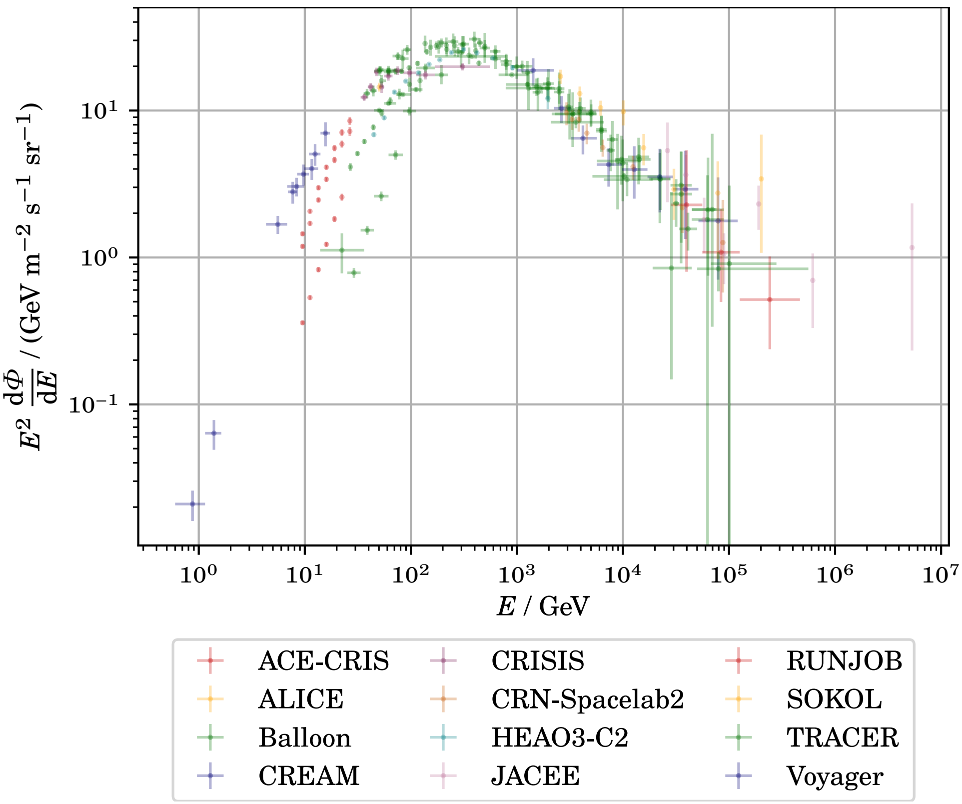

Figure A.1.3: Iron spectrum of measurements with different experiments during the last decades. The experiments contributing to the spectrum are shown in figure 5.3, more detailed in table A.2.3. A full description of the experiments is available in appendix C.

A.2 Cosmic-Ray Measurements

Table A.2.1: Measurements of cosmic proton with different experiments during the last decades. Each experiment is sensitive to specific energies and together they can cover a wide range. All experiments in this list are balloon or satellite based.

experiment	E_min / GeV	E_max / GeV	observation period
AMS01	0.22	162	1998
Balloon	0.011	109 000	1955-1991
BESS	0.18	398	1993-2002
CAPRICE	0.15	270	1994-1998
CREAM	1000	158 000	2004-2008
Discoverer36	0.011	0.06	1961
E6	0.013	0.585	1974-1977
EPHIN	0.292	1.589	1995-2014
IMAX92	0.18	161	1992
IMP	0.002	0.089	1965-1979
JACEE	8000	620 000	1979-1995
LEAP	0.212	82.84	1987
MASS	0.132	71	1989-1991
MUBEE	10 000	160 000	1975-1987
PAMELA	0.082	8600	2006-2014
Pioneer	0.007	1.25	1968
RICH-II	32.1	84.6	1997
RUNJOB	7080	712 000	1995-1999
SOKOL	2850	170 000	1984-1986
Ulysses-KET	0.035	2.45	1990-1994
Voyager	0.003	0.363	2008-2019

Table A.2.2: Measurements of cosmic helium nuclei with different experiments during the last decades. Each experiment is sensitive to specific energies and together they can cover a wide range. All experiments in this list are balloon or satellite based.

experiment	*E*_min / GeV	*E*_max / GeV	observation period
AMS01	0.298	454.468	1998
Balloon	0.12	40 800	1950-1991
BESS	0.6	1004	1993-2008
CAPRICE	0.8	696	1994-1998
CREAM	1004	252 400	2004-2008
CRRES	0.207	0.442	1990-1991
IMAX92	0.92	500	1992
IMP	0.008	4.648	1963-1979
JACEE	10 120	1 052 000	1979-1995
LEAP	0.82	299	1987
MASS	0.872	196.4	1989-1991
MUBEE	26 480	300 000	1975-1987
OGO	0.048	1.836	1964-1968
PAMELA	200	12 800	2006-2010
Pioneer	0.027	6.56	1968-1988
RICH-II	128.4	800	1997
RUNJOB	14 120	564 000	1995-1999
SMILI-I	0.4	8	1989
SMILI-II	0.556	8.28	1991
SOKOL	2960	144 000	1984-1986
Voyager	0.012	2.646	2008-2019
DAMPE	68	79 270	2021

Table A.2.3: Measurements of cosmic iron nuclei with different experiments during the last decades. Each experiment is sensitive to specific energies and together they can cover a wide range. All experiments in this list are balloon or satellite based.

experiment	*E*_min / GeV	*E*_max / GeV	observation period
ACE-CRIS	9.52	26.6	1997-2010
ALICE	50.288	50.288	1987
Balloon	14	89 600	1968-1991
CREAM	1008	120 624	2005-2006
CRISIS	33.6	560	1977
CRN, Spacelab2	2945.6	87 528	1985
HEAO-3-C2	44.8	1960	1979-1980
JACEE	26 320	5 320 000	1979-1995
RUNJOB	28 168	466 480	1995-1999
SOKOL	2576	201 600	1984-1986
TRACER	44.8	560 000	2003-2006
Voyager	0.599	17.276	2012-2015

A.3 Features of the Random Forests

Mono features

`length`
`width`
`concentration`
`concentration1`
`rms_time`
`rms_time_w`
`num_islands`
`num_single_pixels`
`used_area`
`core_area`
`num_used_pixels`
`num_core_pixels`
`m3_longitudinal`
`m3_transversal`
`conc1`
`asymmetry`

Stereo features

`cherenkov_density`
`max_height`
`cos_bs_angle`
`x_max`

Generated features

`log_size`
`abs_p1_grad`
`log_size_single_pixels`
`log_size_main_island`
`log_inner_size`
`impact_per_cherenkov_radius`
`area`
`width_length_ratio`

A.4 Data-Simulation Comparison

In addition to the features examined in the main part of the thesis in figure 5.5, 5.6 and 5.7, all other features are also investigated. The data-simulation comparisons are shown in figures A.4.1 to A.4.25. Simulations are weighted with the observation time and the assumed spectral indices of the three particle types, proton, helium and iron nuclei.

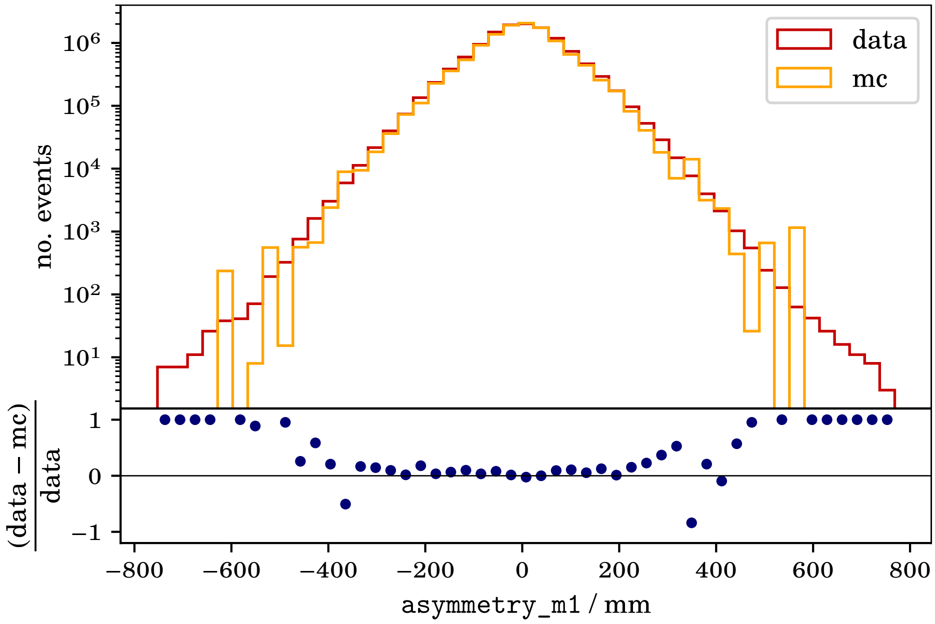

Figure A.4.1: Data-simulation comparison of the feature `asymmetry_m1`. The relative deviation of data and simulations is shown in blue.

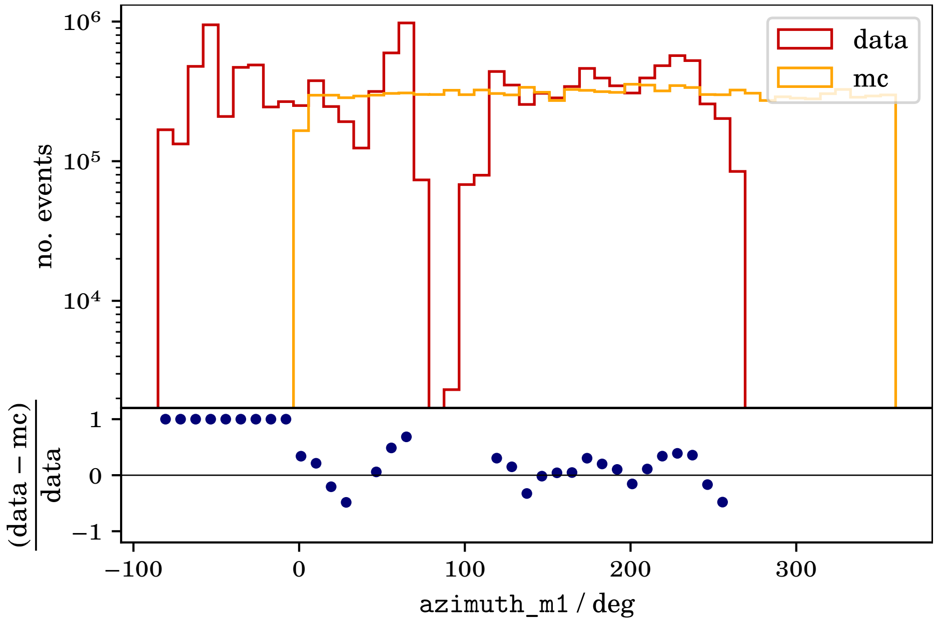

Figure A.4.2: Data-simulation comparison of the feature `azimuth_m1`. The relative deviation of data and simulations is shown in blue.

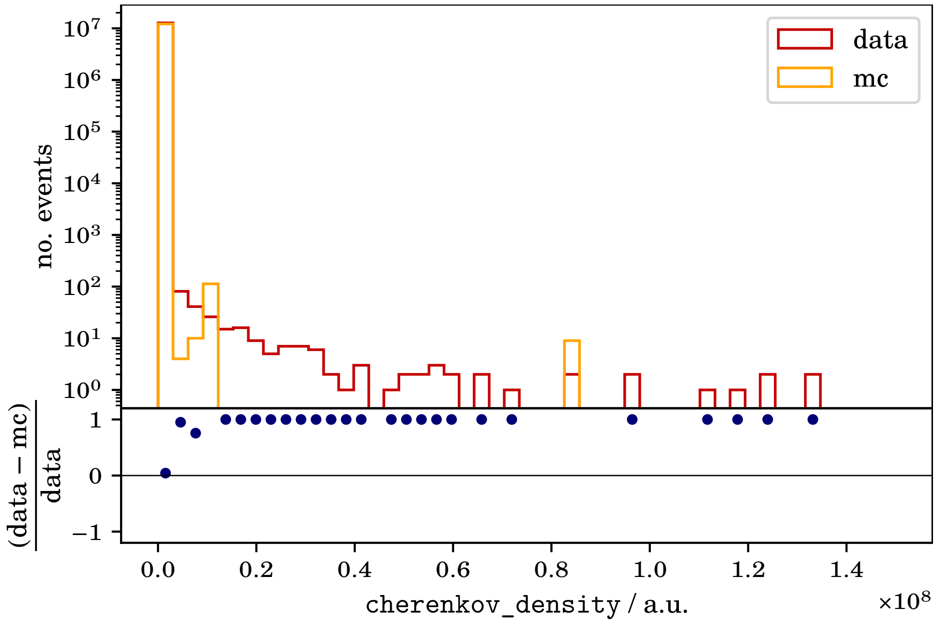

Figure A.4.3: Data-simulation comparison of the feature `cherenkov_density`. The relative deviation of data and simulations is shown in blue.

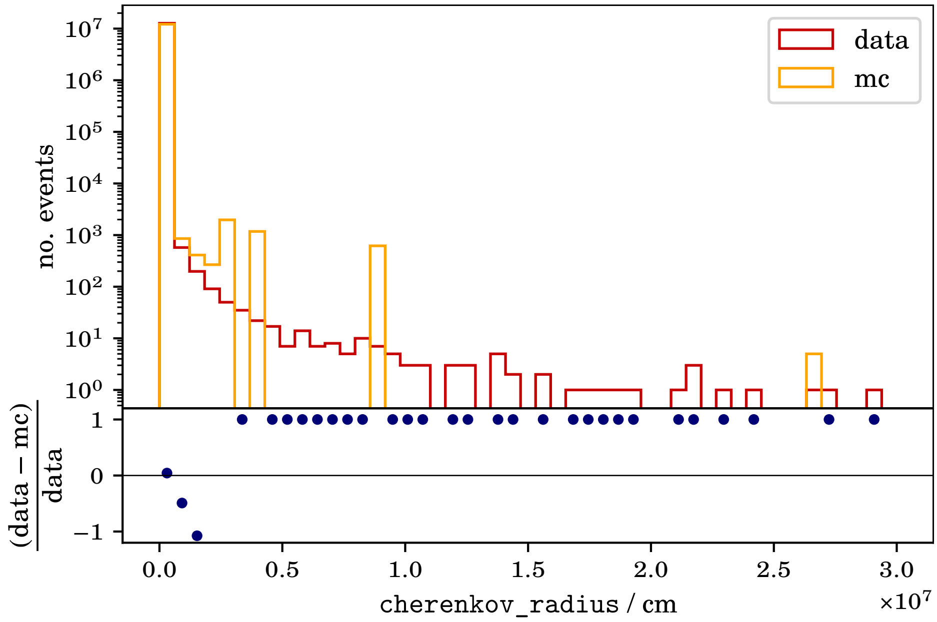

Figure A.4.4: Data-simulation comparison of the feature `cherenkov_radius`. The relative deviation of data and simulations is shown in blue.

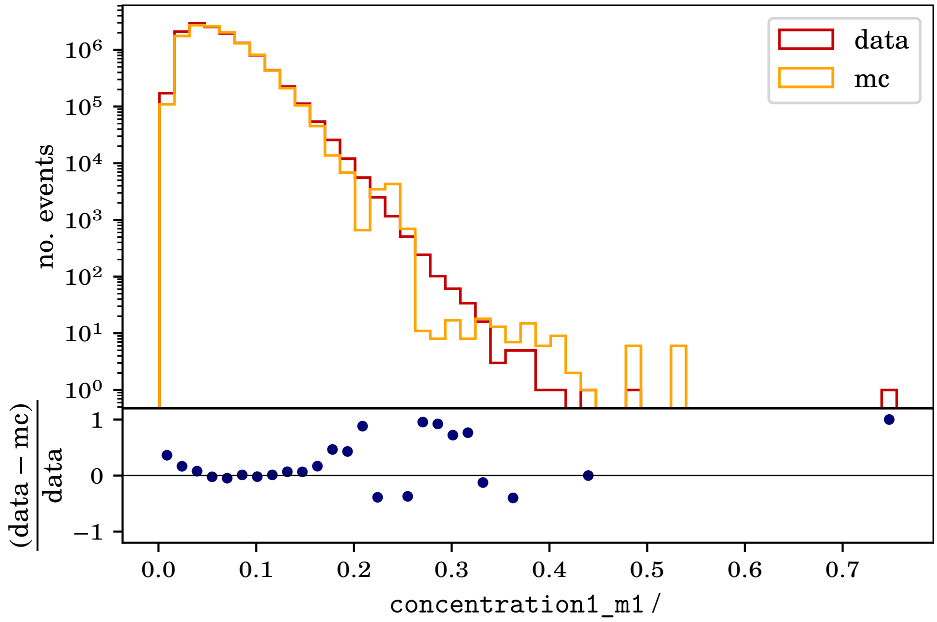

Figure A.4.5: Data-simulation comparison of the feature `concentration1_m1`. The relative deviation of data and simulations is shown in blue.

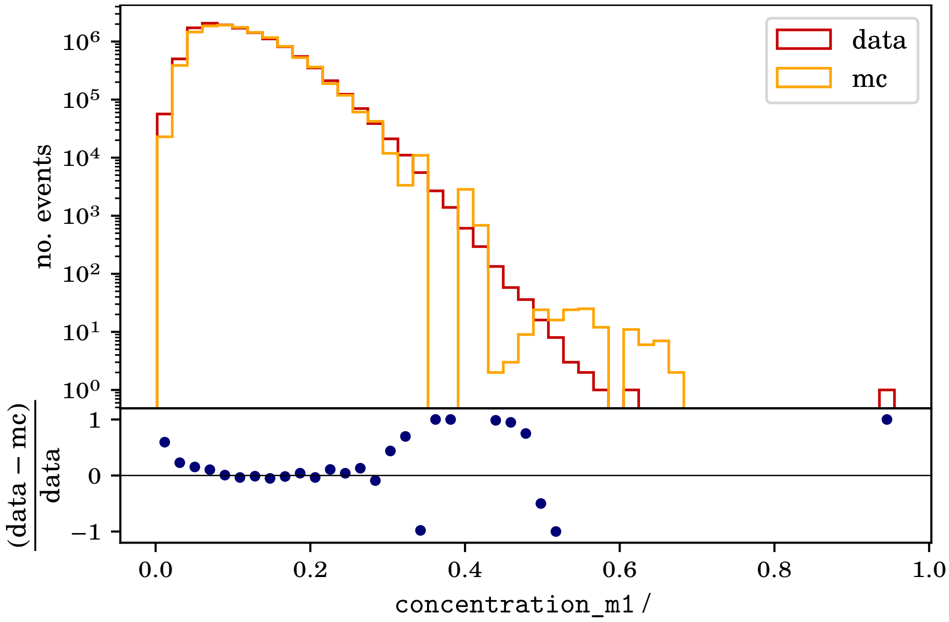

Figure A.4.6: Data-simulation comparison of the feature `concentration_m1`. The relative deviation of data and simulations is shown in blue.

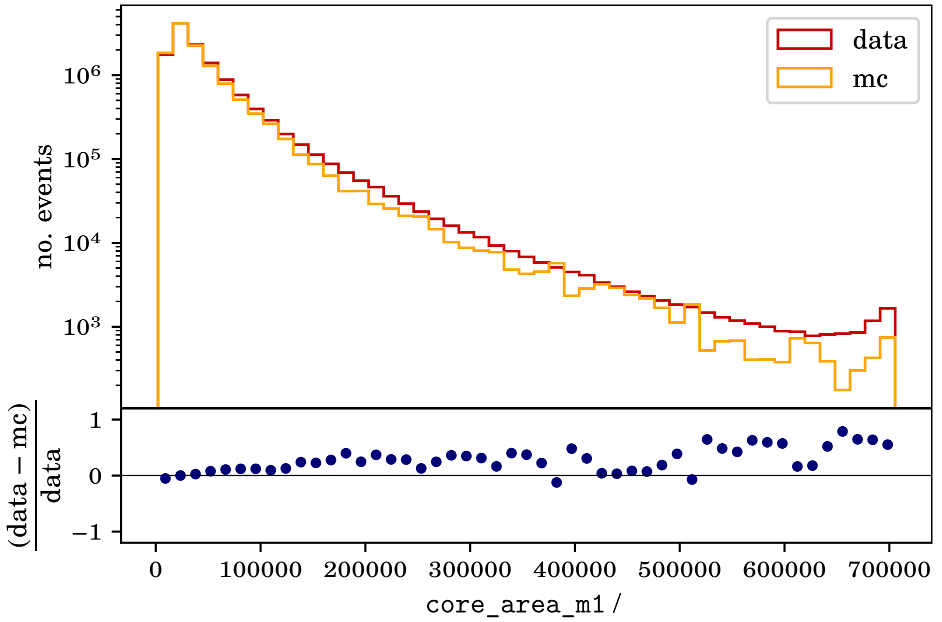

Figure A.4.7: Data-simulation comparison of the feature `core_area_m1`. The relative deviation of data and simulations is shown in blue.

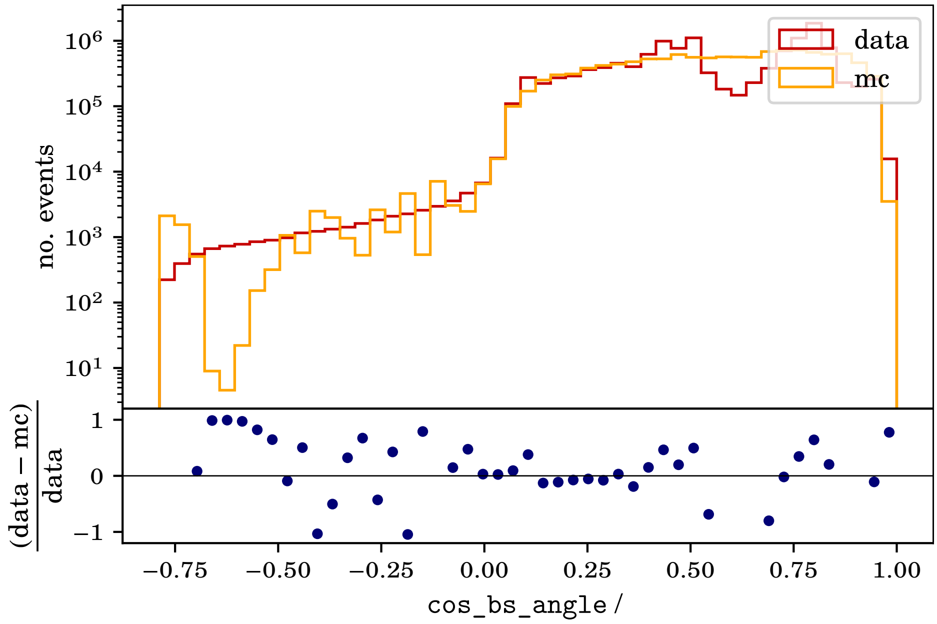

Figure A.4.8: Data-simulation comparison of the feature `cos_bs_angle`. The relative deviation of data and simulations is shown in blue.

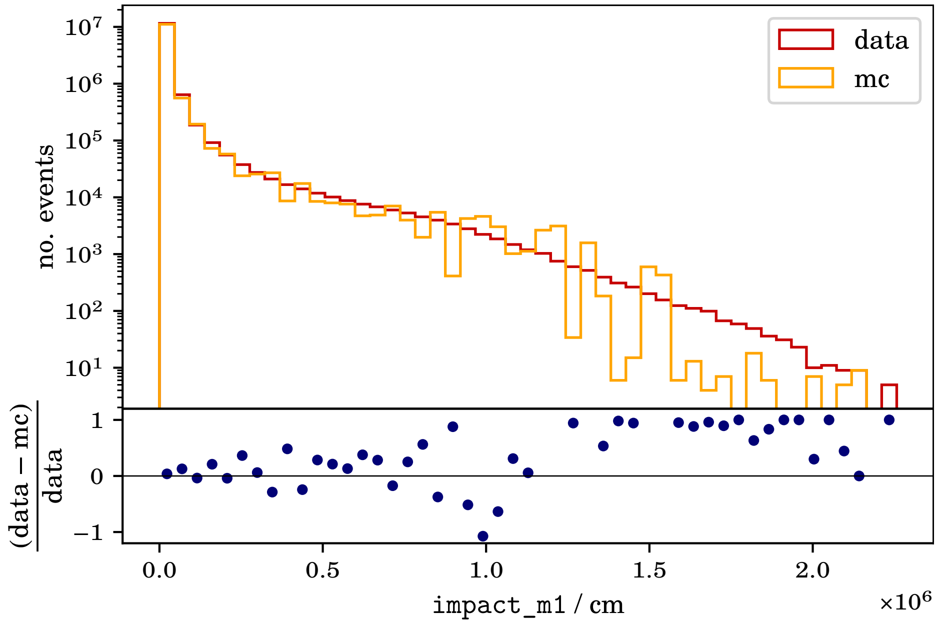

Figure A.4.9: Data-simulation comparison of the feature `impact_m1`. The relative deviation of data and simulations is shown in blue.

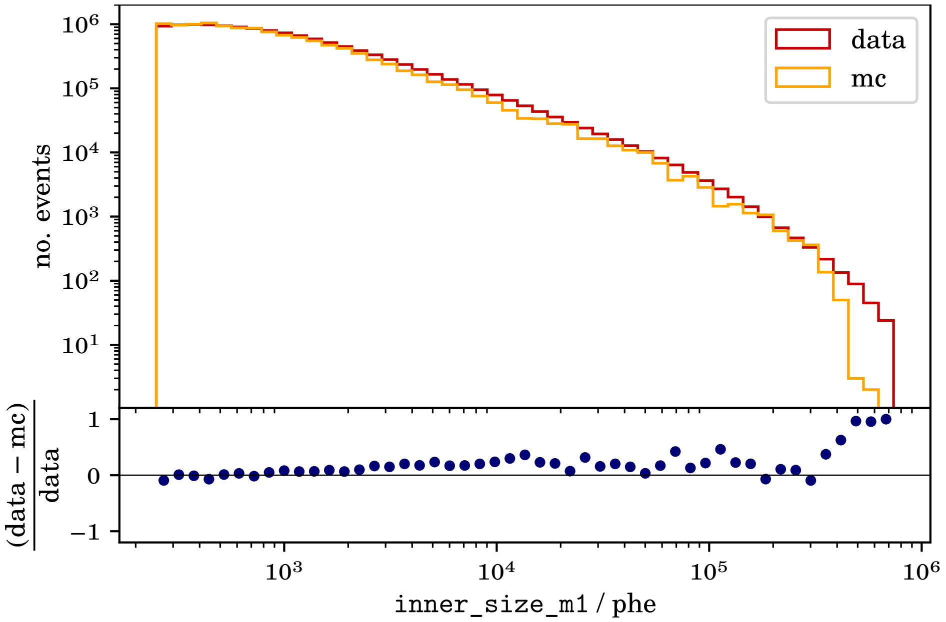

Figure A.4.10: Data-simulation comparison of the feature `inner_size_m1`. The relative deviation of data and simulations is shown in blue.

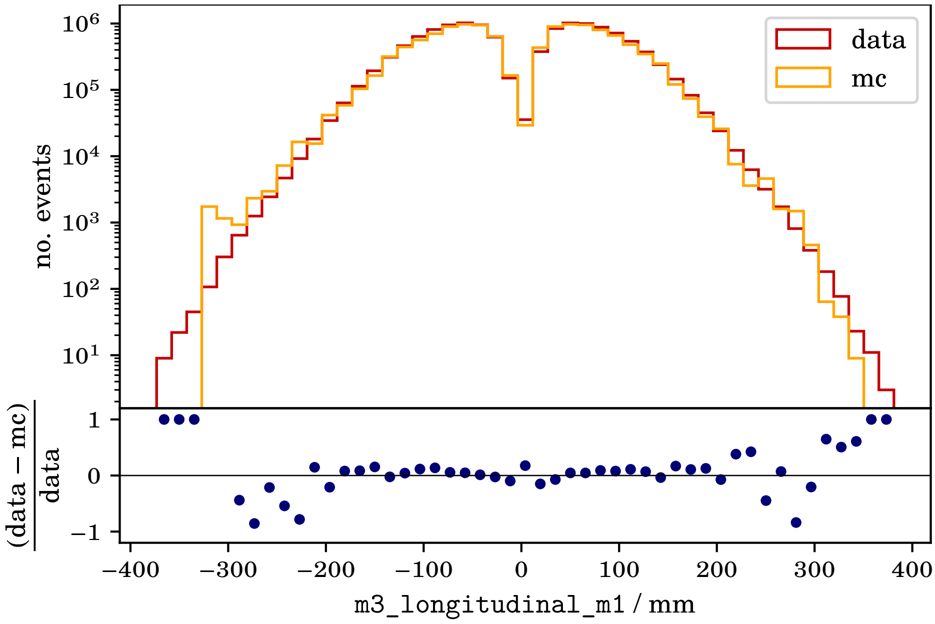

Figure A.4.11: Data-simulation comparison of the feature `m3_longitudinal_m1`. The relative deviation of data and simulations is shown in blue.

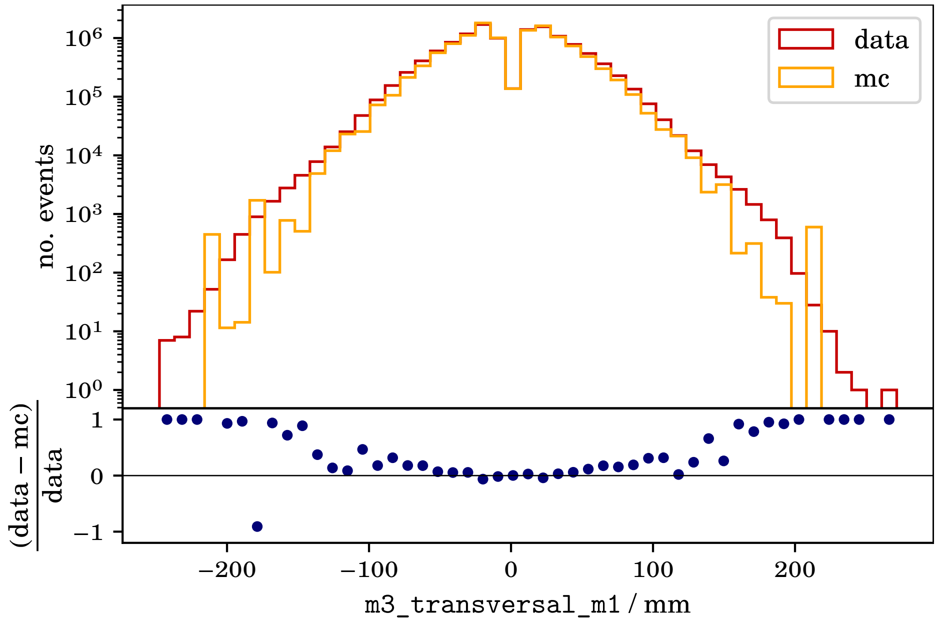

Figure A.4.12: Data-simulation comparison of the feature `m3_transversal_m1`. The relative deviation of data and simulations is shown in blue.

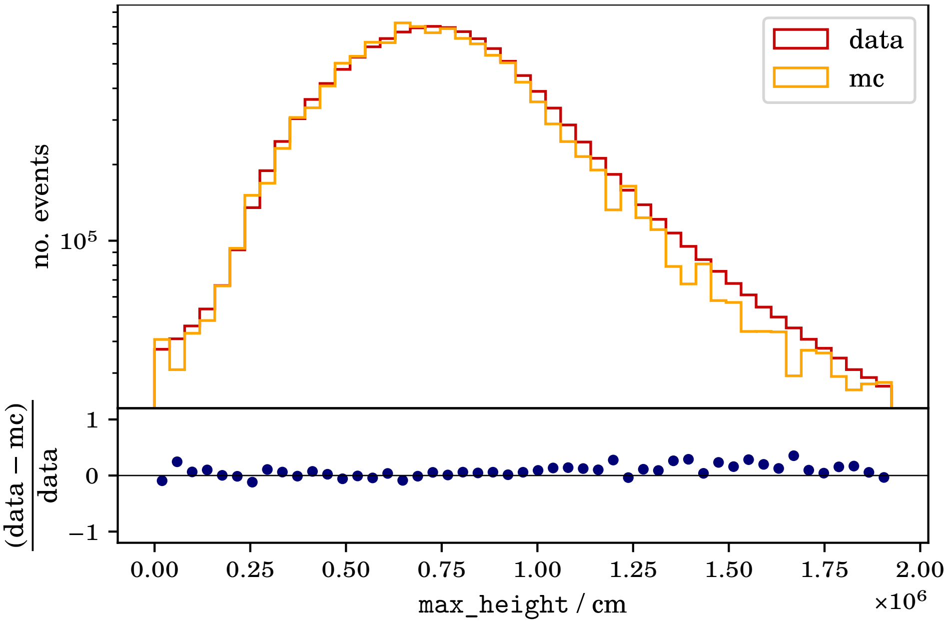

Figure A.4.13: Data-simulation comparison of the feature `max_height`. The relative deviation of data and simulations is shown in blue.

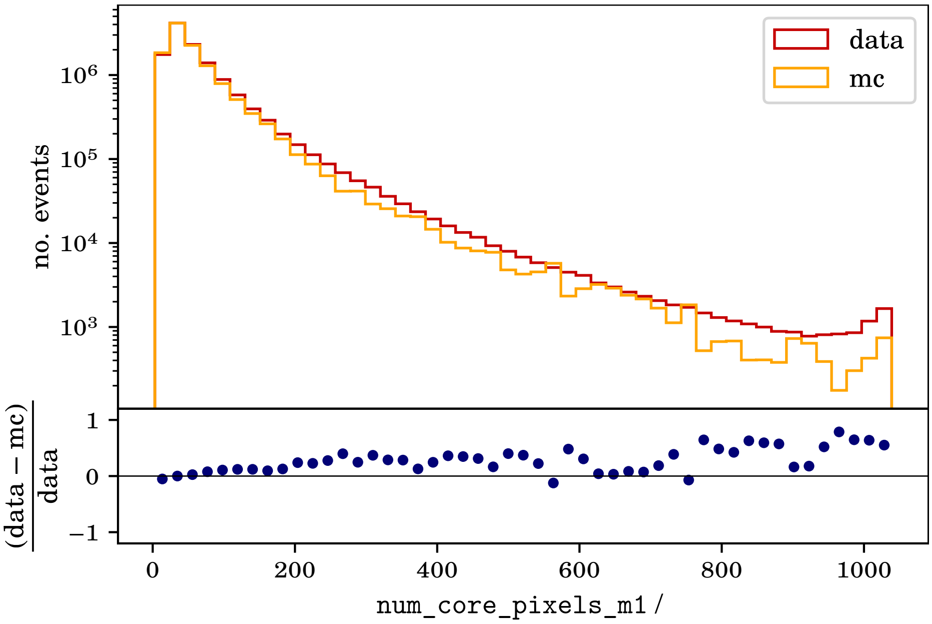

Figure A.4.14: Data-simulation comparison of the feature `num_core_pixels_m1`. The relative deviation of data and simulations is shown in blue.

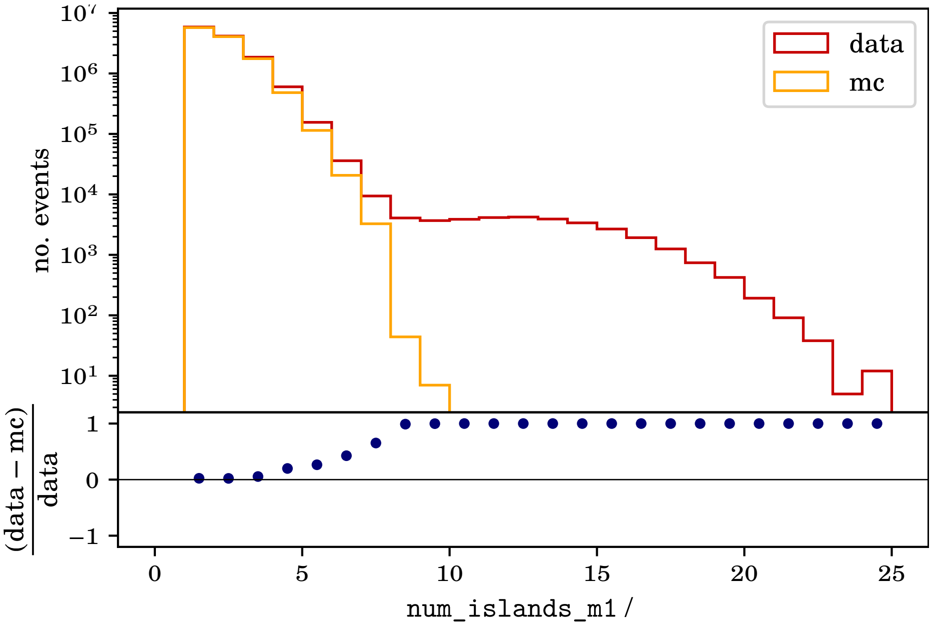

Figure A.4.15: Data-simulation comparison of the feature `num_islands_m1`. The relative deviation of data and simulations is shown in blue.

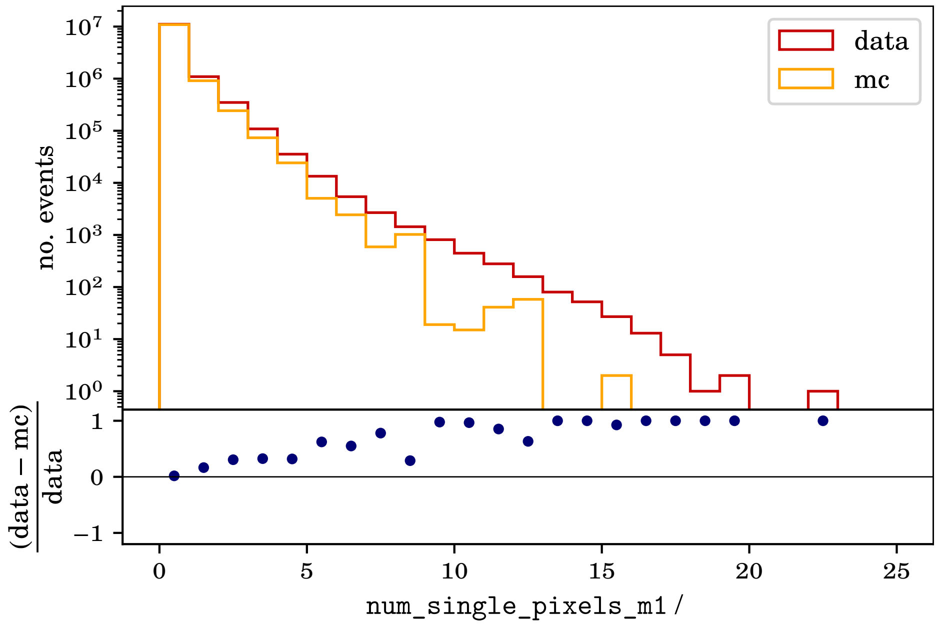

Figure A.4.16: Data-simulation comparison of the feature `num_single_pixels_m1`. The relative deviation of data and simulations is shown in blue.

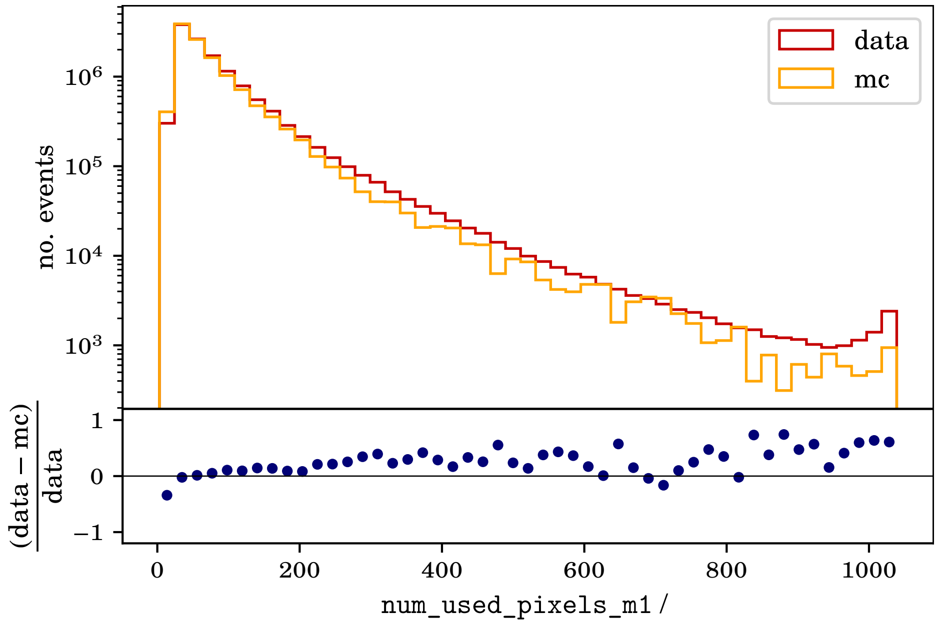

Figure A.4.17: Data-simulation comparison of the feature `num_used_pixels_m1`. The relative deviation of data and simulations is shown in blue.

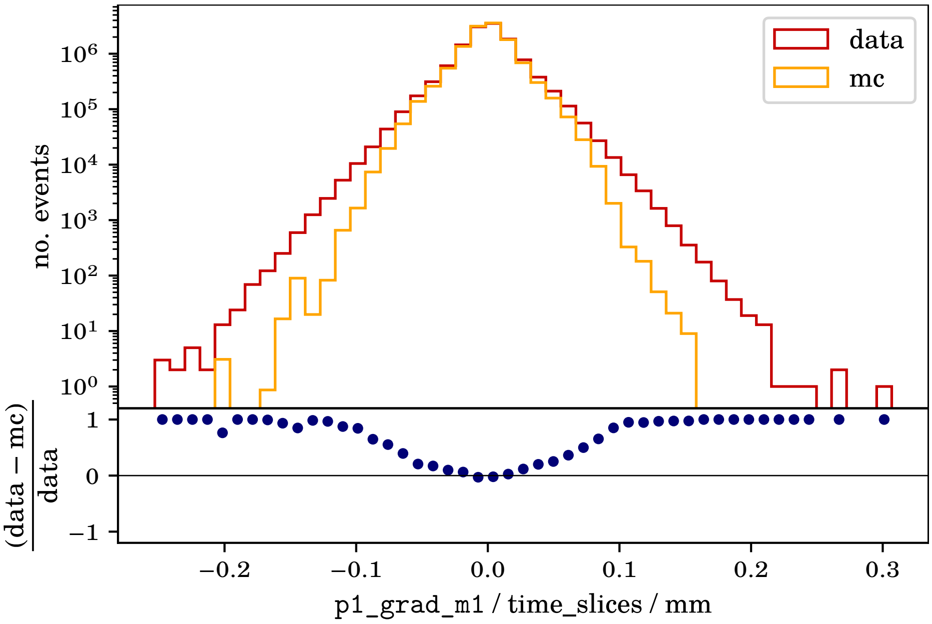

Figure A.4.18: Data-simulation comparison of the feature `p1_grad_m1`. The relative deviation of data and simulations is shown in blue.

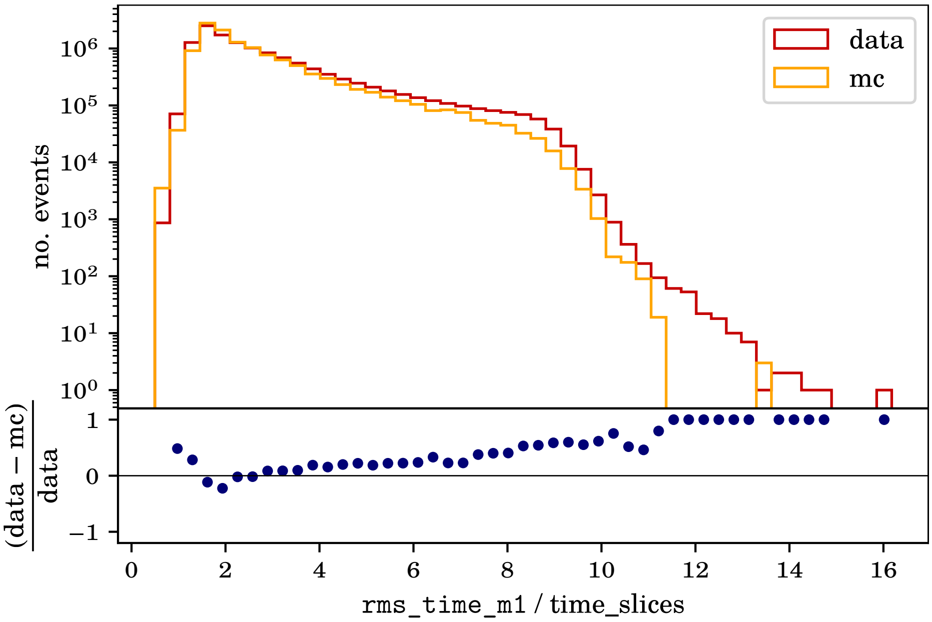

Figure A.4.19: Data-simulation comparison of the feature `rms_time_m1`. The relative deviation of data and simulations is shown in blue.

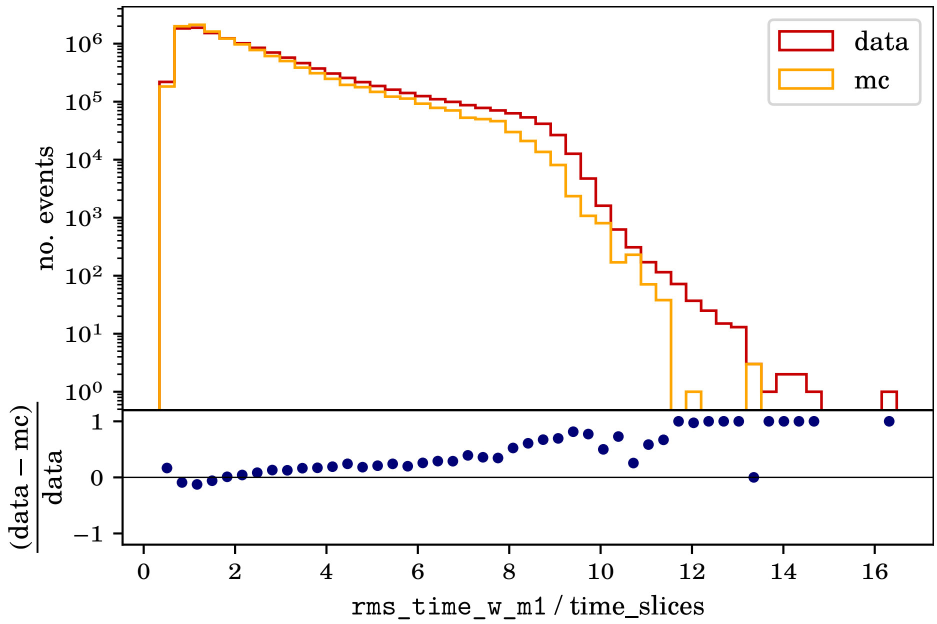

Figure A.4.20: Data-simulation comparison of the feature `ms_time_w_m1`. The relative deviation of data and simulations is shown in blue.

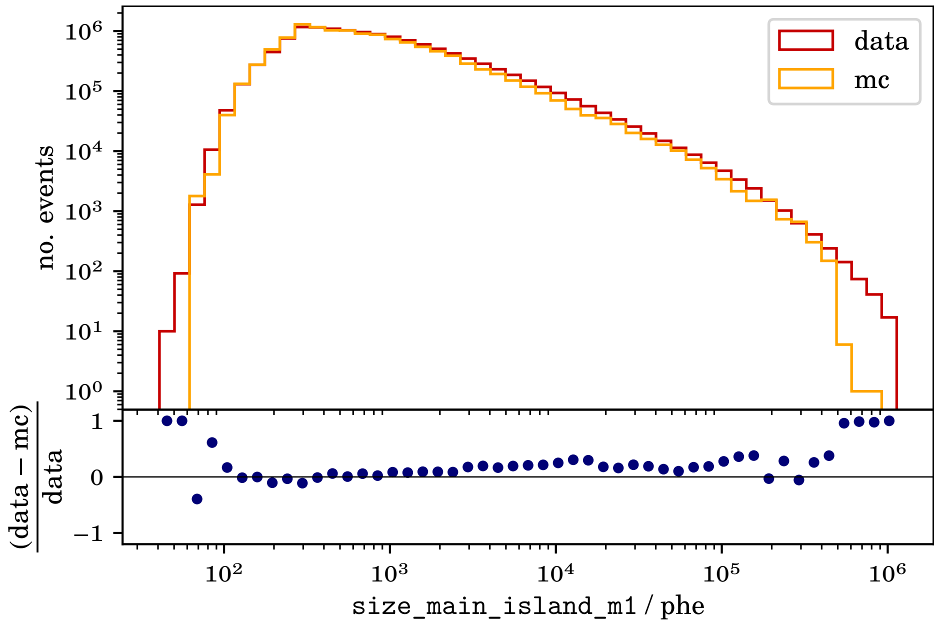

Figure A.4.21: Data-simulation comparison of the feature `size_main_island_m1`. The relative deviation of data and simulations is shown in blue.

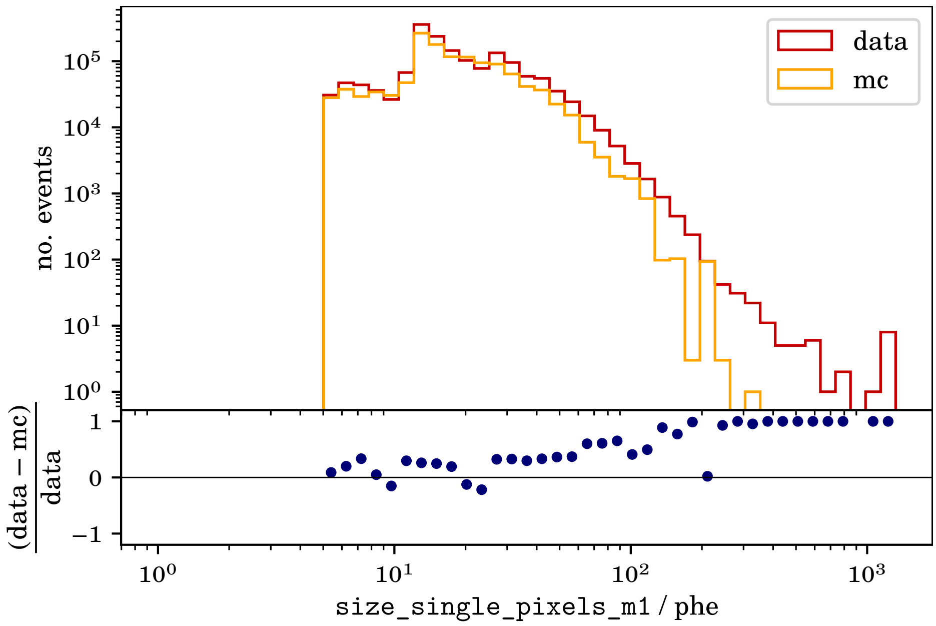

Figure A.4.22: Data-simulation comparison of the feature `size_single_pixels_m1`. The relative deviation of data and simulations is shown in blue.

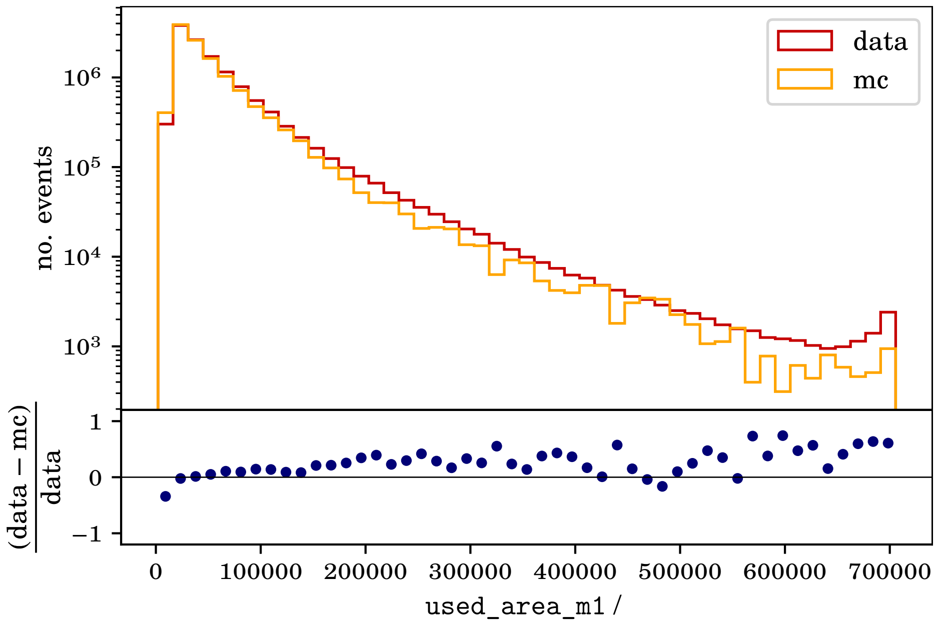

Figure A.4.23: Data-simulation comparison of the feature `used_area_m1`. The relative deviation of data and simulations is shown in blue.

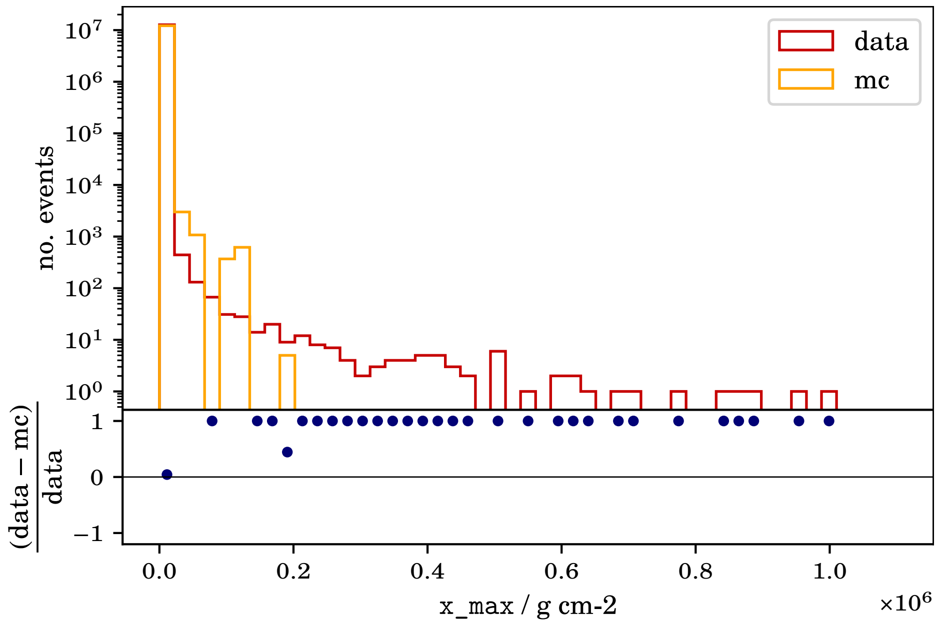

Figure A.4.24: Data-simulation comparison of the feature `x_max`. The relative deviation of data and simulations is shown in blue.

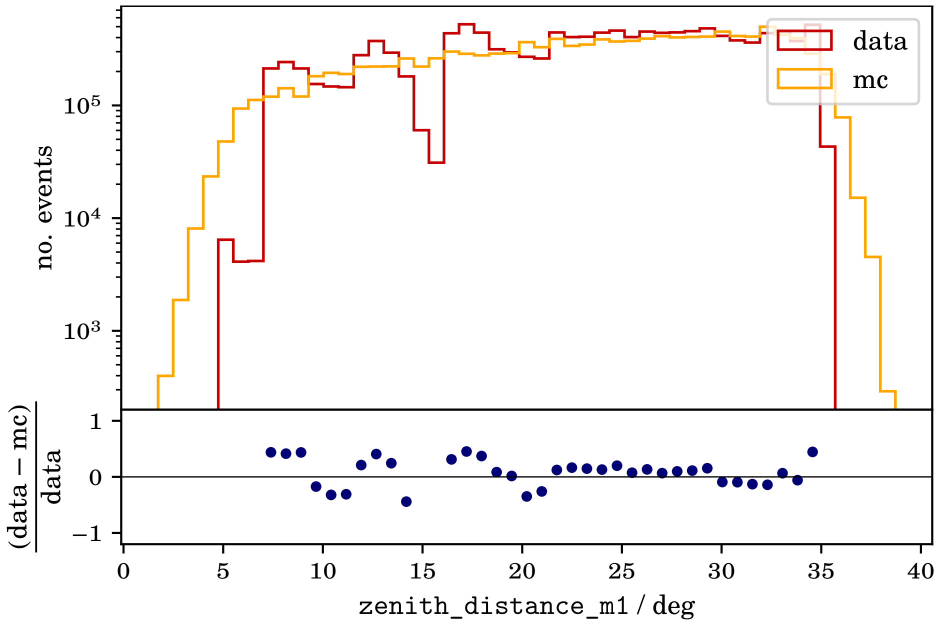

Figure A.4.25: Data-simulation comparison of the feature `zenith_distance_m1`. The relative deviation of data and simulations is shown in blue.

A.5 Classification of all Particle Types

Table A.5.1: Counts of reconstructed events of different datasets. The simulated and **unweighted** proton, helium, and iron sample, as well as the observation data, are listed. The rows show the reconstructed particle types proton, helium, and iron.

	proton sample	helium sample	p + he sample	iron sample	all particles	observation data
iron	1596	4566	6162	18 156	24 318	5 623 417
helium	1151	2695	3846	758	4604	2 377 647
proton	4343	3051	7394	476	7870	7 804 203

Table A.5.2: Fractions of reconstructed events of different datasets. The simulated and **weighted** proton, helium, and iron sample, as well as the observation data, are listed. The rows show the reconstructed particle types proton, helium, and iron.

	proton sample	helium sample	p + he sample	iron sample	all particles	observation data
iron	22 %	44 %	35 %	93 %	66 %	35 %
helium	16 %	26 %	22 %	3 %	12 %	15 %
proton	61 %	29 %	42 %	2 %	21 %	49 %

A.6 Energy Regression of Helium and Iron

With the analysis, it is possible to estimate also the energies of helium and iron. A regressor trained with helium simulations learns to predict the energy of the primary helium nuclei. In principle, it is possible to calculate a helium spectrum from this using the same procedure for producing a proton spectrum. The performance of the regressor is shown in the figures A.6.1, A.6.2 and A.6.3.

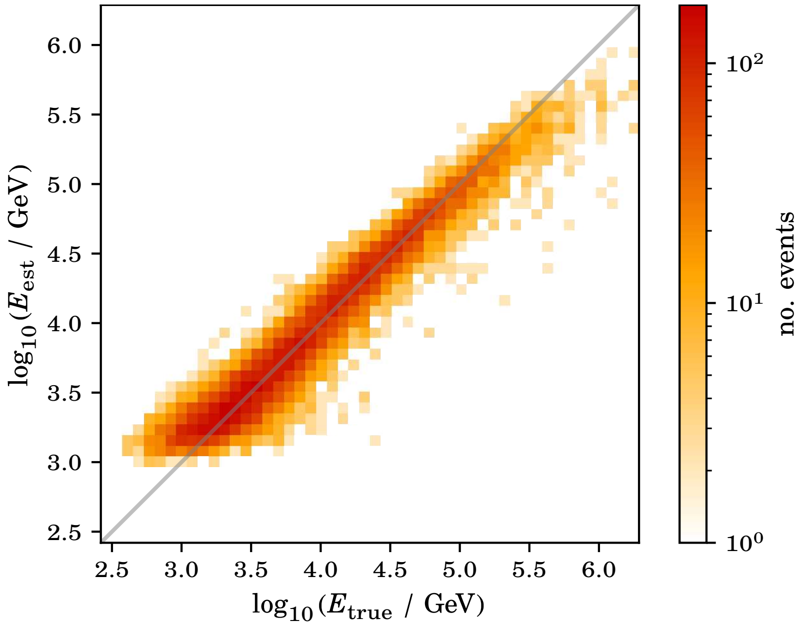

Figure A.6.1: Migration from true primary energy of helium nuclei to reconstructed energy. The characteristic of the regressor is the overestimation at low energies and the underestimation at higher ones.

The energy dispersion in figure A.6.1 shows a clear correlation between true and estimated energy, indicating that the features on which the model was trained contain meaningful information about the primary particle's energy. The bias and resolution in figure A.6.2 exhibit the characteristic properties of a regressor: an overestimation of low energies and an underestimation of high energies. The main features for energy estimation are shown in figure A.6.3. As expected, features that are particularly important for energy estimation contain information about the arrival times and the shower sizes of the events.

Analogously, a model is trained on iron events so that it learns to estimate its energy. The performance of the regressor is shown in the figures A.6.4, A.6.5 and A.6.6. The energy dispersion in figure A.6.4 shows that the model does not estimate the energy of iron nuclei as well as the models of energy estimation for proton and helium nuclei do. The smearing is larger, and the energies of iron nuclei are underestimated over a

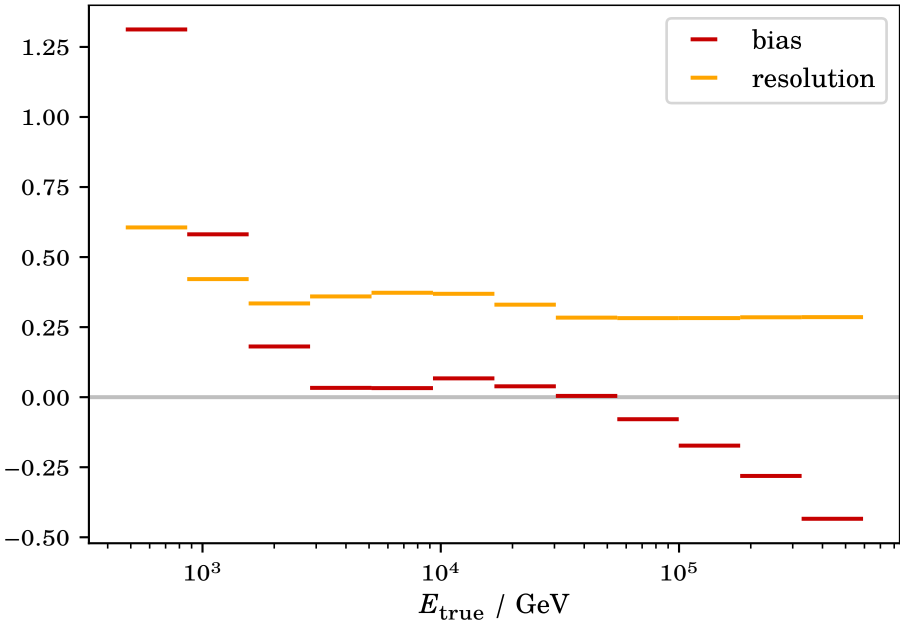

Figure A.6.2: Bias and resolution of the helium energy regressor. The bias shows the random forest overestimating low energies and underestimating high energies. The resolution is constant with an improving tendency towards increasing energy.

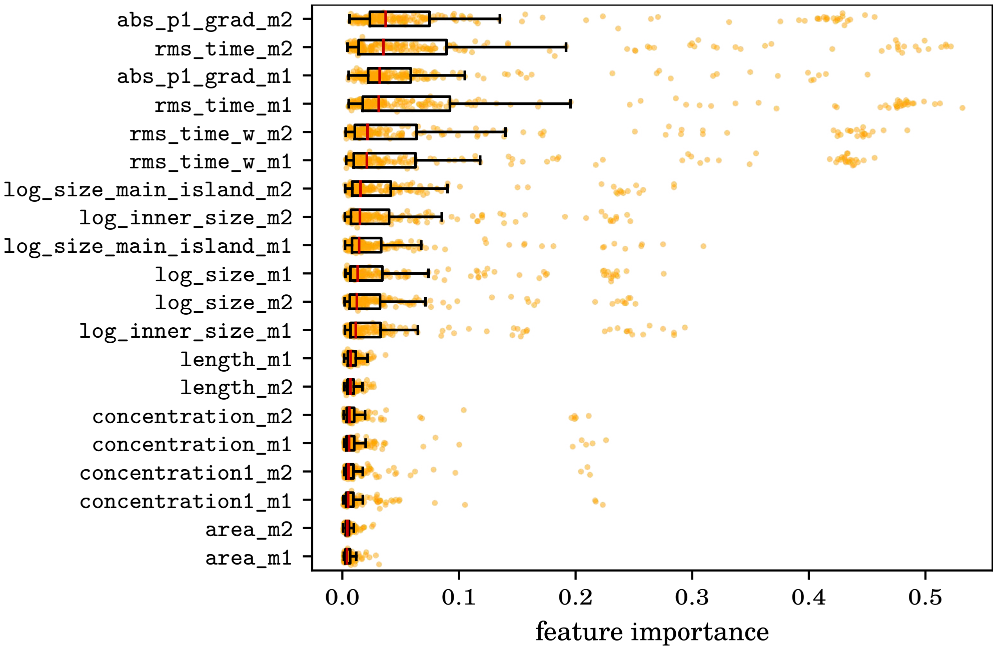

Figure A.6.3: The 20 most important features of the helium energy regressor. The most important features describe features related to the arrival times of the image (`p1_grad` and `rms_time`) and the detected Cherenkov light of the shower (`size`).

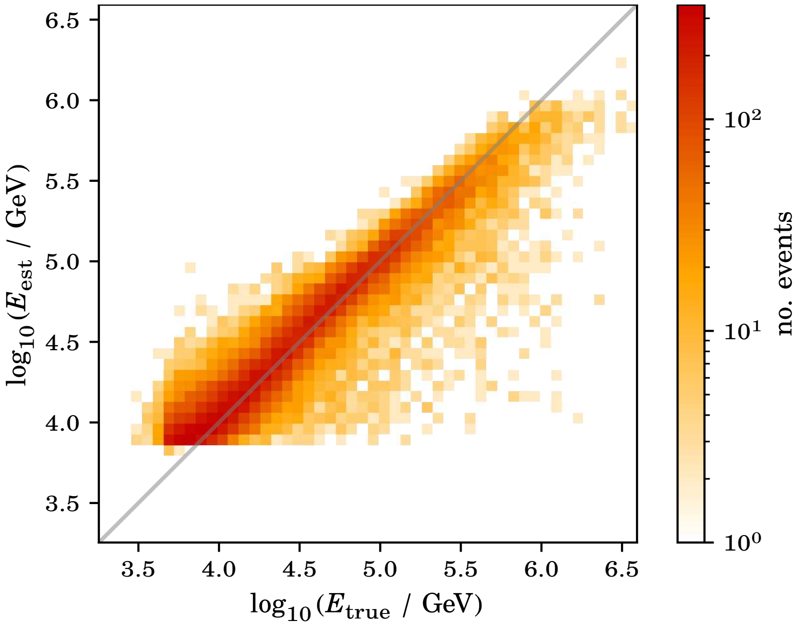

Figure A.6.4: Migration from true primary energy of iron nuclei to reconstructed energy. The characteristic of the regressor is the overestimation at low energies and the underestimation at higher ones.

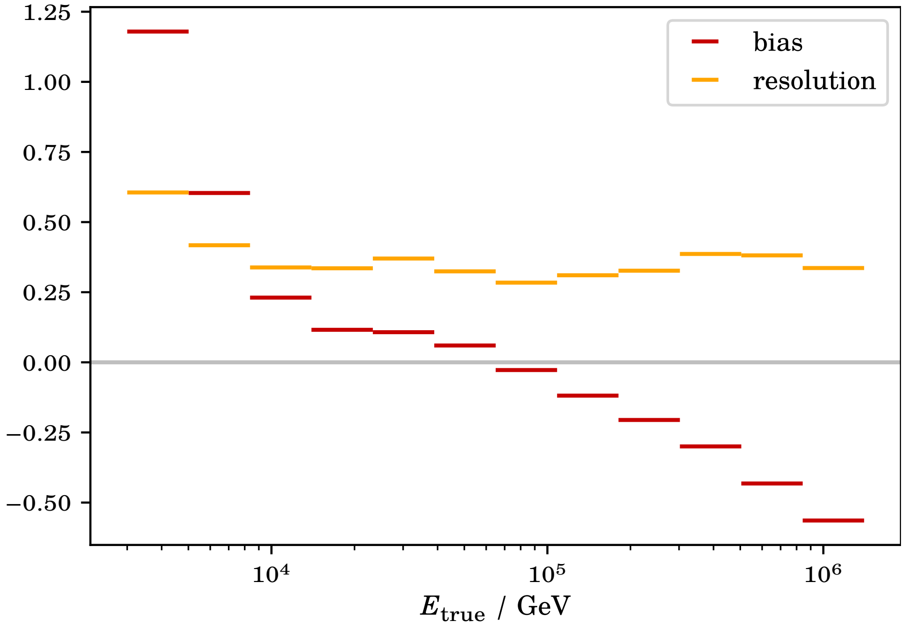

Figure A.6.5: Bias and resolution of the iron energy regressor. The bias shows the random forest overestimating low energies and underestimating high energies. The resolution is constant with an improving tendency towards increasing energy.

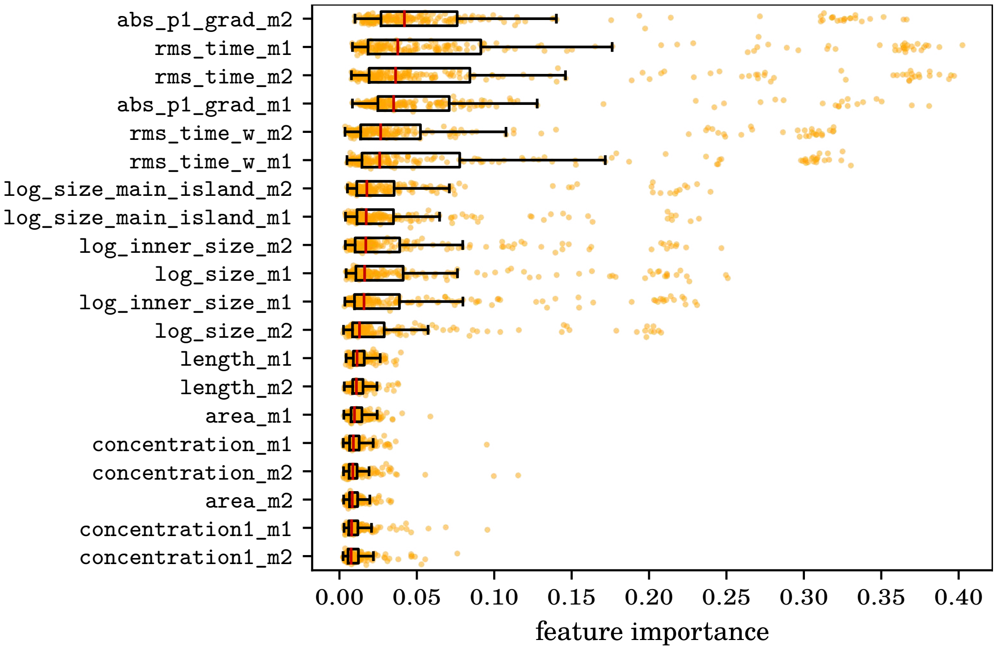

Figure A.6.6: The 20 most important features of the iron energy regressor. The most important features describe features related to the arrival times of the image (`rms_time`) and the detected Cherenkov light of the shower (`size`).

wide range. Nevertheless, the figure shows a correlation between true and estimated energies, which proves that the model does not randomly guess. The bias and resolution in figure A.6.5 show the behavior from the energy dispersion. The energies are underestimated until the characteristics of the regressor lead to overestimation. However, this behavior of strong overestimation and underestimation would have been eliminated by unfolding in an analysis of the iron spectrum. The figure A.6.6 shows the 20 most important features for energy estimation. As expected and comparable with the proton and helium energy regressor, features that retain arrival time and shower size information are essential here.

B Observation Data

20160430_05052838_S_PG1553+113-W0.40+000.root
20160430_05052839_S_PG1553+113-W0.40+180.root
20160430_05052840_S_PG1553+113-W0.40+090.root
20160430_05052841_S_PG1553+113-W0.40+270.root
20160501_05052880_S_RBS0970-W0.40+270.root
20160501_05052886_S_PG1553+113-W0.40+180.root
20160501_05052888_S_PG1553+113-W0.40+270.root
20160501_05052891_S_PG1553+113-W0.40+090.root
20160501_05052897_S_PG1553+113-W0.40+000.root
20160501_05052898_S_PG1553+113-W0.40+180.root
20160502_05052923_S_RBS0970-W0.40+000.root
20160502_05052927_S_PG1553+113-W0.40+090.root
20160502_05052928_S_PG1553+113-W0.40+270.root
20160503_05052959_S_RBS0970-W0.40+000.root
20160503_05052965_S_PG1553+113-W0.40+090.root
20160503_05052966_S_PG1553+113-W0.40+270.root
20160503_05052967_S_PG1553+113-W0.40+000.root
20160504_05052997_S_RBS0970-W0.40+090.root
20160504_05053003_S_RBS0970-W0.40+000.root
20160505_05053035_S_GRB160504-W0.40+090.root
20160505_05053036_S_GRB160504-W0.40+270.root
20160505_05053040_S_RBS0970-W0.40+270.root
20160526_05053334_S_RBS0970-W0.40+090.root
20160526_05053341_S_PG1553+113-W0.40+180.root
20160527_05053358_S_RBS0970-W0.40+000.root
20160527_05053360_S_RBS0970-W0.40+090.root
20160529_05053404_S_RBS0970-W0.40+180.root
20160529_05053405_S_RBS0970-W0.40+090.root
20160529_05053421_S_PSRJ2032-W0.40+321.root
20160529_05053422_S_PSRJ2032-W0.40+051.root
20160530_05053439_S_RBS0970-W0.40+180.root
20160530_05053440_S_RBS0970-W0.40+090.root
20160530_05053445_S_RBS0970-W0.40+270.root
20160530_05053446_S_RBS0970-W0.40+000.root
20160531_05053472_S_RBS0970-W0.40+270.root
20160531_05053473_S_RBS0970-W0.40+000.root
20160531_05053494_S_Cyg-X1-W0.40+270.root
20160531_05053495_S_Cyg-X1-W0.40+000.root
20160531_05053496_S_Cyg-X1-W0.40+180.root
20160601_05053506_S_RBS0970-W0.40+180.root
20160601_05053507_S_RBS0970-W0.40+090.root
20160601_05053516_S_RBS0970-W0.40+180.root
20160601_05053530_S_Cyg-X1-W0.40+090.root
20160601_05053532_S_Cyg-X1-W0.40+000.root
20160601_05053533_S_Cyg-X1-W0.40+180.root
20160601_05053534_S_Cyg-X1-W0.40+090.root
20160601_05053535_S_Cyg-X1-W0.40+270.root
20160602_05053546_S_RBS0970-W0.40+000.root
20160602_05053561_S_PG1553+113-W0.40+090.root
20160602_05053563_S_PG1553+113-W0.40+000.root
20160602_05053571_S_Cyg-X1-W0.40+000.root
20160602_05053572_S_Cyg-X1-W0.40+180.root
20160602_05053573_S_Cyg-X1-W0.40+090.root
20160602_05053574_S_Cyg-X1-W0.40+270.root
20160602_05053575_S_Cyg-X1-W0.40+000.root
20160602_05053576_S_Cyg-X1-W0.40+180.root
20160603_05053586_S_RBS0970-W0.40+000.root
20160603_05053593_S_PG1553+113-W0.40+000.root
20160603_05053594_S_PG1553+113-W0.40+180.root
20160603_05053595_S_PG1553+113-W0.40+090.root
20160603_05053608_S_Cyg-X1-W0.40+090.root
20160603_05053609_S_Cyg-X1-W0.40+270.root
20160603_05053611_S_Cyg-X1-W0.40+180.root
20160603_05053612_S_Cyg-X1-W0.40+090.root
20160603_05053613_S_Cyg-X1-W0.40+270.root
20160604_05053623_S_RBS0970-W0.40+180.root
20160604_05053625_S_RBS0970-W0.40+270.root
20160604_05053626_S_RBS0970-W0.40+000.root
20160604_05053647_S_Cyg-X1-W0.40+000.root
20160604_05053648_S_Cyg-X1-W0.40+180.root
20160604_05053651_S_Cyg-X1-W0.40+000.root
20160604_05053652_S_Cyg-X1-W0.40+180.root
20160605_05053663_S_RBS0970-W0.40+090.root
20160605_05053665_S_RBS0970-W0.40+000.root
20160605_05053666_S_RBS0970-W0.40+180.root
20160605_05053667_S_RBS0970-W0.40+090.root
20160606_05053695_S_RBS0970-W0.40+270.root
20160606_05053696_S_RBS0970-W0.40+000.root
20160606_05053697_S_RBS0970-W0.40+180.root
20160606_05053698_S_RBS0970-W0.40+090.root
20160607_05053731_S_RBS0970-W0.40+270.root
20160607_05053732_S_RBS0970-W0.40+000.root
20160607_05053733_S_RBS0970-W0.40+180.root
20160607_05053734_S_RBS0970-W0.40+090.root
20160607_05053735_S_RBS0970-W0.40+270.root
20160607_05053736_S_RBS0970-W0.40+000.root
20160608_05053767_S_RBS0970-W0.40+180.root
20160608_05053768_S_RBS0970-W0.40+090.root
20160608_05053769_S_RBS0970-W0.40+270.root
20160608_05053770_S_RBS0970-W0.40+000.root
20160608_05053771_S_RBS0970-W0.40+180.root
20160609_05053802_S_M87-W0.40+180.root

20160609_05053803_S_M87-W0.40+180.root
20160609_05053806_S_M87-W0.40+000.root
20160610_05053836_S_PG1553+113-W0.40+270.root
20160610_05053837_S_PG1553+113-W0.40+000.root
20160610_05053838_S_PG1553+113-W0.40+180.root
20160610_05053839_S_PG1553+113-W0.40+090.root
20160610_05053847_S_PG1553+113-W0.40+090.root
20160610_05053848_S_PG1553+113-W0.40+270.root
20160610_05053849_S_PG1553+113-W0.40+000.root
20160615_05053989_S_M15-W0.40+000.root
20160615_05053990_S_M15-W0.40+180.root
20160615_05053991_S_M15-W0.40+090.root
20160701_05054260_S_PG1553+113-W0.40+180.root
20160701_05054261_S_PG1553+113-W0.40+090.root
20160707_05054482_S_M15-W0.40+180.root
20160707_05054483_S_M15-W0.40+090.root
20160707_05054484_S_M15-W0.40+270.root
20160707_05054485_S_M15-W0.40+000.root
20160707_05054486_S_M15-W0.40+180.root
20160707_05054487_S_M15-W0.40+090.root
20160709_05054513_S_M15-W0.40+180.root
20160709_05054514_S_M15-W0.40+090.root
20160709_05054515_S_M15-W0.40+270.root
20160709_05054516_S_M15-W0.40+000.root
20160709_05054517_S_M15-W0.40+180.root
20160709_05054518_S_M15-W0.40+090.root
20160709_05054519_S_M15-W0.40+270.root
20160710_05054524_S_PG1553+113-W0.40+000.root
20160710_05054525_S_PG1553+113-W0.40+180.root
20160710_05054526_S_PG1553+113-W0.40+090.root
20160710_05054534_S_DarkPatch40.root
20160710_05054541_S_DarkPatch39.root
20160710_05054548_S_M15-W0.40+270.root
20160710_05054549_S_M15-W0.40+000.root
20160710_05054550_S_M15-W0.40+180.root
20160710_05054551_S_M15-W0.40+090.root
20160710_05054554_S_M15-W0.40+090.root
20160710_05054555_S_M15-W0.40+270.root
20160711_05054578_S_M15-W0.40+000.root
20160711_05054579_S_M15-W0.40+180.root
20160711_05054580_S_M15-W0.40+180.root
20160711_05054581_S_M15-W0.40+090.root
20160711_05054582_S_M15-W0.40+270.root
20160711_05054583_S_M15-W0.40+000.root
20160711_05054584_S_M15-W0.40+180.root
20160711_05054585_S_M15-W0.40+090.root
20160713_05054610_S_PSRJ2032-W0.40+051.root
20160713_05054612_S_PSRJ2032-W0.40+141.root
20160713_05054615_S_M15-W0.40+090.root
20160713_05054616_S_M15-W0.40+270.root
20160713_05054617_S_M15-W0.40+000.root
20160713_05054618_S_M15-W0.40+180.root
20160713_05054619_S_M15-W0.40+090.root
20160713_05054620_S_M15-W0.40+270.root
20160713_05054621_S_M15-W0.40+000.root
20160714_05054651_S_M15-W0.40+180.root
20160714_05054652_S_M15-W0.40+090.root
20160714_05054653_S_M15-W0.40+270.root
20160714_05054654_S_M15-W0.40+000.root
20160714_05054655_S_M15-W0.40+180.root
20160714_05054656_S_M15-W0.40+090.root
20160715_05054674_S_M15-W0.40+180.root
20160717_05054732_S_M15-W0.40+180.root
20160717_05054733_S_M15-W0.40+090.root
20160717_05054734_S_M15-W0.40+270.root
20160729_05055071_S_M15-W0.40+180.root
20160729_05055072_S_M15-W0.40+090.root
20160729_05055073_S_M15-W0.40+090.root
20160729_05055074_S_M15-W0.40+270.root
20160729_05055077_S_M15-W0.40+000.root
20160729_05055078_S_M15-W0.40+000.root
20160730_05055125_S_M15-W0.40+180.root
20160730_05055126_S_M15-W0.40+180.root
20160730_05055127_S_M15-W0.40+090.root
20160730_05055128_S_M15-W0.40+270.root
20160730_05055129_S_M15-W0.40+270.root
20160730_05055130_S_M15-W0.40+000.root
20160730_05055131_S_M15-W0.40+000.root
20160730_05055132_S_M15-W0.40+180.root
20160730_05055133_S_M15-W0.40+090.root
20160730_05055134_S_M15-W0.40+270.root
20160730_05055139_S_M15-W0.40+270.root
20160801_05055222_S_TXS2241-W0.40+180.root
20160801_05055225_S_M15-W0.40+180.root
20160801_05055226_S_M15-W0.40+090.root
20160801_05055227_S_M15-W0.40+270.root
20160802_05055253_S_M15-W0.40+270.root
20160802_05055254_S_M15-W0.40+000.root
20160802_05055255_S_M15-W0.40+180.root
20160802_05055256_S_M15-W0.40+090.root
20160802_05055257_S_M15-W0.40+270.root
20160802_05055258_S_M15-W0.40+000.root
20160802_05055259_S_M15-W0.40+180.root
20160802_05055260_S_M15-W0.40+090.root
20160802_05055261_S_M15-W0.40+270.root
20160802_05055262_S_M15-W0.40+000.root
20160802_05055263_S_M15-W0.40+180.root
20160803_05055287_S_M15-W0.40+000.root

20160803_05055288_S_M15-W0.40+180.root
20160803_05055289_S_M15-W0.40+090.root
20160803_05055290_S_M15-W0.40+270.root
20160803_05055291_S_M15-W0.40+000.root
20160803_05055292_S_M15-W0.40+180.root
20160803_05055293_S_M15-W0.40+090.root
20160803_05055294_S_M15-W0.40+270.root
20160803_05055295_S_M15-W0.40+000.root
20160803_05055296_S_M15-W0.40+180.root
20160803_05055297_S_M15-W0.40+090.root
20160804_05055310_S_PG1553+113-W0.40+270.root
20160804_05055311_S_PG1553+113-W0.40+000.root
20160804_05055312_S_PG1553+113-W0.40+180.root
20160804_05055313_S_PG1553+113-W0.40+090.root
20160804_05055320_S_M15-W0.40+090.root
20160804_05055321_S_M15-W0.40+270.root
20160804_05055324_S_M15-W0.40+090.root
20160804_05055325_S_M15-W0.40+270.root
20160804_05055326_S_M15-W0.40+000.root
20160807_05055429_S_M15-W0.40+090.root
20160807_05055430_S_M15-W0.40+270.root
20160807_05055431_S_M15-W0.40+000.root
20160808_05055446_S_PG1553+113-W0.40+000.root
20160808_05055447_S_PG1553+113-W0.40+180.root
20160808_05055448_S_PG1553+113-W0.40+090.root
20160808_05055449_S_PG1553+113-W0.40+270.root
20160808_05055458_S_M15-W0.40+180.root
20160808_05055459_S_M15-W0.40+090.root
20160808_05055460_S_M15-W0.40+270.root
20160808_05055461_S_M15-W0.40+000.root
20160808_05055462_S_M15-W0.40+180.root
20160808_05055463_S_M15-W0.40+090.root
20160808_05055464_S_M15-W0.40+270.root
20160808_05055465_S_M15-W0.40+000.root
20160808_05055466_S_M15-W0.40+180.root
20160808_05055467_S_M15-W0.40+090.root
20160808_05055468_S_M15-W0.40+270.root
20160808_05055471_S_3FGLJ2346+07-W0.40+270.root
20160808_05055472_S_3FGLJ2346+07-W0.40+000.root
20160808_05055474_S_3FGLJ2346+07-W0.40+090.root
20160808_05055475_S_3FGLJ2346+07-W0.40+270.root
20160809_05055489_S_M15-W0.40+090.root
20160809_05055490_S_M15-W0.40+270.root
20160809_05055491_S_M15-W0.40+000.root
20160809_05055492_S_M15-W0.40+180.root
20160809_05055493_S_M15-W0.40+090.root
20160809_05055494_S_M15-W0.40+270.root
20160809_05055495_S_M15-W0.40+000.root
20160809_05055496_S_M15-W0.40+180.root
20160809_05055502_S_3FGLJ2346+07-W0.40+270.root
20160809_05055504_S_3FGLJ2346+07-W0.40+180.root
20160809_05055505_S_3FGLJ2346+07-W0.40+180.root
20160809_05055506_S_3FGLJ2346+07-W0.40+090.root
20160809_05055507_S_3FGLJ2346+07-W0.40+270.root
20160813_05055624_S_M15-W0.40+180.root
20160813_05055625_S_M15-W0.40+090.root
20160813_05055629_S_3FGLJ2346+07-W0.40+000.root
20160813_05055632_S_3FGLJ2346+07-W0.40+270.root
20160813_05055633_S_3FGLJ2346+07-W0.40+000.root
20160813_05055635_S_3FGLJ2346+07-W0.40+090.root
20160814_05055655_S_PSRJ2032-W0.40+051.root
20160814_05055659_S_3FGLJ2346+07-W0.40+180.root
20160814_05055660_S_3FGLJ2346+07-W0.40+090.root
20160814_05055663_S_S30218+35-W0.40+180.root
20160814_05055664_S_S30218+35-W0.40+090.root
20160814_05055665_S_S30218+35-W0.40+270.root
20160814_05055666_S_S30218+35-W0.40+000.root
20160814_05055667_S_S30218+35-W0.40+180.root
20160814_05055670_S_3FGLJ2346+07-W0.40+090.root
20160814_05055671_S_3FGLJ2346+07-W0.40+270.root
20160814_05055672_S_3FGLJ2346+07-W0.40+000.root
20160824_05055772_S_Cyg-X3-W0.40+141.root
20160824_05055774_S_Cyg-X3-W0.40+051.root
20160824_05055776_S_Cyg-X3-W0.40+141.root
20160824_05055777_S_Cyg-X3-W0.40+321.root
20160826_05055860_S_PSRJ2032-W0.40+321.root
20160826_05055861_S_PSRJ2032-W0.40+051.root
20160826_05055862_S_PSRJ2032-W0.40+231.root
20160827_05055877_S_Cyg-X3-W0.40+231.root
20160827_05055878_S_Cyg-X3-W0.40+141.root
20160827_05055879_S_Cyg-X3-W0.40+321.root
20160829_05055953_S_Cyg-X3-W0.40+051.root
20160829_05055955_S_Cyg-X3-W0.40+141.root
20160829_05055956_S_Cyg-X3-W0.40+321.root
20160829_05055970_S_M15-W0.40+180.root
20160829_05055971_S_M15-W0.40+090.root
20160829_05055980_S_TriangulumII-W0.40+148.root
20160829_05055981_S_TriangulumII-W0.40+328.root
20160830_05055995_S_Cyg-X3-W0.40+231.root
20160830_05056027_S_TriangulumII-W0.40+148.root
20160830_05056030_S_TriangulumII-W0.40+148.root
20160830_05056031_S_TriangulumII-W0.40+328.root
20160830_05056032_S_TriangulumII-W0.40+148.root
20160830_05056033_S_TriangulumII-W0.40+328.root
20160830_05056034_S_TriangulumII-W0.40+148.root
20160831_05056043_S_Cyg-X3-W0.40+051.root
20160831_05056044_S_Cyg-X3-W0.40+231.root
20160831_05056045_S_Cyg-X3-W0.40+141.root

20160831_05056046_S_Cyg-X3-W0.40+321.root
20160831_05056049_S_Cyg-X3-W0.40+141.root
20160831_05056050_S_Cyg-X3-W0.40+321.root
20160831_05056051_S_Cyg-X3-W0.40+051.root
20160831_05056052_S_Cyg-X3-W0.40+231.root
20160831_05056053_S_Cyg-X3-W0.40+141.root
20160831_05056061_S_M15-W0.40+270.root
20160831_05056071_S_S30218+35-W0.40+270.root
20160831_05056072_S_S30218+35-W0.40+000.root
20160831_05056077_S_TriangulumII-W0.40+328.root
20160831_05056078_S_TriangulumII-W0.40+148.root
20160831_05056079_S_TriangulumII-W0.40+328.root
20160831_05056080_S_TriangulumII-W0.40+148.root
20160831_05056081_S_TriangulumII-W0.40+328.root
20160831_05056082_S_TriangulumII-W0.40+148.root
20160831_05056083_S_TriangulumII-W0.40+328.root
20160901_05056089_S_Cyg-X3-W0.40+141.root
20160901_05056090_S_Cyg-X3-W0.40+321.root
20160901_05056094_S_Cyg-X3-W0.40+321.root
20160901_05056095_S_Cyg-X3-W0.40+051.root
20160901_05056097_S_Cyg-X3-W0.40+141.root
20160901_05056098_S_Cyg-X3-W0.40+321.root
20160901_05056105_S_M15-W0.40+090.root
20160901_05056106_S_M15-W0.40+270.root
20160901_05056117_S_S30218+35-W0.40+270.root
20160901_05056118_S_S30218+35-W0.40+000.root
20160901_05056123_S_TriangulumII-W0.40+148.root
20160901_05056125_S_TriangulumII-W0.40+148.root
20160901_05056126_S_TriangulumII-W0.40+328.root
20160901_05056127_S_TriangulumII-W0.40+148.root
20160901_05056128_S_TriangulumII-W0.40+328.root
20160902_05056153_S_S30218+35-W0.40+090.root
20160902_05056156_S_S30218+35-W0.40+180.root
20160903_05056171_S_Cyg-X3-W0.40+321.root
20160903_05056172_S_Cyg-X3-W0.40+051.root
20160903_05056175_S_Cyg-X3-W0.40+141.root
20160903_05056176_S_Cyg-X3-W0.40+321.root
20160903_05056177_S_Cyg-X3-W0.40+051.root
20160903_05056178_S_Cyg-X3-W0.40+231.root
20160903_05056181_S_Cyg-X3-W0.40+051.root
20160903_05056182_S_Cyg-X3-W0.40+231.root
20160903_05056183_S_Cyg-X3-W0.40+141.root
20160903_05056184_S_Cyg-X3-W0.40+321.root
20160903_05056185_S_Cyg-X3-W0.40+051.root
20160903_05056186_S_Cyg-X3-W0.40+231.root
20160903_05056187_S_Cyg-X3-W0.40+141.root
20160903_05056188_S_Cyg-X3-W0.40+321.root
20160903_05056197_S_TriangulumII-W0.40+328.root
20160903_05056199_S_TriangulumII-W0.40+328.root
20160903_05056200_S_TriangulumII-W0.40+148.root
20160903_05056202_S_TriangulumII-W0.40+148.root
20160904_05056207_S_Cyg-X3-W0.40+051.root
20160904_05056211_S_Cyg-X3-W0.40+051.root
20160904_05056212_S_Cyg-X3-W0.40+231.root
20160904_05056213_S_Cyg-X3-W0.40+141.root
20160904_05056214_S_Cyg-X3-W0.40+321.root
20160904_05056217_S_Cyg-X3-W0.40+141.root
20160904_05056218_S_Cyg-X3-W0.40+321.root
20160904_05056219_S_Cyg-X3-W0.40+051.root
20160904_05056220_S_Cyg-X3-W0.40+231.root
20160904_05056221_S_Cyg-X3-W0.40+141.root
20160904_05056222_S_Cyg-X3-W0.40+321.root
20160904_05056223_S_Cyg-X3-W0.40+051.root
20160904_05056224_S_Cyg-X3-W0.40+231.root
20160904_05056227_S_LSI+61303-W0.40+000.root
20160904_05056228_S_LSI+61303-W0.40+180.root
20160904_05056229_S_LSI+61303-W0.40+090.root
20160904_05056230_S_LSI+61303-W0.40+270.root
20160904_05056231_S_LSI+61303-W0.40+000.root
20160904_05056235_S_FRB121102.root
20160904_05056236_S_FRB121102.root
20160904_05056237_S_FRB121102.root
20160905_05056243_S_Cyg-X3-W0.40+141.root
20160905_05056244_S_Cyg-X3-W0.40+321.root
20160905_05056247_S_Cyg-X3-W0.40+141.root
20160905_05056248_S_Cyg-X3-W0.40+321.root
20160905_05056249_S_Cyg-X3-W0.40+051.root
20160905_05056254_S_Cyg-X3-W0.40+231.root
20160905_05056255_S_Cyg-X3-W0.40+141.root
20160905_05056257_S_Cyg-X3-W0.40+051.root
20160905_05056258_S_Cyg-X3-W0.40+231.root
20160905_05056259_S_Cyg-X3-W0.40+141.root
20160905_05056260_S_Cyg-X3-W0.40+321.root
20160905_05056263_S_LSI+61303-W0.40+000.root
20160905_05056264_S_LSI+61303-W0.40+180.root
20160905_05056265_S_LSI+61303-W0.40+090.root
20160905_05056266_S_LSI+61303-W0.40+270.root
20160905_05056267_S_LSI+61303-W0.40+000.root
20160905_05056271_S_TriangulumII-W0.40+328.root
20160905_05056272_S_TriangulumII-W0.40+148.root
20160905_05056273_S_TriangulumII-W0.40+328.root
20160906_05056284_S_Cyg-X3-W0.40+321.root
20160906_05056289_S_Cyg-X3-W0.40+051.root
20160906_05056291_S_Cyg-X3-W0.40+141.root
20160906_05056293_S_Cyg-X3-W0.40+051.root
20160906_05056294_S_Cyg-X3-W0.40+231.root
20160906_05056295_S_Cyg-X3-W0.40+141.root
20160906_05056298_S_LSI+61303-W0.40+000.root

20160906_05056299_S_LSI+61303-W0.40+180.root
20160906_05056300_S_LSI+61303-W0.40+090.root
20160906_05056301_S_LSI+61303-W0.40+270.root
20160906_05056302_S_LSI+61303-W0.40+000.root
20160906_05056303_S_LSI+61303-W0.40+180.root
20160907_05056314_S_Cyg-X3-W0.40+321.root
20160907_05056319_S_Cyg-X3-W0.40+051.root
20160907_05056320_S_Cyg-X3-W0.40+231.root
20160907_05056323_S_Cyg-X3-W0.40+051.root
20160907_05056324_S_Cyg-X3-W0.40+231.root
20160907_05056325_S_Cyg-X3-W0.40+141.root
20160907_05056326_S_Cyg-X3-W0.40+321.root
20160907_05056327_S_Cyg-X3-W0.40+051.root
20160907_05056328_S_Cyg-X3-W0.40+231.root
20160907_05056329_S_Cyg-X3-W0.40+141.root
20160907_05056330_S_Cyg-X3-W0.40+321.root
20160907_05056333_S_S30218+35-W0.40+090.root
20160907_05056336_S_S30218+35-W0.40+180.root
20160907_05056340_S_LSI+61303-W0.40+270.root
20160907_05056341_S_LSI+61303-W0.40+000.root
20160907_05056342_S_LSI+61303-W0.40+180.root
20160907_05056343_S_LSI+61303-W0.40+090.root
20160908_05056367_S_M15-W0.40+000.root
20160908_05056368_S_M15-W0.40+180.root
20160908_05056369_S_M15-W0.40+090.root
20160908_05056383_S_S30218+35-W0.40+270.root
20160908_05056386_S_LSI+61303-W0.40+270.root
20160908_05056387_S_LSI+61303-W0.40+000.root
20160908_05056390_S_LSI+61303-W0.40+270.root
20160909_05056417_S_TriangulumII-W0.40+148.root
20160909_05056418_S_TriangulumII-W0.40+328.root
20160909_05056421_S_TriangulumII-W0.40+148.root
20160909_05056422_S_TriangulumII-W0.40+328.root
20160909_05056424_S_TriangulumII-W0.40+328.root
20160909_05056425_S_TriangulumII-W0.40+148.root
20160909_05056426_S_TriangulumII-W0.40+328.root
20160910_05056454_S_S30218+35-W0.40+000.root
20160910_05056455_S_S30218+35-W0.40+180.root
20160910_05056456_S_S30218+35-W0.40+090.root
20160910_05056465_S_TriangulumII-W0.40+328.root
20160910_05056466_S_TriangulumII-W0.40+148.root
20160910_05056468_S_TriangulumII-W0.40+148.root
20160911_05056499_S_TriangulumII-W0.40+328.root
20160911_05056500_S_TriangulumII-W0.40+148.root
20160911_05056501_S_TriangulumII-W0.40+328.root
20160911_05056510_S_TriangulumII-W0.40+148.root
20160911_05056511_S_TriangulumII-W0.40+328.root
20160911_05056512_S_TriangulumII-W0.40+148.root
20160911_05056513_S_TriangulumII-W0.40+328.root
20160912_05056546_S_S30218+35-W0.40+270.root
20160912_05056547_S_S30218+35-W0.40+000.root
20160912_05056550_S_TriangulumII-W0.40+148.root
20160912_05056551_S_TriangulumII-W0.40+328.root
20160913_05056580_S_TriangulumII-W0.40+328.root
20160913_05056581_S_TriangulumII-W0.40+148.root
20160913_05056582_S_TriangulumII-W0.40+328.root
20160913_05056583_S_TriangulumII-W0.40+148.root
20160913_05056584_S_TriangulumII-W0.40+328.root
20160921_05056821_S_Cyg-X3-W0.40+231.root
20160921_05056822_S_Cyg-X3-W0.40+141.root
20160921_05056823_S_Cyg-X3-W0.40+321.root
20160921_05056824_S_Cyg-X3-W0.40+051.root
20160922_05056857_S_Cyg-X3-W0.40+141.root
20160922_05056858_S_Cyg-X3-W0.40+321.root
20160922_05056859_S_Cyg-X3-W0.40+051.root
20160923_05056906_S_M15-W0.40+270.root
20160923_05056907_S_M15-W0.40+000.root
20160924_05056943_S_M15-W0.40+270.root
20160924_05056944_S_M15-W0.40+000.root
20160924_05056945_S_M15-W0.40+180.root
20160924_05056946_S_M15-W0.40+090.root
20160925_05056972_S_BLLac-W0.40+180.root
20160925_05056975_S_M15-W0.40+270.root
20160925_05056981_S_M15-W0.40+180.root
20160925_05056982_S_M15-W0.40+090.root
20160925_05056983_S_M15-W0.40+270.root
20160925_05056986_S_S30218+35-W0.40+000.root
20160925_05056988_S_S30218+35-W0.40+180.root
20160925_05056991_S_TriangulumII-W0.40+328.root
20160925_05057004_S_NGC1068-W0.40+000.root
20160925_05057005_S_NGC1068-W0.40+180.root
20160925_05057006_S_NGC1068-W0.40+090.root
20160925_05057007_S_NGC1068-W0.40+270.root
20160927_05057044_S_1ES2037+521-W0.40+090.root
20160927_05057045_S_1ES2037+521-W0.40+270.root
20160927_05057047_S_1ES2037+521-W0.40+180.root
20160927_05057051_S_TriangulumII-W0.40+328.root
20160927_05057056_S_TriangulumII-W0.40+148.root
20160927_05057057_S_TriangulumII-W0.40+328.root
20160927_05057061_S_1ES0229+200-W0.40+180.root
20160928_05057075_S_GRB160927-W0.40+000.root
20160928_05057077_S_GRB160927-W0.40+090.root
20160928_05057078_S_GRB160927-W0.40+270.root
20160928_05057085_S_1ES2037+521-W0.40+090.root
20160928_05057086_S_1ES2037+521-W0.40+270.root
20160928_05057087_S_1ES2037+521-W0.40+000.root
20160928_05057089_S_1ES2037+521-W0.40+090.root
20160928_05057090_S_1ES2037+521-W0.40+270.root

20160928_05057093_S_1ES2037+521-W0.40+090.root
20160928_05057099_S_TriangulumII-W0.40+328.root
20160928_05057100_S_TriangulumII-W0.40+148.root
20160928_05057101_S_TriangulumII-W0.40+328.root
20160928_05057102_S_TriangulumII-W0.40+148.root
20160929_05057117_S_1ES2037+521-W0.40+180.root
20160929_05057118_S_1ES2037+521-W0.40+090.root
20160929_05057119_S_1ES2037+521-W0.40+270.root
20160929_05057121_S_1ES2037+521-W0.40+180.root
20160929_05057123_S_1ES2037+521-W0.40+270.root
20160929_05057124_S_1ES2037+521-W0.40+000.root
20160929_05057125_S_1ES2037+521-W0.40+180.root
20160929_05057126_S_1ES2037+521-W0.40+090.root
20160929_05057127_S_1ES2037+521-W0.40+270.root
20160929_05057128_S_1ES2037+521-W0.40+000.root
20160929_05057129_S_1ES2037+521-W0.40+180.root
20160929_05057130_S_1ES2037+521-W0.40+090.root
20160929_05057131_S_1ES2037+521-W0.40+270.root
20160929_05057136_S_TriangulumII-W0.40+328.root
20160929_05057140_S_TriangulumII-W0.40+328.root
20160930_05057161_S_1ES2037+521-W0.40+090.root
20160930_05057162_S_1ES2037+521-W0.40+270.root
20160930_05057163_S_1ES2037+521-W0.40+000.root
20160930_05057164_S_1ES2037+521-W0.40+180.root
20160930_05057172_S_S30218+35-W0.40+000.root
20160930_05057173_S_S30218+35-W0.40+180.root
20160930_05057178_S_TriangulumII-W0.40+328.root
20160930_05057179_S_TriangulumII-W0.40+148.root
20160930_05057183_S_1ES0229+200-W0.40+180.root
20160930_05057185_S_1ES0229+200-W0.40+270.root
20161001_05057206_S_1ES2037+521-W0.40+090.root
20161001_05057207_S_1ES2037+521-W0.40+270.root
20161001_05057208_S_1ES2037+521-W0.40+000.root
20161001_05057209_S_1ES2037+521-W0.40+180.root
20161001_05057210_S_1ES2037+521-W0.40+090.root
20161001_05057211_S_1ES2037+521-W0.40+270.root
20161001_05057219_S_S30218+35-W0.40+090.root
20161001_05057225_S_TriangulumII-W0.40+148.root
20161001_05057228_S_1ES0229+200-W0.40+090.root
20161001_05057229_S_1ES0229+200-W0.40+270.root
20161002_05057257_S_1ES2037+521-W0.40+000.root
20161002_05057258_S_1ES2037+521-W0.40+180.root
20161002_05057259_S_1ES2037+521-W0.40+090.root
20161002_05057260_S_1ES2037+521-W0.40+270.root
20161002_05057261_S_1ES2037+521-W0.40+000.root
20161002_05057271_S_TriangulumII-W0.40+328.root
20161004_05057347_S_1ES2037+521-W0.40+000.root
20161004_05057348_S_1ES2037+521-W0.40+180.root
20161004_05057356_S_TriangulumII-W0.40+328.root
20161008_05057419_S_1ES2037+521-W0.40+000.root
20161008_05057421_S_1ES2037+521-W0.40+180.root
20161008_05057422_S_1ES2037+521-W0.40+090.root
20161008_05057423_S_1ES2037+521-W0.40+270.root
20161008_05057424_S_1ES2037+521-W0.40+000.root
20161008_05057425_S_1ES2037+521-W0.40+180.root
20161009_05057518_S_0748+333-W0.40+000.root
20161009_05057519_S_0748+333-W0.40+180.root
20161011_05057602_S_TriangulumII-W0.40+328.root
20161011_05057603_S_TriangulumII-W0.40+148.root
20161029_05057892_S_3FGLJ2346+07-W0.40+180.root
20161029_05057894_S_3FGLJ2346+07-W0.40+270.root
20161030_05057952_S_Dragonfly-W0.70+305.root
20161030_05057955_S_1ES0229+200-W0.40+000.root
20161030_05057956_S_1ES0229+200-W0.40+180.root
20161031_05057987_S_Dragonfly-W0.70+305.root
20161031_05057988_S_Dragonfly-W0.70+125.root
20161031_05057989_S_Dragonfly-W0.70+305.root
20161031_05057997_S_TriangulumII-W0.40+328.root
20161101_05058034_S_Dragonfly-W0.70+125.root
20161101_05058035_S_Dragonfly-W0.70+305.root
20161101_05058036_S_Dragonfly-W0.70+125.root
20161101_05058037_S_Dragonfly-W0.70+305.root
20161101_05058047_S_1ES0229+200-W0.40+180.root
20161101_05058061_S_TriangulumII-W0.40+328.root
20161101_05058062_S_TriangulumII-W0.40+148.root
20161102_05058086_S_Dragonfly-W0.70+125.root
20161102_05058087_S_Dragonfly-W0.70+305.root
20161102_05058094_S_TriangulumII-W0.40+148.root
20161205_05058855_S_3FGLJ2346+07-W0.40+180.root
20161205_05058856_S_3FGLJ2346+07-W0.40+090.root
20161205_05058857_S_3FGLJ2346+07-W0.40+270.root
20161205_05058858_S_3FGLJ2346+07-W0.40+000.root
20161205_05058859_S_3FGLJ2346+07-W0.40+180.root
20161205_05058863_S_NGC1068-W0.40+090.root
20161205_05058864_S_NGC1068-W0.40+270.root
20161205_05058865_S_NGC1068-W0.40+000.root
20161205_05058895_S_FRB121102.root
20161205_05058896_S_FRB121102.root
20161205_05058897_S_FRB121102.root
20161205_05058898_S_FRB121102.root
20161205_05058899_S_FRB121102.root
20161206_05058910_S_NGC1068-W0.40+270.root
20161206_05058913_S_NGC1068-W0.40+270.root
20161206_05058916_S_NGC1068-W0.40+000.root
20161206_05058917_S_NGC1068-W0.40+180.root
20161206_05058918_S_NGC1068-W0.40+090.root
20161206_05058919_S_NGC1068-W0.40+270.root
20161206_05058920_S_NGC1068-W0.40+000.root

20161206_05058951_S_DarkPatch32.root
20161206_05058957_S_DarkPatch32.root
20161228_05059330_S_DarkPatch32.root
20161228_05059331_S_DarkPatch32.root
20161228_05059332_S_DarkPatch32.root
20161228_05059333_S_DarkPatch32.root
20170101_05059480_S_NGC1068-W0.40+180.root
20170101_05059484_S_TriangulumII-W0.40+148.root
20170116_05059731_S_LSI+61303-W0.40+000.root
20170116_05059732_S_LSI+61303-W0.40+180.root
20170116_05059733_S_LSI+61303-W0.40+090.root
20170116_05059734_S_LSI+61303-W0.40+270.root
20170117_05059772_S_1ES0229+200-W0.40+000.root
20170119_05059864_S_1ES0229+200-W0.40+180.root
20170119_05059865_S_1ES0229+200-W0.40+090.root
20170123_05059980_S_TriangulumII-W0.40+148.root
20170123_05059982_S_TriangulumII-W0.40+148.root
20170124_05060056_S_TON396-W0.40+180.root
20170125_05060079_S_TriangulumII-W0.40+148.root
20170125_05060080_S_TriangulumII-W0.40+328.root
20170125_05060115_S_TON396-W0.40+270.root
20170126_05060142_S_TriangulumII-W0.40+148.root
20170127_05060176_S_TriangulumII-W0.40+148.root
20170127_05060177_S_TriangulumII-W0.40+328.root
20170127_05060178_S_TriangulumII-W0.40+148.root
20170127_05060200_S_TON396-W0.40+270.root
20170127_05060206_S_TON396-W0.40+180.root
20170127_05060215_S_PG1553+113-W0.40+000.root
20170127_05060216_S_PG1553+113-W0.40+180.root
20170127_05060217_S_PG1553+113-W0.40+090.root
20170127_05060218_S_PG1553+113-W0.40+270.root
20170128_05060226_S_1ES0229+200-W0.40+090.root
20170128_05060272_S_PG1553+113-W0.40+180.root
20170128_05060273_S_PG1553+113-W0.40+090.root
20170128_05060275_S_PG1553+113-W0.40+000.root
20170129_05060284_S_TriangulumII-W0.40+148.root
20170129_05060312_S_TON396-W0.40+270.root
20170129_05060324_S_M87-W0.40+090.root
20170130_05060332_S_1ES0229+200-W0.40+090.root
20170130_05060361_S_TON396-W0.40+090.root
20170130_05060373_S_PG1553+113-W0.40+090.root
20170130_05060374_S_PG1553+113-W0.40+270.root
20170130_05060375_S_PG1553+113-W0.40+000.root
20170130_05060376_S_PG1553+113-W0.40+180.root
20170130_05060377_S_PG1553+113-W0.40+090.root
20170201_05060413_S_TXS0210+515-W0.40+180.root
20170201_05060414_S_TXS0210+515-W0.40+090.root
20170201_05060439_S_TON396-W0.40+180.root
20170202_05060459_S_TXS0210+515-W0.40+090.root
20170203_05060519_S_DarkPatch08.root
20170203_05060520_S_DarkPatch08.root
20170203_05060521_S_DarkPatch08.root
20170203_05060522_S_DarkPatch08.root
20170203_05060523_S_DarkPatch08.root
20170203_05060524_S_DarkPatch08.root
20170205_05060602_S_OJ287-W0.40+090.root
20170205_05060603_S_OJ287-W0.40+270.root
20170205_05060604_S_OJ287-W0.40+000.root
20170206_05060665_S_PG1553+113-W0.40+090.root
20170206_05060667_S_PG1553+113-W0.40+000.root
20170206_05060669_S_PG1553+113-W0.40+090.root
20170216_05061040_S_OJ287-W0.40+000.root
20170216_05061042_S_OJ287-W0.40+180.root
20170216_05061043_S_OJ287-W0.40+090.root
20170216_05061044_S_OJ287-W0.40+270.root
20170216_05061054_S_FRB121102.root
20170216_05061055_S_FRB121102.root
20170225_05061495_S_3c279-W0.40+180.root
20170225_05061496_S_3c279-W0.40+180.root
20170225_05061497_S_3c279-W0.40+090.root
20170225_05061500_S_PG1553+113-W0.40+180.root
20170225_05061501_S_PG1553+113-W0.40+090.root
20170225_05061502_S_PG1553+113-W0.40+270.root
20170225_05061503_S_PG1553+113-W0.40+000.root
20170226_05061542_S_3c279-W0.40+090.root
20170226_05061543_S_3c279-W0.40+270.root
20170226_05061544_S_3c279-W0.40+000.root
20170226_05061545_S_3c279-W0.40+180.root
20170226_05061548_S_2HWCJ1309-05-W0.40+000.root
20170226_05061549_S_2HWCJ1309-05-W0.40+180.root
20170226_05061550_S_2HWCJ1309-05-W0.40+090.root
20170226_05061551_S_2HWCJ1309-05-W0.40+270.root
20170227_05061599_S_DarkPatch08.root
20170227_05061600_S_DarkPatch08.root
20170227_05061601_S_DarkPatch08.root
20170227_05061602_S_DarkPatch08.root
20170228_05061650_S_2HWCJ1309-05-W0.40+105.root
20170228_05061652_S_2HWCJ1309-05-W0.40+105.root
20170228_05061653_S_2HWCJ1309-05-W0.40+285.root
20170228_05061656_S_PG1553+113-W0.40+270.root
20170228_05061657_S_PG1553+113-W0.40+000.root
20170228_05061658_S_PG1553+113-W0.40+180.root
20170228_05061659_S_PG1553+113-W0.40+090.root
20170304_05061830_S_DarkPatch09.root
20170304_05061831_S_DarkPatch09.root
20170305_05061854_S_M87-W0.40+000.root
20170305_05061855_S_M87-W0.40+180.root
20170305_05061863_S_2HWCJ1309-05-W0.40+285.root

20170305_05061864_S_2HWCJ1309-05-W0.40+105.root
20170305_05061865_S_2HWCJ1309-05-W0.40+285.root
20170321_05062097_S_M87-W0.40+180.root
20170321_05062098_S_M87-W0.40+090.root
20170321_05062099_S_M87-W0.40+270.root
20170322_05062139_S_M87-W0.40+000.root
20170322_05062142_S_M87-W0.40+000.root
20170322_05062143_S_M87-W0.40+180.root
20170322_05062144_S_M87-W0.40+090.root
20170322_05062145_S_M87-W0.40+270.root
20170324_05062224_S_3c279-W0.40+090.root
20170325_05062261_S_M87-W0.40+180.root
20170330_05062461_S_M87-W0.40+270.root
20170402_05062603_S_M87-W0.40+000.root
20170402_05062614_S_3c279-W0.40+270.root
20170403_05062653_S_M87-W0.40+180.root
20170403_05062676_S_PG1553+113-W0.40+000.root
20170403_05062682_S_PG1553+113-W0.40+000.root
20170405_05062758_S_M87-W0.40+270.root
20170405_05062761_S_M87-W0.40+270.root
20170405_05062764_S_M87-W0.40+000.root
20170405_05062765_S_M87-W0.40+180.root
20170419_05063050_S_PG1553+113-W0.40+180.root
20170420_05063075_S_TON396-W0.40+180.root
20170420_05063087_S_M87-W0.40+180.root
20170421_05063118_S_TON396-W0.40+180.root
20170422_05063165_S_TON396-W0.40+090.root
20170422_05063186_S_PG1553+113-W0.40+180.root
20170422_05063187_S_PG1553+113-W0.40+090.root
20170423_05063255_S_GammaCygni-W0.60+225.root
20170423_05063256_S_GammaCygni-W0.60+045.root
20170423_05063257_S_GammaCygni-W0.60+225.root
20170424_05063264_S_TON396-W0.40+270.root
20170425_05063321_S_M87-W0.40+090.root
20170425_05063324_S_PG1553+113-W0.40+180.root
20170425_05063325_S_PG1553+113-W0.40+090.root
20170425_05063327_S_PG1553+113-W0.40+270.root
20170425_05063343_S_GammaCygni-W0.60+225.root
20170425_05063344_S_GammaCygni-W0.60+045.root
20170425_05063345_S_GammaCygni-W0.60+225.root
20170501_05063353_S_TON396-W0.40+000.root
20170503_05063460_S_GammaCygni-W0.60+045.root
20170503_05063461_S_GammaCygni-W0.60+225.root
20170518_05063693_S_3c279-W0.40+000.root
20170518_05063710_S_PG1553+113-W0.40+000.root
20170519_05063760_S_GammaCygni-W0.60+045.root
20170519_05063761_S_GammaCygni-W0.60+225.root
20170521_05063812_S_3c279-W0.40+000.root
20170521_05063813_S_3c279-W0.40+180.root
20170521_05063816_S_3c279-W0.40+000.root
20170522_05063850_S_3c279-W0.40+000.root
20170522_05063851_S_3c279-W0.40+180.root
20170522_05063858_S_M87-W0.40+180.root
20170522_05063861_S_M87-W0.40+000.root
20170524_05063940_S_M87-W0.40+270.root
20170525_05063975_S_DarkPatch34.root
20170528_05064055_S_M87-W0.40+090.root
20170528_05064060_S_1ES1218+304-W0.40+090.root
20170529_05064099_S_1ES1218+304-W0.40+090.root
20170529_05064100_S_1ES1218+304-W0.40+270.root
20170529_05064104_S_DarkPatch11.root
20170530_05064144_S_M87-W0.40+180.root
20170530_05064145_S_M87-W0.40+090.root
20170530_05064146_S_M87-W0.40+270.root
20170531_05064174_S_PG1553+113-W0.40+090.root
20170531_05064180_S_DarkPatch34.root
20170531_05064181_S_DarkPatch34.root
20170531_05064182_S_DarkPatch34.root
20170603_05064326_S_3C345-W0.40+270.root

C Experiment List

ACE-CRIS Cosmic Ray Isotope Spectrometer (CRIS) measured cosmic rays during the solar minimum period in 2009-2010 onboard the *Advanced Composition Explorer* [79]. Previous CRIS measurements in the solar minimum period in 1997-1998 and the solar maximum period in 2001-2003 provide additional data.

ALICE A Large Isotopic Composition Experiment (ALICE) is a spectrometer that measured cosmic-ray data for 14.7 h in a balloon at an altitude of 36.6 km in the year 1987 [47].

AMS01 The alpha-magnetic spectrometer (AMS01) belongs to the CERN collaboration. It was launched on the Space Shuttle *Discovery* in 1998 to measure cosmic rays, particularly antihelium. Due to no detection, it was able to give an upper limit on the antihelium flux. Its successor, AMS02, was launched to the ISS on Space Shuttle *Endeavour* in 2011. AMS02, like its predecessor, is an antimatter detector and searches for evidence of dark matter in the universe.

Balloon Many less-known balloon experiments between 1950 and 1991 contribute to the dataset used in this work: these experiments are collected under the label **Balloon**
Fifties: [49, 50, 93, 92]
Sixties: [91, 46, 12, 45, 95, 133, 51, 53, 34, 48, 42, 20, 52, 106]
Seventies: [116, 114, 59, 115, 86, 81, 118]
Eighties: [96, 135, 131]
Nineties: [40, 72, 29, 75].

BESS The Balloon-borne Experiment with a Superconducting Spectrometer (BESS) was built to measure cosmic protons and helium nuclei in the lower GeV range. It was launched on different balloons, for example, at Lynn Lake, Canada [120] and balloon flights over Antarctica [5].

CAPRICE The Cosmic AntiParticle Ring Imaging Cherenkov Experiment (CAPRICE) is a setup of several detectors. In 1998, a balloon with the CAPRICE setup on board flew for about 21 h at an altitude of 36 to 38.2 km from Fort Sumner, New Mexico to Heber, Arizona. The detector setup consists of a gas Ring Imaging CHerenkov detector (Gas RICH), a time-of-flight device (TOF), a superconducting magnet spectrometer (Tracking system), and a silicon-tungsten imaging calorimeter [28].

CREAM The Cosmic Ray Energetics And Mass (CREAM) experiment consists of several subdetectors: a tungsten/scintillating fiber calorimeter, a dual-layer silicon charge detector, a Cherenkov camera, a Cherenkov detector, and a timing charge detector. The balloon with the CREAM detectors had missions in the Antarctic season (December/January): in the years 2004-2005, 2007-2008, lighter particles like protons and helium nuclei were measured [142], and in the year 2005-2006 data from heavier particles like iron nuclei were taken [11].

CRISIS Cosmic Ray Isotope Instrument System (CRISIS) is a balloon-borne experiment that consists of two scintillators, two Cerenkov radiators, a spark chamber, a block of nuclear emulsions, and a penetration scintillator. In 1977, the balloon flew over South Dakota, USA, for 56 h and 41 min [143].

CRN-*Spacelab2* The Cosmic Ray Nuclei (CRN) detector has two plastic scintillators, two gas Cerenkov counters, and a transition radiation detector. It detects heavier nuclei from carbon to iron nuclei and was launched on the *Spacelab-2* mission of the Space Shuttle *Challenger* in 1985 [100].

CRRES The *Combined Release and Radiation Effects Satellite* (*CRRES*) was launched in 1990. 1991 the mission was aborted after its main power system failed, most likely due to design problems [43]. The instrument consisted of charged particle position-sensing detectors, which during a solar maximum in 1990/1991, measured helium [32].

CTA The Cherenkov Telescope Array (CTA) project forms the next generation in high-energy and very high-energy gamma-ray astronomy: an array of over 100 telescopes at two different sites will enable new sensitivities, unprecedented sky, and energy coverage to be achieved. Three types of telescopes form the array: the Large, Medium, and Small-Sized Telescopes (LSTs, MSTs, and SSTs), each focusing on a different energy range. Two sites offer a unique view of the universe: the Paranal site of the European Southern Observatory (ESO) in Chile and the ORM at the Roque de los Muchachos site on La Palma, Canary Islands, allowing simultaneous observations of the northern and southern skies [127]. The prototype Large-Sized Telescope (LST-1) at the CTA-North site started astronomical observations in November 2019 in the lowest energy range of CTA above 20 GeV. Three additional LSTs (LST-2 to LST-4) are currently under construction at the same site [109].

DAMPE The *Dark Matter Particle Explorer* (*DAMPE*) is a satellite of the Chinese Academy of Sciences (CAS) that started its mission in 2015 from the Gobi Desert, Inner Mongolia. It contains the following detector setup: It contains a plastic scintillator detector (charge measurement), a silicon tungsten tracker-converter (trajectory measurement), a Bi3Ge4O12 electromagnetic calorimeter (energy measurement and electron-hadron discrimination), and a neutron detector (electron-hadron discrimination). The helium data used in this work is in the energy range of 70 GeV to 80 TeV [16].

Discoverer36 The *Discoverer36* is a counter telescope launched into polar orbit by satellite in 1961. The telescope was made of nine AuSi solid-state detectors and could collect data of cosmic protons until the vehicle transmitter failed [123].

E6 The two space probes *HELIOS A* and *HELIOS B* were sent into an orbit around the sun in 1974 and 1976. Experiment 6 (E6) on its board comprised five silicon semiconductor detectors and one Sapphire Cherenkov detector surrounded by a plastic anti-coincidence detector. It collected proton data until 1977 [88].

EPHIN The *SOlar and Heliospheric Observatory* (*SOHO*) is a satellite launched in 1995 and is expected to be on a mission until 2025. On its board, the Electron Proton Helium INstrument (EPHIN) collects proton dat (proton data published from 1995 to 2014). The experiment comprises six silicon solid-state detectors surrounded by a scintillator as anti-coincidence [78].

***Fermi*-LAT** The *Fermi* spacecraft was launched by NASA on 2008 and carried the high-energy gamma-ray instrument, the *Fermi* Large Area Telescope (*Fermi*-LAT). The air-conversion telescope with a precision tracker and calorimeter is sensitive to gamma rays from 0.02 GeV to 300 GeV and has a large field of view of 2.4 sr. The *Fermi*-LAT provides a source catalog of several thousand high-energy sources from all-sky surveys and the spectra of hundreds of them, GRBs alerts for the GCN network and works on additional interesting science projects like measure the diffuse isotropic gamma-ray background and search for dark matter [19].

HAWC The High Altitude Water Cherenkov (HAWC) Observatory is located at the flanks of the Sierra Negra volcano in Puebla, Mexico. 300 tanks cover a surface area of 22 000 squarem, each filled with 190 m^3 water. Inside the tanks, PMTs are installed to measure Cherenkov light from gamma-induced air showers. High Altitude Water Cherenkov (HAWC) is sensitive to gamma rays with energies from 0.5 TeV to 100 TeV [6].

***HEAO-3*-C2** The experiment C2 onboard the *HEAO-3* satellite was launched in 1979 by NASA. The experiment consists of five Cherenkov counters and provides cosmic-ray data from heavy particles like iron nuclei measured between 1979 and 1980 [44].

HEGRA The High Energy Gamma-Ray Astronomy (HEGRA) telescope array was located on the ORM in La Palma, Spain, and operated from 1987 to 2002: it consisted of several types of detectors such as scintillation counters and AIROBICC and several stereoscopic-operating IACTs. The HEGRA experiment was sensitive from 1 TeV to several PeV [17].

H. E. S. S. The High Energy Stereoscopic System (H. E. S. S.) is a gamma-ray instrument located in the Khomas Highland of Namibia. The four identical 12 m-diameter IACTs form a square with a side length of 120 m operating stereoscopically since 2004. Since 2012, a fifth telescope with a reflective area of 28 m in diameter completes the array [10]. High Energy Stereoscopic System (H. E. S. S.) has an energy threshold of 100 GeV [69].

IceCube IceCube observatory is a high-energy neutrino detector at the South Pole, completed in 2011. It consists of the main neutrino detector, the IceCube in-ice array, a denser core in the center of the array, DeepCore, and a cosmic-ray instrument on the arctic surface, IceTop. The main detector includes 1 km^3 of ice volume between 1450 m and 2450 m below the surface laced with 86 vertical strings. On each string, 60 digital optical modules measure Cherenkov light from propagating leptons, products of neutrino interactions with the arctic ice. DeepCore has an average inter-string spacing of 72 m, making it able to lower the energy threshold of neutrino detections. The surface detector IceTop includes 162 tanks filled with ice and carrying PMTs to detect Cherenkov light from cosmic-ray-induced air showers [1].

IMAX92 The Isotope Matter-Antimatter Experiment (IMAX) was a detector system installed on a balloon flight from Lynn Lake, Manitoba, Canada, in 1992. The system included a superconducting magnet spectrometer with scintillators, a time-of-flight system, and Cherenkov detectors. The IMAX92 was designed for the search of antiprotons and light isotopes, like protons and helium nuclei [94].

IMP The Interplanetary Monitoring Platform (IMP) is a satellite program managed by NASA. The *IMP-1* satellite, also known as *Explorer 18*, is the first satellite of the series launched with a rocket on 27 November 1963 from Cape Canaveral Air Force Station (CCAFS), Florida, to the Earth's orbit. *IMP-1* has several onboard detectors; a four-element charged particle telescope was used to measure protons and helium nuclei [64]. Data from the missions of the *IMP-1* and its successors *IMP-3* [21], *IMP-5* [90], *IMP-7* [58] and *IMP-8* [24] are used in this work.

JACEE The Japanese-American Cooperative Emulsion Experiment (JACEE) is a balloon-borne detector designed to measure cosmic rays in the TeV range. Between 1979 and 1995, 12 flights were performed: several over Texas and Antarctica and several long-duration flights from Australia to South America. The detector consists of a fine-grained emulsion chamber with about a hundred track-sensitive nuclear emulsion plates and a three-dimensional emulsion/X-ray film/lead plate calorimeter [18].

KASCADE Karlsruhe Shower Core and Array Detector (KASCADE), later KASCADE-Grande, was an experiment inaugurated in 1996, and its successor was dismantled in 2013. It was located at KIT-Karlsruhe, Germany, and consisted of two types of haptic detectors: a large scintillator array for detecting charged particles and an array of shielded scintillation counters for counting muons. It was able to measure cosmic rays with energies from the ten PeV to EeV range [54].

LEAP The Low Energy Antiproton (LEAP) balloon flight experiment was built to detect protons and helium nuclei between 200 MeV and 100 GeV per nucleon. It was launched in 1987 from Prince Albert, Canada, during a solar minimum. The detector included a superconducting magnet spectrometer with a multiwire proportional counter-tracking system, a time-of-flight (TOF) system, and a liquid Cherenkov detector [119].

LHAASO Large High Altitude Air Shower Observatory (LHAASO) was inaugurated in Daocheng, Sichuan, China in 2019. Different detector types together form the experiment: a 1.3 km^2 array (KM2A) of electromagnetic particle detectors (ED) and muon detectors (MD), a water Cherenkov detector array (WCDA) with a total active area of 78 000 m^2, 18 wide-field-of-view air Cherenkov telescopes (WFCTA), and a newly proposed 10 000 m^2 electron-neutron detector array (ENDA). Large High Altitude Air Shower Observatory (LHAASO) detects cosmic-ray-induced air showers of primary particles between tens of TeV and hundreds of PeV and gamma-ray-induced air showers above 30 TeV [30].

LIGO The Laser Interferometer Gravitational-Wave Observatory (LIGO) is a gravitational-wave detector and consists of two modified Michelson interferometers, one located in Hanford, WA (H1) and one in Livingston, LA (L1). Both interferometers have two orthogonally aligned 4-km optical cavities, the arms, measuring the gravitational-wave strain as a difference of length. The first detection by LIGO was the black hole merger in a binary system in 2015 [4].

MAGIC Major Atmospheric Gamma-Ray Imaging Cherenkov (MAGIC) is a stereo-operating system of IACTs observing gamma-ray sources in the GeV to TeV range. The first Cherenkov telescope was inaugurated in 2004, and the second one followed in 2012 [14].

MASS The Matter Antimatter Superconducting Spectrometer (MASS) was launched by a balloon in 1989 from Prince Albert, Canada, for a 4 h flight. In addition to the measurement of antiprotons, measurements of protons and helium spectra also yielded [136]. A second balloon flight with MASS onboard occurred in 1991 in Fort Sumner, New Mexico [26].

MUBEE The Moscow University Balloon Emulsion Experiment (MUBEE) was a ballon-borne detector on ten flights between 1975 and 1987. The emulsion chamber experiment contained 25 layers of lead absorber interleaved with X-ray and nuclear emulsion films and delivered proton and helium nuclei measurements [145].

OGO The Orbiting Geophysical Observatory (OGO) was launched with the first satellite, *OGO-I*, in 1964, which collected data for a year during a minimum of solar activity. It detected helium and heavier cosmic ions in the MeV energy range [33], [21]. The successful *OGO-I* mission was followed by mission *OGO-II* to *OGO-VI* until the year 1969. Also, data from the *OGO-V* satellite is used in this work [126].

PAMELA The Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics (PAMELA) detector is designed for cosmic-ray antimatter measurements. It has been installed onboard the Russian *Resurs-DK1* satellite since June 2006. It is a high-resolution magnetic spectrometer providing data of the proton spectrum measured from July 2006 to December 2009 [9]. From January 2010 to February 2014 [89] and also both data of the proton and helium spectra [8].

Pierre Auger Experiment The Pierre Auger Observatory is, with its 3000 km^2, the largest cosmic-ray observatory of the world. It was inaugurated in 2008 in western Argentina and measures cosmic rays above 10×10^{17} eV. The experiment consists of several detector types. An array of 1660 water Cherenkov particle detectors is spread over the entire area. Additionally, 24 air fluorescence telescopes cover the detector surface. Finally, three high-elevation fluorescence telescopes overlook a 23.5 km^2 [129].

Pioneer The spacecraft *Pioneer 8*, launched in 1967 on a DELTA rocket, went on an interplanetary mission to the inner solar orbit. During this mission, proton and helium nuclei were measured [80]. Additional helium nuclei data was taken with the high energy telescope (HET) onboard the space probe *Pioneer 10*, launched in 1972 in an Altas-Centaur-rocket on a mission to Jupiter [134].

RICH-II The large Ring-Imaging CHerenkov (RICH-II) telescope was on a high-altitude balloon mission launched from Fort Sumner, New Mexico, in 1997. During the mission, measurements of light cosmic-ray nuclei like protons and helium nuclei in the GeV energy range were taken [41].

RUNJOB The RUssia-Nippon JOint Balloon (RUNJOB) carried emulsion chambers on board ten long-duration balloon flights at an altitude of ~32 km. The flights took place between 1995 and 1999 and started from Kamchatka. Light particles such as protons and helium nuclei were measured, as well as heavier particles like iron nuclei in the TeV energy range [39].

SMILI-I The Superconducting Magnet Instrument for Light Isotopes (SMILI) was sent on a balloon flight for 19 h in 1989 to detect cosmic-ray helium nuclei. The detector included a superconducting magnet for the element type identification, three time-of-flight scintillation counters, and one Cherenkov counter for charge and velocity measurements [25].

SMILI-II The second balloon flight of the Superconducting Magnet Instrument for Light Isotopes (SMILI) was one day in 1991 and provided additional data of cosmic-ray helium nuclei [138].

SOKOL The Russian *COSMOS* satellite carried the SOKOL instruments, which measured high energy cosmic rays above 1 TeV with a setup of two Cherenkov charge detectors and one ionization calorimeter. Between 1984 and 1986, data of protons and helium and iron nuclei were taken [73].

***Swift*-BAT** The *Swift* MIDEX spacecraft was launched in 2004 and carries three detectors: the *Swift* Burst Alert Telescope (*Swift*-BAT), the x-ray telescope (XRT), and the ultraviolet and optical telescope (UVOT). The *Swift* Burst Alert Telescope (*Swift*-BAT) is a coded-aperture gamma-ray telescope operating in the lower gamma-ray range. With its large field of view to monitor 80 % to 90 % of the sky per day and the burst trigger, *Swift*-BAT successfully searches for GRBs. When the instrument detects a transient source, it provides the position with a $2.5'$ accuracy to XRT and UVOT and the multimessenger network GCN [77, 23].

TRACER The instrument called Transition Radiation Array for Cosmic Energetic Radiation (TRACER) was designed to detect heavy cosmic particles like iron nuclei with plastic scintillators and a Cerenkov counter. An array of thin-walled single-wire proportional tubes made measurements of specific ionization and transition radiation possible. A first one-day test flight in 1999 delivered already useful iron data [55], and two long-duration flights in 2003 and 2006 extended this dataset [105].

***Ulysses*-KET** After a solar maximum, the *Ulysses* spacecraft was launched in 1990 onboard the Kiel Electron Telescope (KET). From 1990 to the end of 1994, KET measured low energy protons above a few MeV [113].

VERITAS The IACT-system Very Energetic Radiation Imaging Telescope Array System (VERITAS) is located near Tucson in southern Arizona, operating since 2007. The array of four telescopes triggers stereoscopically in the search for cosmic gamma-ray emission between 100 GeV and 30 GeV [70].

Voyager The space probe *Voyager 1* launched in 1977, 16 days after its identically designed sister probe *Voyager 2*. Both went on Jupiter and Saturn missions; since then, both probes have left the solar system, making them, along with Pioneer 10, the furthest artificial objects from Earth. Both probes keep sending data to the ground station. We are expected to lose contact with the probes in the 2030s due to distance. New publications of proton and helium measurements continue to be made, both from *Voyager 1* [132, 122, 35] and *Voyager 2* [124].

Whipple The Whipple telescope was the first successfully operating IACT and a pioneer in gamma-ray astronomy. It was installed on Mount Hopkins in southern Arizona in 1968. The 10 m parabolic mirror reflected Cherenkov light from air showers into a camera of 37 PMTs. A few years later, the first gamma-ray sources were detected: the galactic SNR Crab Nebula [137] and the extragalactic AGN Markarian 421 [110].

Acknowledgements

Zunächst möchte ich mich bei Prof. Dr. Dr. Wolfgang Rhode bedanken. Vielen Dank für die Möglichkeit, in der Astrophysik zu promovieren und für die Unterstützung und das Vertrauen in meine Arbeit, das ich selbst nicht immer aufbringen konnte. Auch Dir, Dr. Dominik Elsässer, möchte ich danken, Dein profundes Wissen ist eine große Bereicherung für unsere Arbeitsgruppe. Prof. Dr. Westphal, ich danke Ihnen herzlich für die Begutachtung meiner Arbeit.

Liebe Andrea, ich danke dir unglaublich, dass du mit deiner fürsorglichen und verständnisvollen Art immer ein offenes Ohr und einen guten Rat für uns hast. Ein großes Dankeschön geht an meinen Mentor Dr. Nicolas Wöhrl, der mich über 2 Jahre in die Kunst der Wissenschaftskommunikation eingeführt hat und mir mit Rat und Tat zur Seite stand. Danke, dass du mich begleitet und mir so viel beigebracht hast.

Danke an alle, die es mir ermöglicht haben, immer wieder zu den Teleskopen nach La Palma zu reisen. Es war mir eine Ehre, an und mit diesen beeindruckenden Instrumenten zu arbeiten. Ich möchte mich auch bei meinen Schichtteams bedanken, ihr habt jede Reise einzigartig gemacht.

Ich danke der MAGIC Arbeitsgruppe am DESY Zeuthen, dass sie mich für ein paar Monate adoptiert haben und dir, liebe Charlie, dass du mir einen Platz in deiner WG angeboten und damit den Stein ins Rollen gebracht hast. Du bist die beste Mitbewohnerin, die ich mir vorstellen kann. Chiara und Konstantin, ihr seid meine wunderbare Zeuthen-Crew, unsere Mittagspausen an der Spree werde ich nie vergessen.

Ich möchte mich auch herzlich bei der `aict`-Crew bedanken: Jens, Kai und Max, ihr seid tolle Programmierer und Menschen sowieso. Dominik B., dir danke ich für die zeitintensive und nervenaufreibende Produktion der Monte Carlos. Max, danke, dass du dein grenzenloses IT- und Programmierwissen mit uns teilst und Struktur in meine Skripte und Arbeit gebracht hast.

Noah, ich danke dir unglaublich für die tolle Rad- und Bürogemeinschaft. Ich habe den Endspurt mit dir sehr genossen und glaub mir, ich habe noch nie den Endspurt von irgendetwas genossen. Ich würde heute (ja, heute!) noch an der Arbeit sitzen, wenn du nicht gewesen wärst, und das sicher nicht um 8 Uhr morgens.

Danke auch an mein wunderbares ”altes” Büro: Lena, du machst aus jedem Elefanten wieder eine Mücke und bist mir mit deiner hilfsbereiten und gleichzeitig trockenen Art ein großes Vorbild. Simone, mit dir kann ich alles hinterfragen und diskutieren, der Tellerrand ist für uns keine Grenze! Und Jan-Lukas, du bist mir so ans Herz gewachsen, mit dir würde ich mich nach einer durchzechten Nacht jederzeit wieder auf den Roque quälen.

Danke auch an die Truppe vom Flur im ersten Stock, dass die letzten Monate trotz Isolation richtig Spaß gemacht haben. Anno, danke für deine Unterstützung in allen Designfragen!

Danke an alle fleißigen Korrekturleser*innen: Lena, Noah, Simone, Jan-Lukas, Kevin, Stefan, Leonora und Finya, eure konstruktiven Anmerkungen waren super hilfreich. Vielen Dank für eure Zeit und Mühe!

Dann möchte ich mich noch bei meiner lieben Medizinphysik-Crew bedanken, mit euch hat alles angefangen und wir haben viel zusammen durchgestanden. Wir hatten eine unvergessliche Zeit und dafür danke ich euch sehr.

Zu guter Letzt möchte ich mich bei meiner Familie und meinen Freunden bedanken, die mich in all den Jahren begleitet und unterstützt haben. Danke, dass ihr mich so nehmt wie ich bin und ich mich auf euch verlassen kann!